



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

LETICIA DE OLIVEIRA SANTOS

**USO DE APRENDIZADO DE MÁQUINA NO DESENVOLVIMENTO DE MODELOS
DE PREVISÃO DA TEMPERATURA DE OPERAÇÃO DE CÉLULAS
FOTOVOLTAICAS**

FORTALEZA

2021

LETICIA DE OLIVEIRA SANTOS

USO DE APRENDIZADO DE MÁQUINA NO DESENVOLVIMENTO DE MODELOS DE
PREVISÃO DA TEMPERATURA DE OPERAÇÃO DE CÉLULAS FOTOVOLTAICAS

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 Carvalho

Coorientador: Prof. Dr. Clodoaldo Oliveira Carvalho Filho

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S236u Santos, Leticia de Oliveira.

Uso de aprendizado de máquina no desenvolvimento de modelos de previsão da temperatura de operação de células fotovoltaicas / Leticia de Oliveira Santos. – 2021.

78 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Mecânica, Fortaleza, 2021.

Orientação: Prof. Dr. Paulo Cesar Marques Carvalho.

Coorientação: Prof. Dr. Clodoaldo Oliveira Carvalho Filho.

1. Temperatura de Operação da Célula FV. 2. Tecnologia Fotovoltaica. 3. Aprendizado de Máquina. I. Título.

CDD 620.1

LETICIA DE OLIVEIRA SANTOS

USO DE APRENDIZADO DE MÁQUINA NO DESENVOLVIMENTO DE
MODELOS DE PREVISÃO DA TEMPERATURA DE OPERAÇÃO DE CÉLULAS
FOTOVOLTAICAS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica, do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial para a obtenção do Título de Mestre em Engenharia Mecânica. Área de Concentração: Processos, Equipamentos e Sistemas para Energias Renováveis.

Aprovada em 20/12/2021

BANCA EXAMINADORA

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

Prof. Dr. Clodoaldo de Oliveira Carvalho Filho (Coorientador)
Universidade Federal do Ceará (UFC)

Prof.^a Dra. Carla Freitas de Andrade
Universidade Federal do Ceará (UFC)

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

Prof. Dr. Douglas Bressan Riffel
Universidade Federal de Sergipe (UFS)

AGRADECIMENTOS

Aos Professores Dr. Paulo Cesar Marques Carvalho e Dr. Clodoaldo Oliveira Carvalho Filho pela orientação e coorientação durante o meu mestrado.

Às colegas, Renata Imaculada Soares Pereira e Ivonne Montero Dupont do Laboratório de Energias Alternativas (LEA-UFC) onde este trabalho foi desenvolvido, pelos recursos disponibilizados.

Aos membros da banca, Prof.^a Dra. Carla Freitas de Andrade, Prof. Dr. Paulo Alexandre Costa Rocha e Prof. Dr. Douglas Bressan Riffel pelo interesse e disponibilidade.

Ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

Aos meus pais, irmãos, sobrinhos e marido, que nos momentos de minha ausência dedicados ao estudo superior, sempre fizeram entender que o futuro é feito a partir da constante dedicação no presente!

Agradeço a todos os professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender.

E à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo financiamento da pesquisa de mestrado via bolsa de estudos.

“Os poetas reclamam que a ciência tira a beleza das estrelas - meros globos de átomos de gás. Mas eu também posso vê-las de noite no deserto e senti-las. Vejo menos ou mais?”

(Richard Feynman)

RESUMO

Projetar um sistema Fotovoltaico (FV) envolve estimar a sua produção de energia elétrica, que é influenciada pelas condições ambientais e pela temperatura de operação da célula FV (T_c). Por isso, é necessário um modelo para estimar T_c . A presente pesquisa tem como objetivo o desenvolvimento de um modelo para determinação de T_c baseado em técnicas de Aprendizado de Máquina (*Machine Learning* - ML) aplicadas a um conjunto de dados de irradiação solar, temperatura ambiente, velocidade do vento, temperatura da célula, dia, ano e hora coletados no Laboratório de Energias Alternativas (LEA) da Universidade Federal do Ceará (UFC). É apresentado inicialmente um estudo dos modelos recentes para determinação de T_c por meio de uma revisão da literatura. 33 equações encontradas na literatura para estimar T_c são compiladas em apenas 3 formulações gerais. Posteriormente, o conjunto de dados coletados entre 2018 e 2020 no LEA é tratado e usado na aplicação de técnicas de aprendizagem por regressão. As técnicas aplicadas são de Regressão Linear (RL), *Support Vector Machine* (SVM), *Gaussian Process Regression* (GPR) e *Neural Network* (NN). Os resultados obtidos por meio de RL, SVM, GPR e NN são comparados, tais modelos apresentam uma magnitude média do erro, o RMSE, de 3,5377; 2,8374; 2,651; e 2,9706 graus no conjunto de teste, respectivamente. O melhor modelo de ML desenvolvido, o modelo GPR, é comparado com alguns modelos convencionais presentes na literatura para previsão de T_c e apresenta a melhor performance entre estes.

Palavras-chave: Temperatura de Operação da Célula FV. Tecnologia Fotovoltaica. Aprendizado de Máquina. Aprendizagem por Regressão.

ABSTRACT

Designing a photovoltaic (PV) system involves estimating its electricity production, which is influenced by the environmental conditions and the PV cell operating temperature (T_c). Therefore, a model is needed to estimate T_c . This research aimed to develop a model for determining T_c based on Machine Learning (ML) techniques applied to a dataset of solar irradiation, ambient temperature, wind speed, cell temperature, day, year and hour collected at the Laboratory of Alternative Energy (LEA) of the Federal University of Ceará (UFC). Initially, a study of the recent models for determining T_c through a literature review is presented. 33 equations found in the literature to estimate T_c are summarized in just 3 general forms. Subsequently, the dataset collected between 2018 and 2020 in the LEA is treated and used in the application of regression learning techniques. The techniques applied are Linear Regression (RL), Support Vector Machine (SVM), Gaussian Process Regression (GPR) and Neural Network (NN). The results obtained by the application of RL, SVM, GPR, and NN are compared, such models presented the average magnitude of the error, the RMSE, of 3.5377; 2.8374; 2,651; and 2.9706 degrees in the test set, respectively. The best ML model developed, the GPR model, is compared with some conventional models present in the literature for forecasting T_c and presents the best performance between those.

Keywords: PV Cell Operating Temperature. Photovoltaic technology. Machine Learning. Regression Learning.

LISTA DE FIGURAS

Figura 1 – Processos térmicos em um módulo FV	19
Figura 2 – Imagem térmica infravermelha de módulos FV instalados no LEA.	33
Figura 3 – Representação da célula FV mostrando a posição ideal da medição de T_c	33
Figura 4 – Estrutura de um problema de aprendizagem de máquina supervisionada.	37
Figura 5 – Representação da escolha de coeficientes pela função de custo.	37
Figura 6 – Mapa conceitual do estudo dos modelos para determinação de T_c	47
Figura 7 – Sistema FV instalado no LEA-UFC, composto de 6 módulos, 1 anemômetro, 1 piranômetro e sensores de temperatura. Vista superior e inferior dos painéis.	48
Figura 8 – Fluxo de trabalho para treinamento de modelos de regressão no <i>Regression Learner App</i>	50
Figura 9 – Histogramas de frequência e correlações entre os pares de variáveis do banco de dados do LEA-UFC	58
Figura 10 – Gráficos de caracterização de V_w , T_a , T_c e G : (a) ao longo do dia (b) ao longo do ano (c) estatísticas descritivas.	59
Figura 11 – Valores de T_c previstos pelo modelo GPR e medidos ao longo de cinco dias vs. o número de registro.	62
Figura 12 – Gráficos Predito vs. Real: (a) modelo RL (b) modelo SVM (c) modelo GPR (d) modelo RN.	63
Figura 13 – Resposta prevista em amarelo sob os valores do conjunto de teste ao longo do ano: (a) modelo RL (b) modelo SVM (c) modelo GPR (d) modelo RN.	64
Figura 14 – Resposta prevista em amarelo sob os valores do conjunto de teste ao longo do dia: (a) modelo RL (b) modelo SVM (c) modelo GPR (d) modelo RN.	65
Figura 15 – Valores de T_c previstos pelo modelos comparativos e GPR e medidos ao longo de dois dias vs. o número de registro.	67

LISTA DE TABELAS

Tabela 1 – Parametro de Ross para vários tipos de instalação	21
Tabela 2 – Equações para determinação de T_c encontradas na literatura a partir do ano 2000 incluindo comentários pertinentes	24
Tabela 3 – Componentes das correlações analisadas (\checkmark significa que se trata de um componente da correlação, \times caso contrário).	28
Tabela 4 – Modelos Isc-Voc para determinação de T_c	34
Tabela 5 – Tipos de modelos de RL e suas definições	39
Tabela 6 – Exemplos de funções de kernel semidefinidas	42
Tabela 7 – Exemplos de funções de RN semidefinidas	45
Tabela 8 – Modelos aplicados aos dados do LEA no RL-App.	52
Tabela 9 – Principais configurações dos modelos	52
Tabela 10 – Termos das correlações lineares quando expressas pela equação 5.1	55
Tabela 11 – Termos das correlações não-lineares quando expressas pela equação 5.2.	56
Tabela 12 – Termos das correlações não-lineares Duffies e Beckman II e Mattei quando expressas pela equação 5.3.	56
Tabela 13 – Principais resultados dos modelos de ML do 1 até 4	61
Tabela 14 – Resultados do modelos GPR e modelos comparativos	66

LISTA DE ABREVIATURAS E SIGLAS

FV	Fotovoltaico
ABSOLAR	Associação Brasileira de Energia Solar Fotovoltaica
ML	<i>Machine Learning</i>
LEA	Laboratório de Energias Alternativas
UFC	Universidade Federal do Ceará
RL	Regressão Linear
SVM	<i>Support Vector Machine</i>
GPR	<i>Gaussian Process Regression</i>
RN	Redes Neurais
BIPV	<i>Building-Integrated Photovoltaics</i>
NOCT	<i>Nominal Operating Cell Temperature</i>
NTE	<i>Nominal Terrestrial Environment</i>
CFD	<i>Computational Fluid dynamics</i>
STC	<i>Standard Test Conditions</i>
a-Si	Silício amorfo
HCPV	<i>High Concentrating Photovoltaic</i>
EVA	<i>Ethylene vinyl acetate</i>
RMSE	<i>root-mean-square error</i>
PG	Processo Gaussiano
NEPAU	Núcleo de Ensino e Pesquisa em Agricultura Urbana
REMS	<i>Renewable Energy Monitoring System</i>
INMET	Instituto Nacional de Meteorologia
RL-App	<i>Regression Learner App</i>

LISTA DE SÍMBOLOS

T_c	Temperatura de operação da célula
V_{oc}	Tensão de circuito aberto
η	Eficiência elétrica
FF	Fator de forma
P_{max}	Potência máxima
I_{sc}	Corrente de curto-circuito
G	Irradiância solar
T_a	Temperatura ambiente
V_w	Velocidade do vento
G_{NOCT}	Irradiância solar nas condições NOCT
$T_{a,NOCT}$	Temperatura ambiente nas condições NOCT
$V_{w,NOCT}$	V_w médio nas condições NOCT
T_b	Temperatura da parte traseira do módulo
G_{ref}	Irradiância solar nas Condições Padrão de Teste
r	Coefficiente de Ross
U_L	Coefficiente de perda de calor
T_r	Temperatura de referência em Kelvin
N_s	Número de células em série no módulo
τ	Transmissividade
α	Absorbância

SUMÁRIO

1	INTRODUÇÃO	15
2	MODELOS PARA DETERMINAÇÃO DA TEMPERATURA DE OPERAÇÃO DA CÉLULA FV	18
2.1	Principais componentes dos modelos	27
2.1.1	<i>Variáveis Meteorológicas</i>	27
2.1.1.1	<i>Irradiância efetiva</i>	27
2.1.1.2	<i>Velocidade do vento</i>	29
2.1.1.3	<i>Temperatura ambiente</i>	30
2.1.2	<i>Propriedades dependentes do material e/ou configuração do sistema</i>	31
2.1.2.1	<i>Eficiência elétrica FV</i>	31
2.1.2.2	<i>Absorção solar e transmitância da camada de vidro</i>	31
2.1.2.3	<i>Outros fatores</i>	32
2.2	Validação do modelo: determinação de T_c	32
3	TÉCNICAS DE APRENDIZADO DE MÁQUINA	36
3.1	Regressão Linear (RL)	38
3.2	Support Vector Machine (SVM)	39
3.3	Gaussian Process Regression (GPR)	42
3.4	Redes Neurais (RN)	44
4	METODOLOGIA	46
4.1	Etapa 1: Estudo dos modelos presentes na literatura para determinação de T_c	46
4.2	Etapa 2 - Parte 1: Obtenção e Tratamento dos dados	47
4.2.1	<i>Obtenção de dados</i>	47
4.2.2	<i>Tratamento dos dados</i>	49
4.3	Etapa 2 - Parte 2: Aplicação dos modelos no Regression Learner App	49
4.3.1	<i>Regression Learner App</i>	49
4.3.2	<i>Aplicação dos modelos no RL-App</i>	50
5	RESULTADOS	53
5.1	Etapa 1: Estudo dos modelos para determinação de T_c	53
5.1.1	<i>Forma geral das correlações para determinação de T_c</i>	53

5.1.2	<i>Modelos comparativos</i>	54
5.2	Etapa 2: Parte 1 - Análise dos dados coletados	57
5.2.1	<i>Irradiância solar</i>	57
5.2.2	<i>Temperatura ambiente</i>	59
5.2.3	<i>Velocidade do vento</i>	60
5.2.4	<i>Temperatura da célula</i>	60
5.3	Etapa 2: Parte 2 -Resultados da aplicação da técnicas de ML	60
5.3.1	<i>Modelos Comparativos</i>	66
6	CONCLUSÃO	68
7	PUBLICAÇÕES ATÉ A PRESENTE DATA	70
	REFERÊNCIAS	71

1 INTRODUÇÃO

A geração fotovoltaica (FV) é considerada uma forma "limpa" de se produzir energia elétrica a partir da conversão direta de energia solar. Consequentemente, tem se popularizado no Brasil e no Mundo. De acordo com a Associação Brasileira de Energia Solar Fotovoltaica (ABSOLAR), a potência instalada no Brasil, somando a geração centralizada e a distribuída, passou de 7 MW em 2012, para 10 859 MW até setembro de 2021, um crescimento exponencial que gerou investimentos, empregos, arrecadação de tributos, redução do custo de geração, além de cerca de 12,4 milhões de toneladas de CO₂ evitadas (ABSOLAR, 2021). A nível mundial o cenário não é muito diferente, a produção de eletricidade a partir da fonte solar FV aumentou de 4 TWh em 2005, para 681 TWh em 2019 (IEA, 2021). Tal crescimento leva a necessidade de estudos do comportamento e produção de sistemas FV's.

Projetar um sistema FV envolve estimar a sua produção de energia elétrica. Portanto, deve ser considerado o desempenho do sistema FV, que é influenciado pelas condições ambientais e pela temperatura de operação da célula FV (T_c) (TRINURUK *et al.*, 2009). Com isso, se faz necessário um modelo térmico para estimar T_c . Projetar, implementar e monitorar efetivamente o desempenho de usinas FV é uma tarefa difícil devido à influência de fatores relacionados à física das células solares e ao meio ambiente. Os desempenhos dos modelos de previsão são afetados por muitos elementos de incertezas e nem sempre fica claro como a metodologia de desenvolvimento dos modelos e os componentes do mesmo contribuem para o erro final (GIGONI *et al.*, 2017). Espera-se que um modelo de desempenho identifique e quantifique a influência de todos os fatores significativos, incluindo efeitos elétricos, térmicos, espectrais, solares e ópticos (KING *et al.*, 2004).

Os modelos para estimativa de T_c podem ser classificados basicamente por duas abordagens: modelos dinâmicos e de estado estacionário. A principal diferença entre as metodologias é que na abordagem de estado estacionário todos os parâmetros são considerados independentes do tempo, enquanto na abordagem dinâmica alguns parâmetros são dependentes do tempo (JAKHRANI *et al.*, 2011).

Os modelos de estado estacionário são relativamente simples, caracterizados por baixo tempo computacional; no entanto, esses modelos não são tão flexíveis se comparado aos modelos dinâmicos (SHAFIEIAN *et al.*, 2019). A intensidade da irradiância solar incidente e outros parâmetros que afetam o desempenho dos módulos FV, em um curto período de tempo, é considerada constante. Por exemplo, se a variação da transferência de calor entre o módulo

FV e a vizinhança for baixa, essa pode ser considerada uma condição de estado estacionário; temperaturas em diferentes posições do módulo FV também são consideradas constantes para a condição de estado estacionário (COSKUN *et al.*, 2017).

Modelos dinâmicos podem expressar resultados mais realistas e precisos considerando a flutuação da irradiância solar em um curto período (COSKUN *et al.*, 2017); esta técnica é baseada na determinação de T_c usando um balanço de energia (TRINURUK *et al.*, 2009). Portanto, os modelos dinâmicos descrevem T_c e seu processo térmico em mais detalhes em comparação com os modelos de estado estacionário. No entanto, esses modelos são relativamente complexos, exigindo mais tempo computacional, custo e esforço (SHAFIEIAN *et al.*, 2019).

Nos últimos anos, esforços consideráveis têm sido dedicados por parte da comunidade científica ao desenvolvimento e melhoria de abordagens para o cálculo de T_c a partir de diversos parâmetros. Nesse processo, o uso de modelos de estado estacionário para determinação de T_c tornou-se comum. No entanto, ainda são escassas pesquisas para as condições climáticas de áreas semiáridas do mundo, como o Nordeste brasileiro. Com o intuito de suprir tal lacuna, a presente pesquisa tem como objetivo geral estudar modelos para determinação de T_c presentes na literatura e desenvolver um modelo de estado estacionário acurado para determinação de T_c baseado em técnicas de Aprendizado de Máquina (*Machine Learning (ML)*) aplicadas a dados coletados no Laboratório de Energias Alternativas (LEA) da Universidade Federal do Ceará (UFC) na cidade de Fortaleza, Ceará, localizada na região Nordeste do Brasil, próxima à Linha do Equador. Tendo como objetivos específicos:

- Estudar os modelos para determinação de T_c presentes na literatura a partir do ano 2000, buscando as melhores praticas adotadas;
- Tratar e preparar os dados coletados entre 2018 e 2020;
- Aplicar técnicas de Regressão Linear (RL) aos dados tratados para previsão de T_c ;
- Aplicar técnicas de *Support Vector Machine (SVM)* aos dados tratados para previsão de T_c ;
- Aplicar técnicas de *Gaussian Process Regression (GPR)* aos dados tratados para previsão de T_c ;
- Aplicar técnicas de Redes Neurais (RN) aos dados tratados para previsão de T_c ;
- Apresentar e caracterizar os resultados obtidos para as diferente técnicas, bem como fazer o comparativo dos resultados.

- Comparar a precisão do melhor modelo proposto com modelos convencionais presentes na literatura para previsão de T_c .

O presente trabalho foi dividido em duas etapas, inicialmente são apresentados os resultados referentes ao primeiro objetivo específico acima. É apresentada uma revisão de modelos recentes de T_c presentes na literatura com o intuito de estudar as melhores práticas de escolha de componentes dos modelos e arranjos experimentais mais adequados para compor um modelo preciso. Além disso, 33 equações encontradas na literatura para estimar T_c são resumidas em apenas 3 formas gerais. Posteriormente, foi trabalhada a segunda etapa de processamento e tratamento dos dados, incluindo transformação e limpeza dos mesmos - padronização e tratamento dos desvios e inconsistências- e a aplicação das técnicas de ML. Os resultados obtidos por meio de RL, SVM, GPR e RN são apresentados e comparados.

A presente dissertação está estruturada em 7 capítulos. O primeiro capítulo contém uma introdução com a motivação do trabalho e os objetivos. O capítulo 2 apresenta uma revisão da literatura relacionada a elaboração de modelos para determinação de T_c . Este capítulo também mostra as melhores práticas experimentais apresentadas na literatura. O capítulo 3 trata de algumas técnicas de aprendizado de máquina aplicadas na presente dissertação no desenvolvimento dos modelos de previsão de T_c . O capítulo 4 descreve a metodologia aplicada nas etapas de estudo dos modelos e de aquisição, extração e transformação dos dados utilizados, e a metodologia executada na aplicação das técnicas de ML. O capítulo 5 contém os resultados obtidos nas etapas 1 e 2 deste trabalho, respectivamente, junto às discussões pertinentes. O capítulo 6 apresenta as conclusões. Finalmente, o capítulo 7 apresenta as publicações a partir do presente estudo até o momento.

2 MODELOS PARA DETERMINAÇÃO DA TEMPERATURA DE OPERAÇÃO DA CÉLULA FV

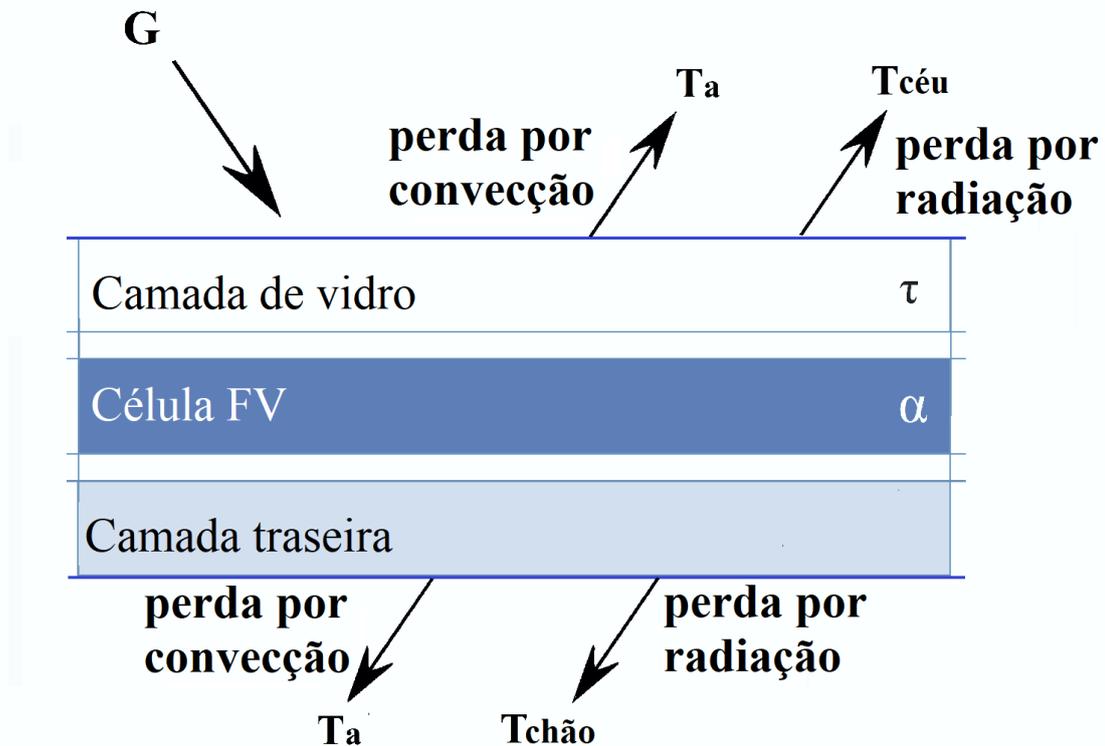
A T_c é fundamental na caracterização do comportamento do módulo, porém, não é um parâmetro facilmente disponível (MORA *et al.*, 2015). Na literatura científica, um número significativo de contribuições demonstrando um efeito adverso de um aumento em T_c na performance de módulos FV tem sido encontrado (MORA *et al.*, 2016). Este efeito adverso resulta de uma queda no *bandgap* da célula FV com o aumento de T_c , implicando em uma mudança no comportamento elétrico da célula FV: tensão de circuito aberto (V_{oc}), eficiência (η), fator de forma (FF) e potência máxima (P_{max}) diminuem; e há um aumento desprezível da corrente de curto-circuito (I_{sc}) (CUCE *et al.*, 2017).

O ambiente térmico que estabelece o valor instantâneo de T_c é um tanto complexo. Um processo que leva a estimativa de T_c é o balanço energético relevante do módulo, que deve considerar ambos processos internos, que ocorrem durante o bombardeamento de fótons no material semicondutor, resultando na produção de eletricidade e liberação na forma de calor da energia não convertida por mecanismos padrões de transferência de calor como convecção e radiação (Fig. 1) (SKOPLAKI; PALYVOS, 2009b). Na maioria dos casos, esses mecanismos afetam a frente e a traseira do módulo, já que em instalações típicas é deixada uma distância entre o módulo e o telhado/ laje parra facilitar a liberação de calor, e o módulo poder operar de forma mais eficiente. Se tratando de painéis instalados livremente (*free-standing*), isto é, sem nenhuma estrutura abaixo dos painéis, também se deve levar em consideração a condução pela estrutura de suporte: o calor é transportado para as superfícies por condução, e das superfícies para o ambiente por convecção e radiação (SALA, 1989).

Em resumo, T_c depende de fatores tais como: propriedades de absorção das células; materiais constituintes dos módulos (semicondutores, células, camadas, encapsulantes, dentre outros); dissipação térmica para o meio ambiente; instalação e condições ambientais (localização, irradiância solar (G), temperatura ambiente (T_a) e velocidade do vento (V_w)) (GARCÍA; BALENZATEGUI, 2004; MARION *et al.*, 1999).

A determinação de T_c é o último passo antes de alcançar a possibilidade de previsão da geração FV (DENOIX *et al.*, 2014). As correlações para T_c encontradas na literatura geralmente descrevem arranjos FV independentes, coletores FV/térmicos e instalações FV integradas em edifícios (*Building-Integrated Photovoltaics* (BIPV)) (JAKHRANI *et al.*, 2011; DUBEY *et al.*, 2013). Essas correlações expressam T_c em função de variáveis meteorológicas relevantes e

Figura 1 – Processos térmicos em um módulo FV



Fonte: Adaptado de (SKOPLAKI; PALYVOS, 2009b)

incluem propriedades e parâmetros dependentes do material e do sistema, dependendo do tipo de arranjos/esquemas de montagem. Algumas correlações expressam o efeito adverso de um aumento de T_c na eficiência do módulo FV (FERNÁNDEZ *et al.*, 2013; SIEFER; BETT, 2014). Métodos baseados em parâmetros elétricos consideram que parâmetros como V_{oc} variam com a temperatura, permitindo estimar T_c com o auxílio de medições de dados elétricos, utilizando diferentes procedimentos (RODRIGO *et al.*, 2014).

Grande parte das correlações inclui um estado de referência e valores correspondentes das variáveis significativas. Um método para formular T_c envolve o uso da *Nominal Operating Cell Temperature* (NOCT); (CASTAÑER *et al.*, 2003), definido como uma temperatura do dispositivo no Ambiente Terrestre Nominal (*Nominal Terrestrial Environment* (NTE)): irradiância solar de 800 W/m^2 (G_{NOCT}); T_a de 20°C ($T_{a,NOCT}$); V_w médio de 1 m/s ($V_{w,NOCT}$); carga elétrica zero (circuito aberto); estrutura de montagem independente orientada "normal ao meio-dia solar" (STULTZ; WEN, 1977; ASTM E1036M-96e2, 1996).

As condições padrões são empregadas para a 'classificação' ou 'especificação'. Os parâmetros de desempenho associados são geralmente as classificações (especificações) da placa de identificação do fabricante ou resultados de testes de um laboratório de teste de módulo. A

precisão dessas especificações de desempenho é fundamental para o projeto de usinas FV, o ponto de referência é fornecido a partir do qual o desempenho em todas as outras condições operacionais é provido (DOWS; GOUGH, 1995). Idealmente, o NOCT deve ser exatamente o mesmo, independentemente do laboratório de teste, local, mês ou estação (KUITCHE *et al.*, 2011). Em Koehl *et al.* (2011) é apresentada uma proposta de modificação do método de obtenção de NOCT para calcular a chamada 'Temperatura Nominal do Módulo Realista', mais representativa, visto que em condições de NOCT as condições de exposição são "circuito aberto", que não contam para a retirada de energia elétrica (KOEHL *et al.*, 2011).

Se não estiverem disponíveis nos dados fornecidos pelo fabricante do módulo FV, os parâmetros necessários podem ser medidos durante testes externos em condições reais de operação (KING *et al.*, 1997). Em plantas BIPV, os módulos são instalados a uma distância ideal da fachada do edifício, por isso, o balanço de energia não se limita às camadas dos módulos. Neste caso, ambos os lados dos módulos estão sob condições ambientais bastante diferentes e o modelo NOCT simples pode subestimar a T_c em até 20°C (DAVIS *et al.*, 2001). Em tais aplicações é necessário um sistema de três equações simultaneamente, um balanço de energia é feito em cada uma das três camadas - módulo FV, a camada de ar entre o módulo e a parede, e a parede - resultando nas respectivas equações, caracterizando cada temperatura. Esses balanços consideram a transferência de calor entre as camadas, ambiente e espaço interior. A metodologia envolve uma abordagem de análise concentrada, adotando condições uniformes ao longo da lacuna; em estudos mais detalhados são empregados modelos dinâmicos e métodos de dinâmica de fluidos computacional (*Computational Fluid dynamics (CFD)*).

Do ponto de vista matemático, os modelos para determinação de T_c podem ser explícitos, fornecendo T_c diretamente, ou implícitos, envolvendo variáveis que dependem de T_c . No segundo caso, é necessário um procedimento iterativo (MOHAMMED *et al.*, 2019). Os modelos implícitos são baseados nas propriedades térmicas do módulo e nos mecanismos de transferência de calor. T_c é determinada iterativamente a partir do balanço de energia aplicado ao módulo. Os métodos explícitos calculam T_c usando parâmetros conhecidos (MORA *et al.*, 2015). T_c pode ser associado à temperatura da parte traseira do módulo (T_b). A diferença entre ambas temperaturas depende dos materiais do substrato do módulo e dos níveis de G . Um modelo térmico empírico implícito simples relacionando T_c e T_b foi aplicado com sucesso em vários arranjos de módulos FV, fornecendo o valor T_c esperado com uma precisão de cerca de $\pm 5^\circ\text{C}$ (Eq. (2.1)). Esta magnitude de incertezas em T_c resulta num efeito de menos de 3% na saída de

potência do módulo (KING *et al.*, 2004).

$$T_c = T_b + \frac{G}{G_{ref}} \Delta T \quad (2.1)$$

G é a irradiância solar no módulo FV (W/m^2), G_{ref} é a irradiância solar nas Condições Padrão de Teste (*Standard Test Conditions* (STC)) (1000 W/m^2) e ΔT é a diferença de temperatura entre T_c e T_b na G_{ref} .

A equação explícita mais simples encontrada liga T_c a T_a e G (Eq.(2.2)) (ROSS, 1976).

$$T_c = T_a + rG \quad (2.2)$$

Essa expressão linear não considera o vento ou a carga elétrica, mas um parâmetro dimensional r , o chamado coeficiente de Ross. Os valores reportados para r variam e podem ser categorizados qualitativamente de acordo com o nível de integração e o espaço por trás dos módulos (BURESH, 1984; NORDMANN; CLAVADETSCHER, 2003). Alguns estudos associam valores estimados de r a diferentes tipos de arranjos/esquemas de montagem (SKOPLAKI; PALYVOS, 2009b) (Tabela 1).

Tabela 1 – Parametro de Ross para vários tipos de instalação

Tipo de arranjo FV	r ($\text{K m}^2/\text{W}$)
Independente	0,021
Teto plano	0,026
Teto inclinado bem ventilado	0,020
Teto inclinado não tão bem ventilado	0,034
Telhado inclinado altamente integrado, mal ventilado	0,056
FV transparente integrado a fachada	0,046
FV opaco integrado a fachada	0,054

Fonte: Adaptado de (SKOPLAKI *et al.*, 2008).

Sob condições NOCT, a Eq.(2.3) assume o mesmo T_a para ambos os lados do módulo, resultado do fato de que a diferença de temperatura ($T_c - T_a$) é praticamente independente de T_a , mas linearmente proporcional a G (DUFFIE; BECKMAN, 2013). Além disso, o coeficiente de perda de calor (U_L) é considerado constante nesta equação, este coeficiente inclui as perdas de calor por convecção, radiação e condução. Portanto, o balanço de energia para uma área unitária do módulo FV é:

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

Em condições NOCT, não há operação de carga, ou seja, $\eta = 0$ (DUFFIE; BECKMAN, 2013). Como η é uma função de T_c , Eq. (2.3) é uma equação implícita de T_c , uma correlação particularmente adequada para módulos instalados de forma independente.

Medições diretas da célula incluem parâmetros de performance básicos (I_{sc} e corrente no ponto de máxima potência, V_{oc} e tensão no ponto de máxima potência) na condição padrão (referência), bem como outros coeficientes (BARANGE; SHARMA, 2019). Métodos baseados em medidas diretas têm se mostrado mais precisos do que métodos baseados em parâmetros meteorológicos, o último apresenta incertezas associadas às fontes de dados e ao próprio modelo (RODRIGO *et al.*, 2014). Porém, é uma ferramenta muito útil dada a vantagem de que T_c pode ser estimado em qualquer local usando dados meteorológicos (PÉREZ-HIGUERAS *et al.*, 2012).

Entre os métodos baseados em medições diretas, alguns deles apresentam a vantagem de os parâmetros poderem ser obtidos através da literatura científica ou por dados do fabricante (MULLER *et al.*, 2011; IEC, 2011; JU *et al.*, 2013; YANDT *et al.*, 2012). No entanto, valores de parâmetro precisos são necessários, porque os métodos fornecidos são altamente sensíveis aos valores escolhidos. Supõe-se que a influência espectral, coeficientes de temperatura do módulo e perdas ópticas, estão disponíveis nos resultados dos testes com módulos individuais. Este tipo de modelo matemático inclui perdas internas associadas à resistência da fiação e incompatibilidade do módulo, valores difíceis de prever ou determinar explicitamente. Outros métodos desse tipo apresentam a desvantagem de que os parâmetros devem ser ajustados através de testes experimentais ao ar livre -por exemplo, pode exigir um módulo com um sensor interno de temperatura da célula - ou a partir de medições internas em um simulador de Sol.

Considerando um conjunto de módulos FV, o modelo térmico geralmente o trata como um grande módulo. Em geral, o efeito de incompatibilidade e perdas de resistência é pequeno (<5%) em relação ao desempenho esperado das classificações da placa de identificação dos módulos individuais (KING *et al.*, 2004). Idealmente, o desempenho das medições está disponível na escala do arranjo; o que pode melhorar ainda mais a precisão do modelo. Uma vez desenvolvido, o modelo pode ser aplicado durante o projeto de desempenho para usinas FV autônomas. Tais sistemas se comparados a sistemas conectados à rede são mais complexos, pois incluem baterias para armazenamento de energia e gerador como fonte de energia auxiliar.

Equações para determinação de T_c encontradas na literatura a partir do ano 2000 são apresentadas na Tabela 2, incluindo comentários. Essas equações foram desenvolvidas considerando uma geometria de montagem específica ou um nível de integração de construção.

Portanto, deve-se ter cuidado ao aplicar uma expressão qualquer para determinação de T_c . Dependendo da aplicação específica, alguns métodos podem ser mais adequados do que outros. Escolher o método mais adequado não é fácil e depende principalmente da disponibilidade de informações do módulo, da precisão necessária e de questões técnicas.

Tabela 2 – Equações para determinação de T_c encontradas na literatura a partir do ano 2000 incluindo comentários pertinentes

Modelos	Formulação	Comentários
Fernández <i>et al.</i> ¹ (FERNÁNDEZ <i>et al.</i> , 2013)	$T_c = \frac{(V_{oc} - c_1 G - c_3)}{c_2}$	c_1 , c_2 e c_3 são coeficientes lineares
Durisch <i>et al.</i> (DURISCH <i>et al.</i> , 2007; TOPI <i>et al.</i> , 2007)	$T_c = T_a + kG$	$k = \frac{\Delta T_c}{\Delta G}$: 0,02 - 0,04°Cm ² /W
Nordmann (NORDMANN; CLAVADTSCHER, 2003)	$T_c = T_a + kG$	0,02 < k < 0,056 para situações BIPV
Krauter (KRAUTER, 2004)	$T_c = T_a + kG$	$k = 0,03, 0,012, 0,0058$ para sistemas domésticos especificado em (KRAUTER, 2004)
Mondol I (MONDOL <i>et al.</i> , 2005)	$T_c = T_a + 0,031G$	Para V_w acima de 1m/s com um U_L constante
Hove (HOVE, 2000)	$T_c = T_a + G \left(\frac{\tau\alpha - \eta}{U_L} \right)$	$\frac{\tau\alpha}{U_L}$ determinado experimentalmente
Tiwari (TIWARI, 2002)	$T_c = T_a + G \left(\frac{\tau\alpha}{U_L} \right) \left[1 - \left(\frac{\eta}{\tau\alpha} \right) \right]$	$\frac{\tau\alpha}{U_L}$ assumido como constante
Eicker (EICKER, 2006)	$T_c = T_a + G \left(\frac{\alpha}{U_L} \right) \left[1 - \frac{\eta}{\alpha} \right]$	$U_L = h_{radn} + h_{conv}$
Standard (MARKVART, 2000)	$T_c = T_a + \left(\frac{G}{G_{NOCT}} \right) (T_{c,NOCT} - T_{a,NOCT})$	Modelo de estado estacionário para um módulos de silício cristalino e células de filmes finos
Davis (DAVIS <i>et al.</i> , 2001)	$T_c = T_a + \left(\frac{G}{G_{NOCT}} \right) (T_{c,NOCT} - T_{a,NOCT}) \left[1 - \left(\frac{\eta}{\tau\alpha} \right) \right]$	Assume que U_L é constante
Mondol II (MONDOL <i>et al.</i> , 2007)	$T_c = T_a + 0,031G - 0,058$	Para V_w acima de 1m/s com U_L constante
Tselepis (TSELEPIS; TRIPANAGNOSTOPOULOS, 2002)	$T_c = 30 + 0,0175(G - 150) + 1,14(T_a - 25)$	Estima T_c para um módulo FV de silício amorfo (Silício amorfo (a-Si))

Tiwari I (TIWARI <i>et al.</i> , 2006)	$T_c = \frac{\rho G(\tau\alpha - \eta) + (U_i T_a + U_T T_b)}{(U_i + U_T)}$	U_i, U_T : coeficientes de transferência de calor
Tiwari II (TIWARI <i>et al.</i> , 2006)	$T_c = \frac{\tau[\alpha_c p + \alpha_\gamma(1 - \beta_c)G - \eta_c G \beta_c + U_i T_a + U_T T_b]}{(U_i + U_T)}$	T_b função do produto ηT_c
Almonacid ¹ (ALMONACID <i>et al.</i> , 2012)	$T_c = T_a + d_1 G + d_2 V_w$	d_1, d_2 : parâmetros de regressão multilinear
Markvart (MARKVART, 2000)	$T_c = 0,943T_a + 4,3 + 0,028G - 1,528V_w$	
Muzathik (MUZATHIK, 2014)	$T_c = 0,943T_a + 0,3529 + 0,0195G - 1,528V_w$	
Akyuz <i>et al.</i> (AKYUZ <i>et al.</i> , 2012)	$T_c = 0,95T_a + 3,1 + 0,025G - 0,3V_w$	
NOCT-1p (MORA <i>et al.</i> , 2015)	$T_c = T_a + \left(\frac{G}{G_{NOCT}}\right) (T_{c,NOCT} - T_{a,NOCT}) + a(V_w - V_{w,NOCT})$	a : parâmetro determinado por ajuste de dados
NOCT-2p (MORA <i>et al.</i> , 2015)	$T_c = T_a + b \left(\frac{G}{G_{NOCT}}\right) (T_{c,NOCT} - T_{a,NOCT}) + c(V_w - V_{w,NOCT})$	b e c : parâmetros empíricos
ISFOC ¹ (RUBIO <i>et al.</i> , 2008)	$T_c = T_{h-s} + \left(\eta C_g \sum \frac{L_i}{\lambda_i}\right) G$	i, L_i e λ_i : camada, espessura e condutividade térmica do material atrás da célula, respectivamente
Faiman (FAIMAN, 2008)	$T_c = T_a + \frac{G}{U_0 + U_1 V_w}$	U_0, U_1 : especificado em (KOEHL <i>et al.</i> , 2011) para células FV selecionadas
Skoplaki e Palyvos (SKOPLAKI; PALYVOS, 2009b)	$T_c = T_a + \frac{0,32}{8,91 + 2V_w} G$	Estima T_c para um módulo FV a-Si
Skoplaki <i>et al.</i> I (SKOPLAKI; PALYVOS, 2009b)	$T_c = T_a + \frac{0,25}{5,7 + 3,8V_w} G$	
Duffie e Beckman I (DUFFIE; BECKMAN, 2013)	$T_c = T_a + \left(\frac{G}{G_{NOCT}}\right) \left(\frac{9,5}{(5,7 + 3,8V_w)}\right) (T_{c,NOCT} - T_{a,NOCT}) \left[1 - \left(\frac{\eta}{\tau\alpha}\right)\right]$	onde $\tau\alpha \approx 0,9$
Skoplaki <i>et al.</i> II (SKOPLAKI <i>et al.</i> , 2008)	$T_c = T_a + \frac{G}{G_{NOCT}} (T_{c,NOCT} - T_{a,ref}) \left[\frac{h_{NOCT}}{h} \left\{1 - \frac{\eta_{STC}}{\tau\alpha} (1 + \beta_{STC} T_{STC})\right\}\right]$	o valor de $h = a + bV_w$ é detalhado em (SKOPLAKI <i>et al.</i> , 2008)

Chenni *et al.* (CHENNI *et al.*, 2007) $T_c = T_a + 0,0138G(1 + 0,031T_a)(1 - 0,042V_w)$

Kurtz *et al.* (KURTZ *et al.*, 2009) $T_c = T_a + Ge^{-3,473-0,0594V_w}$

Hornung *et al.*¹ (HORNUNG *et al.*, 2012) $T_c = T_a + m \left[e^{\left(\frac{-0,5V_w}{V_{w0}}\right)} + c \right] G$

Coskun (COSKUN *et al.*, 2017) $T_c = 1,4T_a + 0,01(G - 500) - V_w^{0,8}$

Sandia (KING *et al.*, 2004) $T_c = G(e^{a+b.V_w}) + T_a + \left(\frac{G}{G_{ref}}\right) \Delta T$

Duffie e Beckman II (DUFFIE; BECKMAN, 2013) $T_c = \frac{T_a + \frac{G}{G_{NOCT}}(T_{c,NOCT} - T_{a,ref}) \left\{ 1 - \frac{\eta_{STC}}{\tau\alpha} (1 + \beta_{STC} T_{STC}) \right\}}{1 - \frac{\beta_{STC} \eta_{STC}}{\tau\alpha} \frac{G}{G_{ref,NTE}} (T_{c,NOCT} - T_{a,ref})}$

Mattei (MATTEI *et al.*, 2006) $T_c = \frac{UT_a + G(\tau\alpha - \eta_{STC} - \beta_{STC} \eta_{STC} T_{STC})}{U - \beta_{STC} \eta_{STC} G}$

Estima T_c para um módulo FV Policristalino

Estima T_c para diferentes tecnologias FV

m , V_{w0} e c : coeficientes obtidos por regressão multilinear, detalhados em (HORNUNG *et al.*, 2012)

Estima T_c para um módulo FV Policristalino

$\Delta T = T_c - T_b$ no G_{ref} Adaptado para módulos *High Concentrating Photovoltaic* (HCPV) em (RODRIGO *et al.*, 2014)

$U = 26,6 + 2,3V_w$ or $U = 24,1 + 2,9V_w$, detalhado em (MATTEI *et al.*, 2006). $\tau\alpha = 0,81$

¹ Aplicado à HCPV.

Fonte: Adaptado de (SKOPLAKI *et al.*, 2008).

2.1 Principais componentes dos modelos

As propriedades térmicas dos materiais e as condições ambientais têm grande influência em T_c , pois parte da irradiação solar é convertida em calor (HERSCH; ZWEIBEL, 1982). T_c depende fortemente de G e T_a e é muito sensível a V_w (GRIFFITH *et al.*, 1981). Nesta seção, é feita uma análise dos componentes, vantagens e dificuldades técnicas dos modelos para o cálculo de T_c . Dentre os fatores que influenciam T_c foram destacados aqueles que exercem maior influência segundo a literatura, visto que aparecem em maior número de correlações. Considerando as 33 correlações analisadas, os componentes G ; T_a ; V_w ; η ; absorção solar; a transmitância da camada de vidro são as mais frequentes (Tabela 3), aparecendo em 100%, 93,9%, 54,5%, 33,3%, 30,3% e 24,2% das correlações, respectivamente. Esses componentes são agrupados de acordo com a seguinte classificação: variáveis meteorológicas e propriedades dependentes do material e/ou configuração do sistema.

2.1.1 Variáveis Meteorológicas

Geralmente, os modelos para determinação de T_c usam dados solares que dependem da localização, dados meteorológicos obtidos através de bases de dados reconhecidas ou modelos meteorológicos (ANON, 1995; METEOTEST, 2003). As estimativas de valores médios por hora são geralmente aplicadas em modelos térmicos para prever o T_c associado. Se não, os métodos baseados em parâmetros atmosféricos requerem uma campanha experimental ao ar livre para obter os valores G , T_a e V_w . Estimar dados e/ou médias gera incerteza relacionada com os valores tabulados e o modelo térmico. Quando utilizados instrumentos de medição, os modelos também não estão isentos de incertezas, pois, dependendo de como o arranjo experimental é elaborado, pode haver influências externas ou falhas na medição e na precisão do instrumento. Algumas dificuldades técnicas são discutidas, como um sistema que bloqueia a luz solar (MULLER *et al.*, 2011), a necessidade de um simulador solar (PEHARZ *et al.*, 2011), a necessidade de toda a curva I-V do módulo (YANDT *et al.*, 2014) e a necessidade de conhecimento avançado em algumas técnicas específicas.

2.1.1.1 Irradiância efetiva

A potência de saída da célula FV é diretamente proporcional a G incidente em sua superfície (MORA *et al.*, 2016). Portanto, medições ou estimativas de G em um local específico

Tabela 3 – Componentes das correlações analisadas (✓ significa que se trata de um componente da correlação, × caso contrário).

Correlação												Outros componen- tes (aparecem em cinco ou menos das correlações estudadas)
vs	G	T_a	V_w	η	α	τ	G_{NOCT}	$T_{c,NOCT}$	$T_{a,NOCT}$	β	U	
componentes												
Fernández <i>et al.</i>	✓	×	✓	×	×	×	×	×	×	×	×	V_{oc}, c_1, c_2, c_3
Durisch <i>et al.</i>	✓	✓	×	×	×	×	×	×	×	×	×	k
Nordmann	✓	✓	×	×	×	×	×	×	×	×	×	k
Krauter	✓	✓	×	×	×	×	×	×	×	×	×	k
Mondol I	✓	✓	×	×	×	×	×	×	×	×	×	
Hove	✓	✓	×	✓	✓	✓	×	×	×	×	✓	
Tiwari	✓	✓	×	✓	✓	✓	×	×	×	×	✓	
Eicker	✓	✓	×	✓	✓	×	×	×	×	×	✓	
Standard	✓	✓	×	×	×	×	✓	✓	✓	×	×	
Davis	✓	✓	×	✓	✓	✓	✓	✓	✓	×	×	
Mondol II	✓	✓	×	×	×	×	×	×	×	×	×	
Tselepis	✓	✓	×	×	×	×	×	×	×	×	×	
Tiwari I	✓	✓	×	✓	✓	✓	×	×	×	×	×	p, T_b, U_t, U_T
Tiwari II	✓	✓	×	✓	✓	✓	×	×	×	✓	×	p, T_b, U_t, U_T
Almonacid	✓	✓	✓	×	×	×	×	×	×	×	×	d_1, d_2
Markvart	✓	✓	✓	×	×	×	×	×	×	×	×	
Muzathik	✓	✓	✓	×	×	×	×	×	×	×	×	
Akyuz <i>et al.</i>	✓	✓	✓	×	×	×	×	×	×	×	×	
NOCT-1p	✓	✓	✓	×	×	×	✓	✓	✓	×	×	$a, V_w, NOCT$
NOCT-2p	✓	✓	✓	×	×	×	✓	✓	✓	×	×	$b, c, V_w, NOCT$
ISFOC	✓	×	×	✓	×	×	×	×	×	×	×	T_b, C_g, L_i, λ_i
Faiman	✓	✓	✓	×	×	×	×	×	×	×	×	U_0, U_1
Skoplaki and Palyvos	✓	✓	✓	×	×	×	×	×	×	×	×	
Skoplaki <i>et al.</i> I	✓	✓	✓	×	×	×	×	×	×	×	×	
Duffie e Beckman I	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	
Skoplaki <i>et al.</i> II	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	h, h_{NOCT}, T_{STC}
Chenni <i>et al.</i>	✓	✓	✓	×	×	×	×	×	×	×	×	
Kurtz <i>et al.</i>	✓	✓	✓	×	×	×	×	×	×	×	×	
Hornung <i>et al.</i>	✓	✓	✓	×	×	×	×	×	×	×	×	m, V_{w0}, c
Coskun	✓	✓	✓	×	×	×	×	×	×	×	×	
Sandia	✓	✓	✓	×	×	×	×	×	×	×	×	G_{ref}, e, a, b
Duffie e Beckman II	✓	✓	×	✓	✓	×	✓	✓	✓	✓	×	T_{STC}
Mattei	✓	✓	✓	✓	✓	✓	×	×	×	✓	✓	T_{STC}

Fonte: Elaborado pelo autor.

são essenciais para prever o desempenho e a eficiência dos sistemas FV (AL-ADDOUS *et al.*, 2017; MA *et al.*, 2016). G produz transferência de calor por radiação, que é o fator mais influente nas variações de T_c (BARDHI *et al.*, 2012). Para um aumento de 100 W/m^2 , aumentos entre $1,8 - 4,93^\circ\text{C}$ em T_c são encontrados na literatura, esses valores não são universais e dependem do tipo de módulo FV (RAHMAN *et al.*, 2015). Alguns dispositivos são projetados especificamente para medir G , como o piranômetro (SEAWARD, 2019; PANDEY; KATIYAR, 2013). Além

disso, ferramentas de baixo custo são encontradas na literatura para estimar G (AL-TAANI; ARABASI, 2018). Frequentemente, a fonte mais importante de erro na determinação da potência FV está relacionada com o procedimento e instrumento escolhido para quantificar G (KING *et al.*, 2004). Isso se deve às diferentes influências sistemáticas nos resultados do teste: os módulos FV respondem a apenas parte do espectro solar, enquanto os dispositivos podem responder a uma faixa diferente ou a todos os comprimentos de onda solares; a orientação e o ângulo de aceitação óptico do módulo e o ângulo de visão do sensor de G podem ser diferentes, a resposta do módulo e do sensor de G varia significativamente dependendo do ângulo de incidência (OTA *et al.*, 2019). Além disso, em caso de sombreamento ou poeira, o valor detectado pelo instrumento de medição e o valor de G na célula podem ser diferentes (CHAVES *et al.*, 2019; SINHA *et al.*, 2014; LIAO *et al.*, 2019).

O conceito de ' G efetiva' fornece um método para reduzir a dificuldade e a incerteza relacionadas com os testes de campo de plantas FV e para lidar com as influências sistemáticas mencionadas. A G efetiva é a G no plano do módulo ao qual as células FV realmente respondem, sob as influências das perdas ópticas e da variação espectral solar devido ao ângulo de incidência solar e à sujeira do módulo. Dependendo da precisão necessária e dos dados medidos disponíveis, diferentes abordagens podem ser usadas para determinar o G efetivo, conforme demonstrado em (AL-TAANI; ARABASI, 2018), onde os autores propõem uma configuração acessível auxiliada por smartphone para estimar a irradiação solar em uma determinada localização. Em Ota *et al.*, os autores medem G usando um arranjo de piranômetros móvel com cinco piranômetros (OTA *et al.*, 2019).

2.1.1.2 Velocidade do vento

A V_w torna-se uma variável ambiental relevante, favorecendo a perda de calor por convecção. Os resultados de uma análise de sensibilidade de um módulo FV policristalino 125G-2 são reportados em Tina e Abate (2008): uma queda em T_c de $3,3^\circ\text{C}$ é apresentada quando V_w aumenta de 1 para 2 m/s; esse valor não é universal e depende do tipo de módulo FV e materiais utilizados (TINA; ABATE, 2008). Monitorar e/ou estabelecer condições uniformes para medições relevantes do vento, especialmente em campo, não é uma tarefa simples. Portanto, muitas correlações usando coeficientes de transferência de calor relacionados a V_w podem ser encontradas na literatura científica nos últimos anos (PALYVOS, 2008).

A baixa disponibilidade de dados locais dificulta garantir resultados precisos para

avaliar a real eficiência de operação em situações onde T_c desempenha um papel fundamental (ROMARY *et al.*, 2011). Consequentemente, alguns dos simuladores de eficiência do sistema FV que baseiam T_c em variáveis ambientais geralmente desconsideram V_w . Uma opção para substituir as medições de campo são os dados locais de modelos numéricos de previsão do clima (SCHWINGSHACKL *et al.*, 2013; FINAMORE *et al.*, 2019).

A prática meteorológica padrão para registro de V_w e da direção do vento, localiza o dispositivo de medição, anemômetro, a uma altura de 10 metros em uma área com um número mínimo de estruturas ou edifícios que possam obstruir a movimentação do ar (KING *et al.*, 2004). No entanto, em alguns estudos, após a instalação do sistema, o modelo térmico pode ser "ajustado" determinando coeficientes que compensam as influências e instalações dependentes da localização de anemômetros que são diferentes da prática meteorológica padrão (SILVA *et al.*, 2017).

Os coeficientes determinados empiricamente também variam para diferentes tipos de módulos e configurações de montagem (ABIOLA-OGEDENGBE *et al.*, 2015). Em alguns casos, um coeficiente genérico é apresentado para módulos FV de placa plana típicos de fabricantes distintos. No entanto, o comportamento térmico dos módulos concentradores pode variar significativamente, dependendo do projeto do módulo (MULLER *et al.*, 2011; PEHARZ *et al.*, 2011). Portanto, os coeficientes dos módulos concentradores devem ser determinados empiricamente para cada projeto.

A direção do vento pode ter uma influência pequena, mas perceptível em T_c . Entretanto, a incorporação do efeito da direção do vento no modelo térmico é desnecessariamente complexa (SCHWINGSHACKL *et al.*, 2013). Portanto, na maioria dos estudos, a influência da direção do vento em T_c é desconsiderada ou considerada como uma influência aleatória, adicionando incerteza ao modelo térmico.

2.1.1.3 Temperatura ambiente

A T_a tem se mostrado uma variável importante nos modelos, influenciando η e T_c (PEREIRA *et al.*, 2019). Para um módulo FV policristalino, em média T_c aumenta linearmente 1°C com o aumento de 1°C de T_a mais uma diferença fixa relacionada a G (Equação 2.2), esse aumento depende do tipo de módulo FV (TINA; ABATE, 2008). Quase todos os modelos para determinação de T_c requerem T_a como entrada. Monitorar T_a não representa um grande desafio, considerando os aparelhos de alta precisão disponíveis no mercado, com boa faixa de operação

e resolução a um custo aceitável. Além disso, técnicas precisas de monitoramento T_a já são conhecidas e estabelecidas (ZIMMERMAN; HARTWICK, 1991). Globalmente, a literatura científica tem mostrado que a taxa de desempenho de módulos FV aumenta com a altitude devido à baixa temperatura. Regiões com grande altitude, como o sul dos Andes, a região do Himalaia e da Antártica, têm demonstrado os maiores potenciais FV (KAWAJIRI *et al.*, 2011).

2.1.2 Propriedades dependentes do material e/ou configuração do sistema

2.1.2.1 Eficiência elétrica FV

A eficiência (η) decai com o aumento de T_c (SINGH; RAVINDRA, 2012). Para módulos FV cristalinos sob uma irradiância de 1000 W/m^2 , η diminui em torno de 0,03-0,06% com aumento de 1°C em T_c (RAHMAN *et al.*, 2015). Portanto, η é um componente presente em um grande número de correlações e depende dos parâmetros elétricos da célula FV, como V_{oc} , FF , I_{sc} e G (SINGH; RAVINDRA, 2012). Normalmente o fabricante fornece os valores de η , V_{oc} , I_{sc} e FF em condições padrão, uma vez que esses componentes estão diretamente ligados à produção da célula/módulo FV. Porém, essas variáveis podem ser obtidas experimentalmente (SALAS *et al.*, 2006), cada uma delas fornecendo informações sobre a física do dispositivo (DUPRÉ *et al.*, 2017).

2.1.2.2 Absorção solar e transmitância da camada de vidro

A energia solar incidente em um corpo pode ser absorvida, refletida e/ou transmitida através do material. Esta característica pode ser observada em todos os materiais semitransparentes, representados pelos conhecidos coeficientes de transmissividade (τ), refletividade e absorvância (α) (INCROPERA *et al.*, 2000). Esses coeficientes são dependentes de G e influenciam a taxa de irradiância convertida pela célula FV (CUCE *et al.*, 2017). Normalmente, as células são fabricadas para maximizar a absorção dos comprimentos de onda (OSPINA, 2016). A irradiância absorvida pela camada de vidro é representada pela Eq. (2.4).

$$Q_g = G\alpha_g = G(1 - \tau_a) \quad (2.4)$$

τ_a é a transmitância do vidro considerando a perda apenas por absorção (CHOW, 2003).

α aparece em correlações de T_c tanto de forma independente quanto por meio do

produto $\tau\alpha$ (SKOPLAKI; PALYVOS, 2009a). Modelos matemáticos relacionando parâmetros de fabricação à eficiência elétrica e térmica dos módulos FV demonstram ganhos ao se considerar o produto $\tau\alpha$ (OSPINA, 2016). A influência das perdas ópticas (refletância) para módulos de placa plana é geralmente insignificante para ângulos de incidência G abaixo de 55° . Essas perdas ópticas para uma superfície de módulo não orientada perpendicularmente ao caminho da luz solar são adicionais à perda típica do 'cosseno'. O efeito cumulativo (perda) ao longo do ano deve ser considerado no projeto do sistema. Para módulos que rastreiam o Sol com precisão, não há perda óptica (KING *et al.*, 2004).

2.1.2.3 Outros fatores

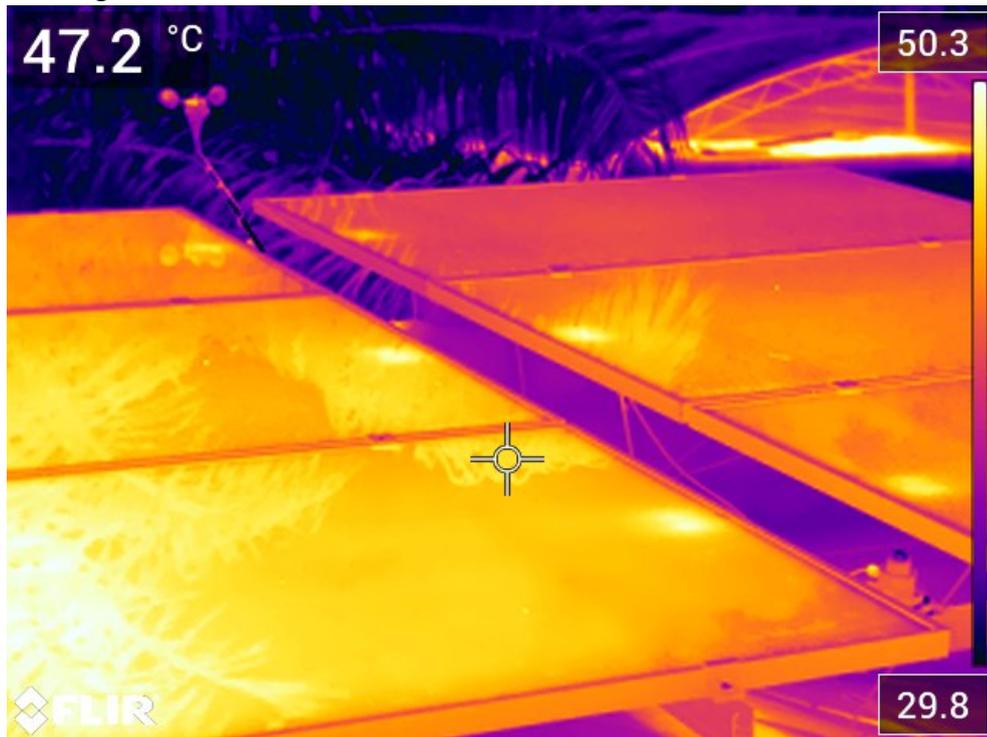
Vários coeficientes aparecem na literatura relacionados com T_c , influenciados pela construção do módulo, a configuração do conjunto e a localização e altura em que V_w é medido. Existem também os chamados fatores que "introduzem influências aleatórias" em T_c , como a capacitância térmica do módulo e os transientes térmicos causados por nuvens, sombreamento, umidade e poeira (CHAVES *et al.*, 2019; ARAÚJO *et al.*, 2019). Esses efeitos aleatórios são calculados diariamente ou anualmente.

2.2 Validação do modelo: determinação de T_c

As células FV operam em temperaturas ligeiramente diferentes em um módulo (FAIMAN, 2008). Para registrar tal comportamento, a termografia permite um monitoramento confiável da temperatura de plantas FV (GALLARDO-SAAVEDRA *et al.*, 2018), conforme pode ser visto na Fig. 2. Por causa dessa diferença de temperatura, os pesquisadores mencionam a "temperatura média da célula", ou apenas "temperatura da célula", T_c (RODRIGO *et al.*, 2014). A célula FV é cercada por materiais encapsulantes; portanto, uma medição direta de T_c , valor exato, não é possível em geral. O conceito de T_c não é claramente definido na literatura científica de modo geral, dependendo da metodologia e técnica utilizadas em cada trabalho.

A medição real de T_c é feita dentro da estrutura do módulo FV, entre a célula FV e o revestimento de *Ethylene vinyl acetate* (EVA) (NISHIOKA *et al.*, 2018) (Fig. 3). T_c pode ser obtido empiricamente, um trabalho não trivial, ou estimado através de uma relação matemática com T_b , com base na condução térmica unidimensional através dos materiais do módulo. T_c é então calculado usando T_b e uma diferença entre a temperatura da célula e da superfície posterior.

Figura 2 – Imagem térmica infravermelha de módulos FV instalados no LEA.



Fonte: Elaborado pelo autor

Figura 3 – Representação da célula FV mostrando a posição ideal da medição de T_c .



Fonte: Elaborado pelo autor

Na maioria dos casos, os autores consideram a temperatura do módulo como a temperatura de uma célula representativa ou o valor médio de várias células. Esta temperatura pode ser medida por um sensor acoplado na superfície traseira do módulo, supondo que o valor obtido seja a temperatura do arranjo. Medir a temperatura por um sensor fixado na parte traseira do módulo FV é uma prática comum para estimar T_c (MORA *et al.*, 2015). Mas geralmente, T_c é estimada conectando vários sensores de temperatura (termopares) fixados na parte de trás

do módulo; os sensores detectam a temperatura da camada inferior do módulo FV. T_c pode ser superior a T_b em alguns graus, esta diferença depende dos materiais do substrato do módulo e de G . Para módulos planos em um arranjo aberto, T_c é menos sensível a V_w do que T_b , pois a célula está dentro da estrutura do módulo, enquanto a superfície traseira está diretamente exposta ao vento (NISHIOKA *et al.*, 2018). Para módulos planos nos quais a superfície traseira é isolada termicamente, pode-se presumir que a diferença de temperatura entre T_c e T_b é nula.

A influência de V_w , direção do vento, componentes de suporte estrutural, caixas de junção e estruturas de módulo podem causar distribuições de temperatura não uniformes em toda a superfície do módulo. Normalmente, essas diferenças espaciais de temperatura variam em cerca de 5°C. As células centrais no módulo têm temperaturas mais altas do que aquelas mais próximas das bordas do módulo (TINA; ABATE, 2008). Portanto, o posicionamento cuidadoso dos termopares ajuda a obter um valor preciso para a temperatura média do módulo. A variação espacial da temperatura no módulo pode ser compensada pela média de várias medidas fornecidas pelos sensores de temperatura, possibilitando determinar uma boa média para T_b .

Correlações baseadas em parâmetros elétricos consideram que V_{oc} varia com a temperatura. Modelos Isc-Voc podem ser utilizados para extrair T_c da curva I-V e resultam em medições úteis para validação de T_c . Esses modelos não são usados para "modelar" T_c a partir de medições climáticas, e sim expressar T_c em função de parâmetros elétricos do módulo/célula FV, e podem ser classificados com uma classe distinta de modelos não-lineares. Não é possível comparar a forma funcional dos modelos Isc-Voc com os modelos da Tabela 2, visto que os parâmetros são distintos. Alguns modelos Isc-Voc estão representados na Tabela 4.

Tabela 4 – Modelos Isc-Voc para determinação de T_c

Correlação	Formulação	Comentários
Muller ¹ (MULLER <i>et al.</i> , 2011)	$T_c = T_{h,0} + \frac{(V_{\max} - V_{oc(t)})}{(N_s \beta)}$	$\beta = -0.0045$ (V/C/cell). N_s : Número de células em série no módulo. $T_{h,0}$: Temperatura do dissipador de calor em condição especificada em (MULLER <i>et al.</i> , 2011)
Voc-Isc ¹ (KING <i>et al.</i> , 2004; JU <i>et al.</i> , 2013)	$T_c = \frac{V_{oc} - V_{ocr} + \beta(C_g)T_r}{\frac{nk}{q} \ln \frac{I_{sc}}{I_{scr}} + \beta(C_g)}$	$T_r = 298.15$ Temperatura de referência em Kelvin
Voc-Isc melhorado (KING <i>et al.</i> , 2004; JU <i>et al.</i> , 2013)	$T_c = \frac{V_{oc} - V_{ocr} - \frac{nkT_r}{q} \ln(C_g)}{\frac{nk}{q} \ln(C_g) + \beta} + T_r$	Similar ao modelo apresentado por (KING <i>et al.</i> , 2004) para $C_g = 1$
IEC 60904-5 (IEC, 2011)	$T_c = \frac{\beta T_{c,ref} + V_{oc} - V_{ocr} + \frac{nk}{q} \ln\left(\frac{G_{ref}}{G}\right) 273 N_s}{\beta - \frac{nk}{q} \ln\left(\frac{G_{ref}}{G}\right) N_s}$	$T_{c,ref}$ é T_c nas condições STC

¹ Aplicado à HCPV.

T_c também pode ser estimado com base em parâmetros ambientais. No entanto, esta estimativa agrega incertezas, que afetam a precisão do modelo de desempenho (KING *et al.*, 2004). Outros métodos requerem uma campanha experimental ao ar livre com um módulo com um sensor interno de temperatura da célula. Este tipo de dispositivo nem sempre está disponível e o procedimento leva tempo para realizar as medições. Estudos comparando a temperatura medida com um sensor interno inserido logo abaixo da célula com a temperatura obtida usando modelos baseados no cálculo do fluxo de calor usando T_b , mostram uma precisão de $\pm 1^\circ\text{C}$ (NISHIOKA *et al.*, 2018). Como resultado, a aplicação de modelos que associam T_b a T_c tem se mostrado atrativa uma vez que T_b pode ser facilmente obtida, enquanto a medição de temperatura usando um módulo FV com um sensor interno, embora precisa, é difícil e dispendiosa (RODRIGO *et al.*, 2014). Deve-se considerar que para um projeto confiável e uma estimativa de desempenho de usinas FV, modelos precisos e fáceis de implementar são necessários (CUCE *et al.*, 2017).

3 TÉCNICAS DE APRENDIZADO DE MÁQUINA

São conhecidas duas definições de ML, Arthur Samuel o descreveu como: "o campo de estudo que dá aos computadores a capacidade de aprender sem serem explicitamente programados"(WIEDERHOLD; MCCARTHY, 1992). Esta é uma definição mais antiga e informal. Tom Mitchell posteriormente fornece outra definição: "Diz-se que um programa de computador aprende com a experiência 'E' com relação a alguma classe de tarefas 'T' e medida de desempenho 'P', se seu desempenho nas tarefas em 'T', conforme medido por 'P', melhora com a experiência 'E'"(MITCHELL *et al.*, 1997).

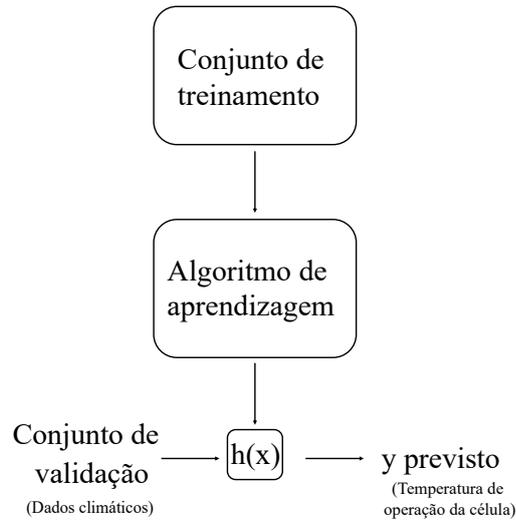
Em geral, qualquer problema de ML pode ser atribuído a uma de duas classificações gerais: aprendizagem supervisionada e aprendizagem não supervisionada. Na aprendizagem supervisionada, é utilizado um conjunto de dados e já se sabe como deve ser a resposta, neste caso T_c , tendo a ideia de que existe uma relação entre o *input* e o *output*. O aprendizado não supervisionado nos permite abordar problemas com pouca ou nenhuma ideia de como devem ser os resultados. Há também problemas de aprendizagem híbrida, ou seja, nestes casos os problemas não são unicamente classificados como supervisionados ou não supervisionados. Os problemas de aprendizagem híbrida podem ser classificados como aprendizagem semi-supervisionada (Semi-Supervised Learning), auto-supervisionada (Self-Supervised Learning) e aprendizagem em várias instâncias (Multi-Instance Learning).

Os problemas de aprendizagem supervisionada são categorizados em problemas de "regressão" e "classificação". Nos problemas de regressão, procura-se prever os resultados em uma saída contínua, o que significa uma tentativa de mapear os preditores de entrada para alguma função contínua. Em um problema de classificação, procura-se prever os resultados em uma saída discreta. Em um problema de classificação, o *output* é classificado em categorias distintas com base nos *inputs*, por exemplo, dado um paciente com tumor, busca-se que prever se o tumor é maligno ou benigno.

Para descrever o problema de aprendizagem supervisionada um pouco mais formalmente, procura-se, dado um conjunto de treinamento, aprender uma função $h : X \rightarrow Y$ de forma que $h(x)$ seja "bom" em estimar o valor correspondente de y . Por razões históricas, essa função h é chamada de hipótese. Este processo é representado na Figura 4.

Pode-se mensurar a precisão da função de hipótese usando uma "função de custo". Isso leva a uma diferença média de todos os resultados da hipótese com entradas de x e a resposta real y , ou a diferença entre o valor previsto e o valor real. Comumente, essa função é também

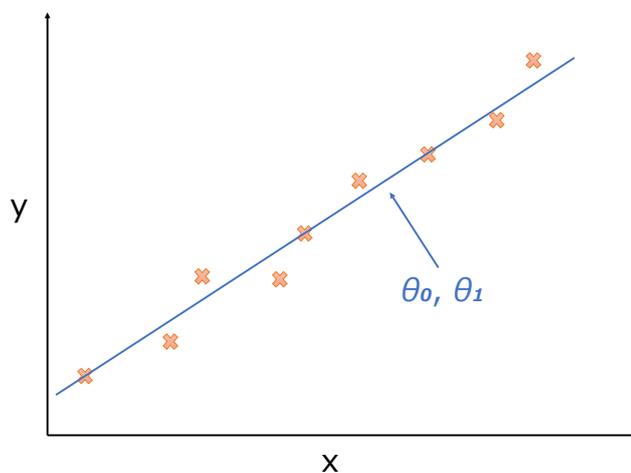
Figura 4 – Estrutura de um problema de aprendizagem de máquina supervisionada.



Fonte: Elaborado pelo autor.

chamada de "Função de erro quadrático" ou "Erro quadrático médio", do inglês *root-mean-square error* (RMSE). Em resumo, o que a função de custo faz é escolher coeficientes θ_i de maneira que a função hipótese, $h_{\theta}(x)$, seja próximo do valor real y , para os valores x dos exemplos de treinamento (Figura 5). No caso dos modelos lineares o objetivo da função custo é obter a melhor reta possível. A melhor reta possível será tal que as distâncias verticais quadradas médias dos pontos dos valores reais ("x" vermelho na Figura 5) espalhados e da reta de previsão da função hipótese (reta azul na Figura 5) serão as menores possíveis. Assim, procura-se minimizar a função de custo.

Figura 5 – Representação da escolha de coeficientes pela função de custo.



Fonte: Elaborado pelo autor.

3.1 Regressão Linear (RL)

A RL é uma técnica de modelagem estatística utilizada para descrever uma variável de resposta contínua em função de uma ou mais variáveis preditoras. Pode ajudar na compreensão e previsão do comportamento de sistemas complexos ou na análise de dados experimentais, financeiros e biológicos. Técnicas de RL são usadas para criar um modelo linear. O modelo descreve a relação entre uma variável resposta y como uma função de um ou mais preditores x_i . A equação geral para um modelo de regressão linear é dada pela Eq. 3.1.

$$y = \beta_0 + \sum \beta_i x_i + \varepsilon_i, \quad (3.1)$$

onde β representa estimativas de parâmetros lineares a serem calculados e ε representa os termos de erro. O intercepto ou coeficiente linear β_0 representa o ponto em que a reta da regressão corta o eixo Y quando $x = 0$. Já os coeficientes β_i são os chamados coeficientes de regressão ou coeficiente angular (MATHWORKS, 2021e).

As três maneiras comuns de ajustar um modelo RL aos dados são: ajuste de mínimos quadrados; ajuste robusto; e *stepwise*. O primeiro constrói um ajuste de mínimos quadrados de um modelo aos dados, este método é melhor quando se está razoavelmente certo da forma do modelo e principalmente precisa encontrar seus parâmetros. Também é útil quando se deseja explorar alguns modelos. No entanto, o método requer um exame manual dos dados para descartar valores discrepantes, embora existam técnicas para ajudar. No ajuste robusto é possível criar um modelo que é pouco afetado por *outliers*. O que evita o trabalho de descartar manualmente os valores discrepantes. O *stepwise* encontra um modelo e ajusta os parâmetros ao mesmo. O *stepwise* começa a partir de um modelo, como uma constante, e adiciona ou subtrai os termos um de cada vez, até que não possa melhorar mais (MATHWORKS, 2021c).

A técnica de regressão também pode ser aplicada de forma não linear, sendo assim uma técnica estatística que ajuda a descrever relações não lineares em dados experimentais. Os modelos de regressão não linear são geralmente considerados paramétricos, onde o modelo é descrito como uma equação não linear. A regressão não linear paramétrica modela a variável resposta em função de uma combinação de parâmetros não lineares e um ou mais preditores. Os preditores podem assumir a forma de uma função exponencial, trigonométrica, de potência ou qualquer outra função não linear. Para determinar as estimativas de parâmetros não lineares, normalmente é usado um algoritmo iterativo (Eq. 3.2).

$$y = f(x, \beta) + \varepsilon, \quad (3.2)$$

onde, β representa estimativas de parâmetros não lineares a serem calculados e ε representa os termos de erro.

Os algoritmos populares para ajustar uma regressão não linear incluem: Algoritmo de Gauss-Newton; Algoritmo do gradiente descendente; e Algoritmo de Levenberg-Marquardt. A Tabela 5 mostra alguns tipos de modelos de RL simples e não linear, dos mais simples aos mais complexos.

Tabela 5 – Tipos de modelos de RL e suas definições

Nome	Tipo de Modelo
Constante	Contém apenas um termo constante (interceptação)
Linear	Contém uma interceptação e termos lineares para cada preditor.
Interações	Contém uma interceptação, termos lineares e todos os produtos de pares de preditores distintos (sem termos quadrados).
Puramente quadrático	Contém uma interceptação, termos lineares e termos quadrados.
Quadrático	Contém uma interceptação, termos lineares, interações e termos quadrados.
Polinomial	É um polinômio com todos os termos até grau i no primeiro preditor, grau j no segundo preditor, etc.

Fonte: Adaptado de (MATHWORKS, 2021c).

3.2 Support Vector Machine (SVM)

A análise SVM é uma ferramenta popular de ML para classificação e regressão, identificada pela primeira vez por Vladimir Vapnik e seus colegas em 1992 (Vapnik, V. T., 1995). Esta se trata de uma técnica não paramétrica porque depende de funções de *kernel*, recursos do SVM úteis por proverem uma ponte entre algoritmos lineares e não lineares, melhor definidas posteriormente nesta seção. Na técnica SVM o conjunto de dados de treinamento inclui variáveis preditoras e valores de resposta observados. O objetivo é encontrar uma função $f(x)$ que se desvia de y_i por um valor não maior que ε para cada ponto de treinamento x , e ao mesmo tempo é o mais plano possível.

Dado um conjunto de dados onde x_n é o vetor do espaço de entrada da amostra de dados, y_n é o valor de resposta e N é o número de pontos. Para encontrar a função linear representada pela Eq. 3.3:

$$f(x) = x'\beta + b, \quad (3.3)$$

e certificar-se de que é a mais plana possível, deve-se encontrar $f(x)$ com o valor mínimo da

norma $(\beta'\beta)$. Isso é formulado como um problema de otimização convexa para minimizar (Eq. 3.4):

$$J(\beta) = \frac{1}{2}(\beta'\beta) \quad (3.4)$$

sujeito a todos os resíduos com valor inferior a ε ; ou, na forma de Eq. 3.5 (Vapnik, V. T., 1995):

$$\forall n : |y_n - (x_n'\beta + b)| \leq \varepsilon \quad (3.5)$$

É possível que tal função $f(x)$ não exista para satisfazer essas restrições para todos os pontos. Para lidar com restrições inviáveis de outra forma, introduz-se as variáveis de folga ξ_n e ξ_n^* para cada ponto. Essa abordagem é semelhante ao conceito de “margem suave” na classificação SVM, porque as variáveis de folga permitem que erros de regressão existam até o valor de ξ_n e ξ_n^* , mas ainda assim satisfazem as condições requeridas.

Incluir variáveis de folga leva à função objetivo, também conhecida como fórmula primária, representada pela Eq. 3.6:

$$J(\beta) = \frac{1}{2}\beta'\beta + C \sum_{n=1}^N (\xi_n + \xi_n^*) \quad (3.6)$$

sujeito as restrições nas Eq. 3.7, 3.8, 3.9 e 3.10:

$$\forall n : y_n - (x_n'\beta + b) \leq \varepsilon + \xi_n \quad (3.7)$$

$$\forall n : (x_n'\beta + b) - y_n \leq \varepsilon + \xi_n^* \quad (3.8)$$

$$\forall n : \xi_n^* \geq 0 \quad (3.9)$$

$$\forall n : \xi_n \geq 0. \quad (3.10)$$

A constante C é a restrição de caixa, um valor numérico positivo que controla a penalidade imposta às observações que estão fora da margem épsilon (ε) e ajuda a prevenir *overfitting* (sobreajuste). Este valor determina a compensação entre a planura de $f(x)$ e a quantidade até a qual desvios maiores que ε são tolerados (Vapnik, V. T., 1995).

A função de perda linear ε -insensível ignora os erros que estão dentro da distância ε do valor observado, tratando-os como iguais a zero. A perda é medida com base na distância entre o valor observado y e o limite ε . Isso é formalmente descrito pela Eq. 3.11:

$$L_\varepsilon = \begin{cases} 0, & \text{se } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon, & \text{caso contrário.} \end{cases} \quad (3.11)$$

O problema de otimização descrito anteriormente é computacionalmente mais simples de resolver em sua formulação dual de Lagrange. A solução para o problema dual fornece um limite inferior para a solução do problema primordial (minimização). Os valores ideais dos problemas primários e duais não precisam ser iguais, e a diferença é chamada de "lacuna de dualidade". Mas quando o problema é convexo e satisfaz uma condição de qualificação de restrição, o valor da solução ótima para o problema primal é dado pela solução do problema dual (MATHWORKS, 2021d).

Para obter a fórmula dual, constrói-se uma função Lagrangiana a partir da função primária introduzindo multiplicadores não negativos α_n e α_n^* para cada observação x_n . Isso leva à fórmula dupla, onde minimiza-se a Eq. 3.12:

$$L(\alpha) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) x'_i x_j + \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i^* - \alpha_i), \quad (3.12)$$

sujeito às restrições nas Eq. 3.13, 3.14 e 3.15:

$$\sum_{n=1}^N (\alpha_n - \alpha_n^*) = 0 \quad (3.13)$$

$$\forall n : 0 \leq a_n \leq C \quad (3.14)$$

$$\forall n : 0 \leq a_n^* \leq C. \quad (3.15)$$

O parâmetro β pode ser completamente descrito como uma combinação linear das observações de treinamento usando a Eq. 3.16

$$\beta = \sum_{n=1}^N (\alpha_n - \alpha_n^*) x_n \quad (3.16)$$

A função usada para prever novos valores depende apenas dos vetores de suporte, Eq. 3.17:

$$f(x) = \sum_{n=1}^N (\alpha_n - \alpha_n^*) (x'_n x) + b \quad (3.17)$$

As condições de complementaridade de Karush-Kuhn-Tucker (KKT) são restrições de otimização necessárias para obter soluções ótimas. Para regressão SVM linear, essas condições são expressas nas Eq. 3.18, 3.19, 3.20 e 3.21:

$$\forall n : a_n(\varepsilon + \xi_n - y_n + x'_n \beta + b) = 0 \quad (3.18)$$

$$\forall n : a_n^*(\varepsilon + \xi_n^* + y_n - x'_n \beta - b) = 0 \quad (3.19)$$

$$\forall n : \xi_n(C - a_n) = 0 \quad (3.20)$$

$$\forall n : \xi_n^*(C - a_n^*) = 0 \quad (3.21)$$

Essas condições indicam que todas as observações estritamente dentro do tubo ε têm multiplicadores de Lagrange $\alpha_n = 0$ e $\alpha_n^* = 0$. Se α_n ou α_n^* não for zero, então a observação correspondente é chamada de vetor de suporte.

Alguns problemas de regressão não podem ser descritos adequadamente usando um modelo linear. Nesse caso, a formulação dual de Lagrange permite que a técnica descrita anteriormente seja estendida para funções não lineares.

Pode-se obter um modelo de regressão SVM não linear substituindo o produto escalar $x'_1 x_2$ por uma função kernel não linear $G(x_1, x_2) = \langle \varphi(x_1), \varphi(x_2) \rangle$, onde $\varphi(x)$ é uma transformação que mapeia x para um espaço de alta dimensão. A tabela 6 fornece algumas funções de kernel semidefinidas (MATHWORKS, 2021d).

Tabela 6 – Exemplos de funções de kernel semidefinidas

Nome da Kernel	Função Kernel
Linear	$G(x_j, x_k) = x'_j x_k$
Gaussiana	$G(x_j, x_k) = \exp(- x_j - x_k ^2)$
Polinomial	$G(x_j, x_k) = (1 + x'_j x_k)^q$ onde q está no conjunto $\{2, 3, \dots\}$.

Fonte: Adaptado de (MATHWORKS, 2021d).

A matriz de G_{ram} é uma matriz quadrada de ordem n que contém os elementos $g_{i,j} = G(x_i, x_j)$. Cada elemento $g_{i,j}$ é igual ao produto interno dos preditores transformado por φ . No entanto, não é preciso saber φ , porque pode-se utilizar a função kernel para gerar a matriz de G_{ram} diretamente. Usando este método, o SVM não linear encontra a função ótima $f(x)$ no espaço do preditor transformado.

3.3 Gaussian Process Regression (GPR)

Os modelos GPR são modelos probabilísticos não paramétricos baseados em kernel. Dado um conjunto de treinamento $\{(x_i, y_i); i = 1, 2, \dots, n\}$, onde $x_i \in \mathbb{R}^d$ e $y_i \in \mathbb{R}$, retirado de uma distribuição desconhecida. Um modelo GPR aborda a questão de prever o valor de uma

variável de resposta y_{novo} , dado o novo vetor de entrada x_{novo} e os dados de treinamento. Um modelo de regressão linear é da forma da Eq. 3.2. Onde $\varepsilon \sim N(0, \sigma^2)$. A variância do erro σ^2 e os coeficientes β são estimados a partir dos dados. Um modelo GPR explica a resposta introduzindo variáveis latentes, $f(x_i), i = 1, 2, \dots, n$, a partir de um Processo Gaussiano (PG) e funções de base explícitas, h . A função de covariância das variáveis latentes captura a suavidade da resposta e as funções de base projetam as entradas x em um espaço de recursos p -dimensional. Um PG é um conjunto de variáveis aleatórias, de modo que qualquer número finito delas tenha uma distribuição gaussiana conjunta. Se $\{f(x), x \in \mathbb{R}^d\}$ é um PG, então dado n observações x_1, x_2, \dots, x_n , a distribuição conjunta das variáveis aleatórias $f(x_1), f(x_2), \dots, f(x_n)$ é gaussiano. Um PG é definido por sua função média $m(x)$ e sua função de covariância, $k(x, x')$. Ou seja, se $\{f(x), x \in \mathbb{R}^d\}$ é um PG, então $E(f(x)) = m(x)$ e $Cov[f(x), f(x')] = E[\{f(x) - m(x)\}\{f(x') - m(x')\}] = k(x, x')$ (MATHWORKS, 2021b).

Seja o modelo apresentado na Eq. 3.22.

$$h(x)^T \beta + f(x), \quad (3.22)$$

Onde $f(x) \sim GP(0, k(x, x'))$, isto é, $f(x)$ são de um PG de média zero com função de covariância, $k(x, x')$. $h(x)$ são um conjunto de funções básicas que transformam o vetor de características original x em \mathbb{R}^d em um novo vetor de características $h(x)$ em \mathbb{R}^p . β é um vetor p -por-1 dos coeficientes da função de base. Este modelo representa um modelo GPR. Uma instância de resposta y pode ser modelada pela Eq. 3.23

$$P(y_i | f(x_i), x_i) \sim N(y_i | h(x_i)^T \beta + f(x_i), \sigma^2) \quad (3.23)$$

Portanto, um modelo GPR é um modelo probabilístico. Há uma variável latente $f(x_i)$ introduzida para cada observação x_i , o que torna o modelo GPR não paramétrico. Na forma vetorial, este modelo é equivalente a Eq. 3.24

$$P(y | f, X) \sim N(y | H\beta + f, \sigma^2 I), \quad (3.24)$$

onde,

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, H = \begin{pmatrix} h(x_1^T) \\ h(x_2^T) \\ \vdots \\ h(x_n^T) \end{pmatrix}, f = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}.$$

A distribuição conjunta de variáveis latentes $f(x_1), f(x_2), \dots, f(x_n)$ no modelo GPR é dado pela Eq. 3.25:

$$P(f|X) \sim N(f|0, K(X, X)), \quad (3.25)$$

perto de um modelo de regressão linear, onde $K(X, X)$ parece com Eq. 3.26:

$$K(X, X) = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{pmatrix} \quad (3.26)$$

A função de covariância $k(x, x')$ geralmente é parametrizada por um conjunto de parâmetros de kernel ou hiperparâmetros, θ . Muitas vezes $k(x, x')$ é escrito como $k(x, x'|\theta)$ para indicar explicitamente a dependência de θ (MATHWORKS, 2021b).

3.4 Redes Neurais (RN)

As RN são um modelo matemático que realiza uma simulação computacional do comportamento do neurônio no cérebro humano, replicando o padrão do cérebro para produzir resultados com base na aprendizagem de um conjunto de dados de treinamento (IZGI *et al.*, 2012). Modelos de RN normalmente têm boa precisão preditiva; no entanto, não são os mais fáceis de interpretar. A rede *feedforward* multicamadas com um algoritmo de aprendizado de retropropagação é uma das arquiteturas de RN mais populares.

Normalmente, uma RN de regressão consiste em três ou mais camadas totalmente conectadas: (1) uma camada de entrada; (2) uma camada de saída; e (3) uma ou mais camadas intermediárias ou ocultas. Cada camada subsequente à primeira tem uma conexão da camada anterior. Cada camada totalmente conectada multiplica a entrada por uma matriz de ponderação e adiciona um vetor polarizador. Uma função de ativação segue cada camada totalmente conectada, excluindo a última. A camada final totalmente conectada produz a saída da rede, ou seja, os valores de resposta previstos. A flexibilidade do modelo aumenta com o tamanho e o número de camadas totalmente conectadas na RN (MATHWORKS, 2021a).

Os vetores de entrada são $\in R^n$ e $D = (X_1, X_2, \dots, X_n)^T$; as saídas dos neurônios q na camada oculta são $Z = (Z_1, Z_2, \dots, Z_n)^T$; e as saídas da camada de saída são $Y \in R^m$, $Y = (Y_1, Y_2, \dots, Y_n)^T$. Supondo que o peso e o limite entre a camada de entrada e a camada oculta

são w_{ij} e y_j , respectivamente, e que o peso e o limiar entre a camada oculta e a camada de saída são w_{jk} e y_k respectivamente, as saídas de cada neurônio em uma camada oculta e a camada de saída são;

$$Z_j = f \left(\sum_{i=1}^n w_{ij} X_i - \theta_j \right) \quad (3.27)$$

$$Y_k = f \left(\sum_{j=1}^q w_{kj} Z_j - \theta_k \right) \quad (3.28)$$

onde $f()$ é uma função de transferência, que é a regra para mapear a entrada somada do neurônio em sua saída e, por uma escolha adequada, é um meio de introduzir uma não linearidade no projeto da rede. Uma das funções mais comumente usadas é a função sigmóide, que é monotônica crescente e varia de 0 a 1 (OLATOMIWA *et al.*, 2015). A tabela 7 fornece algumas funções de RN semidefinidas, o nível de dificuldade de interpretação dos resultados e a flexibilidade destes modelos (MATHWORKS, 2021d).

Tabela 7 – Exemplos de funções de RN semidefinidas

Tipo de modelo de regressão	Definição	Interpretabilidade	Flexibilidade do modelo
RN Estreita	regressão RN com uma camada totalmente conectada de tamanho 10, excluindo a camada final totalmente conectada para regressão	Difícil	Médio - aumenta com a configuração de tamanho da primeira camada
RN Média	regressão RN com uma camada totalmente conectada de tamanho 25, excluindo a camada final totalmente conectada para regressão	Difícil	Médio - aumenta com a configuração de tamanho da primeira camada
Ampla RN	regressão RN com uma camada totalmente conectada de tamanho 100, excluindo a camada final totalmente conectada para regressão	Difícil	Médio - aumenta com a configuração de tamanho da primeira camada
RN de Duas Camadas	regressão RN com duas camadas totalmente conectadas, excluindo a camada final totalmente conectada para regressão	Difícil	Alto - aumenta com as configurações de tamanho da primeira camada e da segunda camada
RN de Três Camadas	regressão RN com três camadas totalmente conectadas, excluindo a camada final totalmente conectada para regressão	Difícil	Alto - aumenta com as configurações de tamanho de primeira camada, tamanho da segunda camada, e tamanho da terceira camada

Fonte: Adaptado de (MATHWORKS, 2021a).

4 METODOLOGIA

A presente dissertação foi trabalhada em duas etapas: a primeira, se trata de um estudo dos modelos para determinação de T_c encontrados na literatura a partir do ano 2000; na segunda, é feito processamento de dados, a fim de aplicar e comparar técnicas de ML buscando desenvolver um modelo explícito para determinação de T_c . T_c influencia a eficiência do módulo FV, o desenvolvimento desse modelo pode trazer contribuições para a projeção de geração de energia elétrica por meio da energia solar e em estudos para a previsão de séries temporais. Neste capítulo é apresentada a metodologia para a obtenção e tratamento dos dados de G , V_w , T_a e T_c , e as estruturas das técnicas de aprendizagem utilizadas.

4.1 Etapa 1: Estudo dos modelos presentes na literatura para determinação de T_c

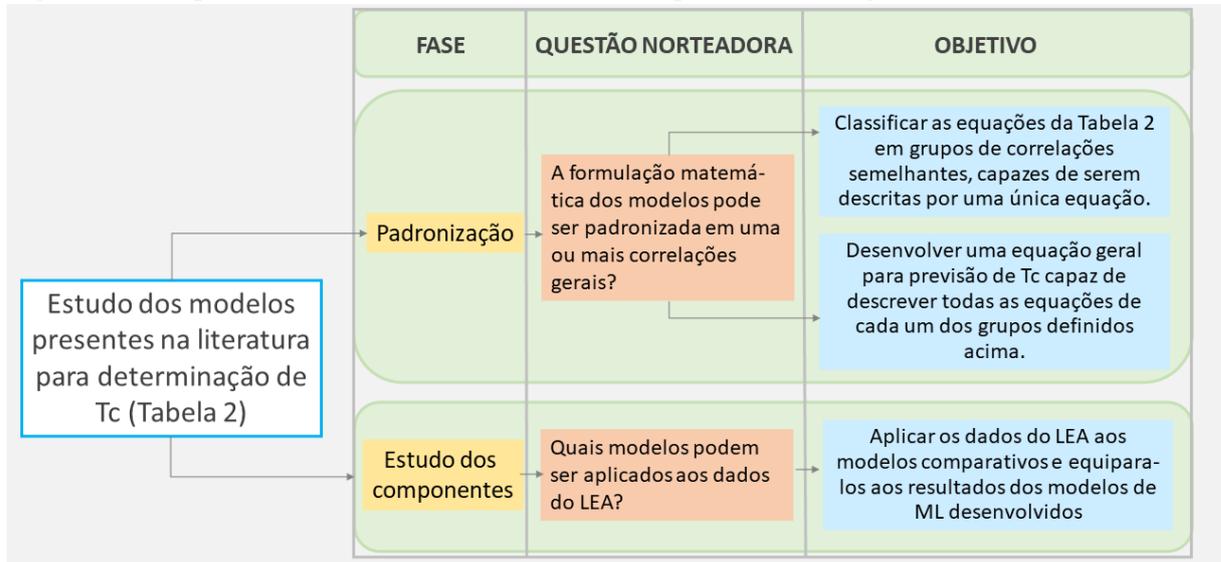
É realizado na etapa 1 desta dissertação, o estudo dos modelos para determinação de T_c encontrados na literatura a partir do ano 2000, ou desenvolvidos anteriormente a essa data mas que continuam sendo aplicados em trabalhos recentes. Os modelos, apresentados na Tabela 2, são estudados em busca de: padronizar as equações utilizadas para previsão de T_c ; e selecionar modelos com componentes presentes no banco de dados do LEA-UFC a fim usá-los para comparação com os modelos desenvolvidos na parte 2 desta dissertação. O estudo foi feito na seguinte ordem:

1. Busca de padrões no equacionamento dos modelos de previsão de T_c e divisão dos mesmos em grupos que seguem um mesmo padrão.
2. Divisão dos modelos da Tabela 2 em três grupos: modelos lineares e dois grupos de não lineares.
3. Desenvolvimento de três correlações, cada uma capaz de descrever os modelos contidos em um dos três grupos de mesmo padrão definidos no item anterior.
4. Separação dos modelos presentes na tabela 2 que utilizem os mesmos preditores dos modelos de ML desenvolvidos na presente dissertação, seção 4.2. Excluindo-se os seguintes modelos: aqueles desenvolvidos utilizando um arranjo experimental distinto do utilizado na presente dissertação, como os modelos para módulos HCPV, ou de sistemas FV integrados. Os modelos selecionados são referidos posteriormente na presente dissertação como "modelos comparativos".

A Figura 6 resume o passo-a-passo acima e facilita a visualização dos estudos dos

modelos.

Figura 6 – Mapa conceitual do estudo dos modelos para determinação de T_c



Fonte: Elaborado pelo autor.

4.2 Etapa 2 - Parte 1: Obteção e Tratamento dos dados

Na presente dissertação foram analisados dados coletados no LEA-UFC. A planta FV do LEA foi conectada à rede elétrica da UFC em julho de 2016 em parceria com o Núcleo de Ensino e Pesquisa em Agricultura Urbana (NEPAU), Instituto Joazeiro de Desenvolvimento Sustentável e a Superintendência de Infraestrutura. Essa seção apresenta os detalhes da coleta e obtenção dos dados utilizados na aplicação das técnicas de ML.

4.2.1 Obtenção de dados

O LEA da UFC é localizado em Fortaleza (latitude: $-03^{\circ} 43'$, longitude: $-38^{\circ} 32'$, elevação: 21 m), Ceará, região Nordeste brasileira, próximo à Linha do Equador. O primeiro semestre do ano nesta cidade é considerado o período chuvoso, já o segundo, o período seco, com maior incidência de ventos e de irradiância solar. Estes dados foram obtidos utilizando-se de um Sistema de Monitoramento de Energia Renovável (*Renewable Energy Monitoring System* (REMS)) desenvolvido em Pereira (2018).

A planta FV do LEA é composta de 6 módulos de 250 Wp, com uma potência total de 1500 Wp, da fabricante Yingli Solar e um inversor de fabricação nacional, necessário para conexão à rede. Os módulos estão orientados para o Norte geográfico com inclinação de 5 graus.

A Figura 7 mostra os módulos FV vistos de cima e de baixo do LEA, assim como os sensores instalados para aquisição dos dados de G , V_w , T_a e T_c .

Figura 7 – Sistema FV instalado no LEA-UFC, composto de 6 módulos, 1 anemômetro, 1 piranômetro e sensores de temperatura. Vista superior e inferior dos painéis.



Fonte: Elaborado pelo autor.

O REMS foi utilizado para coleta de dados do sistema FV (PEREIRA, 2018), foram considerados no presente estudo dados de T_c , G , V_w , T_a , dia e $hora$ no momento da coleta dos dados. É utilizado um piranômetro Hukseflux LP02 no mesmo ângulo de inclinação dos módulos FV para medições de G no plano inclinado. Os dados de medição do piranômetro foram validados com o uso do medidor de irradiância Solar-100 da marca AMPROBE. Um anemômetro NRG #40C também na mesma inclinação dos módulos FV foi instalado com o intuito de medir V_w sobre a superfície dos módulos FV. Para a medição T_a , é utilizado um sensor PT100 com escudo de radiação (*radiation shield*). Para o monitoramento de T_c dos seis módulos FV conectados à rede no LEA-UFC, foram fixados sensores PT100 na parte posterior central de cada módulo, que é o local de maior temperatura, devido ao menor fluxo de ar. A validação dos dados de temperatura é feita utilizando uma câmera térmica infravermelha do modelo FLIR ONE.

Os dados foram coletados de abril de 2018 a novembro de 2020, totalizando 32 meses. Os mesmos foram coletados e armazenados em planilhas eletrônicas com intervalos de 1 minuto.

4.2.2 Tratamento dos dados

A partir do conjunto de dados, foram selecionadas as medições realizadas durante o período de 06:00 horas às 17:00 horas. Os dados também foram tratados em busca de valores atípicos (*outliers*), foram determinados limites superior e inferior de G e T_a com base em informações históricas obtidas no local de estudo, cidade de Fortaleza, por meio do Instituto Nacional de Meteorologia (INMET). A faixa de T_a adotada foi de 21,7°C a 39°C e a faixa de G considerada foi de 0 W/m² a 1250 W/m² (INSTITUTO NACIONAL DE METEOROLOGIA, 2019). O sensor de temperatura acoplado no módulo FV de número 2 apresenta valores descalibrados e com divergência considerável dos valores medidos pelos demais sensores de temperatura. Os valores do sensor de temperatura número 2, assim como todos os casos de *outliers* foram retirados do banco de dados utilizado no estudo. Os valores das medições dos outros 5 sensores de temperatura acoplados aos painéis FV foram utilizados para realização do cálculo da T_{c_med} , valor médio da temperatura de operação dos painéis. O preditor *dia* se refere ao dia do ano em que foi feita a medição, varia de 1 a 365, começando a contagem em 01 de janeiro, a hora (*hr*) foi colocada como numérica no formato de fração de hora variando de zero a um, onde o valor 1 corresponde a hora 24.

Após o processo de limpeza e transformação dos dados o conjunto de dados ficou com 208.841 medições, sendo destes 80% destinados ao treinamento (167073 medições) e 20% (41768 medições) destinados à testagem dos modelos de ML. A partir daí foi possível fazer uma análise estatística e gráfica das variáveis do banco de dados.

4.3 Etapa 2 - Parte 2: Aplicação dos modelos no Regression Learner App

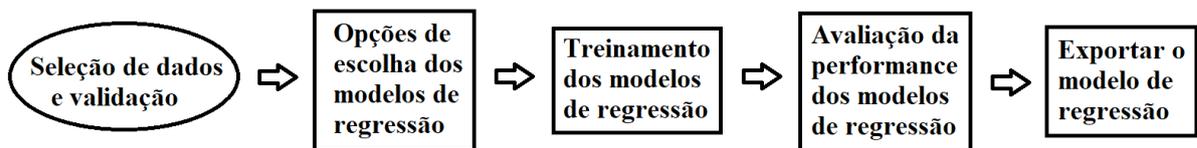
No presente estudo foram aplicadas técnicas de aprendizagem supervisionada de regressão para a modelagem de T_c . Os algoritmos de RL, SVM, GPR e RN são aplicadas usando como componentes os preditores G , V_w , T_a , *dia* e *hr*. Os valores de T_c são atribuídos à variável resposta. Esta seção apresenta os detalhes da ferramenta e técnicas de ML utilizadas para o desenvolvimento dos modelos de previsão de T_c .

4.3.1 Regression Learner App

O *Regression Learner App* (RL-App) do *software Matlab* é capaz de realizar o treinamento de modelos de regressão de ML supervisionado dado um conjunto de dados de entrada

(preditores) e respostas, para treinar um modelo que gera respostas previstas para novos dados por meio de diferentes algoritmos. Também é possível explorar os dados, selecionar preditores, especificar esquemas de validação e avaliar os resultados. Por padrão, o aplicativo protege contra *overfitting* treinando um modelo com um esquema de validação aplicando validação cruzada (*Cross-Validation*) *k*-fold, mas também é possível escolher outras técnicas de validação como o método *holdout*. Ao escolher o método *holdout* o RL-App separa o banco de dados em um conjunto de treinamento e outro de validação, por padrão, é definido 75% dos dados para treinamento e 25% para validação, mas estes valores podem ser ajustados. Posteriormente, o modelo é treinado no conjunto de treinamento e seu desempenho é avaliado com o conjunto de validação. O modelo usado para validação é baseado em apenas uma parte dos dados (conjunto de treinamento), então essa técnica de validação é apropriada apenas para grandes quantidades de dados. Após o treinamento, pode-se comparar o RMSE de cada modelo lado a lado, uma ferramenta útil na escolha do melhor modelo de regressão, incluindo modelos de regressão linear, árvores de regressão, modelos de regressão por processo Gaussiano, SVM e conjuntos de árvores de regressão (MATHWORKS, 2020b). O fluxograma da Fig. 8 mostra a linha de treinamento de modelos de regressão no RL-App.

Figura 8 – Fluxo de trabalho para treinamento de modelos de regressão no *Regression Learner App*



Fonte: Adaptado de Mathworks (MATHWORKS, 2020b).

4.3.2 Aplicação dos modelos no RL-App

Na presente pesquisa os modelos gerados para previsão de T_c utilizando as técnicas de RL, SVM, GPR e RN disponíveis no RL-App são comparados. Os modelos de RL são fáceis de interpretar e rápidos para fazer previsões, capaz de desenvolver modelos em questão de segundos, essas características tornam os modelos de regressão linear modelos populares para serem testados primeiro (MATHWORKS, 2020a). Em seguida foram aplicadas as técnicas de SVM, GPR e RN percorrendo o seguinte passo-a-passo:

- 1. Seleção dos dados:** do conjunto de dados obtidos no LEA foi utilizado para treinamento dos modelos 80% do conjunto de dados, os outros 20% foi reservado para teste dos modelos.
- 2. Especificação dos variáveis de resposta e as usadas como preditores:** a coluna de valores de T_c é atribuída à variável resposta, o valor a ser estimado pelos modelos. As colunas de dia , hr , G , T_a e V_w são definidos como preditores (componentes) dos modelos.
- 3. Escolha do esquema de validação:** Dada a utilização de uma grande quantidade de amostras no conjunto de dados de treinamento, o método de validação *houdout* é escolhido para prevenção de *overfitting* com a escolha de usar 25% do conjunto de dados de treinamento para validação. Ou seja, o conjunto de validação representa 25% dos dados reservados para treinamento (80% do conjunto de dados obtido no LEA). Isto é, 20% do total de dados são utilizados para validação.
- 4. Seção Model Type:** Para decidir a melhor abordagem para cada uma das quatro técnicas, RL, SVM, GPR e RN, inicialmente foi realizada uma etapa de otimização para cada uma das técnicas. Na otimização o RL-App testa diferentes configurações dos modelos e busca aquele que apresente um menor Erro quadrático médio (*Mean Square Error* - MSE). Após realizar a otimização para cada uma das quatro técnicas, são selecionados quatro modelos: RL com Interações Lineares, *Fine Gaussian* SVM, *Exponential* GPR e Wide RN. Os modelos estão detalhados na Tabela 8. As principais configurações utilizadas em cada um dos modelos estão definidas na tabela 9.
- 5. Avaliação dos modelos:** os tipos de modelos selecionados ficam na lista do Histórico. Ao final do treinamento a melhor pontuação RMSE é destacada, também são exibidas as estatísticas do modelo, o R^2 , o MSE e o Erro médio absoluto (*Mean Absolut Error* - MAE). O R^2 é o coeficiente de determinação, quanto mais alto seu valor, melhor o modelo se ajusta aos dados, seu valor varia entre 0 e 1. O MAE é semelhante ao RMSE, mas não penaliza tanto os pontos mais distantes daqueles preditos pelo modelo.
- 6. Dados de teste:** São aplicados os dados de teste aos modelos desenvolvidos utilizando a opção "*Test Data*" do RL-App com o intuito de explorar os resultados dos modelos no conjunto de teste. É feita comparação do desempenho dos mesmos inspecionando os resultados estatísticos e gráficos.
- 7. Escolha do modelo:** O melhor modelo de ML é escolhido, sendo este o que apresente os menores valores de erro (RMSE, MSE e MAE) e um melhor R^2 .
- 8. Dados de teste:** São aplicados os dados de teste nos modelos comparativos selecionados.

9. Avaliação dos modelos nos dados de teste: são explorados os resultados dos modelos comparativos e do melhor modelo desenvolvido no RL-App no conjunto de teste; em seguida é feita a comparação do desempenho dos mesmos inspecionando seus resultados estatísticos.

Tabela 8 – Modelos aplicados aos dados do LEA no RL-App.

Tipo de modelo	Interpretabilidade	Descrição
RL com Interações Lineares	Fácil	Modelo de regressão linear com interceptação, termos lineares e interações entre os termos
<i>Fine Gaussian SVM</i>	Difícil	Modelo SVM que permite variações rápidas na função de resposta. A função kernel é Gaussiana com a escala do kernel definida como $\sqrt{P}/4$, onde P é o número de preditores
<i>Exponential GPR</i>	Difícil	Modelo GPR que utiliza um kernel exponencial
<i>Wide RN</i>	Difícil	Modelo de regressão RN com uma camada totalmente conectada de tamanho 100, excluindo a camada final totalmente conectada

Fonte: Adaptado de (MATHWORKS, 2020a).

Tabela 9 – Principais configurações dos modelos

Modelo	Tipo	Kernel function	Kernel scale	Standardize data	Outras configurações
RL	<i>Interactions Linear</i>	-	-	-	<i>Terms: Quadratic</i>
SVM	<i>Fine Gaussian SVM</i>	<i>Gaussian</i>	0.56	<i>true</i>	<i>Box constraint: Automatic Epsilon: Automatic</i>
GPR	<i>Exponential GPR</i>	<i>Exponential</i>	<i>Automatic</i>	<i>true</i>	<i>Basis function: Constant Use isotropic kernel: true Signal standard deviation: Automatic Sigma: Automatic Optimize numeric parameters: true</i>
RN	<i>Wide Neural Network</i>	-	-	<i>Yes</i>	<i>Activation function: Sigmoid Number of fully connected layers: 1 First layer size: 100 Iteration limit: 1000 Regularization strength (Lambda): 0</i>

Fonte: Elaborado pelo autor.

5 RESULTADOS

A presente dissertação foi dividida em duas etapas, inicialmente são apresentados os resultados referentes ao estudo dos modelos para determinação de T_c encontrados na literatura a partir do ano 2000. Em seguida, os resultados referentes à caracterização dos dados de dia , hr , G , T_a e V_w , para desenvolvimento de modelos para determinação de T_c . E dos modelos de regressão por ML por meio de RL, SVM, GPR e RN. Estes resultados são discutidos e comparados. Em seguida, é feita uma equiparação com os resultados dos modelos comparativos selecionados, já existentes na literatura.

5.1 Etapa 1: Estudo dos modelos para determinação de T_c

Os resultados desta seção são fruto de uma revisão de modelos recentes para estimativa de T_c presentes na literatura, 33 equações encontradas na literatura, apresentadas na tabela 2, para estimar T_c são resumidas em apenas 3 formas gerais. Além disso, são apresentados os resultados da seleção de modelos presentes na tabela 2 que utilizem os mesmos preditores dos modelos de ML desenvolvidos na presente dissertação, os denominados modelos comparativos.

5.1.1 Forma geral das correlações para determinação de T_c

Analisando a forma funcional das equações dos modelos da Tabela 2 os modelos Fernández *et al.*, Durish *et al.*, Nordmann, Krauter, Mondol I, Hove, Tiwari, Eicker, Standard, Davis, Mondol II, Tselepis, Tiwari I e II, Almonacid, Markvart, Muzathik, Akyuz *et al.*, NOCT-1p e NOCT-2p são lineares e podem ser expressados de forma geral pela Eq. (5.1). O modelo do ISFOC é linear se T_{h-s} e os termos entre parênteses forem assumidos constantes.

$$T_c = a_0 + a_1 T_a + a_2 G + a_3 V_w + a_4 T_b \quad (5.1)$$

Os termos a_0 , a_1 , a_2 , a_3 e a_4 são específicos de cada um dos modelos lineares mencionados e estão expostos na Tabela 10.

Os modelos de Faiman, Skoplaki & Palyvos, Skoplaki *et al.* I, Duffie & Beckman I, Skoplaki *et al.* II, Chenni *et al.*, Kurtz *et al.*, Hornung *et al.*, Coskun *et al.*, são não-lineares e podem ser expressados de forma geral pela Eq. (5.2). O modelo Sandia pode ser enquadrado

nesta expressão geral se o $\Delta T = (T_c - T_b)G_{ref}$ for constante.

$$T_c = b_0 + b_1 T_a + C_1 (e^{d_1 + d_2 V_w} + e_1) G + C_2 (e_2 + f_1 T_a)^{g_1} (e_3 + f_2 G)^{g_2} (e_4 + f_3 V_w)^{g_3} \quad (5.2)$$

Os termos $b_0, b_1, C_1, d_1, d_2, e_1, C_2, e_2, f_1, g_1, e_3, f_2, g_2, e_4, f_3$ e g_3 são específicos de cada um dos modelos não lineares mencionados e expostos na Tabela 11. Aparentemente, o modelo de Kurtz *et al.* é um caso particular do modelo de Hornung *et al.*. Os modelos de Duffie e Beckman II e de Mattei são não-lineares e podem ser expressados na forma da Eq. (5.3).

$$T_c = (h_1 T_a + h_2 G)(m_1 + n_1 G)^{-1} \quad (5.3)$$

Os termos h_1, h_2, m_1 e n_1 são específicos dos modelos Duffie and Beckman II e Mattei e estão expostos na Tabela 12 . Resumindo, todos os modelos compilados na Tabela 2 podem ser agrupados na forma de uma das equações definidas acima de acordo com os termos que compõem os mesmos.

5.1.2 Modelos comparativos

Dentre os modelos presentes na Tabela 2, é separado um grupo de modelos que será aplicado no conjunto de dados de teste. Tais modelos serão utilizados para comparação com os modelos desenvolvidos na etapa 2 deste trabalho.

Para filtragem dos modelos, exclui-se inicialmente aqueles que consideram preditores que não estão presentes no banco de dados do LEA utilizado neste trabalho ou que não estão presentes no *datasheet* dos módulos FV utilizados, como $V_{oc}, U_L, \alpha, \tau$, dentre outros. Também exclui-se os modelos desenvolvidos utilizando um arranjo experimental distinto do arranjo utilizado neste trabalho, como no caso dos modelos para módulos HCPV, ou de sistemas FV integrados.

Após o processo de filtragem os modelos selecionados, denominados "modelos comparativos" são: Akyuz *et al.*; Coskun; Standard; NOCT-1p; NOCT-2p; e Sandia.

Tabela 10 – Termos das correlações lineares quando expressas pela equação 5.1

Correlação	Formulação	a_0	a_1	a_2	a_3	a_4
Fernández <i>et al.</i>	$T_c = \frac{(V_{oc}-c_1G-c_3)}{c_2}$	$\frac{(V_{oc}-c_3)}{c_2}$	0	$\frac{-c_1}{c_2}$	0	0
Durisch <i>et al.</i>	$T_c = T_a + kG$	0	1	k	0	0
Nordmann	Igual o de cima	0	1	k	0	0
Krauter	Igual o de cima	0	1	k	0	0
Mondol I	$T_c = T_a + 0.031G$	0	1	0.031	0	0
Hove	$T_c = T_a + G \left(\frac{\tau\alpha-\eta}{U_L} \right)$	0	1	$\frac{\tau\alpha-\eta}{U_L}$	0	0
Tiwari	$T_c = T_a + G \left(\frac{\tau\alpha}{U_L} \right) \left[1 - \left(\frac{\eta}{\tau\alpha} \right) \right]$	0	1	$\left(\frac{\tau\alpha}{U_L} \right) \left[1 - \left(\frac{\eta}{\tau\alpha} \right) \right]$	0	0
Eicker	$T_c = T_a + G \left(\frac{\alpha}{U_L} \right) \left[1 - \frac{\eta}{\alpha} \right]$	0	1	$\left(\frac{\alpha}{U_L} \right) \left[1 - \frac{\eta}{\alpha} \right]$	0	0
Standard	$T_c = T_a + \left(\frac{G}{G_{NOCT}} \right) (T_{c,NOCT} - T_{a,NOCT})$	0	1	$\frac{(T_{c,NOCT}-T_{a,NOCT})}{G_{NOCT}}$	0	0
Davis	$T_c = T_a + \left(\frac{G}{G_{NOCT}} \right) (T_{c,NOCT} - T_{a,NOCT}) \left[1 - \left(\frac{\eta}{\tau\alpha} \right) \right]$	0	1	$\left(\frac{T_{c,NOCT}-T_{a,NOCT}}{G_{NOCT}} \right) \left[1 - \left(\frac{\eta}{\tau\alpha} \right) \right]$	0	0
Mondol II	$T_c = T_a + 0.031G - 0.058$	-0.058	1	0.031	0	0
Tselepis	$T_c = 30 + 0.0175(G - 150) + 1.14(T_a - 25)$	$[30 - 0.0175(150) - 1.14T_a + 1.14(25)]$	1	0.0175G	0	0
Tiwari I	$T_c = \frac{pG(\tau\alpha-\eta)+(U_iT_a+U_TT_b)}{(U_i+U_T)}$	0	$\frac{U_i}{(U_i+U_T)}$	$\frac{p(\tau\alpha-\eta)}{(U_i+U_T)}$	0	$\frac{U_T}{(U_i+U_T)}$
Tiwari II	$T_c = \frac{\tau[\alpha_C p + \alpha_\gamma(1-\beta_c)G - \eta_c G \beta_c + U_i T_a + U_T T_b]}{(U_i+U_T)}$	$\frac{\tau\alpha_C p}{(U_i+U_T)}$	$\frac{\tau U_i}{(U_i+U_T)}$	$\frac{\tau\alpha_\gamma(1-\beta_c-\eta_c\beta_c)}{(U_i+U_T)}$	0	$\frac{\tau U_T}{(U_i+U_T)}$
Almonacid	$T_c = T_a + d_1 G + d_2 V_w$	0	1	d_1	d_2	0
Markvart	$T_c = 0.943T_a + 4.3 + 0.028G - 1.528V_w$	4.3	0.943	0.028	-1.528	0
Muzathik	$T_c = 0.943T_a + 0.3529 + 0.0195G - 1.528V_w$	0.3529	0.943	0.0195	-1.528	0
Akyuz <i>et al.</i>	$T_c = 0.95T_a + 3.1 + 0.025G - 0.3V_w$	3.1	0.95	0.025	-0.3	0
NOCT-1p	$T_c = T_a + \left(\frac{G}{G_{NOCT}} \right) (T_{c,NOCT} - T_{a,NOCT}) + a(V_w - V_{w,NOCT})$	$-aV_{w,NOCT}$	1	$\left(\frac{T_{c,NOCT}-T_{a,NOCT}}{G_{NOCT}} \right)$	a	0
NOCT-2p	$T_c = T_a + b \left(\frac{G}{G_{NOCT}} \right) (T_{c,NOCT} - T_{a,NOCT}) + c(V_w - V_{w,NOCT})$	$-cV_{w,NOCT}$	1	$b \left(\frac{T_{c,NOCT}-T_{a,NOCT}}{G_{NOCT}} \right)$	c	0
ISFOC	$T_c = T_b + \left(\eta C_g \sum \frac{L_i}{\lambda_i} \right) G$	0	0	$\left(\eta C_g \sum \frac{L_i}{\lambda_i} \right)$	0	1

Fonte: Elaborado pelo autor.

Tabela 11 – Termos das correlações não-lineares quando expressas pela equação 5.2.

Correlação	Formulação	b_0	b_1	C_1	d_1	d_2	e_1	C_2	e_2	f_1	g_1	e_3	f_2	g_2	e_4	f_3	g_3
Faiman	$T_c = T_a + \frac{G}{U_0 + U_1 V_w}$	0	1	0	-	-	-	1	1	0	1	0	1	1	U_0	U_1	-1
Skoplaki e Palyvos	$T_c = T_a + \frac{0.32}{8.91 + 2V_w} G$	0	1	0	-	-	-	1	1	0	1	0	0.32	1	8.91	2	-1
Skoplaki <i>et al.</i> I	$T_c = T_a + \frac{0.25}{5.7 + 3.8V_w} G$	0	1	0	-	-	-	1	1	0	1	0	0.25	1	5.7	3.8	-1
Duffie e Beckman I	$T_c = T_a + \left(\frac{G}{G_{NOCT}}\right) \left(\frac{9.5}{(5.7 + 3.8V_w)}\right) (T_{c,NOCT} - T_{a,NOCT}) \left[1 - \left(\frac{\eta}{\tau\alpha}\right)\right]$	0	1	0	-	-	-	1	1	0	1	0	$\frac{9.5(T_{c,NOCT} - T_{a,NOCT})}{\tau\alpha G_{NOCT}(\tau\alpha - \eta)^{-1}}$	1	5.7	3.8	-1
Skoplaki <i>et al.</i> II	$T_c = T_a + \frac{G}{G_{NOCT}} (T_{c,NOCT} - T_{a,ref}) \left[\frac{h_{NOCT}}{h} \left\{ 1 - \frac{\eta_{STC}}{\tau\alpha} (1 + \beta_{STC} T_{STC}) \right\} \right]$	0	1	0	-	-	-	1	1	0	1	0	$\frac{h_{NOCT}(T_{c,NOCT} - T_{a,ref})(\tau\alpha)^{-1}}{G_{NOCT}[\tau\alpha - \eta_{STC}(1 + \beta_{STC} T_{STC})]^{-1}}$	1	a	b	-1
Chenni <i>et al.</i>	$T_c = T_a + 0.0138G(1 + 0.031T_a)(1 - 0.042V_w)$	0	1	0	-	-	-	0.0138	1	0.031	1	0	1	1	1	-0.042	1
Kurtz <i>et al.</i>	$T_c = T_a + Ge^{-3.473 - 0.0594V_w}$	0	1	1	-3.473	-0.0594	0	0	-	-	-	-	-	-	-	-	-
Hornung <i>et al.</i>	$T_c = T_a + m \left[e^{\left(\frac{-0.5V_w}{V_{w0}}\right)} + c \right] G$	0	1	m	0	$\frac{-0.5}{V_{w0}}$	c	0	-	-	-	-	-	-	-	-	-
Coskun <i>et al.</i>	$T_c = 1.4T_a + 0.01(G - 500) - V_w^{0.8}$	-0.01	1.4	0.01	0	0	0	-1	1	0	1	1	0	1	0	1	0.8
Sandia	$T_c = T_b + \left(\frac{G}{G_{ref}}\right) \Delta T$ and $T_b = G(e^{a+bV_w}) + T_a$	0	1	$\frac{\Delta T}{G_{ref}}$	a	b	0	0	-	-	-	-	-	-	-	-	-

Se $C_1 = 0$, os termos d_1 , d_2 e e_1 que dependem de C_1 são desconsiderados e representados por '-'. O mesmo ocorre com os termos e_2 , f_1 , g_1 , e_3 , f_2 , g_2 , e_4 , f_3 e g_3 quando $C_2 = 0$.

Fonte: Elaborado pelo autor.

Tabela 12 – Termos das correlações não-lineares Duffies e Beckman II e Mattei quando expressas pela equação 5.3.

Correlação	Formulação	h_1	h_2	m_1	n_1
Duffie e Beckman II	$T_c = \frac{T_a + \frac{G}{G_{NOCT}} (T_{c,NOCT} - T_{a,ref}) \left\{ 1 - \frac{\eta_{STC}}{\tau\alpha} (1 + \beta_{STC} T_{STC}) \right\}}{1 - \frac{\beta_{STC} \eta_{STC}}{\tau\alpha} \frac{G}{G_{ref,NTE}} (T_{c,NOCT} - T_{a,ref})}$	1	$\frac{(T_{c,NOCT} - T_{a,ref})}{G_{NOCT}} \left\{ 1 - \frac{\eta_{STC}}{\tau\alpha} (1 + \beta_{STC} T_{STC}) \right\}$	1	$-\frac{\beta_{STC} \eta_{STC}}{\tau\alpha} \frac{(T_{c,NOCT} - T_{a,ref})}{G_{ref,NTE}}$
Mattei	$T_c = \frac{UT_a + G(\tau\alpha - \eta_{STC} - \beta_{STC} \eta_{STC} T_{STC})}{U - \beta_{STC} \eta_{STC} G}$	U	$(\tau\alpha - \eta_{STC} - \beta_{STC} \eta_{STC} T_{STC})$	U	$-\beta_{STC} \eta_{STC}$

Fonte: Elaborado pelo autor.

5.2 Etapa 2: Parte 1 - Análise dos dados coletados

Os resultados referentes à caracterização dos dados de G , T_a , T_c e V_w obtidos no LEA-UFC estão expostos a seguir. Na cidade de Fortaleza, Ceará, localizada em zona equatorial, há basicamente duas estações no ano, a estação seca e a chuvosa. A estação chuvosa ocorre no primeiro semestre do ano que é marcado por chuvas, mais nebuloso e com menor intensidade de vento e radiação solar. Segundo o Atlas Eólico e Solar do Ceará, a estação seca ocorre no segundo semestre onde prevalecem maiores velocidades do vento e índices de radiação, em conjunto com menor nebulosidade e precipitação (CAMARGO SCHUBERT ENGENHEIROS ASSOCIADOS et al, 2019).

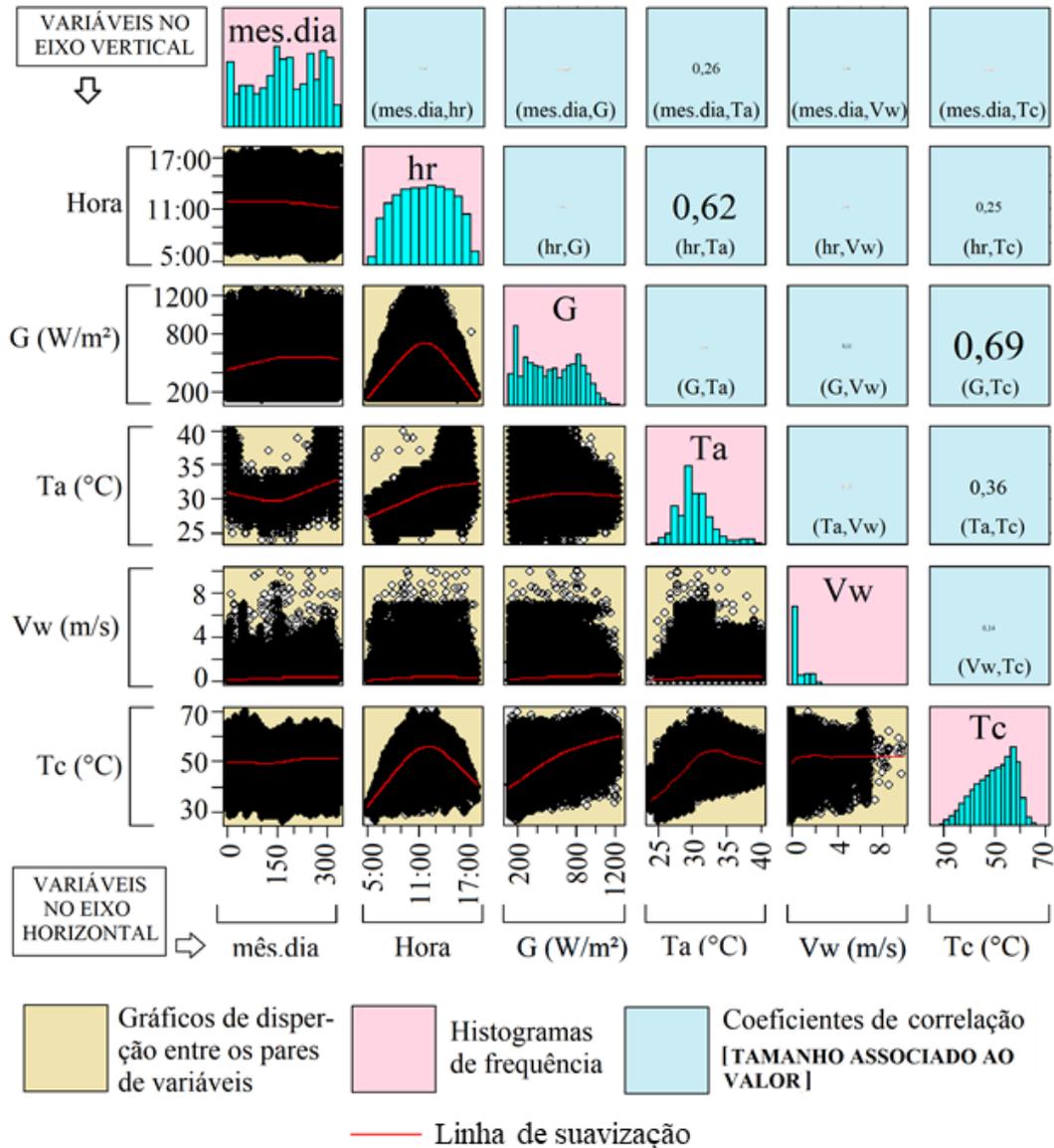
A Figura 9 expõe as correlações entre as diferentes variáveis utilizadas no banco de dados obtidos no LEA. Na diagonal principal estão histogramas de frequência para visualizar a distribuição dos dados em cada variável. Abaixo da diagonal principal estão os gráficos de dispersão para cada par de variáveis. A linha em vermelho nos gráficos é uma linha de suavização, permitindo avaliar relação entre os pares de variáveis. Acima da diagonal principal estão os coeficientes das correlações, o tamanho dos números está condicionado ao valor da correlação, quanto maior o número, maior a correlação entre este par. Se os preditores forem muito correlacionados, as inferências baseadas no modelo de regressão podem ser errôneas ou pouco confiáveis. Normalmente, procura-se utilizar preditores fortemente relacionados com a variável a ser prevista, neste caso T_c , e com o mínimo de correlação entre si.

Analisando a figura 9 pode-se perceber que algumas das correlações entre os pares de variáveis, como a correlação entre T_c e G , são visivelmente lineares. Já correlações como T_c e hr se mostram de forma exponencial, o que deve ser levado em consideração no desenvolvimento dos modelos. Um modelo para previsão de T_c unicamente linear não irá demonstrar de forma eficiente a correlação entre T_c e hr . Além disso, olhando para a última coluna azul, onde estão os valores das correlações com T_c , é possível ver quais as variáveis possuem maior influência em T_c . Pode-se observar forte correlação com G , seguida por T_a e hr , a correlação com as demais variáveis, V_w , ano e dia , é tão pequena que os valores se tornam ilegíveis.

5.2.1 Irradiância solar

As horas com Sol em Fortaleza não variam significativamente durante o ano, cerca de 20 minutos a mais ou a menos de 12 horas no ano inteiro. Já a irradiância passa por variações

Figura 9 – Histogramas de frequência e correlações entre os pares de variáveis do banco de dados do LEA-UFC



Fonte: Elaborado pelo autor.

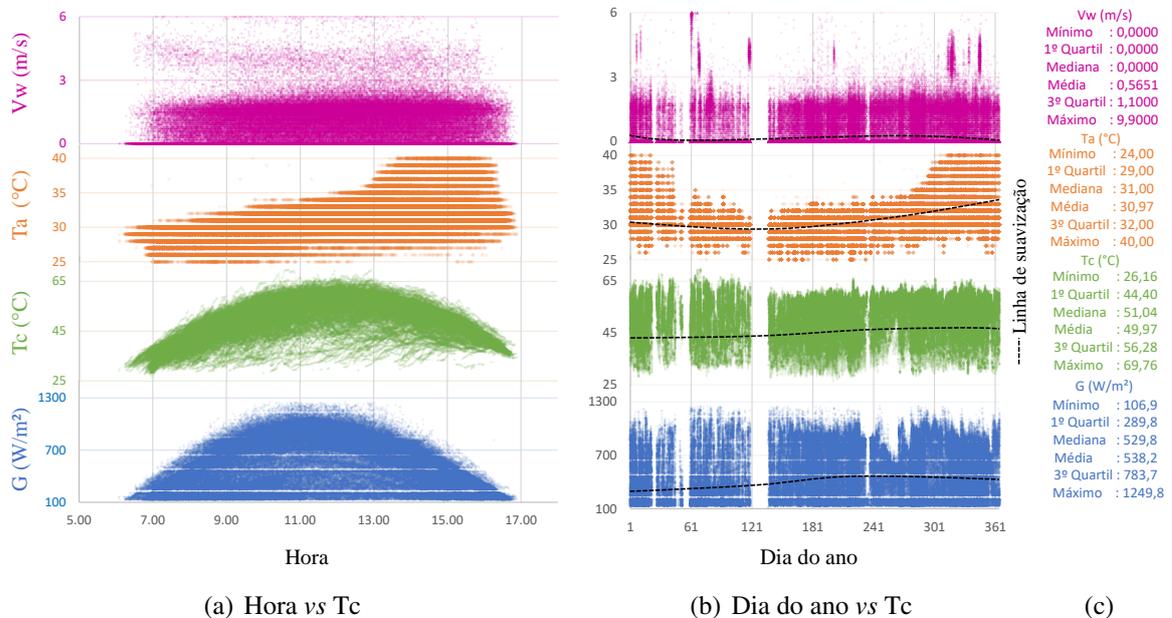
sazonais moderadas ao longo do ano. Segundo o Atlas Eólico e Solar do Ceará, o período com mais G é durante a estação seca, especialmente no mês de outubro. O período com menos G se dá durante a estação chuvosa (CAMARGO SCHUBERT ENGENHEIROS ASSOCIADOS et al, 2019). Essa diferença de G ao longo do ano, do dia 1 ao 365, a contar de 01 de janeiro, pode ser vista na Figura 10(b) na cor azul. A partir das 6:00hr a irradiância é crescente até em torno das 12:00hr, tendo os valores mais elevados por volta das 10 - 14:00hr. A Figura 10(a) também na cor azul mostra uma caracterização das irradiâncias horárias para o período em estudo, de abril de 2018 a novembro de 2020, o eixo horizontal indica a hora e o eixo vertical indica a G medida. A Figura 10(c) na cor azul apresenta as estatísticas descritivas de G : os valores mínimo e

máximo; a média e mediana; e os valores do 1º e 3º quartis. O 1º quartil, ou quartil inferior, é o valor que corta os primeiros 25% dos dados quando são classificados em ordem crescente. O terceiro quartil, ou quartil superior, é o valor que corta os primeiros 75%.

5.2.2 Temperatura ambiente

Devido à proximidade com a Linha do Equador, no Ceará há baixa variação de T_a ao longo do ano. Os valores mínimos mensais são observados na estação chuvosa, com médias cerca de $25,7^\circ\text{C}$, sendo o mês de junho a apresentar as menores temperaturas. Já os valores máximos mensais ocorrem na estação seca, de acordo com o Atlas Eólico e Solar do Ceará, as médias chegam a $27,4^\circ\text{C}$, com pico em novembro, devido ao clima seco e aos maiores índices de radiação (CAMARGO SCHUBERT ENGENHEIROS ASSOCIADOS et al, 2019). Essa diferença de T_a ao longo do ano pode ser vista na Figura 10(b) na cor laranja. A Figura 10(a) também na cor laranja mostra uma caracterização das temperaturas horárias. As temperaturas são crescentes até em torno das 16:30hr, sendo as temperaturas mais elevadas por volta das 12 - 16:30hr. A Figura 10(c) na cor laranja apresenta as estatísticas descritivas de T_a .

Figura 10 – Gráficos de caracterização de V_w , T_a , T_c e G : (a) ao longo do dia (b) ao longo do ano (c) estatísticas descritivas.



Fonte: Elaborado pelo autor utilizando os dados obtidos no LEA-UFC no período de 2018-2020.

5.2.3 *Velocidade do vento*

Sazonalmente, o Ceará apresenta maiores velocidades do vento no segundo semestre do ano em relação à primeira metade. Segundo o Atlas Eólico e Solar do Ceará, os ventos na região exibem altas velocidades médias durante o período seco, predominantemente de setembro a dezembro, e menores médias mensais no ápice do período chuvoso, de março a abril (CAMARGO SCHUBERT ENGENHEIROS ASSOCIADOS et al, 2019). Na Figura 10(b) na cor roxo é possível observar uma diferença entre os valores de V_w medidos no primeiro e segundo semestre do ano no LEA-UFC. O anemômetro do LEA é instalado próximo e com a mesma inclinação dos painéis, além disso, o laboratório é circundado de árvores o que reduz a intensidade de V_w . A Figura 10(a) também na cor roxa mostra uma caracterização das V_w horárias. A Figura 10(c) na cor roxa apresenta as estatísticas descritivas de V_w .

5.2.4 *Temperatura da célula*

Os valores de T_c assim como de T_a são mais elevados na estação seca em Fortaleza. A estação chuvosa apresenta valores levementes inferiores nas medições T_c . Essa diferença de T_c ao longo do ano pode ser vista na Figura 10(b) na cor verde, é possível ver que os valores de T_c não oscilam tanto como os de T_a ao longo do ano. A Figura 10(a) mostra uma caracterização de T_c horária para o período de abril de 2018 a novembro de 2020. As temperaturas são crescentes até em torno das 12:00hr, sendo as temperaturas mais elevadas por volta das 10 - 14:00hr. Vale notar que o pico de T_c é semelhante ao de G no quesito horário, T_a tem o pico com temperaturas mais elevadas algumas horas mais tarde. A Figura 10(c) na cor verde apresenta as estatísticas descritivas de T_c .

5.3 **Etapa 2: Parte 2 -Resultados da aplicação da técnicas de ML**

Os modelos RL, SVM, GPR e RN foram treinados utilizando os dados de treinamento e validação (80% do total de dados obtidos) com base em cinco entradas: dia , hr , G , T_a e V_w . Após o processo de treinamento, os modelos gerados foram testados no conjunto de dados de teste (20% do total de dados obtidos).

A T_c estimada é representada nas Figuras 12(a), 12(b), 12(c) e 12(d) expondo os resultados no conjunto de teste dos modelos RL, SVM, GPR e RN respectivamente na forma de um gráfico de dispersão Predito vs. Real. Este gráfico ajuda a entender o quão bem o modelo de

regressão faz previsões para diferentes valores de resposta. A distância vertical da reta a qualquer ponto é o erro da previsão para aquele ponto. Nos 4 gráficos, os pontos geralmente seguem um padrão linear com resultados distribuídos ao longo da reta de correlação entre os valores preditos e os reais, com poucas amostras com comportamentos anômalos (*outliers*). Pode-se destacar na Figura 12 os gráficos 12(b) e 12(c), referente aos modelos SVM e GPR, por apresentarem um comportamento mais simétrico e próximos da reta de correlação.

As Figuras 13(a), 13(b), 13(c) e 13(d) exibem a previsão da T_c pelas técnicas RL, SVM, GPR e RN respectivamente. A resposta prevista é mostrada em amarelo sobrepondo os valores de T_c do conjunto de teste em azul *versus* o dia do ano. Comparando os gráficos na Figura 13 é possível observar que os modelos SVM e GPR, Figura 13(b) e 13(c), apresentam uma cobertura melhor dos dados de teste. Os modelos RL na Figura 13(a), e RN na Figura 13(d) ficam limitados para temperaturas acima de 60°C, sendo o modelo RL destacadamente o pior nessa faixa de temperatura.

Nas Figuras 14(a), 14(b), 14(c) e 14(d) estão expostos os resultados de previsão dos modelos por hora pelas técnicas RL, SVM, GPR e RN respectivamente. A resposta prevista é mostrada em amarelo sobrepondo os valores de T_c do conjunto de teste em azul *versus* a hora. Comparando os quatro gráficos da Figura 13 é possível observar que os modelos SVM e GPR 14(b), 14(c) apresentam uma cobertura melhor dos dados de teste. Os modelos RL na Figura 14(a), e RN na Figura 14(d) ficam limitados para temperaturas acima de 60°C, entre as 9:00 e 15:00. Além disso, o modelo RL também mostra uma tendência a superestimar os valores de T_c abaixo de 40°C nesta mesma faixa de horário.

A fim de avaliar o desempenho dos modelos propostos, a Tabela 13 mostra os principais resultados dos modelos no conjunto de treinamento e de teste. O RMSE, o R^2 , o MSE e o MAE serviram para avaliar as diferenças entre os valores previstos e reais dos modelos.

Tabela 13 – Principais resultados dos modelos de ML do 1 até 4

Modelo	Preditores	Conjunto	RMSE	R^2	MSE	MAE
1	RL com Interações Lineares	Treinamento	3,5782	0,78	12,803	2,6573
		Teste	3,5377	0,78	12,515	2,6405
2	<i>Fine Gaussian SVM</i>	Treinamento	2,9041	0,85	8,4339	1,9273
		Teste	2,8374	0,86	8,0507	1,8851
3	<i>Exponential GPR</i>	Treinamento	2,7216	0,87	7,4073	1,8343
		Teste	2,651	0,88	7,0278	1,787
4	<i>Wide RN</i>	Treinamento	3,0004	0,84	9,0023	2,0969
		Teste	2,9706	0,85	8,8245	2,0817

Fonte: Elaborado pelo autor.

Pode-se observar por meio da Figura 12, 13 e 14 que os modelos SVM e GPR têm melhores habilidades de previsão para T_c do que os métodos RL e RN. Os resultados da Tabela 13 indicam que o modelo GPR tem os melhores recursos para estimar T_c , apresentando os menores valores de erro (RMSE, MSE e MAE) e um melhor R^2 . O RMSE do modelo GPR de 2,7216 no conjunto de treinamento e de 2,651 no conjunto de teste nos diz que a magnitude média do erro do modelo em estimar T_c é de 2,7216°C e 2,651°C nos conjuntos de treinamento e teste, respectivamente. O R^2 de 0,87 para o grupo de treinamento e 0,88 para o grupo de teste, indica que o modelo GPR pode prever 87% e 88% da variância de T_c nos grupos de treinamento e teste respectivamente. A Figura 11 mostra os valores previstos pelo modelo GPR e os valores medidos de T_c ao longo dos cinco primeiros dias do conjunto de teste, pode-se observar que o modelo GPR segue os padrões de mudança de T_c ao longo do intervalo de horas estudado, de 6:00 as 17:00 horas.

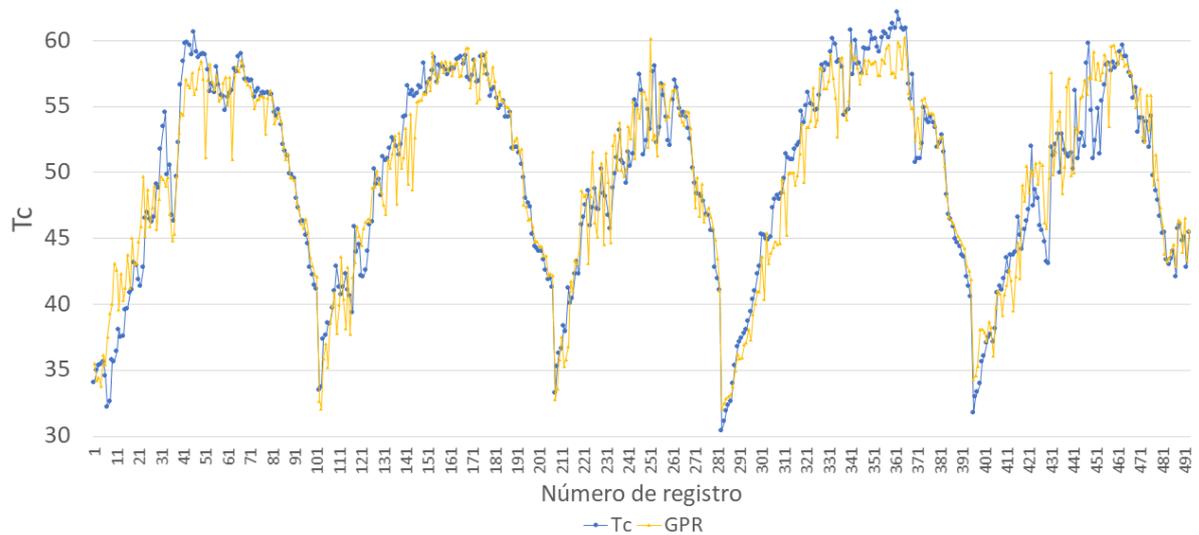
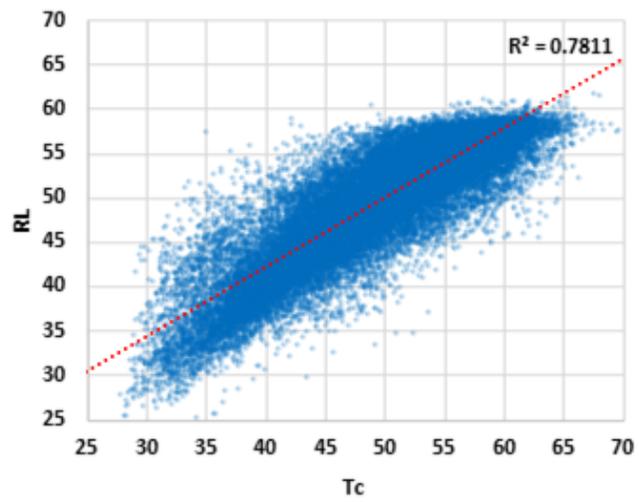
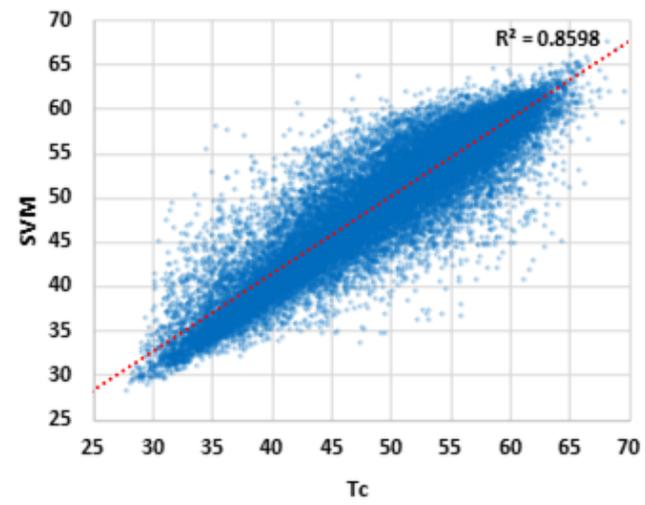


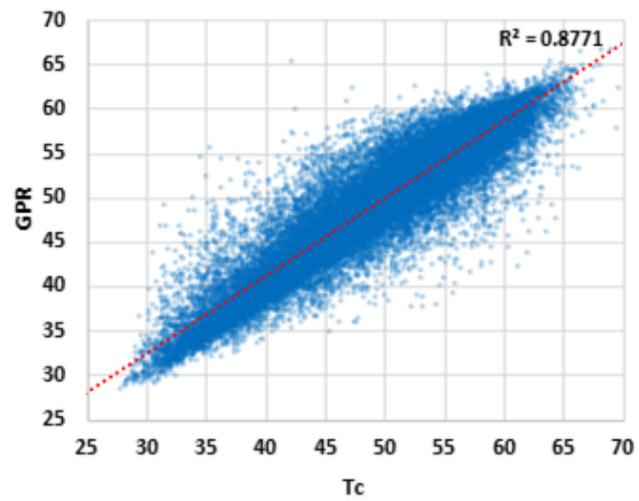
Figura 11 – Valores de T_c previstos pelo modelo GPR e medidos ao longo de cinco dias vs. o número de registro.



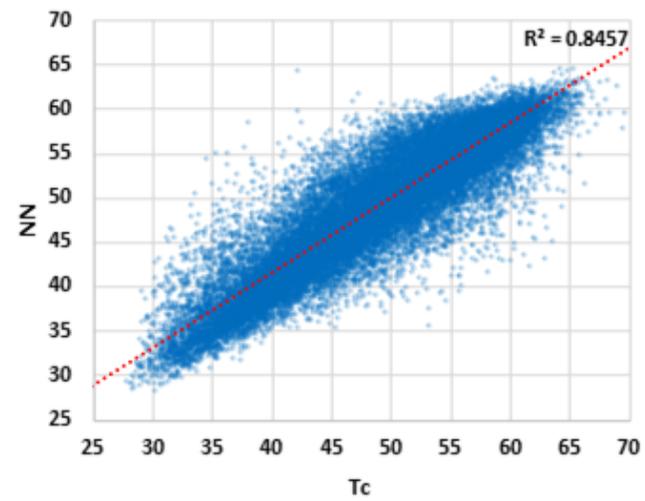
(a)



(b)

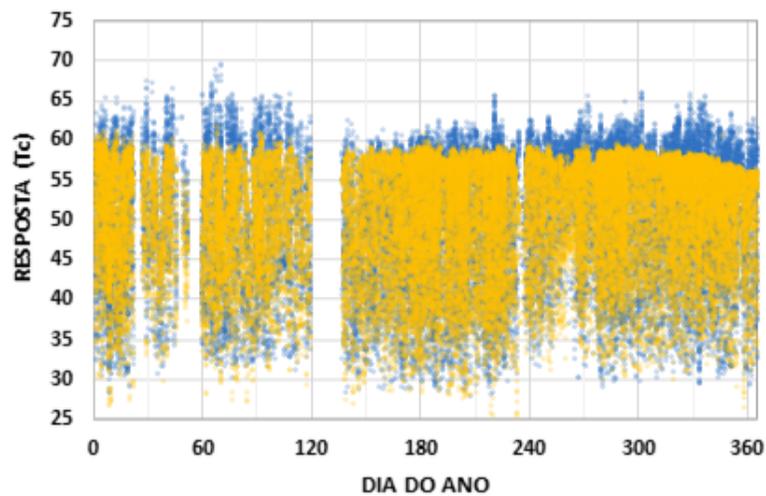


(c)

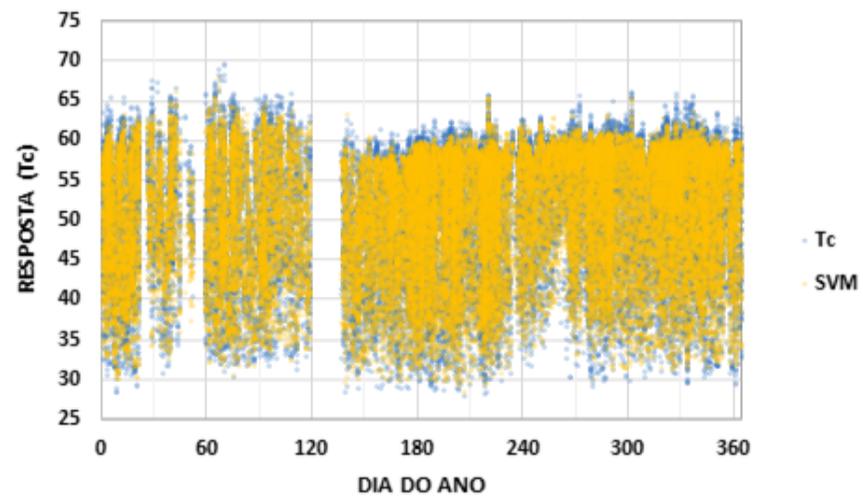


(d)

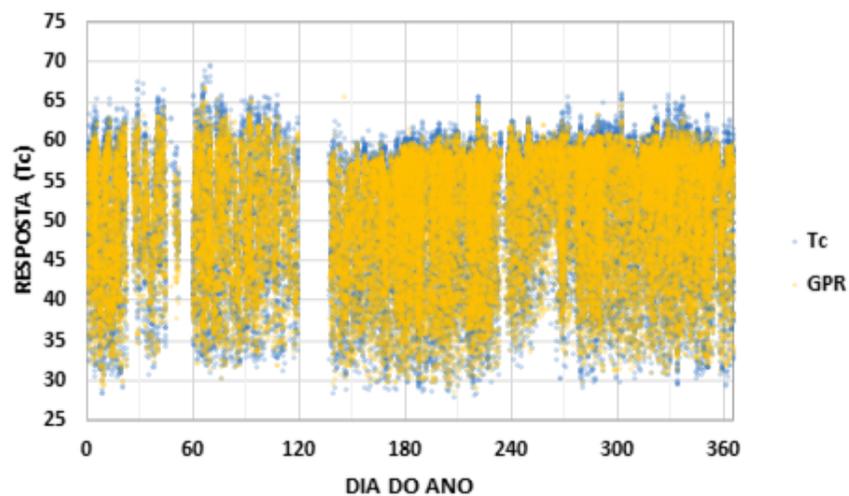
Figura 12 – Gráficos Predito vs. Real: (a) modelo RL (b) modelo SVM (c) modelo GPR (d) modelo RN.



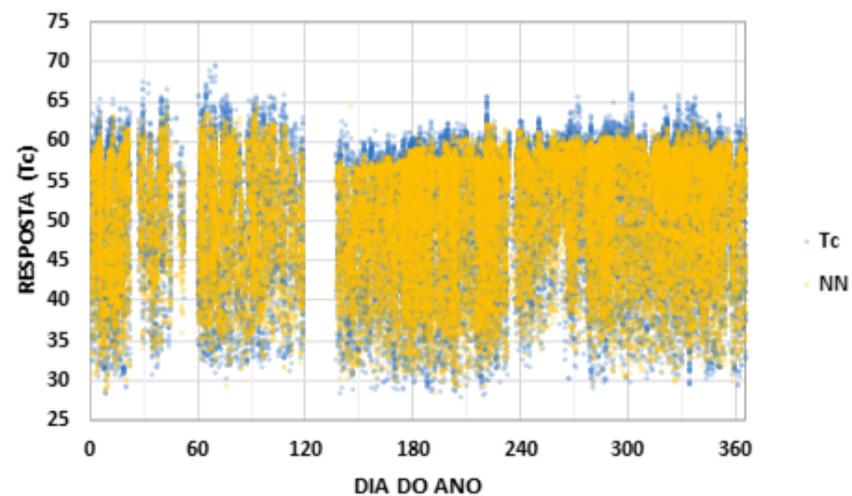
(a)



(b)

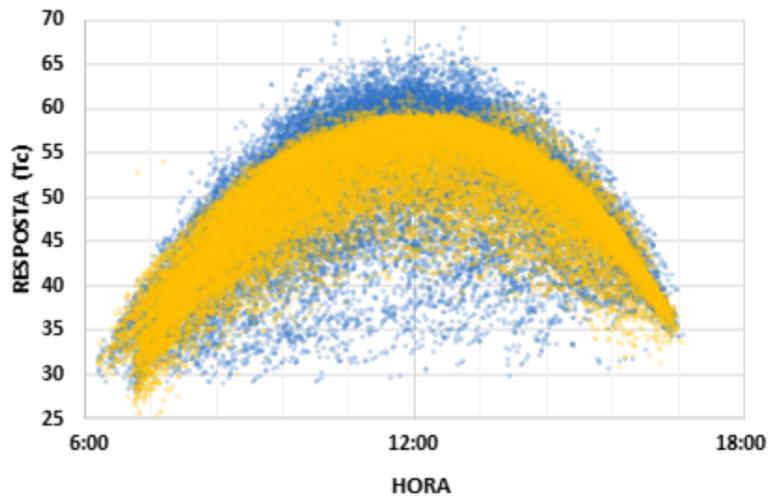


(c)

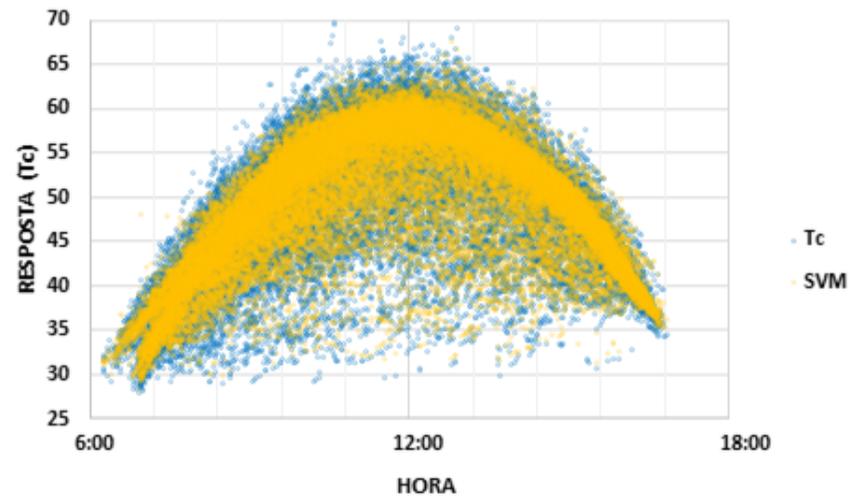


(d)

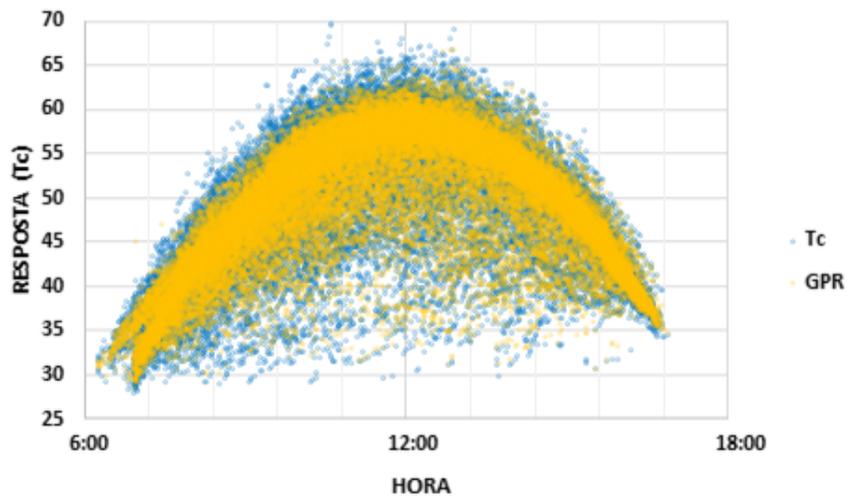
Figura 13 – Resposta prevista em amarelo sob os valores do conjunto de teste ao longo do ano: (a) modelo RL (b) modelo SVM (c) modelo GPR (d) modelo RN.



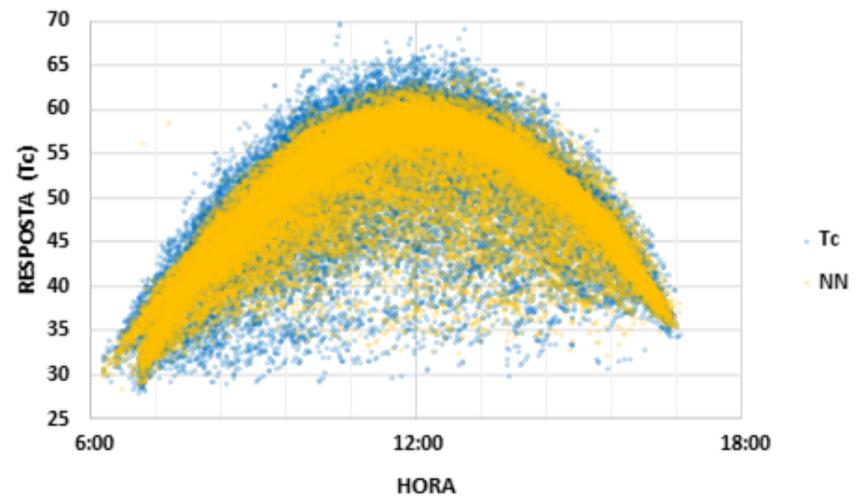
(a)



(b)



(c)



(d)

Figura 14 – Resposta prevista em amarelo sob os valores do conjunto de teste ao longo do dia: (a) modelo RL (b) modelo SVM (c) modelo GPR (d) modelo RN.

5.3.1 Modelos Comparativos

A precisão do melhor modelo proposto, o modelo GPR, também é equiparada com os resultados dos modelos comparativos, modelos convencionais de previsão de T_c presentes na literatura, conforme apresentado na Tabela 14. Para isso são aplicados os dados de teste aos modelos Akyuz *et al.*, Coskun, Standard, NOCT-1p, NOCT-2p e Sandia expressos na Tabela 2.

Tabela 14 – Resultados do modelos GPR e modelos comparativos

Modelo	RMSE	MSE	R ²	MAE
GPR	2,651	7,028	0,88	1,787
Standard	6,325	40,006	0,57	4,676
NOCT-1P	6,391	40,840	0,53	4,713
Sandia	6,454	41,649	0,56	4,783
Akyuz et al	7,202	51,868	0,55	5,649
NOCT-2P	7,701	59,312	0,57	6,249
Coskun	8,773	76,958	0,45	7,4670

Fonte: Elaborado pelo autor.

De acordo com a tabela, o modelo GPR proposto fornece resultados mais precisos do que os modelos convencionais previamente propostos para os dados do LEA-UFC apresentando os menores valores de erro (RMSE, MSE e MAE) e um melhor R². O RMSE do modelo GPR de 2,651 no conjunto de teste nos diz que a magnitude média do erro do modelo em estimar T_c é de 2,651°C. Dentre os seis modelos comparativos, o que apresentou melhores resultados foi o modelos Standard, com um RMSE de 6,325, ou seja, uma apresenta uma magnitude média do erro na T_c de 6,325°C, mais que o dobro do valor apresentado pelo modelo GPR desenvolvido. A Figura 15 mostra os valores previstos pelo modelo GPR e pelos modelos comparativos e os valores medidos de T_c ao longo dos dois primeiros dias do conjunto de teste, pode-se observar que o modelo GPR segue os padrões de variação de T_c próximo a linha dos valores medidos.

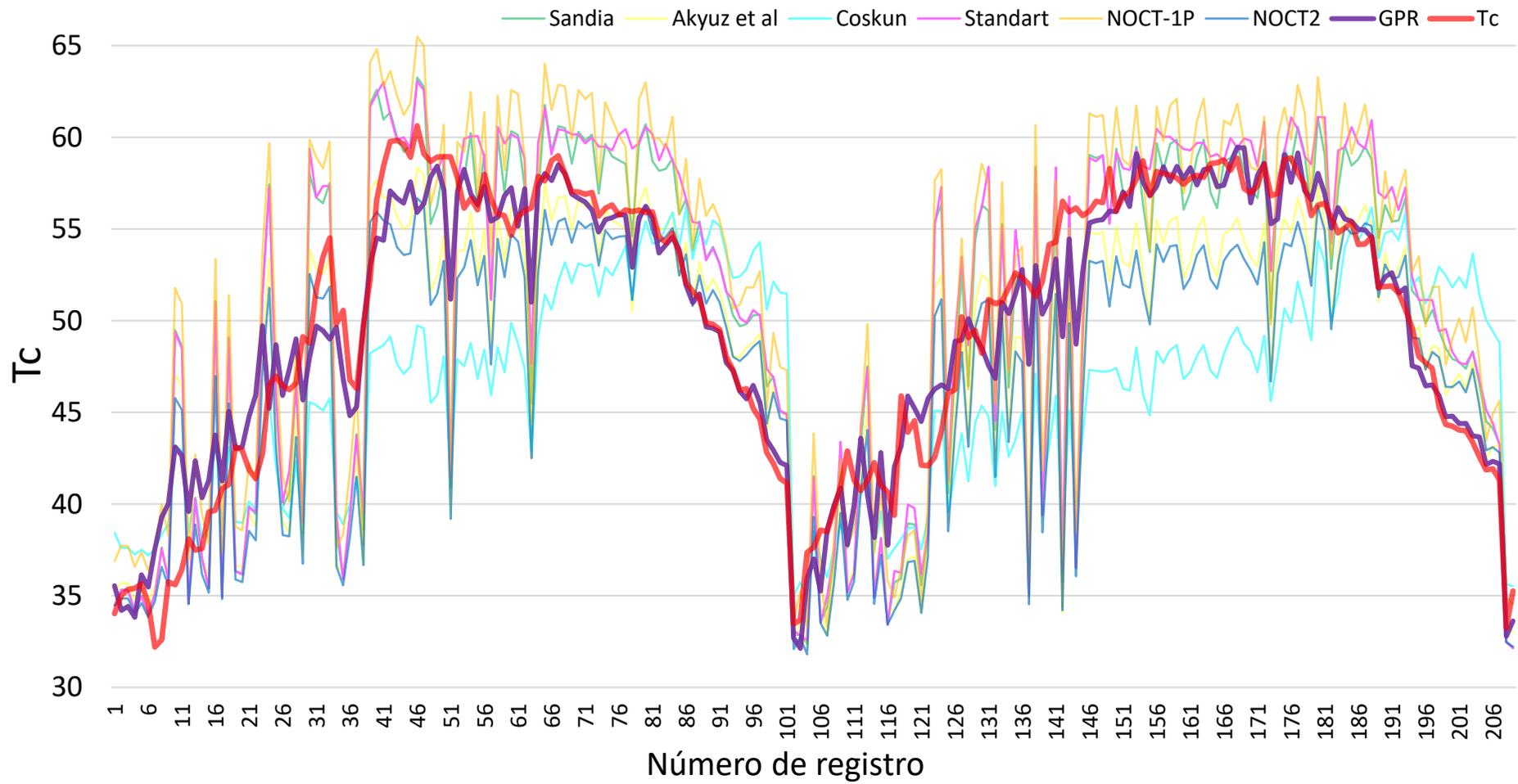


Figura 15 – Valores de T_c previstos pelo modelos comparativos e GPR e medidos ao longo de dois dias vs. o número de registro.

6 CONCLUSÃO

A presente pesquisa busca estudar modelos para determinação de T_c presentes na literatura e desenvolver um modelo baseado em técnicas de ML aplicadas aos dados coletados no LEA-UFC na cidade de Fortaleza, região Nordeste do Brasil. Foram estudados os componentes dos modelos de previsão de T_c presentes na literatura. Dentre os fatores que influenciam T_c foram destacados aqueles que exercem maior influência, visto que aparecem em maior número de correlações. Considerando as 33 correlações analisadas, os componentes $G; T_a; V_w; \eta$; absorção solar; a transmitância da camada de vidro são as mais frequentes, aparecendo em 100%, 93,9%, 54,5%, 33,3%, 30,3% e 24,2% das correlações, respectivamente. Esses componentes são agrupados em variáveis meteorológicas e propriedades dependentes do material e/ou configuração do sistema. É feita uma análise de cada um desses componentes, incluindo melhores práticas e dificuldades técnicas de aquisição de dados. Esta fase do estudo tem a finalidade de levantar as melhores práticas de escolha de componentes dos modelos e arranjos experimentais mais adequados para compor um modelo preciso.

A presente dissertação foi dividida em duas etapas, uma voltada ao estudo dos modelos presentes na literatura e outra à preparação de dados e aplicação das técnicas de ML. Inicialmente, na etapa 1, são apresentados os resultados referentes à revisão de modelos recentes de T_c presentes na literatura. São estudadas 33 equações encontradas na literatura para estimar T_c , que foram classificadas em 3 grupos, um linear e dois não lineares. A partir da classificação são propostas 3 formas gerais para determinação de T_c , cada uma capaz de descrever todas as equações de cada um dos 3 grupos aqui classificados, que podem convenientemente representar a maioria dos modelos estudados de acordo com seus parâmetros. Como possível aplicação, as formas gerais propostas podem auxiliar no desenvolvimento de algoritmos de ajuste contínuo e automático de modelos ao longo do tempo, baseados em ML. Além disso, foram selecionados entre os modelos estudados, modelos possíveis de serem testados no conjunto de dados do LEA, sendo estes denominados modelos comparativos: Akyuz et al.; Coskun; Standard; NOCT-1p; NOCT-2p; e Sandia. Tais modelos são utilizados para comparação com os modelos desenvolvidos na etapa 2 deste trabalho.

Na etapa 2 os dados coletados entre 2018 e 2020 foram tratados e foi feita a aplicação nos modelos RL, SVM, GPR e RN. O modelo RL é o modelo com interações entre os termos e termos quadráticos, apresenta um RMSE de 3,5782 e 3,5377 nos conjuntos de treinamento e teste, respectivamente. O modelo SVM é o modelo do tipo ‘*Fine Gaussian*’, apresenta um RMSE de

2,9041 e 2,8374 nos conjuntos de treinamento e teste, respectivamente. O modelo GPR se trata de um modelo exponencial, apresentando um RMSE de 2,7216 e 2,651 nos conjuntos de treinamento e teste, respectivamente. E o modelo RN se trata de um modelo do tipo '*Wide RN*', apresentando um RMSE de 3,0004 e 2,9706 nos conjuntos de treinamento e teste, respectivamente. Sendo assim, o modelo GPR foi escolhido como o melhor modelo de ML desenvolvido na presente dissertação com a magnitude média do erro do modelo em estimar T_c de 2,7216°C e 2,651°C nos conjuntos de treinamento e teste, respectivamente.

Ainda na etapa 2, o modelo GPR é comparado com os modelos comparativos aplicados aos dados do conjunto de teste do LEA. O RMSE nestes modelos teve valores superiores ao do modelo GPR (RMSE = 2,652), com resultados de 6,325; 6,391; 6,454; 7,202; 7,701; e 8,773 nos modelos Standard, NOCT-1p, Sandia, Akyuz et al., NOCT-2p e Coskun, respectivamente. Dentre os seis modelos comparativos, o que apresentou melhores resultados foi o modelo Standard, com um RMSE de 6,325, ou seja, apresenta uma magnitude média do erro na T_c de 6,325°C, mais que o dobro do valor apresentado pelo modelo GPR desenvolvido. Fica demonstrado assim um melhor poder preditivo do modelo GPR desenvolvido com relação aos demais quando aplicado aos dados coletados no LEA-UFC. O desenvolvimento desse modelo pode trazer contribuições para a projeção de geração de energia elétrica por meio da energia solar e em estudos para a previsão de séries temporais.

Fica como principal sugestão para futuros trabalhos, o teste das 3 correlações gerais desenvolvidas na presente dissertação. As correlações são capazes de descrever diversas equações de previsão de T_c desenvolvidas em diferentes partes do mundo, mas ainda é necessário um estudo do seu desempenho e a determinação de seus termos. Também é tópico interessante para ser desenvolvido em futuros trabalhos a testagem do modelo GPR, que apresentou bons resultados em dados obtidos no LEA-UFC, em dados obtidos em outras regiões a fim de alcançar uma generalização do método.

7 PUBLICAÇÕES ATÉ A PRESENTE DATA

1. **SANTOS, L. O.**; CARVALHO, P. C. M.; CARVALHO FILHO, C. O., "Photovoltaic Cell Operating Temperature Models: A Review of Correlations and Parameters,"in IEEE Journal of Photovoltaics, doi: 10.1109/JPHOTOV.2021.3113156. (Early Access).
2. Ana Patricia Fontenele Barros; **Leticia de Oliveira Santos**; Renata Imaculada Soares Pereira; Paulo Carvalho. Comparison of temperature forecasting models of photovoltaic modules operating under tropical climate. In: **International Congress of Mechanical Engineering, COBEM 2021**. [S. l.:s. n.], 2021. COBEM2021-2039.
3. Eliezer da Silva Holanda Neto; **Leticia de Oliveira Santos**; Carla Freitas de Andrade; Paulo Carvalho; Paulo Alexandre Costa Rocha. The influence of wind speed on the PV plant operating temperature and performance: A case study for Fortaleza. In: **International Congress of Mechanical Engineering, COBEM 2021**. [S. l.:s. n.], 2021. COBEM2021-1663.
4. BUSSON, B. O.; **SANTOS, L. O.**; CARVALHO, P. C. M.; CARVALHO FILHO, C. O., Experimental Assessment and Modeling of a Floating Photovoltaic Module with Heat Bridges. **IEEE Latin America Transactions**. Fator de Impacto(2020 JCR): 0,7290, v.19, p.2079 - 2086, 2021. doi: 10.1109/TLA.2021.9480150.
5. **SANTOS, L. O.**; CARVALHO, P. C. M.; CARVALHO FILHO, C. O. Modelos para a temperatura de operação de módulos fotovoltaicos: Uma revisão das correlações e variáveis pertinentes. In: **Anais do VIII Congresso Brasileiro de Energia Solar 2020**. [S. l.:s. n.], 2020. v. 1.
6. FREITAS, B.B.; TELES, C.F.; LINS, D.R.; LOAYZA, D.M.J; RAMALHO, D.C.L; QUEIROZ, D.M.B; FREITAS, D.M.; SILVA, F.E.M.; SILVA, F.J.R.; PALACIO, G.B.A; RIBEIRO, J.C.C; UCHÔA, K.M.; **SANTOS, L. O.**; *et al.*. Cenários para a matriz de geração de eletricidade do Ceará em 2050. **CIÊNCIA E NATUREA**. , v.42, p.e63 - , 2020.

REFERÊNCIAS

- ABIOLA-OGEDENGBE, A.; HANGAN, H.; SIDDIQUI, K. Experimental investigation of wind effects on a standalone photovoltaic (PV) module. **Renew. Energy**, Elsevier, v. 78, p. 657–665, 2015.
- ABSOLAR. **Energia solar fotovoltaica no Brasil: Infográfico ABSOLAR**. 2021. Atualizado em 01/10/2021, no. 36. Disponível em: <https://www.absolar.org.br/mercado/infografico/>.
- AKYUZ, E.; COSKUN, C.; OKTAY, Z.; DINCER, I. A novel approach for estimation of photovoltaic exergy efficiency. **Energy**, Elsevier, v. 44, n. 1, p. 1059–1066, 2012.
- AL-ADDOUS, M.; DALALA, Z.; CLASS, C. B.; ALAWNEH, F.; AL-TAANI, H. Performance analysis of off-grid PV systems in the Jordan Valley. **Renew. Energy**, Elsevier, v. 113, p. 930–941, 2017.
- AL-TAANI, H.; ARABASI, S. Solar irradiance measurements using smart devices: A cost-effective technique for estimation of solar irradiance for sustainable energy systems. **Sustainability**, Multidisciplinary Digital Publishing Institute, v. 10, n. 2, p. 508, 2018.
- ALMONACID, F.; PÉREZ-HIGUERAS, P. J.; FERNÁNDEZ, E. F.; RODRIGO, P. Relation between the cell temperature of a HCPV module and atmospheric parameters. **Sol. Energy Mater. Sol. Cells**, Elsevier, v. 105, p. 322–327, 2012.
- ANON. **National Solar Radiation Database, 1961-1990, Vol.2**. NSRDB, 1995. Disponível em: <https://nsrdb.nrel.gov/>.
- ARAÚJO, D. N.; CARVALHO, P. C. M. de; DUPONT, I. M. Efeitos da acumulação de sujeira sobre o desempenho de módulos fotovoltaicos. **Rev. Tecnol.**, v. 40, n. 2, 2019.
- ASTM E1036M-96e2. **Method for determining the nominal operating temperature (NOCT) of an array or module, Annex A.1. p.544**. ASTM International, West Conshohocken, PA, 1996. Disponível em: www.astm.org.
- BARANGE, A.; SHARMA, V. Maximum Power Point Tracking for Photovoltaic Systems. **Int. Res. J. Eng. Technol.**, v. 6, p. 604–608, 2019.
- BARDHI, M.; GRANDI, G.; TINA, G. M. Comparison of PV cell temperature estimation by different thermal power exchange calculation methods. In: **Proc. Int. Conf. Renew. Energies Power Qual. (ICREPQ'12), Santiago Compost. Spain**. [S. l.: s. n.], 2012. p. 28–30.
- BURESH, M. Photovoltaic Energy Systems, Design and Installation (ed. MacGraw Hill, New York) 1983. Barra, L., Catalanotti, S., Fontana, F. and Lavorante, F. **Sol. Energy**, v. 33, p. 509, 1984.
- CAMARGO SCHUBERT ENGENHEIROS ASSOCIADOS et al. **Atlas Eólico e Solar: Ceará**. [S. l.]: Curitiba: Camargo Schubert; Fortaleza: ADECE, FIEC, SEBRAE, 2019.
- CASTAÑER, L.; BERMEJO, S.; MARKVART, T.; FRAGAKI, K. Energy production by a PV array. In: **Pract. Handb. Photovoltaics**. [S. l.]: Elsevier, 2003. p. 517–529.
- CHAVES, M. P.; DUPONT, I. M.; CARVALHO, P. C. M.; ARAUJO, D. N. Estudo sobre sombreamento em planta fotovoltaica localizada em zona urbana de Fortaleza-CE. **Rev. Tecnol.**, v. 40, n. 1, 2019.

- CHENNI, R.; MAKHLOUF, M.; KERBACHE, T.; BOUZID, A. A detailed modeling method for photovoltaic cells. **Energy**, Elsevier, v. 32, n. 9, p. 1724–1730, 2007.
- CHOW, T. T. Performance analysis of photovoltaic-thermal collector by explicit dynamic model. **Sol. Energy**, Elsevier, v. 75, n. 2, p. 143–152, 2003.
- COSKUN, C.; TOYGAR, U.; SARPDAG, O.; OKTAY, Z. Sensitivity analysis of implicit correlations for photovoltaic module temperature: a review. **J. Clean. Prod.**, Elsevier, v. 164, p. 1474–1485, 2017.
- CUCE, E.; CUCE, P. M.; KARAKAS, I. H.; BALI, T. An accurate model for photovoltaic (PV) modules to determine electrical characteristics and thermodynamic performance parameters. **Energy Convers. Manag.**, Elsevier, v. 146, p. 205–216, 2017.
- DAVIS, M. W.; FANNEY, A. H.; DOUGHERTY, B. P. Prediction of building integrated photovoltaic cell temperatures. **J. Sol. Energy Eng.**, v. 123, n. 3, p. 200–210, 2001.
- DENOIX, T.; SECHILARIU, M.; LOCMENT, F. Experimental comparison of photovoltaic panel operating cell temperature models. In: IEEE. **IECON 2014-40th Annu. Conf. IEEE Ind. Electron. Soc.** [S. l.], 2014. p. 2089–2095.
- DOWS, R. N.; GOUGH, E. J. **PVUSA procurement, acceptance, and rating practices for photovoltaic power plants.** [S. l.], 1995.
- DUBEY, S.; SARVAIYA, J. N.; SESHADRI, B. Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world—a review. **Energy Procedia**, Elsevier, v. 33, p. 311–321, 2013.
- DUFFIE, J. A.; BECKMAN, W. A. **Solar engineering of thermal processes, Fourth Edition.** [S. l.]: Wiley New York, 2013.
- DUPRÉ, O.; VAILLON, R.; GREEN, M. A. Temperature coefficients of photovoltaic devices. In: **Therm. Behav. Photovolt. Devices.** [S. l.]: Springer, 2017. p. 29–74.
- DURISCH, W.; BITNAR, B.; MAYOR, J.; KIESS, H.; LAM, K.; CLOSE, J. Efficiency model for photovoltaic modules and demonstration of its application to energy yield estimation. **Sol. energy Mater. Sol. cells**, Elsevier, v. 91, n. 1, p. 79–84, 2007.
- EICKER, U. **Solar technologies for buildings.** [S. l.]: John Wiley & Sons, 2006.
- FAIMAN, D. Assessing the outdoor operating temperature of photovoltaic modules. **Prog. Photovoltaics Res. Appl.**, Wiley Online Library, v. 16, n. 4, p. 307–315, 2008.
- FERNÁNDEZ, E. F.; SIEFER, G.; ALMONACID, F.; LOUREIRO, A. J. G.; PÉREZ-HIGUERAS, P. A two subcell equivalent solar cell model for III–V triple junction solar cells under spectrum and temperature variations. **Sol. Energy**, Elsevier, v. 92, p. 221–229, 2013.
- FINAMORE, A.; CALDERARO, V.; GALDI, V.; PICCOLO, A.; CONIO, G. A Day-ahead Wind Speed Prediction based on Meteorological Data and the Seasonality of Weather Fronts. In: IEEE. **2019 IEEE PES GTD Gd. Int. Conf. Expo. Asia (GTD Asia).** [S. l.], 2019. p. 915–920.

GALLARDO-SAAVEDRA, S.; HERNÁNDEZ-CALLEJO, L.; DUQUE-PEREZ, O. Technological review of the instrumentation used in aerial thermographic inspection of photovoltaic plants. **Renew. Sustain. Energy Rev.**, Elsevier, v. 93, p. 566–579, 2018.

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

GIGONI, L.; BETTI, A.; CRISOSTOMI, E.; FRANCO, A.; TUCCI, M.; BIZZARRI, F.; MUCCI, D. Day-ahead hourly forecasting of power generation from photovoltaic plants. **IEEE Trans. Sustain. Energy**, IEEE, v. 9, n. 2, p. 831–842, 2017.

GRIFFITH, J. S.; RATHOD, M. S.; PASLASKI, J. Some tests of flat plate photovoltaic module cell temperatures in simulated field conditions. **pvsp**, p. 822–830, 1981.

HERSCH, P.; ZWEIBEL, K. **Basic photovoltaic principles and methods**. [S. l.], 1982.

HORNUNG, T.; STEINER, M.; NITZ, P. Estimation of the influence of Fresnel lens temperature on energy generation of a concentrator photovoltaic system. **Sol. Energy Mater. Sol. Cells**, Elsevier, v. 99, p. 333–338, 2012.

HOVE, T. A method for predicting long-term average performance of photovoltaic systems. **Renew. Energy**, Elsevier, v. 21, n. 2, p. 207–229, 2000.

IEA. **Key World Energy Statistics 2021**. 2021. IEA, Paris. Disponível em: <https://www.iea.org/reports/key-world-energy-statistics-2021>.

IEC. **Photovoltaic devices – Part 5: Determination of the equivalent cell temperature (ECT) of photovoltaic (PV) devices by the open-circuit voltage method (IEC 60904-5:2011 (EQV))**. International Electrotechnical Commission, 2011. Disponível em: <https://webstore.iec.ch/publication/3877>.

INCROPERA, F. P.; DEWITT, D. P.; BERGMAN, T. L. **Fundamentos de Transferência de Calor E de Massa**. [S. l.]: Grupo Gen-LTC, 2000.

IZGI, E.; ÖZTOPAL, A.; YERLI, B.; KAYMAK, M. K.; ŞAHIN, A. D. Short–mid-term solar power prediction by using artificial neural networks. **Solar Energy**, Elsevier, v. 86, n. 2, p. 725–733, 2012.

JAKHRANI, A. Q.; OTHMAN, A.; RIGIT, A. R. H.; SAMO, S. R. Determination and comparison of different photovoltaic module temperature models for Kuching, Sarawak. In: IEEE. **2011 IEEE Conf. Clean Energy Technol.** [S. l.], 2011. p. 231–236.

JU, X.; VOSSIER, A.; WANG, Z.; DOLLET, A.; FLAMANT, G. An improved temperature estimation method for solar cells operating at high concentrations. **Sol. Energy**, Elsevier, v. 93, p. 80–89, 2013.

KAWAJIRI, K.; OOZEKI, T.; GENCHI, Y. Effect of temperature on PV potential in the world. **Environ. Sci. Technol.**, ACS Publications, v. 45, n. 20, p. 9030–9035, 2011.

KING, D. L.; KRATOCHVIL, J. A.; BOYSON, W. E. Temperature coefficients for PV modules and arrays: measurement methods, difficulties, and results. In: IEEE. **Conf. Rec. twenty sixth IEEE Photovolt. Spec. Conf.** [S. l.], 1997. p. 1183–1186.

KING, D. L.; KRATOCHVIL, J. A.; BOYSON, W. E. **Photovoltaic array performance model**. [S. l.]: United States. Department of Energy, 2004.

KOEHL, M.; HECK, M.; WIESMEIER, S.; WIRTH, J. Modeling of the nominal operating cell temperature based on outdoor weathering. **Sol. Energy Mater. Sol. Cells**, Elsevier, v. 95, n. 7, p. 1638–1646, 2011.

KRAUTER, S. C. W. Development of an integrated solar home system. **Sol. energy Mater. Sol. cells**, Elsevier, v. 82, n. 1-2, p. 119–130, 2004.

KUITCHE, J.; OH, J.; BRUNGER, A.; INOUE, T.; MULLER, M.; BAUERDICK, C.; ALTHAUS, J.; KIEHN, S.; FENG, V.; THERHAAG, U.; OTHERS. One year NOCT Round-Robin testing per IEC 61215 standard. In: IEEE. **2011 37th IEEE Photovolt. Spec. Conf.** [S. l.], 2011. p. 2380–2385.

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 Photovolt. Spec. Conf.** [S. l.], 2009. p. 2399–2404.

LIAO, W.; HEO, Y.; XU, S. Simplified vector-based model tailored for urban-scale prediction of solar irradiance. **Sol. Energy**, Elsevier, v. 183, p. 566–586, 2019.

MA, W. W.; RASUL, M. G.; LIU, G.; LI, M.; TAN, X. H. Climate change impacts on techno-economic performance of roof PV solar system in Australia. **Renew. Energy**, Elsevier, v. 88, p. 430–438, 2016.

MARION, B.; KROPOSKI, B.; EMERY, K.; Del Cueto, J.; MYERS, D.; OSTERWALD, C. **Validation of a photovoltaic module energy ratings procedure at NREL**. [S. l.], 1999.

MARKVART, T. **Solar electricity**. [S. l.]: John Wiley & Sons, 2000. v. 6.

MATHWORKS. **Choose Regression Model Options**. 2020. Disponível em: <https://www.mathworks.com/help/stats/choose-regression-model-options.html>.

MATHWORKS. **Regression Learner App**. 2020. Disponível em: <https://www.mathworks.com/help/stats/regressionlearner-app.html>.

MATHWORKS. **Fitrnet**. 2021. Disponível em: https://nl.mathworks.com/help/stats/fitrnet.html#mw_9c45a4fd-114b-4724-8dd9-fa78079cf2d9.

MATHWORKS. **Gaussian Process Regression Models**. 2021. Disponível em: <https://nl.mathworks.com/help/stats/gaussian-process-regression-models.html>.

MATHWORKS. **Linear regression**. 2021. Disponível em: https://nl.mathworks.com/help/stats/linear-regression-model-workflow.html?searchHighlight=linear%20regression&s_tid=srchtitle.

MATHWORKS. **Understanding Support Vector Machine Regression**. 2021. Disponível em: <https://nl.mathworks.com/help/stats/understanding-support-vector-machine-regression.html#buyrzay>.

MATHWORKS. **What is linear regression?** 2021. Disponível em: https://nl.mathworks.com/discovery/linear-regression.html?s_tid=srchtitle.

MATTEI, M.; NOTTON, G.; CRISTOFARI, C.; MUSELLI, M.; POGGI, P. Calculation of the polycrystalline PV module temperature using a simple method of energy balance. **Renew. energy**, Elsevier, v. 31, n. 4, p. 553–567, 2006.

METEOTEST, M. **Global Meteorological Database for Solar Energy and Applied Meteorology**. 2003.

MITCHELL, T. M. *et al.* **Machine learning**. 1997. [S. l.]: Burr Ridge, IL: McGraw Hill, 45(37):870–877, 1997.

MOHAMMED, H.; GUPTA, R.; SASTRY, O.; MAGARE, D. Assessment of different correlations to estimate distinct technology PV module operating temperature for Indian site. **Energy Sci. Eng.**, Wiley Online Library, v. 7, n. 3, p. 1032–1041, 2019.

MONDOL, J. D.; YOHANIS, Y. G.; NORTON, B. The effect of low insolation conditions and inverter oversizing on the long-term performance of a grid-connected photovoltaic system. **Prog. Photovoltaics Res. Appl.**, Wiley Online Library, v. 15, n. 4, p. 353–368, 2007.

MONDOL, J. D.; YOHANIS, Y. G.; SMYTH, M.; NORTON, B. Long-term validated simulation of a building integrated photovoltaic system. **Sol. Energy**, Elsevier, v. 78, n. 2, p. 163–176, 2005.

MORA, S. P.; CARRETERO, J.; CARDONA, M. Sidrach-de. Models to predict the operating temperature of different photovoltaic modules in outdoor conditions. **Prog. Photovoltaics Res. Appl.**, Wiley Online Library, v. 23, n. 10, p. 1267–1282, 2015.

MORA, S. P.; PILIOUGINE, M.; CARRETERO, J.; NOFUENTES, G.; CARDONA, M. Sidrach-de. Empirical model to predict the operating temperature of the modules of a photovoltaic system. **Int. J. Smart Grid Clean Energy**, 2016.

MULLER, M.; DELINE, C.; MARION, B.; KURTZ, S.; BOSCO, N. Determining outdoor CPV cell temperature. In: AMERICAN INSTITUTE OF PHYSICS. **AIP Conf. Proc.** [S. l.], 2011. v. 1, p. 331–335.

MUZATHIK, A. M. Photovoltaic modules operating temperature estimation using a simple correlation. **Int. J. Energy Eng.**, World Academic Publishing LTD, v. 4, n. 4, p. 151, 2014.

NISHIOKA, K.; MIYAMURA, K.; OTA, Y.; AKITOMI, M.; CHIBA, Y.; MASUDA, A. Accurate measurement and estimation of solar cell temperature in photovoltaic module operating in real environmental conditions. **Jpn. J. Appl. Phys.**, IOP Publishing, v. 57, n. 8S3, p. 08RG08, 2018.

NORDMANN, T.; CLAVADETSCHER, L. Understanding temperature effects on FV system performance. In: **Proc. third world Conf. Photovolt. energy conversion, Osaka, Japan, May 11–18**. [S. l.: s. n.], 2003. p. 4.

OLATOMIWA, L.; MEKHILEF, S.; SHAMSHIRBAND, S.; MOHAMMADI, K.; PETKOVIĆ, D.; SUDHEER, C. A support vector machine–firefly algorithm-based model for global solar radiation prediction. **Solar Energy**, Elsevier, v. 115, p. 632–644, 2015.

OSPINA, L. M. R. **Modelagem e análise de um coletor fotovoltaico térmico**. Dissertação (Mestrado) – Universidade Federal de Pernambuco, 2016.

OTA, Y.; MASUDA, T.; ARAKI, K.; YAMAGUCHI, M. A mobile multipyranometer array for the assessment of solar irradiance incident on a photovoltaic-powered vehicle. **Sol. Energy**, Elsevier, v. 184, p. 84–90, 2019.

PALYVOS, J. A. A survey of wind convection coefficient correlations for building envelope energy systems' modeling. **Appl. Therm. Eng.**, Elsevier, v. 28, n. 8-9, p. 801–808, 2008.

PANDEY, C. K.; KATIYAR, A. K. Solar radiation: Models and measurement techniques. **J. Energy**, Hindawi Publishing Corporation, v. 2013, 2013.

PEHARZ, G.; RODRIGUEZ, J. P. F.; SIEFER, G.; BETT, A. W. A method for using CPV modules as temperature sensors and its application to rating procedures. **Sol. Energy Mater. Sol. Cells**, Elsevier, v. 95, n. 10, p. 2734–2744, 2011.

PEREIRA, R. I. S. **Rede de monitoramento IoT aplicada em plantas de microgeração fotovoltaica**. [S. l.]: 20 f. Tese (Doutorado em Engenharia Elétrica)-Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2018.

PEREIRA, R. I. S.; CAMBOIM, M. M.; VILLARIM, A. W. R.; SOUZA, C. P.; JUCÁ, S. C. S.; CARVALHO, P. C. M. On harvesting residual thermal energy from photovoltaic module back surface. **AEU-International J. Electron. Commun.**, Elsevier, v. 111, p. 152878, 2019.

PÉREZ-HIGUERAS, P. J.; RODRIGO, P.; FERNÁNDEZ, E. F.; ALMONACID, F.; HONTORIA, L. A simplified method for estimating direct normal solar irradiation from global horizontal irradiation useful for CPV applications. **Renew. Sustain. Energy Rev.**, Elsevier, v. 16, n. 8, p. 5529–5534, 2012.

RAHMAN, M. M.; HASANUZZAMAN, M.; RAHIM, N. A. Effects of various parameters on PV-module power and efficiency. **Energy Convers. Manag.**, Elsevier, v. 103, p. 348–358, 2015.

RODRIGO, P.; FERNÁNDEZ, E. F.; ALMONACID, F.; PÉREZ-HIGUERAS, P. J. Review of methods for the calculation of cell temperature in high concentration photovoltaic modules for electrical characterization. **Renew. Sustain. Energy Rev.**, Elsevier, v. 38, p. 478–488, 2014.

ROMARY, F.; CALDEIRA, A.; JACQUES, S.; SCHELLMANN, A. Thermal modelling to analyze the effect of cell temperature on PV modules energy efficiency. In: IEEE. **Proc. 2011 14th Eur. Conf. Power Electron. Appl.** [S. l.], 2011. p. 1–9.

ROSS, R. G. Interface design considerations for terrestrial solar cell modules. In: IEEE. **Proc. 12th IEEE Photovolt. Spec. Conf. Bat. Rouge, LA, Novemb. 15–18.** [S. l.], 1976. p. 801–6.

RUBIO, F.; MARTINEZ, M.; CORONADO, R.; PACHÓN, J. L.; BANDA, P. Deploying CPV power plants-ISFOC experiences. In: IEEE. **2008 33rd IEEE Photovolt. Spec. Conf.** [S. l.], 2008. p. 1–4.

SALA, G. Cooling of solar cells. In: **Sol. Cells Opt. Photovolt. Conc. ed. Luque, A. Adam Hilger, Bristol.** [S. l.: s. n.], 1989. p. 239—267 Chp. 8.

SALAS, V.; OLIAS, E.; BARRADO, A.; LAZARO, A. Review of the maximum power point tracking algorithms for stand-alone photovoltaic systems. **Sol. energy Mater. Sol. cells**, Elsevier, v. 90, n. 11, p. 1555–1578, 2006.

SCHWINGSHACKL, C.; PETITTA, M.; WAGNER, J. E.; BELLUARDO, G.; MOSER, D.; CASTELLI, M.; ZEBISCH, M.; TETZLAFF, A. Wind effect on PV module temperature: Analysis of different techniques for an accurate estimation. **Energy Procedia**, Elsevier, v. 40, p. 77–86, 2013.

SEAWARD. **Solar survey 100\200R series - The ultimate solar site survey tool**. [S. l.], 2019. Disponível em: <http://www.seaward.com/gb/support/download/371>.

SHAFIEIAN, A.; KHIADANI, M.; NOSRATI, A. Theoretical modelling approaches of heat pipe solar collectors in solar systems: A comprehensive review. **Solar Energy**, Elsevier, v. 193, p. 227—243, 2019.

SIEFER, G.; BETT, A. W. Analysis of temperature coefficients for III–V multi-junction concentrator cells. **Prog. Photovoltaics Res. Appl.**, Wiley Online Library, v. 22, n. 5, p. 515–524, 2014.

SILVA, A. F. G.; ZAPAROLI, E. L.; ANDRADE, C. R. Least Squares Fitting of Computational Fluid Dynamics Results to Measured Vertical Wind Profiles. **J. Sol. Energy Eng.**, American Society of Mechanical Engineers Digital Collection, v. 139, n. 3, 2017.

SINGH, P.; RAVINDRA, N. M. Temperature dependence of solar cell performance—an analysis. **Sol. energy Mater. Sol. cells**, Elsevier, v. 101, p. 36–45, 2012.

SINHA, P.; HAYES, W.; LITTMANN, B.; NGAN, L.; ZNAIDI, R. Environmental variables affecting solar photovoltaic energy generation in Morocco. In: IEEE. **2014 Int. Renew. Sustain. Energy Conf.** [S. l.], 2014. p. 230–234.

SKOPLAKI, E.; BOUDOUVIS, A. G.; PALYVOS, J. A. A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting. **Sol. energy Mater. Sol. cells**, Elsevier, v. 92, n. 11, p. 1393–1402, 2008.

SKOPLAKI, E.; PALYVOS, J. A. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. **Sol. energy**, Elsevier, v. 83, n. 5, p. 614–624, 2009.

SKOPLAKI, E. P. J. A.; PALYVOS, J. A. Operating temperature of photovoltaic modules: A survey of pertinent correlations. **Renew. energy**, Elsevier, v. 34, n. 1, p. 23–29, 2009.

STULTZ, J. W.; WEN, L. C. Thermal performance testing and analysis of photovoltaic modules in natural sunlight. **LSA Task Rep.**, v. 5101, p. 31, 1977.

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

TIWARI, A.; SODHA, M. S.; CHANDRA, A.; JOSHI, J. C. Performance evaluation of photovoltaic thermal solar air collector for composite climate of India. **Sol. energy Mater. Sol. cells**, Elsevier, v. 90, n. 2, p. 175–189, 2006.

TIWARI, G. N. **Solar energy: fundamentals, design, modelling and applications**. [S. l.]: Alpha Science Int'l Ltd., 2002.

TOPI, M.; BRECL, K.; SITES, J. Effective efficiency of PV modules under field conditions. **Prog. Photovoltaics Res. Appl.**, Wiley Online Library, v. 15, n. 1, p. 19–26, 2007.

TRINURUK, P.; SORAPIPATANA, C.; CHENVIDHYA, D. Estimating operating cell temperature of BIPV modules in Thailand. **Renew. energy**, Elsevier, v. 34, n. 11, p. 2515–2523, 2009.

TSELEPIS, S.; TRIPANAGNOSTOPOULOS, Y. Economic analysis of hybrid photovoltaic/thermal solar systems and comparison with standard PV modules. In: **Proc. Int. Conf. PV Eur.** [*S. l.: s. n.*], 2002. v. 7, p. 4.

Vapnik, V. T. **The Nature of Statistical Learning Theory.** [*S. l.*]: Springer, New York, NY, USA, 1995.

WIEDERHOLD, G.; MCCARTHY, J. Arthur samuel: pioneer in machine learning. **IBM Journal of Research and Development**, IBM, v. 36, n. 3, p. 329–331, 1992.

YANDT, M. D.; COOK, J. P. D.; KELLY, M.; SCHRIEMER, H.; HINZER, K. Dynamic real-time I–V curve measurement system for indoor/outdoor characterization of photovoltaic cells and modules. **IEEE J. Photovoltaics**, IEEE, v. 5, p. 337–343, 2014.

YANDT, M. D.; WHEELDON, J. F.; COOK, J.; BEAL, R.; WALKER, A. W.; THÉRIAULT, O.; SCHRIEMER, H.; HALL, T. J.; HINZER, K. Estimating cell temperature in a concentrating photovoltaic system. In: AMERICAN INSTITUTE OF PHYSICS. **AIP Conf. Proc.** [*S. l.*], 2012. v. 1477, p. 172–175.

ZIMMERMAN, E. S.; HARTWICK, L. R. **Ambient temperature monitoring technique.** [*S. l.*]: Google Patents, 1991.