

# OTIMIZAÇÃO DE MODELOS LINEARES PARA ESTIMATIVA DA IRRADIÂNCIA SOLAR GLOBAL POR MEIO DE PRODUTOS DE SENSORIAMENTO REMOTO EM BOTUCATU (SP) - BRASIL

Marcus Vinícius Contes Calça - marcus.calca@unesp.br

Daniele Cristina Lopes Mariano

José Rafael Franco

Matheus Rodrigues Raniero

Carlos Gabriel dos Santos Modesto

Enzo Dal Pai

Alexandre Dal Pai

Universidade Estadual Paulista - UNESP, Faculdade de Ciências Agrônomicas, Botucatu (SP) - Brasil

**Resumo.** É consenso que o conhecimento sobre os níveis de irradiância solar obtidos na superfície terrestre por meio de sensores automáticos representa a situação mais desejável. Porém, a construção de estações solarimétricas apresenta aporte financeiro elevado, o que dificulta a implantação deste cenário em muitas localidades no Brasil. O objetivo deste estudo foi criar três modelos de regressão linear, com combinações de parâmetros de entrada diferentes, para estimar a irradiância solar global horária na superfície terrestre, utilizando informações obtidas pelo Global Land Data Assimilation System 2.1. Foi realizada uma avaliação para determinar qual combinação de variáveis independentes de regressão linear apresentaria o melhor ajuste as medidas coletadas na Faculdade de Ciências Agrônomicas (UNESP) de Botucatu (SP) - Brasil, durante o período de 2020-2022. Foram criados, portanto, em Python três modelos de regressão linear usando a irradiância solar global, fornecida pelo produto de sensoriamento remoto, índice de transmissividade atmosférica e elevação solar, como variáveis independentes, em escala temporal horária. A avaliação dos modelos de regressão linear se deu a partir dos indicadores estatísticos MBE, rMBE, RMSE, rRMSE e  $R^2$ , usando como referência medidas obtidas por um piranômetro na superfície terrestre. A primeira combinação de ajuste linear produziu um  $R^2$  de 0,83, com um rMBE de 18,13% e um rRMSE de 26,63%. Assim como, a segunda combinação de ajuste linear gerou um  $R^2$  de 0,88, com um rMBE de 15,69% e um rRMSE de 26,31%. Por fim, a terceira combinação de ajuste linear apresentou um  $R^2$  de 0,90, com um rMBE de 15,52% e um rRMSE de 20,75%. Usar a irradiância solar fornecida pelo produto de sensoriamento remoto, índice de transmissividade atmosférica e elevação solar, como variáveis independentes, criou o modelo de regressão linear que melhor compreendeu os processos atmosféricos e astronômicos ocorridos, permitindo obter estimativas de irradiância solar na superfície terrestre com maior precisão.

**Palavras-chave:** Modelagem da Radiação Solar. Previsão da Energia Solar. Regressão Linear Multivariada.

## 1. INTRODUÇÃO

O conhecimento sobre os níveis de irradiância solar disponíveis na superfície terrestre é crucial para entidades públicas, instituições acadêmicas e organizações não governamentais criarem adequadamente projetos relacionados a geração de energia elétrica, arquitetura sustentável, produção de biomassa, conforto térmico e investigarem as mudanças climáticas, no Brasil (Pashiardis & Kalogirou, 2016; Pereira *et al.*, 2017). Esse conhecimento pode ser obtido a partir de medições realizadas na superfície terrestre, com ampla cobertura temporal e espacial, por meio de estações de radiometria solar ou meteorológicas. Neste sentido, o fluxo de irradiância solar que atinge um determinado local do planeta Terra, normalmente, é mensurado por meio de piranômetros ou pireliômetros e pode ser classificado como direto/direta, quando incide diretamente do disco solar, difuso/difusa, quando sofre influência dos fenômenos físicos de reflexão ou espalhamento na atmosfera terrestre, e global/total, quando é a soma dos dois anteriores, e representa a densidade total de energia solar incidente sob uma determinada área.

A obtenção de informações sobre os níveis de irradiância solar disponíveis na superfície terrestre pode ocorrer a partir de medições realizadas por sensores automáticos instalados no local de interesse, ou por meio de estimativas, geradas por modelos estatísticos, físicos ou estocásticos. É consenso que o conhecimento sobre o fluxo de irradiância solar obtido a partir de sensores automáticos representa a situação mais desejável. No entanto, o investimento em instrumentação e a capacitação de recursos humanos para sua operacionalização apresenta aporte financeiro elevado, o que dificulta a implantação desse cenário em muitas localidades ao redor do Brasil (Flumignan *et al.*, 2018). Neste sentido, o objetivo deste estudo foi criar três modelos de regressão linear, utilizando três combinações diferentes de parâmetros de entrada, para estimar a irradiância solar global horária disponível na superfície terrestre, a partir de valores fornecidos por um produto de sensoriamento remoto (Global Land Data Assimilation System - GLDAS 2.1), para Botucatu (São Paulo) - Brasil. Foi avaliada, também, qual combinação de regressão linear gerou o melhor ajuste às medidas locais.

## 2. MATERIAL E MÉTODOS

### 2.1 Local de Estudo

Este estudo foi elaborado utilizando as medidas obtidas pelo Laboratório de Agrometeorologia e Radiometria Solar (22°54'S de latitude, 48°27'O de longitude e 786m de altitude) da Faculdade de Ciências Agrônomicas da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP de Botucatu, São Paulo - Brasil. Situado na região centro-sul do estado de São Paulo, com 149 mil habitantes e uma área territorial de 1.482,642 km<sup>2</sup>, Botucatu possui um bioma caracterizado por uma vegetação de Cerrado e Mata Atlântica (IBGE, 2022), com fortes atividades agrícolas, que compreendem grandes plantações de cana-de-açúcar e eucalipto, assim como atividades industriais e comerciais moderadas (Codato *et al.*, 2008; Dal Pai *et al.*, 2016). Sua formação geográfica compreende um relevo de colinas não simétricas caracterizado por dois níveis distintos de altitudes, na faixa de 500 a 400 metros na região mais baixa, e de 700 a 900 metros na região serrana, que é a mais alta. Esta formação geográfica provoca diferenças climáticas consideráveis no município (Silva *et al.*, 2017).

O clima de Botucatu - SP é identificado como Aw, de acordo com a classificação climática de Köppen (Franco, *et al.*, 2023), compreendendo um inverno seco (junho a agosto) e um verão quente e chuvoso (dezembro a fevereiro). De acordo com a série histórica de 1991 a 2020 da Estação Meteorológica do Departamento de Engenharia Rural e Socioeconomia da Faculdade de Ciências Agrônomicas da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP de Botucatu (SP):

- Fevereiro é o mês mais quente, com temperatura média do ar de 23,80 °C.
- Agosto é o mês mais frio, com temperatura média do ar de 18,28 °C.
- Janeiro é o mês mais úmido, com umidade relativa média do ar de 75,76%.
- Agosto é o mês menos úmido, com umidade relativa média do ar de 61,87%.
- Janeiro é o mês mais chuvoso, com uma média pluviométrica de 315mm.
- Agosto é o mês menos chuvoso, com uma média pluviométrica de 38mm.

### 2.2 Aquisição de Medidas de Radiação Solar

#### • Medidas Obtidas na Superfície Terrestre

A irradiância solar global na superfície terrestre foi medida por um Piranômetro da Eppley - PSP, em um plano horizontal não sombreado, instalado em uma área rural com grama verde e curta a 1,5 metros do solo. Um Micrologger CR3000 da Campbell Scientific foi utilizado para coletar, pré-processar e gravar no local as medidas a cada 1 minuto (média das observações obtidas a cada 5 segundos) no período de 2020 a 2022 (Fig. 1). Após a coleta, as medições eram transmitidas para um servidor de dados a partir de um módulo de radiofrequência da Campbell Scientific - RF450, onde eram estocadas no Sistema de Gerenciamento de Banco de Dados Relacional - MySQL, para posterior uso. Uma rotina regular de manutenção foi realizada, no intuito de fornecer as condições ideais de monitoramento da irradiância solar, de acordo com as diretrizes da Organização Mundial de Meteorologia (WMO, 2012). O Piranômetro da Eppley - PSP, usado para obter a irradiância solar global, foi calibrado uma vez por ano, a partir de outro Piranômetro Eppley - PSP de referência, acondicionado de forma íntegra em um ambiente protegido e utilizado somente para esta finalidade.



Figura 1 - Instrumentos da Estação de Radiometria Solar da Faculdade de Ciências Agrônomicas da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP de Botucatu (SP) - Brasil.

Medir a irradiância solar não é uma tarefa fácil, em virtude das incertezas contidas no processo, que afetam diretamente a qualidade dos valores obtidos, como o efeito do cosseno, seletividade espectral, excremento de aves,

nivelamento incorreto da instrumentação, gotas de orvalho e poeira na cúpula do sensor. Portanto, além da consistente rotina de manutenção e calibração dos instrumentos, um método de identificação de outliers (Tab. 1), que são valores anômalos, foi executado nas medidas de irradiância solar, na escala temporal instantânea (1 minuto) e horária (1 hora).

Tabela 1 - Método de identificação de outliers da irradiância solar global em W/m<sup>2</sup> de acordo com Calça *et al.* (2019).

MÉTODOS DE IDENTIFICAÇÃO DE OUTLIERS			
MEDIDAS DE 1 MINUTO		MEDIDAS MÉDIAS DE 1 HORA	
$0 \leq G \leq (1,20 G_o)$	(1)	$0 \leq K_T \leq 1$	(2)

**Legenda:**

G - Irradiância solar global.

G<sub>o</sub> - Irradiância solar extraterrestre.

Neste sentido, o método de identificação de outliers na escala temporal instantânea (1 minuto) em W/m<sup>2</sup>, teve como objetivo identificar valores considerados fisicamente impossíveis, que ocorrem quando a medida da irradiância solar global é menor do que zero ou maior do que a extraterrestre mais 20% do seu valor (Tab. 1 - Eq. 1). Portanto, os valores apontados como fisicamente impossíveis foram removidos do conjunto de dados (Tab. 2), onde o ano de 2022 foi o que apresentou a maior ocorrência (5,70 %), seguido de 2020 (1,87 %) e por último de 2021 (1,13 %).

Tabela 2 - Resultado da identificação de outliers nas medidas em escala temporal instantânea (1 minuto) em W/m<sup>2</sup>.

ANO	IRRADIÂNCIA SOLAR GLOBAL – INSTÂNTANEA (1 MINUTO)		
	QUANTIDADE TOTAL DE MEDIDAS DISPONÍVEIS	OUTLIERS REMOVIDOS	
		TOTAL	PERCENTUAL
2020	259477	4854	1,87 %
2021	258713	2919	1,13 %
2022	249740	14245	5,70 %
TOTAL	767930	22018	2,87 %

Os valores das efemérides geoespaciais (declinação solar, ângulo horário, cosseno do ângulo zenital e excentricidade da órbita terrestre), necessários para a obtenção da irradiância solar extraterrestre, que foi utilizada para identificar os outliers da irradiância solar, foram calculados a partir das equações de IQBAL (1983). Após a remoção das medidas consideradas fisicamente impossíveis na escala temporal instantânea (1 minuto) em W/m<sup>2</sup>, foram obtidos os valores médios da irradiância solar global para o período das 05:00 às 06:00, das 08:00 às 09:00, das 11:00 às 12:00, das 14:00 às 15:00 e das 17:00 às 18:00. Portanto, aplicou-se também o método de identificação de outliers, que neste caso são considerados anomalias quando o valor do índice de transmissividade atmosférica (Tab. 1 - Eq. 2) se encontra fora do intervalo de 0 a 1 nas medidas da irradiância solar em escala temporal horária (1 hora). O índice de transmissividade atmosférica (K<sub>T</sub>), usado também para classificar a cobertura de céu (Tab. 3), foi calculado a partir da razão entre a irradiância solar global medida na superfície terrestre (G) e a extraterrestre (G<sub>o</sub>) calculada (G/G<sub>o</sub>).

Tabela 3 - Intervalos de classificação da cobertura de céu de acordo com Escobedo *et al.* (2009).

CLASSIFICAÇÃO DA COBERTURA DE CÉU	INTERVALO DO ÍNDICE DE TRANSMISSIVIDADE ATMOSFÉRICA
Nublado	$0 \leq K_T < 0,35$
Parcialmente Nublado	$0,35 \leq K_T < 0,55$
Aberto	$0,55 \leq K_T < 0,65$
Parcialmente Aberto	$0,65 \leq K_T < 1$

Ao contrário do método de identificação de outliers aplicado à escala temporal instantânea (1 minuto) em W/m<sup>2</sup>, na escala temporal horária (1 hora) em W/m<sup>2</sup>, o ano de 2021 teve a maior ocorrência (11,96%) de medidas consideradas

fisicamente impossíveis, seguido por 2022 (10,82%) e, por último, por 2020 (9,61%). Consequentemente, assim como aconteceu anteriormente, os valores medidos de irradiância solar considerados outliers foram removidos do conjunto de dados, a fim de proporcionar condições mais favoráveis para o ajuste dos modelos lineares.

Tabela 4 - Resultado da identificação de outliers nas medidas em escala temporal horária (1 hora) em W/m<sup>2</sup>.

ANO	IRRADIÂNCIA SOLAR GLOBAL - HORÁRIA (1 HORA)		
	QUANTIDADE TOTAL DE MEDIDAS DISPONÍVEIS	OUTLIERS REMOVIDOS	
		TOTAL	PERCENTUAL
2020	1373	132	9,61 %
2021	1388	166	11,96 %
2022	1340	145	10,82 %
<b>TOTAL</b>	4101	443	10,80 %

- **Medidas Obtidas no Google Earth Engine**

O Google Earth Engine é uma plataforma de computação em nuvem que oferece gratuitamente um vasto catálogo de informações de sensoriamento remoto na internet. Composto por vários petabytes de dados, que podem ser acessados usando as linguagens de programação Python ou Java Script. Seu catálogo abrange um extenso repositório de observações provenientes de sistemas de imagens aéreas e de satélite, com comprimentos de onda do visível ao não visível, informações ambientais, previsões meteorológicas, cobertura de solo e topográficas (Gorelick *et al.*, 2017).

Os produtos de sensoriamento remoto são uma excelente alternativa para a obtenção de medidas climáticas em locais onde não se tem a instrumentação necessária. Neste sentido, foram extraídas informações apenas de irradiância solar global em escala temporal horária (6:00, 09:00, 12:00, 15:00 e 18:00) do Global Land Data Assimilation System - Versão 2.1 (GLDAS 2.1) do período de 2020 a 2022, a partir do editor de códigos do Google Earth Engine ([bit.ly/gldas-data-extraction](http://bit.ly/gldas-data-extraction)), no intuito de ajustar os valores obtidos em função de medidas coletadas na superfície terrestre.

O Global Land Data Assimilation System - Versão 2.1 é um sistema global que modela dados observacionais de várias fontes (satélites, estações meteorológicas, e outras fontes terrestres) para fornecer informações sobre processos hidrológicos e climáticos na superfície da Terra, como previsão de tempo, monitoramento de secas, modelagem hidrológica, estudos climáticos e pesquisas em ciências ambientais, com uma ampla cobertura espacial e temporal. É um produto de sensoriamento remoto mantido pela NASA (National Aeronautics and Space Administration - Estados Unidos da América), NCAR (National Center for Atmospheric Research - Estados Unidos da América) e HPC (Hydrological Data Processing Center - Estados Unidos da América), que possui altas resoluções espaciais (2,5° a 1 km) e utiliza quatro modelos de superfície terrestre: NOAH, Catchment, Community Land Model e Variable Infiltration Capacity. Sua versão 2.1 tem disponível medidas de evapotranspiração, irradiância solar de ondas curtas, precipitação, temperatura do ar e outras informações climáticas desde janeiro de 2000, a cada 3 horas (NASA, 2021).

### 2.3 Modelos de Regressão de Linear

A regressão linear é um instrumento estatístico que busca estabelecer e modelar uma relação entre uma variável dependente e uma, ou mais, variáveis independentes, por meio de uma equação linear simples (Eq. 3) ou múltipla (Eq. 4). A equação linear é representada por uma reta, que melhor se ajusta aos dados observados, permitindo prever o comportamento da variável dependente a partir de uma, ou mais, informações de entrada. A regressão linear múltipla é uma extensão da simples que permite incluir mais de uma variável independente (Chein, 2019).

$$y = a + bX \quad (3)$$

$$y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (4)$$

Na Eq. 3,  $y$  é a variável dependente,  $X$  é a variável independente,  $a$  e  $b$  são o intercepto e a inclinação da reta de regressão linear, respectivamente. Na regressão linear múltipla,  $X_1$ ,  $X_2$  e  $X_n$  são as variáveis independentes,  $b_1$ ,  $b_2$ , e  $b_n$  são os coeficientes da reta de regressão, que indicam a relação entre cada variável independente e a variável dependente. A inclusão de mais variáveis independentes no modelo de regressão linear pode aprimorar sua capacidade preditiva e explicativa, neste sentido, foram construídos modelos com 3 combinações diferentes neste estudo (Tab. 5).

Tabela 5 - Informações de entrada e saída dos modelos de regressão linear.

MODELO DE REGRESSÃO LINEAR	VARIÁVEIS INDEPENDENTES (INFORMAÇÕES DE ENTRADA)			VARIÁVEL DEPENDENTE (INFORMAÇÃO DE SAÍDA)
Combinação 1	GS			GS-RL
Combinação 2	GS	KT		
Combinação 3	GS	KT	E	

Legenda:

GS - Irradiância solar global obtida pelo GLDAS 2.1.

KT - Índice de transmissividade atmosférica.

E - Elevação solar.

GS-RL - Irradiância solar global ajustada pelo modelo linear.

No ajuste dos modelos de regressão linear para prever melhor a irradiância solar global na superfície terrestre, a decisão de incorporar o índice de transmissividade atmosférica e a elevação solar como variáveis independentes, além dos valores de irradiância solar fornecidos pelo GLDAS 2.1, permite uma melhor compreensão física dos processos atmosféricos e astronômicos. O índice de transmissividade atmosférica representa a capacidade da atmosfera permitir a passagem da radiação solar e a elevação solar, é crucial na determinação da trajetória da radiação solar incidente ao longo do dia, influenciando diretamente na quantidade de energia solar recebida na superfície terrestre. Neste caso, em ambas as combinações propostas foi realizada a divisão aleatória dos dados, onde 80% foram alocados para o treinamento do modelo e os 20% restantes foram reservados para a validação dos resultados. Portanto, das 3658 linhas disponíveis no conjunto de dados, considerando a remoção de 443 outliers dos valores medidos da irradiância solar na superfície terrestre, 2926 foram utilizadas no treinamento e 732 na validação dos modelos de regressão linear.

Os modelos de regressão linear foram criados utilizando a linguagem de programação Python, versão 3.11.4, a partir do ambiente de desenvolvimento integrado da Microsoft (IDE) Visual Studio Code, versão 1.83.1, com uma extensão que habilitou o uso de células interativas de execução de códigos a partir de notebooks do Jupyter. As bibliotecas Numpy, Pandas, Scikit-Learn, Matplotlib e Seaborn foram utilizadas para realizar a transformação dos dados, treinar e avaliar os modelos de regressão linear e definir a plotagem de gráficos de correlação usando mapas de calor, respectivamente.

#### 2.4 Indicativos Estatísticos

Os modelos lineares ajustados foram avaliados quanto à sua capacidade de estimar os valores da irradiância solar global na superfície terrestre, utilizando as métricas disponíveis no framework Scikit-Learn (2022a). O desempenho foi avaliado por meio de métricas estatísticas obtidas a partir do Erro Médio de Viés (MBE), do Erro Quadrático Médio (RMSE) e do Coeficiente de Determinação ( $R^2$ ). O MBE (Eq. 5) quantifica tendências significativas em superestimar ou subestimar os valores de referência. O RMSE (Eq. 7) estima o erro absoluto considerando as diferenças entre os pares de valores avaliados, atribuindo um peso maior a grandes erros. Os valores relativos de MBE (rMBE), obtidos pela Eq. 6, e de RMSE (rRMSE), obtidos pela Eq. 8, expressam o grau de correspondência percentual em relação à média das observações. Assim como o  $R^2$  (Eq. 9), representa a adequação de um modelo estatístico linear generalizado aos valores observados (Rincón *et al.*, 2018). Para o MBE e o RMSE, valores próximos de zero são desejáveis, enquanto para o  $R^2$ , um valor próximo de um é um indicativo de um bom ajuste do modelo.

$$\mathbf{MBE} = \frac{\sum_{i=1}^n (Y'_i - Y_i)}{n} \quad (5)$$

$$\mathbf{rMBE} (\%) = \frac{\mathbf{MBE}}{\underline{Y}} \times 100 \quad (6)$$

$$\mathbf{RMSE} = \left[ \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n} \right] \quad (7)$$

$$\mathbf{rRMSE} = \left( \frac{\mathbf{RMSE}}{\bar{y}} \right) \times 100 \quad (8)$$

$$\mathbf{R}^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

Nas equações 5, 6, 7, 8 e 9 o  $Y'$  representa os valores geoespaciais de irradiância solar global, obtidos pelo GLDAS 2.1 e fornecidos pelo Google Earth Engine, já o  $Y$  faz referência as medições realizadas na superfície terrestre pelo

Laboratório de Agrometeorologia e Radiometria Solar (referência), de modo que, o  $\bar{Y}$  retrata a média dos valores medidos na superfície terrestre (Scikit-Learn, 2023). As métricas de avaliação do ajuste dos modelos de regressão linear foram calculadas utilizando 20% dos dados disponíveis na escala temporal horária (1 hora).

### 3. RESULTADOS E DISCUSSÃO

Primeiramente, as medidas em escala temporal instantânea (1 minuto), da irradiância solar global na superfície terrestre, foram submetidas ao método de identificação e remoção de outliers, considerando possíveis imprecisões no processo de obtenção das medições, que poderiam resultar em valores incorretos. Posteriormente, os valores foram agregados para a escala temporal horária (1 hora). No entanto, uma análise gráfica exploratória revelou a necessidade de também remover os outliers das medidas em escala horária (1 hora), devido a pontos de dados que se destacavam significativamente dos demais, na série temporal (Fig. 1a). Após a eliminação dessas medidas anômalas, identificou-se que os valores máximos de irradiância solar global na superfície terrestre variam de 613,84 W/m<sup>2</sup> no inverno a 1087,29 W/m<sup>2</sup> no verão, no horário de maior incidência solar (Fig. 1b). De maneira semelhante, para a irradiância solar global fornecida pelo GLDAS 2.1, os valores máximos variam de 612,91 W/m<sup>2</sup> no inverno a 970,79 W/m<sup>2</sup> no verão, também no horário de maior incidência solar (Fig. 1c).

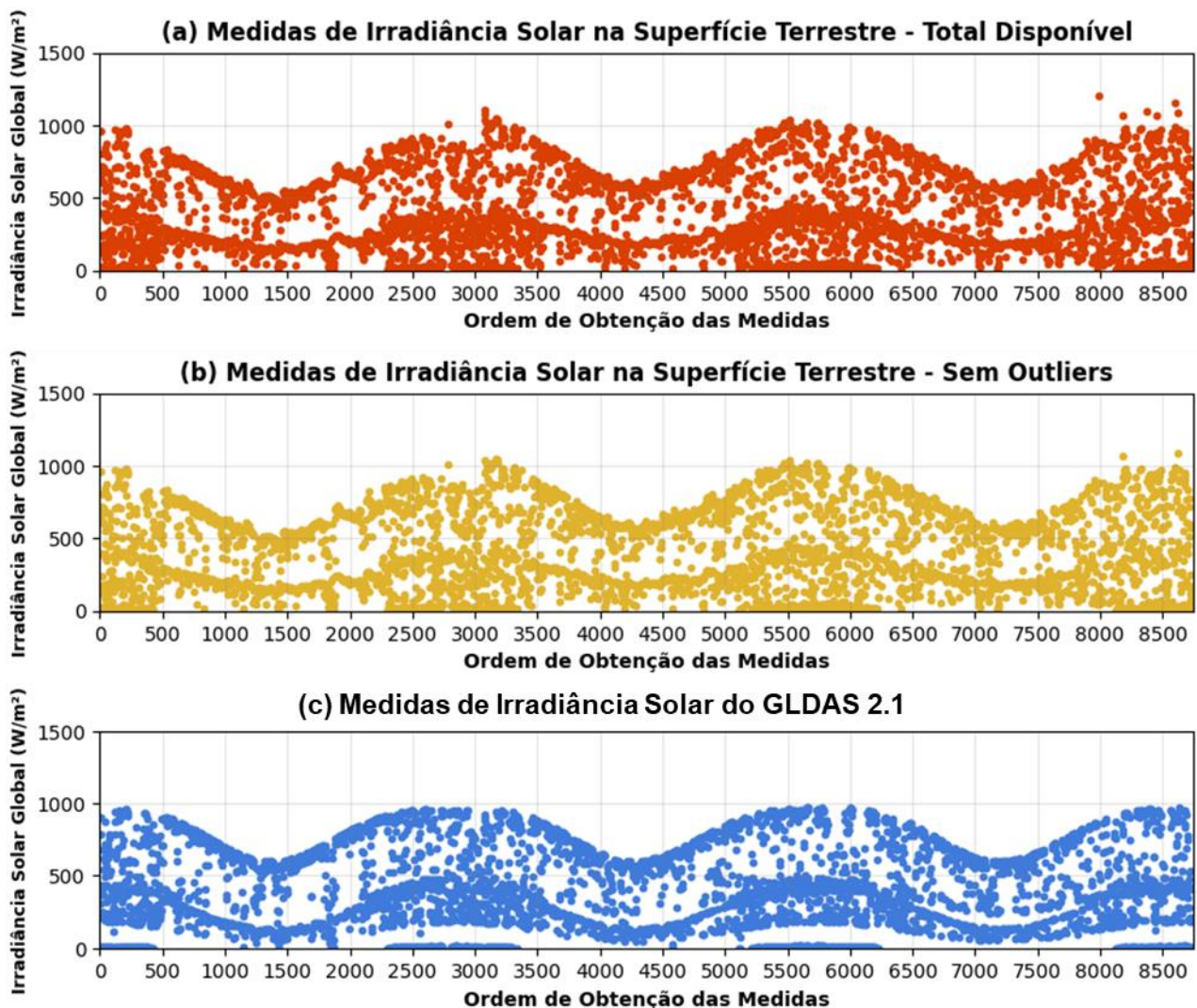


Figura 2 - Série temporal das medidas de irradiância solar global obtidas na superfície terrestre e pelo GLDAS 2.1 durante o período de 2020 a 2022.

Uma análise comparativa entre os valores da irradiância solar global fornecidos pelo GLDAS 2.1 e as medições realizadas em superfície terrestre foi executada. As métricas estatísticas (Tab. 6) indicam que o GLDAS 2.1 tem valores superiores à irradiância solar global medida na superfície terrestre (19,52%), com uma notável dispersão nos erros (28,41%). Porém, o coeficiente de determinação (0,81) sugere que o GLDAS 2.1 tem uma boa capacidade de explicar a variabilidade das medidas reais obtidas na superfície terrestre.

Tabela 6 - Indicativos estatísticos da comparação entre os valores da irradiância solar global fornecidos pelo GLDAS 2.1 e as medições realizadas em superfície terrestre.

INDICATIVO ESTATÍSTICO	VALOR OBTIDO
MBE	80,82 W/m <sup>2</sup>
rMBE	19,52 %
RMSE	117,60 W/m <sup>2</sup>
rRMSE	28,41 %
R <sup>2</sup>	0,81

A primeira combinação de ajuste linear (Fig. 3), usando exclusivamente a irradiância solar fornecida pelo GLDAS 2.1 como variável independente, demonstra que há uma forte correlação positiva entre os valores estimados e as medições reais. Os pontos de dispersão no gráfico estão bem distribuídos, sem aglomerações ou lacunas significativas, o que indica que há variabilidade entre as estimativas produzidas e as medidas de referência, obtidas na superfície terrestre, o que já era esperado, devido à natureza distinta entre os pares de dados em questão.

Os indicativos estatísticos sinalizam um erro médio de viés (MBE) de 77,43 W/m<sup>2</sup>, que demonstra uma tendência significativa em superestimar as medidas de irradiância solar na superfície terrestre em 18,13%, na média. O erro quadrático médio (RMSE) de 113,72 W/m<sup>2</sup> indica que há uma dispersão considerável, de 26,63% em média, entre os valores estimados e os medidos. Apesar do coeficiente de determinação (R<sup>2</sup>) de 0,83 indicar uma boa adequação do modelo de estimativa à realidade das medições reais, o MBE e o RMSE enfatizam a necessidade de um refinamento.

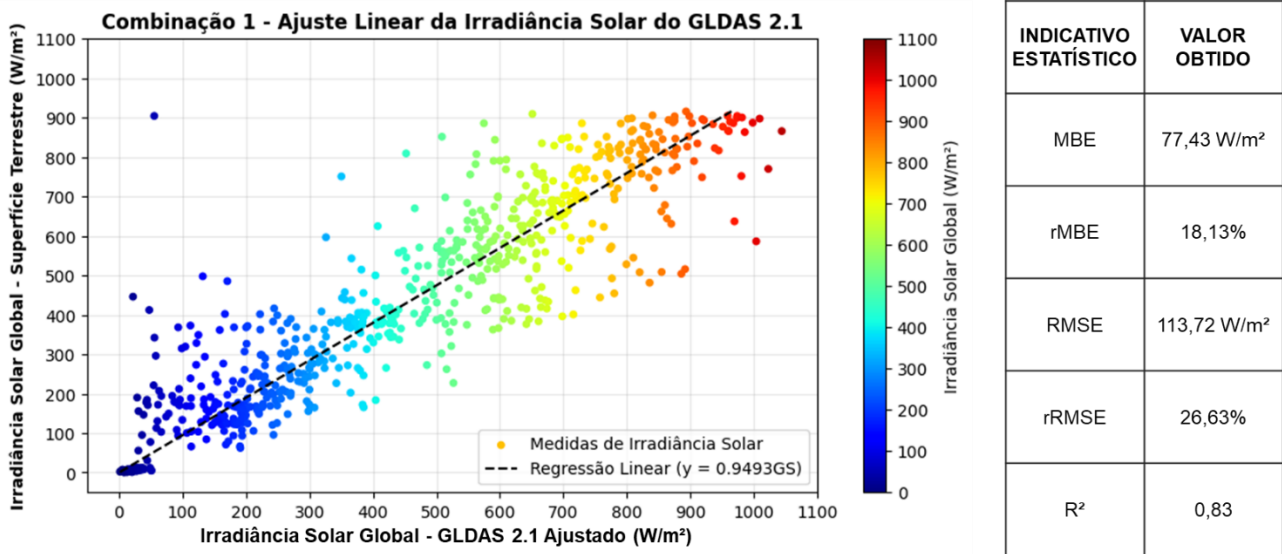


Figura 3 - Gráfico de dispersão e indicadores estatísticos da 1ª combinação de ajuste linear, que usa somente a irradiância solar global do GLDAS 2.1 (GS).

A segunda combinação de ajuste linear (Fig. 4), que utiliza a irradiância solar fornecida pelo GLDAS 2.1 e o índice de transmissividade atmosférica como variáveis independentes, evidencia uma forte correlação positiva entre os valores estimados e as medições reais. A inclusão do índice de transmissividade atmosférica como um parâmetro de entrada do modelo de estimativa, gerou uma melhoria notável. Os pontos de dispersão estão mais próximos da reta de regressão linear, indicando uma concordância maior em comparação a combinação de ajuste linear anterior (Fig. 3).

Os indicadores estatísticos também corroboram a melhoria obtida na segunda combinação de ajuste linear, com um erro médio de viés (MBE) de 67,00 W/m<sup>2</sup>, que ainda indica uma tendência de superestimação média de 15,69%, porém, obteve uma redução de quase 3 pontos percentuais em relação à combinação de ajuste linear anterior. No entanto, o erro quadrático médio (RMSE) de 112,32 W/m<sup>2</sup> revela ainda uma dispersão considerável de 26,31%, em média, entre os valores estimados e as medidas reais, não demonstrando uma melhoria significativa, neste caso.

O coeficiente de determinação (R<sup>2</sup>) apresentou uma melhora evidente, atingindo um valor de 0,88, o que indica uma melhor adaptação do modelo de regressão linear à realidade das medições de referência. A inclusão do índice de transmissividade atmosférica demonstrou ser uma escolha excelente, no sentido de otimizar os resultados obtidos em

relação aos valores estimados, permitindo capturar melhor a natureza dinâmica da irradiância solar medida na superfície terrestre, pois leva em consideração a sua interação com os constituintes atmosféricos.

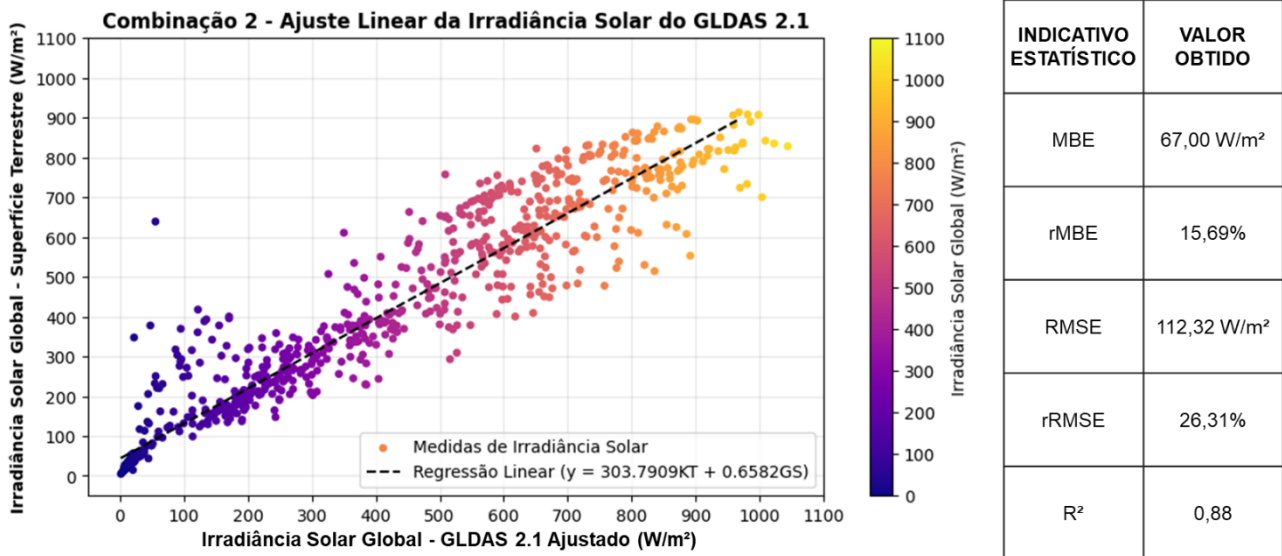


Figura 4 - Gráfico de dispersão e indicadores estatísticos da 2ª combinação de ajuste linear, que usa o índice de transmissividade atmosférica (KT) e a irradiância solar global do GLDAS 2.1 (GS).

A terceira combinação de ajuste linear (Fig. 5), usando a irradiância solar fornecida pelo GLDAS 2.1, índice de transmissividade atmosférica e elevação solar como variáveis independentes, revelou uma forte correlação positiva entre os valores estimados e as medições reais, com uma aproximação maior dos pontos de dispersão a reta de regressão linear. A inclusão do índice de transmissividade atmosférica e da elevação solar demonstrou um aprimoramento ainda maior da capacidade do modelo de estimativa compreender os processos atmosféricos e astronômicos da irradiação solar.

Os indicadores estatísticos indicam uma evolução positiva em relação às combinações anteriores de ajustes lineares. O erro médio de viés (MBE) de 66,26 W/m<sup>2</sup> persiste em indicar uma tendência de superestimação média de 15,69% das medições reais. O erro quadrático médio (RMSE) de 88,58 W/m<sup>2</sup> reflete uma melhora significativa na capacidade do modelo de prever os níveis de irradiância solar na superfície terrestre. Porém, apesar da melhoria de 6 pontos percentuais em relação à combinação anterior, ainda há uma dispersão considerável entre as estimativas e as medidas reais.

O coeficiente de determinação (R<sup>2</sup>), atingindo o valor de 0,90, destaca a melhor capacidade adaptativa do modelo de regressão linear em relação à variabilidade observada na irradiância solar medida na superfície terrestre. Neste sentido, as métricas estatísticas obtidas superam todas as combinações de ajustes lineares anteriores, evidenciando uma precisão mais aprimorada na representação dos padrões de irradiância solar na superfície terrestre. Portanto, a terceira combinação de ajuste linear demonstrou um avanço substancial, proporcionando resultados mais precisos e representativos.

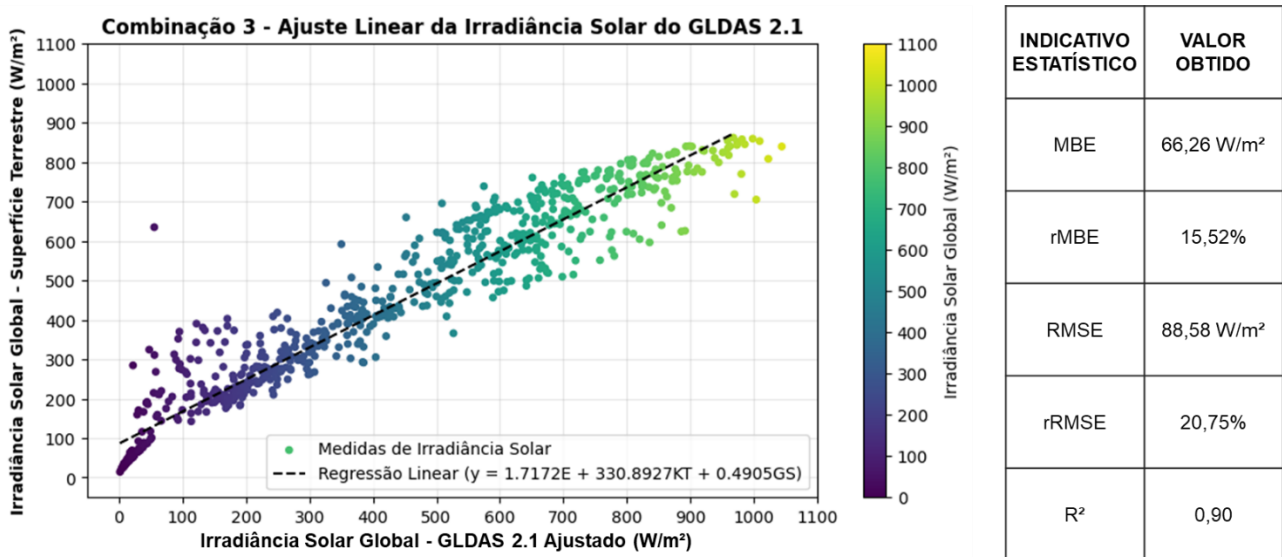


Figura 5 - Gráfico de dispersão e indicadores estatísticos da 3ª combinação de ajuste linear que usa a elevação solar (E), o índice de transmissividade atmosférica (KT) e a irradiância solar global do GLDAS 2.1 (GS).



#### 4. CONCLUSÃO

É consenso que obter informações sobre a irradiância solar incidente na superfície terrestre por meio de sensores automáticos representa a situação mais desejável. No entanto, as limitações financeiras e de mão de obra, em países em desenvolvimento, permitem explorar melhor a obtenção de estimativas a partir de modelos estatísticos, físicos ou estocásticos. Neste sentido, foi possível obter as seguintes conclusões sobre a otimização de modelos lineares para estimar a irradiância solar global na superfície terrestre, em Botucatu (São Paulo) - Brasil, em escala temporal horária:

- A terceira combinação de ajuste linear, que usa a irradiância solar fornecida pelo GLDAS 2.1, o índice de transmissividade atmosférica e a elevação solar, como variáveis independentes, destacou-se como a mais precisa. Apesar de ainda haver uma tendência de superestimação média de 15,69%, a dispersão entre as estimativas e as medidas reais foi consideravelmente reduzida, pois obteve um menor erro quadrático médio (RMSE), em relação as combinações de ajuste linear anteriores. O coeficiente de determinação ( $R^2$ ) de 0,90 indica uma compreensão notável do modelo de regressão linear em relação a variabilidade e natureza dinâmica observada na irradiância solar. Neste sentido, pode-se perceber que a inclusão do índice de transmissividade atmosférica e da elevação solar para se estimar valores da irradiância solar, proporcionou melhorias significativas quanto ao ajuste dos modelos de regressão linear. Isso indica que considerar os fatores atmosféricos e astronômicos foi crucial para se estimar com mais precisão a irradiância solar na superfície terrestre.
- Portanto, sugere-se a utilização do modelo de regressão linear ajustado a partir da combinação da irradiância solar fornecida pelo GLDAS 2.1, que pode ser obtida pelo Google Earth Engine, do índice de transmissividade atmosférica, que deve ser calculado por meio da razão entre a irradiância solar global medida na superfície terrestre e a extraterrestre calculada, e da elevação solar, que pode ser gerada a partir de equações astronômicas, como variáveis independentes, devido a sua melhor precisão. No entanto, é evidente a sua limitação em depender de uma informação que necessita de valores medidos para ser obtida (índice de transmissividade atmosférica). Neste sentido, em situações em que não se tem os valores do índice de transmissividade atmosférica disponíveis, sugere-se, a utilização do modelo de regressão linear ajustado a partir da primeira combinação de variáveis independentes, que considera somente a irradiância solar fornecida pelo GLDAS 2.1, para se estimar a irradiância solar na superfície terrestre.

Em futuros estudos pretende-se avaliar a possibilidade de usar modelos não-lineares para a estimativa da irradiância solar global na superfície terrestre. Os modelos não-lineares possibilitariam a representação de relações mais complexas e não lineares observadas nas medidas reais, permitindo capturar padrões intrincados, lidar com a não homocedasticidade e modelar relações não monotônicas, resultando em possíveis estimativas mais precisas, principalmente em situações em que a regressão linear se apresenta como um método mais limitado.

#### Agradecimentos

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Faculdade de Ciências Agrárias (UNESP) de Botucatu (São Paulo) - Brasil pelo apoio financeiro concedido através das bolsas de estudos (Mestrado e Doutorado).

#### REFERÊNCIAS

- Calça, M. V. C.; Raniero, M. R.; Fernando, D. M. Z.; Rodrigues, S. A.; Dal Pai, A., 2019. Outliers Detection in a Quality Control Procedure for Measurements of Solar Radiation. *IEEE Latin America Transactions*, vol. 17, n. 11, pp. 1815-1822.
- Chein, F., 2019. Introdução aos modelos de regressão linear. Brasília - DF: Escola Nacional de Administração Pública - ENAP. 76 p.
- Codato, G.; Oliveira, A. P.; Soares, J.; Escobedo, J. S.; Gomes, E. N.; Dal Pai, A., 2008. Global and diffuse solar irradiances in urban and rural areas in southeast Brazil. *Theoretical and Applied Climatology*, vol. 93, pp. 57-73.
- Dal Pai, A.; Escobedo, J. F.; Dal Pai, E.; Oliveira, A. P.; Soares, J.; Codato, G., 2016. MEO shadowring method for measuring diffuse solar irradiance: Corrections based on sky cover. *Renewable Energy*, vol. 99, pp. 754-763.
- Escobedo, J. F.; Gomes, E. N.; Oliveira, A. P.; Soares, J., 2009. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil. *Applied Energy*, vol. 86, pp. 299-309.
- Flumignan, D. L.; Rezende, M. K. A.; Comunello, E.; Fietz, C. R., 2018. Empirical methods for estimating reference surface net radiation from solar radiation. *Engenharia Agrícola*, vol. 38, n. 1, pp. 32-37. Disponível em: <https://doi.org/10.1590/1809-4430-Eng.Agric.v38n1p32-37/2018>.
- Franco, J. R.; Dal Pai, E.; Calça, M. V. C.; Raniero, M. R.; Dal Pai, A.; Sarnighausen, V. C. R.; Román, R. M. S., 2023. Atualização da normal climatológica e classificação climática de Köppen para o município de Botucatu - SP. *Irriga - Brazilian Journal of Irrigation and Drainage*, vol. 28, n. 1, pp. 77-92.
- Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, vol. 202, pp. 18-27.
- IBGE., 2022. Panorama da Cidade de Botucatu (SP) - Brasil. Rio de Janeiro: IBGE - Instituto Brasileiro de Geografia e Estatística. Disponível em: <https://cidades.ibge.gov.br/brasil/sp/botucatu/panorama>. Acesso em: 02 ago. 2022.

- IQBAL, M., 1983. An Introduction to Solar Radiation. Ontario: Academic Press Canada.
- NASA., 2021. README document for NASA GLDAS version 2 data products. Washington - EUA: NASA - National Aeronautics and Space Administration. Disponível em: [https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS\\_NOAH025\\_3H.2.1/doc/README\\_GLDAS2.pdf](https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_3H.2.1/doc/README_GLDAS2.pdf). Acesso em: 22 out. 2023.
- Pashiardis, S.; Kalogirou, S. A., 2016. Quality control of solar shortwave and terrestrial longwave radiation for surface radiation measurements at two sites in Cyprus. Renewable Energy, vol. 96, pp. 1015-1033. Disponível em: <https://doi.org/10.1016/j.renene.2016.04.001>.
- Pereira, E. B.; Martins, F. R.; Gonçalves, A. R.; Costa, R. S.; Lima, F. L.; Rüther, R.; Abreu, S. L.; Tiepolo, G. M.; Pereira, S. V.; Souza, J. G., 2017. Atlas Brasileiro de Energia Solar. 2 ed., São José dos Campos: Instituto Nacional de Pesquisas Espaciais. Disponível em: <http://doi.org/10.34024/978851700089>.
- Rincón. A.; Jorba, O.; Frutos, M.; Alvarez, L.; Barrios, F. P.; González, J. A., 2018. Bias correction of global irradiance modelled with weather and research forecasting model over Paraguay. Solar Energy, vol. 170, pp. 201-211.
- Scikit-Learn., 2022. Model Evaluation. Documentation. Disponível em: [https://scikit-learn.org/0.22/modules/model\\_evaluation.html#mean-absolute-error](https://scikit-learn.org/0.22/modules/model_evaluation.html#mean-absolute-error). Acesso em: 13 nov. 2022.
- Silva, M. B. P.; Escobedo, J. F.; Santos, C. M.; Rossi, T. J.; Silva, S. H. M. G., 2017. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. Journal of Atmospheric and Solar-Terrestrial Physics, vol. 160, pp. 11-27.
- WMO., 2012. Guide to Meteorological Instruments and Methods of Observation. Geneva: WMO - World Meteorological Organization, n. 8.

### OPTIMIZATION OF LINEAR MODELS FOR ESTIMATION OF GLOBAL SOLAR IRRADIANCE THROUGH REMOTE SENSING PRODUCTS IN BOTUCATU (SP) - BRAZIL

**Abstract.** *There is a consensus that knowledge about solar irradiance levels obtained on the Earth's surface through automatic sensors represents the most desirable situation. However, the construction of solarimetric stations involves a high financial contribution, which makes it difficult to implement this scenario in many locations in Brazil. The aim of this study was to create three linear regression models, with different input parameter combinations, to estimate hourly global solar irradiance on the Earth's surface, using information obtained by the Global Land Data Assimilation System 2.1. An assessment was carried out to determine which combination of independent linear regression variables would best fit the measurements collected at the School of Agricultural Sciences (UNESP) in Botucatu (SP) - Brazil, during the period 2020-2022. Therefore, three linear regression models were created in Python using global solar irradiance, provided by the remote sensing product, atmospheric transmissivity index and solar elevation, as independent variables, on an hourly time scale. The evaluation of the linear regression models was based on the statistical indicators MBE, rMBE, RMSE, rRMSE and  $R^2$ , using as a reference measurement obtained by a pyranometer on the Earth's surface. The first linear fit combination produced an  $R^2$  of 0.83, with an rMBE of 18.13% and an rRMSE of 26.63%. Likewise, the second linear adjustment combination generated an  $R^2$  of 0.88, with an rMBE of 15.69% and an rRMSE of 26.31%. Finally, the third linear adjustment combination presented an  $R^2$  of 0.90, with an rMBE of 15.52% and an rRMSE of 20.75%. Using the solar irradiance provided by the remote sensing product, atmospheric transmissivity index and solar elevation, as independent variables, created the linear regression model that better understood the atmospheric and astronomical processes that occurred, allowing to obtain estimates of solar irradiance on the Earth's surface with greater precision.*

**Keywords:** *Solar Radiation Modeling. Solar Energy Forecast. Multivariate Linear Regression.*