

MODELAGEM COMPUTACIONAL DA SÉRIE TEMPORAL DA IRRADIAÇÃO GLOBAL HORIZONTAL ESTIMADA POR DADOS DE SATÉLITE

Caio Sóter de Barros Mota – caio.soter@ufpe.br

Universidade Federal de Pernambuco, Departamento de Engenharia Elétrica

André dos Santos Albuquerque

Universidade Federal de Pernambuco, Departamento de Engenharia Elétrica

Manoel Henriques de Sá Campos

Universidade Federal de Pernambuco, Departamento de Energia Nuclear

Chigueru Tiba

Universidade Federal de Pernambuco, Departamento de Energia Nuclear

Resumo. *O item de maior risco de um empreendimento solar é o conhecimento imperfeito da irradiação solar, pelo fato de não haver capilaridade das estações de medições porque seria inviável economicamente a densificação nos níveis desejados. Dessa maneira, é padrão no mercado buscar ou modelar dados de imagens de satélites porque estas possuem abrangência de praticamente todo o globo terrestre e séries temporais longas de mais de 20 anos, entretanto, estão sujeitas a problemas como viés. O presente trabalho tem como objetivo desenvolver modelos para estimar os níveis de irradiância global horizontal GHI de longo prazo na cidade de Patos na Paraíba utilizando dados estimados por satélite e fazendo seu ajuste com uma série temporal medida em uma estação solarimétrica de alto desempenho durante 01 ano. A modelagem foi feita através dos métodos: regressão linear multivariada, Redes Neurais Artificiais (MLP) e machine learning (Support Vector Machine for Regression-SVR), implementados na linguagem Python. Para obter melhores resultados, houve o pré-processamento dos datasets para identificar e suprimir lacunas, retirar outliers sincronizar as séries temporais (satélite e terrestre) e a partição em 03 faixas de Kt. Para evidenciar o sucesso desta implementação, foram utilizadas quatro métricas estatísticas: coeficiente de correlação de Pearson, erro médio absoluto, erro médio quadrático e erro médio. Os seguintes resultados foram obtidos: no intervalo $0,0 < Kt \leq 0,35$ houve melhora em todos os indicadores estatísticos para os 03 métodos e considerando as 04 métricas o melhor método foi o MLP; no intervalo $0,3 < Kt < 0,70$ não houve melhora (as diferenças são erros de arredondamento) e finalmente no intervalo $0,7 \leq Kt < 1,1$ houve discreta melhora no método MLP.*

Palavras-chave: *Modelagem de GHI por satélite, adaptação de sítio, estimativa de irradiância solar.*

1. INTRODUÇÃO

O Brasil é um dos países com maior potencial para prospecção de energia solar, se destacando a região Nordeste que, de acordo com TIBA *et al.* (2000) e PEREIRA *et al.* (2017), possui uma irradiação solar global diária no plano horizontal com média anual entre 5,39 e 5,59 kWh/m². Aliado a isso, há uma preocupação crescente e de abrangência internacional com os fenômenos extremos decorrentes do aquecimento global provocados pela queima de combustíveis fósseis, bem como com a consequente necessidade de realizar a descarbonização e aumentar a segurança de fornecimento diversificando a matriz de geração de energia elétrica. Esse último aspecto é exemplarmente demonstrado em decorrência da crise hidrológica atual no Brasil. Esses são apenas alguns dos fatores que conferem às energias renováveis o rápido e amplo aumento de espaço percentual na matriz energética de diversos países. Considerado de forma isolada, o conhecimento da irradiação solar é o parâmetro de maior impacto e risco de um empreendimento solar. Nesse sentido, quanto melhor for a qualidade e precisão dos dados sob análise dentro de uma série temporal tão longa quanto disponível, menores serão as incertezas sobre as informações solarimétricas associadas que impactam no custo nivelado da energia elétrica (LCOE). As informações mensuradas a nível do solo sobre irradiação solar no mundo são escassas, não havendo frequentemente - e à disposição para livre acesso aos dados medidos - estações solarimétricas com a capilaridade necessária. Por outro lado, há vasta quantidade de dados resultantes de modelagem da irradiação por imagens de satélites com grande abrangência espacial acompanhados de séries com variáveis climatológicas e atmosféricas.

Na Engenharia Solar atual, em todas as fases do empreendimento solar (FV ou heliotérmica), projeto, construção e operação, os requisitos mínimos necessários para as informações sobre a irradiação solar, são: que seja específico do local da futura central solar, tenha representatividade climatológica (série histórica de longa duração, por exemplo 20 anos) e que possa ser sempre atualizada. Então o procedimento padrão da indústria solar hoje em dia é a fusão de dados de satélites de longa duração (15-20 anos) com dados terrestres de alta qualidade de relativamente curta duração (01 ano ou mais) cujo procedimento é denominado ajuste local da irradiação solar. Inúmeros estudos internacionais mostraram a possibilidade de aperfeiçoar as informações de satélites mediante o ajuste local com dados experimental de alta qualidade coletados durante pelo menos 01 ano GUEYMARD *et al.* (2012); AGUIAR *et al.* (2019); PERUCHENA *et al.* (2020); POLO *et al.* (2020); NARVAEZ *et al.* (2021). Considerando as dimensões continentais do Brasil, apesar de existirem alguns estudos já realizados (TIBA *et al.* 2019); (Miranda, Araújo e Vilela, 2020), ainda há muito espaço para novas pesquisas e novas técnicas a serem aprimoradas.

Dentro do contexto apresentado, este trabalho busca o desenvolvimento de um modelo que promova a redução das incertezas relativas às medições de irradiação solar global horizontal (GHI) para um determinado local a partir de uma série temporal de dados de satélite, tendo para comparação medições terrestres efetuadas ao nível do solo e armazenadas em um Banco de Dados terrestre. A abordagem utilizada é comumente citada como *site-adaptation* em artigos científicos e periódicos. Como o próprio nome sugere, o estudo é naturalmente singular e local.

2. MATERIAL E MÉTODOS

Para o estudo foram utilizados dois bancos de dados relativos ao ano de 2015, no qual um deles foi obtido através do Solargis (Satélite), que é um *web service* baseado em um histórico de dados de 10 anos e que possui alta eficiência como visto no trabalho de Šúri *et al.* (2011), formando o banco de dados de satélite utilizado no presente trabalho; o outro banco de dados foi proveniente da estação solarimétrica localizada na cidade de Patos, PB, (latitude: -7,003 S; longitude: -37,178 W). Por meio dessa estação, foi possível fazer a leitura das irradiações através de equipamentos como piranômetro e pireliômetro. Dessa forma, a discussão dos equipamentos de medida em termos de acurácia é um aspecto que merece ênfase. Nesse contexto, o padrão ISO 9060:1990 (*Solar energy - Specification and classification of instruments for measuring hemispherical solar and direct solar radiation*) estabelece dentro das especificações sobre piranômetros, sob a referência 3c, que equipamentos classificados como de primeira ou segunda classe têm uma faixa de erros de ± 20 kWh/m². A classificação designada como padrão secundário, caso do piranômetro CMP-21, fabricado pela Kipp & Zonnen e incorporado à estação solarimétrica, encontra-se dentro de uma acurácia de ± 10 W/m². É importante salientar que sua faixa espectral compreende desde o infravermelho até parte do ultravioleta, percorrendo o intervalo de 285 a 2800nm – e designado como classe A espectralmente plana, segundo o fabricante. Com o apoio de um datalogger, utilizado para as medições/digitalização, armazenamento e transferência dos dados via comunicação serial para o computador, foi criado o banco de dados terrestre. As variáveis utilizadas podem ser vistas na Tab. 1.

Tabela 1 - Parâmetros efetivamente utilizados do Banco de Dados Terrestre e de Satélite.

| Parâmetros Terrestres (médias de hora em hora). | | Parâmetros de Satélite. | |
|--|---------------------|-------------------------------------|----------------------|
| Temperatura média | [°C] | GHI -Irradiância Global horizontal | [W/m ²] |
| Umidade relativa do ar | [%] | DNI -Irradiância normal direta | [W/m ²] |
| Velocidade do vento média | [m/s] | DIF - Irradiância horizontal difusa | [W/m ²] |
| Direção do vento média | [graus] | SE - Ângulo de elevação solar | [graus.] |
| Irradiância total horizontal média | [W/m ²] | SA - Azimute solar | [graus.] |
| Irradiância horizontal difusa média | [W/m ²] | TEMP - Temperatura do ar em 2 m | [°C] |
| Irradiação Direta média | [W/m ²] | AP - Pressão Atmosférica | [hPa] |
| Irradiância coletada(plano de 17 graus norte) média. | [W/m ²] | RH - Umidade Relativa | [%] |
| Precipitação média | [mm] | WS - Velocidade do Vento a 10 m | [m/s] |
| Pressão Atmosférica média | [mbar] | WD - Direção do Vento | [graus.] |
| - | - | PWAT - Água Precipitável | [kg/m ²] |

2.1 Pré-processamento e Coeficiente de Transmitância (Kt)

Para melhorar a qualidade das medições originais, é necessário que os dados sejam tratados pela correção de eventuais tendências estatísticas e *outliers*, assim como pelo tratamento de intervalos com ausência de dados. Para tal, foi preciso primeiramente detectar onde poderia haver lacunas de dados decorrentes de problemas de funcionamento dos equipamentos de medições diante de falta de energia, por exemplo, ou por eventuais paradas para manutenções. A maior falha encontrada foi de 1 semana sem dados no mês de junho; outras lacunas durante o ano foram de aproximadamente 3 a 4 horas em dias pontuais. Os dados destas lacunas não foram preenchidos e foram descartados. Outra providência no pré-tratamento foi a sincronização entre os dados de satélite e de superfície com a transformação das informações para escala horária, isto porque conjunto de dados de satélite foi construído de hora em hora e o conjunto de dados terrestres construído de minuto em minuto. Para atenuar as tendências determinísticas da irradiação solar, as mesmas foram expressas como coeficientes de transmitância.

Optou-se por 3 faixas de coeficientes de transmitância: $0 < Kt \leq 0,35$ com 385 amostras; $0,3 < Kt < 0,7$ com 1799 amostras e $0,7 \leq Kt < 1,1$ com 1851 amostras, conforme metodologia também identificada em outros trabalhos (Martinez e Velásquez, 2018). Adicionalmente, foram excluídas amostras onde a radiação total horizontal GHI ficou abaixo de 30 W/m^2 . Na faixa superior, valores acima da unidade também foram contemplados, uma vez que diversos trabalhos têm sido desenvolvidos mostrando a ocorrência do efeito lente, também referido em alguns trabalhos como “*cloud enhancement*”.

2.2 Métricas estatísticas

Esta pesquisa utilizou 4 métricas distintas. A primeira delas foi o coeficiente de correlação de Pearson (R^2). Eq. (1), que tem o seu valor dentro do intervalo $[-1; 1]$ e indica a intensidade de correlação linear entre duas variáveis quantitativas. Essa métrica de natureza estatística indicará correlação maior na medida que se aproxima de 1 (em módulo); caso seu valor seja 0, isso indica que não existe correlação linear entre as duas variáveis. Outra importante métrica utilizada é a média entre valores absolutos das diferenças entre valores estimados (derivados dos dados de satélite ou dos modelos) e os valores reais (terrestres), ou MAE (*mean absolute error*), Eq. (2). Além das duas anteriores, estudos comparativos comumente utilizam a métrica conhecida por RMSE (root mean square error), que significa a raiz quadrada computada sobre a média dos erros (ou às vezes mais apropriadamente chamados desvios) elevados ao quadrado, Eq (3). Concluindo as métricas, foi também usado o *Mean Bias Error* (MBE), Eq. (4), que é uma métrica para estimar o viés ou desvio médio relativo ao que apresenta um modelo e os valores efetivamente medidos; quanto mais próximo de zero for o valor do MBE, menor será o desvio do modelo, em média, com relação aos dados medidos observados. Por fim, e não menos importante, foi coletado o tempo de processamento para cada modelo, através da biblioteca *Time*, pois é uma métrica importante a ser levada em consideração devido a relevância do esforço computacional para obtenção dos resultados.

$$R^2 = \frac{\sum_{i=1}^N (G_{sat} - G_{satm\u00e9dia})(G_{abs} - G_{absm\u00e9dia})}{\sqrt{(\sum_{i=1}^N (G_{sat} - G_{satm\u00e9dia})^2)(\sum_{i=1}^N (G_{abs} - G_{absm\u00e9dia})^2)}} \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |(G_{sati} - G_{absi})| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N E(t)^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (G_{sat}(t) - G_{abs}(t))^2} \quad (3)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (G_{sati} - G_{absi}) \quad (4)$$

2.3 Modelos de Regressão

Aspectos gerais dos modelos

Como prática usual em modelos que usam Machine Learning (ML) e Redes Neurais Artificiais do tipo MLP - *multilayer perceptron*, precisou-se dividir o conjunto de dados em amostras de treinamento, validação e teste. Para tanto, utilizou-se a função *Train_test_split* da biblioteca *Sklearn* do pacote *linear model* usado na linguagem Python. Escolheu-se um valor de 20% para teste dos modelos e o restante (80%) foi usado para treinamento e validação. No entanto, não houve conjunto de validação para o modelo de regressão multivariada.

Na próxima etapa houve o desenvolvimentos de 3 modelagens distintas: um modelo estatístico de regressão multivariada, um modelo através da técnica de *Machine Learning* fundamentada na teoria do aprendizado estatístico (a SVM aplicada à regressão ou SVR) e finalmente um último modelo através de uma Rede Neural Artificial (RNA) do tipo MLP.

Regressão linear multivariada

A regressão linear multivariada (Hong, 2010) trata-se da expressão matemática descrita na Eq. 5 capaz de realizar inferências através de parâmetros onde a variável de interesse de saída é modelada em função de mais de uma variável causal de entrada, ou seja, que pode influenciar no comportamento final dessa saída ou alvo da modelagem. A Eq. 5 rege o comportamento do modelo da regressão multivariada, utilizando como entrada mais de uma característica ou *feature* (variáveis X_i) para resultar na estimativa de saída GHI, onde Y é a irradiação solar global e os X_i são as variáveis de entrada relacionadas na Tab. 1. Além disso, vale ressaltar que foi utilizado o método *LinearRegression* do pacote *linear_model* da biblioteca *sklearn*, para a arquitetura deste modelo.

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3... + \beta_n * X_n + \varepsilon \quad (5)$$

RNA - Multilayer perceptron.

As Redes Neurais Artificiais (RNA's) têm como objetivo criar um sistema computacional de modelagem capaz de estabelecer estimativas de regressão simulando neurônios biológicos, com capacidade de correlacionar entradas e saídas de modo totalmente diferente da abordagem anterior através de sua essência não-linear. Esta ferramenta utiliza diversos parâmetros de configuração, a saber: função de ativação, otimizadores, arquitetura da rede neural, taxa de aprendizado, momento, etc. Por meio do modelo *Sequential* do pacote *keras.modes* da biblioteca do *Tensor Flow* foi configurada camada após camada. A Tab. 4 ilustra alguns argumentos usados para cada faixa do coeficiente de transmitância.

Tabela 2 - Hiperparâmetros para o *Sequential, Tensor Flow*, para cada faixa de K_t .

| - | $0 < K_t \leq 0,35$ | $0,3 < K_t < 0,7$ | $0,7 \leq K_t < 1,1$ |
|---------------------------|----------------------------|---|---|
| <i>kernel_initializer</i> | 'glorot_uniform' | 'glorot_uniform' | 'glorot_uniform' |
| <i>layers</i> | (120, 70, 20, 1) | (120, 60, 20, 1) | (120, 40, 20, 1) |
| <i>activation</i> | relu e tanh(última camada) | relu e sigmoid(última camada) | relu e sigmoid(última camada) |
| <i>loss</i> | 'mse' | 'mse' | 'mse' |
| <i>optimizer</i> | 'adam' | 'adam' | 'adam' |
| <i>drop_out</i> | - | 0,01(Entre primeira e segunda <i>hidden layer</i>) | 0,01(Entre primeira e segunda <i>hidden layer</i>) |

2.3.4 Support Vector Machine with Grid Search

O SVM aplicado à regressão é um método baseado em uma técnica de aprendizagem supervisionada que é advinda da teoria do aprendizado estatístico e encontra-se disponível para uso em problemas que envolvem classificações (quando é chamado de SVC) ou regressões (quando recebe a designação de SVR). Para a situação proposta do trabalho, como se tem a necessidade de realizar regressões, foi feito o uso da versão SVR (*Support vector Regression*). Para tal, utilizando a biblioteca *Sklearn.svm* foi possível acessar esse tipo de metodologia com os seguintes hiperparâmetros listados na Tab. 3 aplicados à função *Grid Search*:

Tabela 3 - Listas e t-uplas de Hiperparâmetros para o *Support Vector Regression* passadas ao *Grid Search*.

| | |
|--------|--|
| Kernel | ['rbf'] |
| C | [0,01, 0,2, 1, 10, 50] |
| gamma | ('auto', 'scale', 10, 1, 0,1, 0,01, 0,001, 0,0001) |

Cabe ressaltar que existem diversos hiperparâmetros possíveis para se utilizar no modelo, com muitas combinações possíveis e seria um esforço manual muito grande para identificar qual melhor escolha. A ferramenta *Grid Search*, também advinda da biblioteca *Sklearn*, permite a identificação automática da melhor combinação dos hiperparâmetros. É previsível que este método demande um esforço computacional maior que a utilização de uma combinação única de valores de cada uma das duas últimas linhas da Tab. 3, mas espera-se, por outro lado, melhores resultados em relação à busca manual por métodos heurísticos ou tentativas. Por fim, é importante ressaltar que os algoritmos de *Machine Learning*, *Deep learning* e para regressão linear multivariada foram implementados através da linguagem python no ambiente Colab do Google.

3. RESULTADOS

3.1 Métricas iniciais: Satélite versus Medições de superfície

Inicialmente, antes da construção dos modelos, havia um quadro de dispersão retratado pelas métricas, coeficiente de correlação de Pearson, *erro médio quadrático*, *erro médio absoluto* e *erro médio*, como pode ser visto na Tab. 4. Esses dados serão úteis para efeito de comparação ao final dos resultados de cada modelo a ser apresentado.

Tabela 4 - Métricas iniciais (Satélite versus Medições de superfície).

| Métricas | $0 < Kt \leq 0,35$ | $0,3 < Kt < 0,7$ | $0,7 \leq Kt < 1,1$ |
|--------------------------------|--------------------|------------------|---------------------|
| R ² (Coef. Pearson) | 0,68 | 0,88 | 0,95 |
| MAE (W/m ²) | 55,86 | 63,33 | 38,79 |
| RMSE (W/m ²) | 73,65 | 88,42 | 61,18 |
| MBE (W/m ²) | 13,06 | 0,38 | -14,10 |

3.2 Métricas após ajustes

Os resultados obtidos foram desiguais em função da faixa escolhida para o Kt. Para o intervalo inferior do coeficiente de transmitância, como visto na Fig. 1 abaixo, independente dos modelos utilizados houve uma maior dispersão dos resultados de saída quando comparado com as faixas superiores. Com a finalidade de comparar o desempenho do algoritmo, foi colocado uma reta de coeficiente angular de 45°, cruzando o gráfico, sendo possível

observar visualmente a dispersão dos pontos obtidos nas diferentes faixas de Kt, procedimento esse repetido igualmente na Fig. 2 e na Fig. 3, que se referem, respectivamente, às partições para $0,3 < Kt < 0,70$ e para $0,7 \leq Kt < 1,1$.

Por fim, vale ressaltar que, para todos os modelos, foi testado o uso da técnica conhecida como PCA, ou Análise de Componentes Principais, que é uma ferramenta para diminuir a dimensionalidade do conjunto de dados, sem perda da informação mais relevante, reunindo apenas as (novas) variáveis com maior significância para usufruir de vantagens adicionais como eventual redução de ruído, convergência e aumento da velocidade de processamento dos modelos. De modo semelhante, também foi testado o método denominado *One Class SVM* como detector de *outliers* junto aos modelos. Entretanto, embora os dois métodos anteriores sejam utilizados em modelos de regressão, não foram obtidos resultados conclusivos de ganhos proporcionados por essas ferramentas, não justificando assim o uso de ambas.

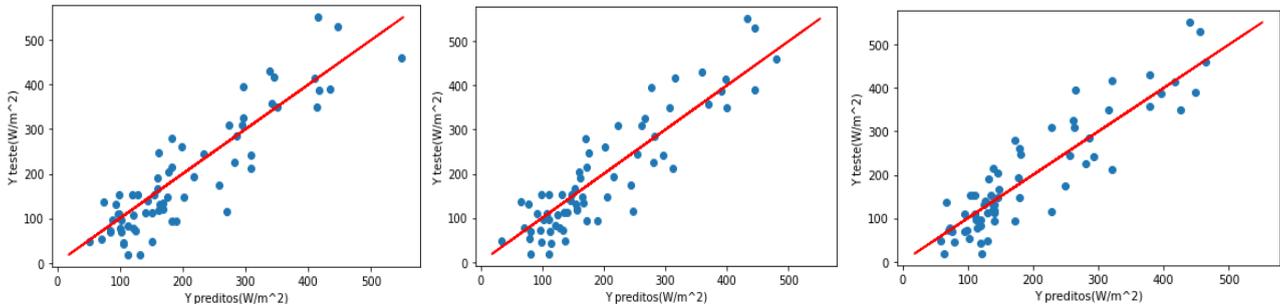


Figura 1: Gráficos da Regressão Multivariada (à esquerda), SVR (ao centro) e MLP (à direita), $0 < Kt \leq 0,35$.

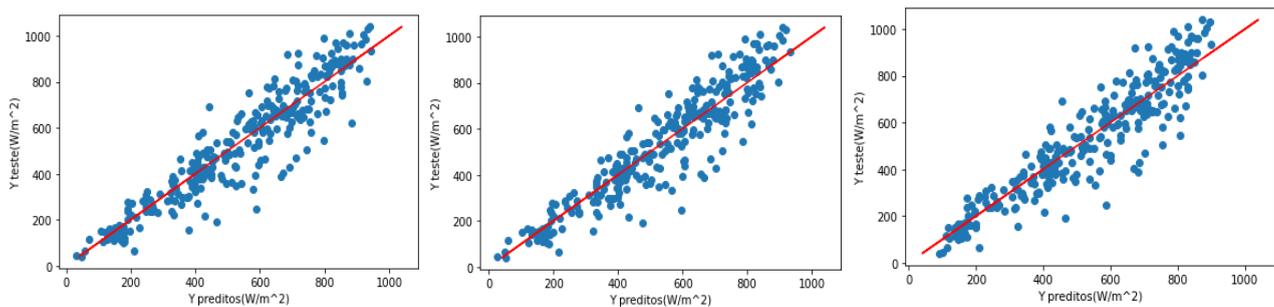


Figura 2: Gráficos da Regressão Multivariada (à esquerda), SVR (ao centro) e MLP (à direita), $0,3 < Kt < 0,7$.

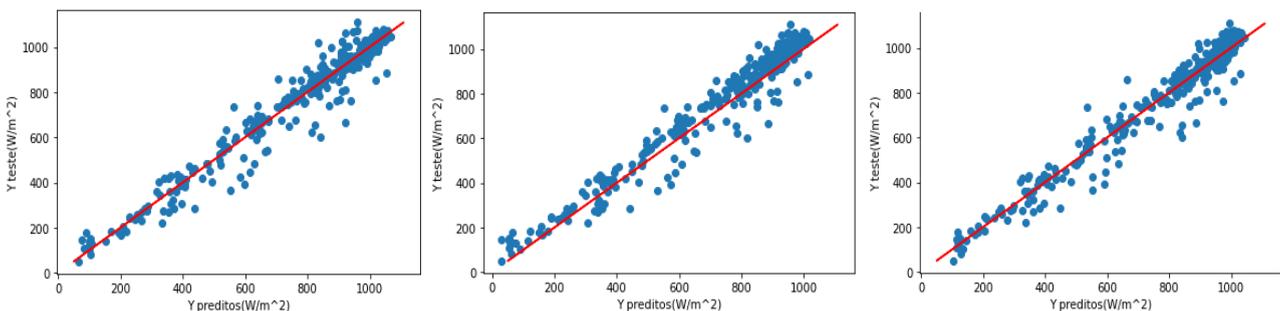


Figura 3: Gráficos da Regressão Multivariada (à esquerda), SVR (ao centro) e MLP (à direita), $0,7 \leq Kt < 1,1$.

A Tab. 5 mostra os resultados por faixas de Kt. Os números em parênteses são da métrica inicial ou seja os dados brutos. Além disso, os números em percentuais representam os ganhos ou perdas de cada método e, em resumo, no intervalo $0,0 < Kt \leq 0,35$ houve melhora em todos os indicadores estatísticos para os 03 métodos. Todavia, no intervalo $0,3 < Kt < 0,70$ não houve melhora (as diferenças são erros de arredondamento). Por fim, embora o MLP (em negrito), para o intervalo de $0,7 \leq Kt < 1,1$, tenha obtido uma discreta melhora, ainda sim, considerando as 04 métricas, foi o que apresen

Tabela 5 - Métricas estatísticas encontradas após a modelagem. Em parênteses a métrica inicial. Em % os ganhos ou perdas de cada métodos

| Métricas | 0 < Kt ≤ 0,35 | | | 0,3 < Kt < 0,7 | | | 0,7 ≤ Kt < 1,1 | | |
|--------------------------------|-----------------------------|------------------|--------------------------------|----------------------------|---------------------|-----------------|----------------------------|------------------|-------------------------------|
| | Reg. Multivar | SVR | MLP | Reg. Multivar | SVR | MLP | Reg. Multivar | SVR | MLP |
| R ² (Coef. Pearson) | 0,80 (0,68) 17,65%↑ | 0,81 19,12%↑ | 0,82 20,59%↑ | 0,88 (0,88) 0,00% | 0,88 0,00% | 0,88 0,00% | 0,95 (0,95) 0,00% | 0,93 -2,11%↓ | 0,95 0,00% |
| MAE (W/m ²) | 46,62 (55,86) 16,54%↓ | 46,82 16,18%↓ | 43,29 22,50%↓ | 63,39 (63,33) 0,09%↑ | 65,50 3,43% ↑ | 64,78 2,29%↑ | 39,22 (38,79) 1,11%↑ | 56,89 46,66%↑ | 38,32 1,21%↓ |
| RMSE (W/m ²) | 58,63 (73,65) 20,39%↓ | 57,08 22,50%↓ | 55,04 25,27%↓ | 87,91 (88,42) 0,58%↓ | 88,40 0,02% ↓ | 88,40 0,02%↓ | 59,20 (61,68) 3,24%↓ | 68,13 11,36%↑ | 57,35 6,26%↓ |
| MBE (W/m ²) | -10,58 (13,06) | -3,94 | 0,49 | -6,47 (0,38) | -0,46 | -1,36 | 2,40 (-14,10) | 33,92 | -0,39 |

4. CONCLUSÕES

Através dos resultados obtidos foi possível constatar que, para a primeira partição do coeficiente de transmitância ($0 < Kt \leq 0,35$), houve uma melhora bastante significativa em todos os modelos em comparação aos dados originariamente obtidos por satélite. Destaca-se, em especial, o modelo MLP, que gerou um ganho de mais de 20% relativamente à correlação inicial dos dados terrestres com os de satélite, além de conseguir uma redução de aproximadamente 25% no RMSE. Em termos gerais, se pôde verificar que os resultados também foram fortemente correlacionados em todos os modelos ($R^2 \geq 80\%$) e as métricas RMSE e MAE em todos os casos foram reduzidas, pelo menos, em 16%.

Para a partição mais alta de Kt's, ($0,7 \leq Kt < 1,1$), onde se poderia pressupor haver uma maior precisão dos dados de satélite, tanto para a Regressão Multivariada quanto para o MLP, ainda foi possível alcançar uma melhora significativa na métrica MBE em relação aos dados de satélite, com aprimoramentos de, respectivamente, 10,7 W/m² e 13,7 W/m². Em valores absolutos foram, respectivamente, 2,40 W/m² e -0,39 W/m², portanto significando ganhos claros quando aplicadas essas duas modelagens. Resta observar que também houve discreta diminuição no que tange às métricas RMSE e MAE entre os dados originais de satélite e os obtidos pelos modelo MLP nessa faixa.

Embora as métricas dentro da partição central, ($0,3 < Kt < 0,7$), não tenham mostrado ganhos expressivos relativos ao conjunto de dados original entregue pelo satélite, é importante observar que as correlações se mantiveram aproximadamente constantes e as melhoras relativas à métrica RMSE foram muito discretas em todos os modelos. Por outro lado, houve também uma discreta degradação sobre a métrica MAE. De fato, essa faixa intermediária de Kt's foi aquela em que houve maior dificuldade em superar de forma ampla todas as métricas oriundas dos valores das estimativas de satélite, o que pode ser indício ou sugerir uma equivalência às estimativas realizadas pelo satélite. O trabalho de Miranda, Araújo e Vilela (2020), que trata da análise do recurso solar para a cidade de São João do Rio do Peixe, estabelecida a 164 km de Patos, também aponta para a dificuldade de melhora significativa de resultados a partir de dados satelitais, o que corrobora uma eventual hipótese de a região nesse entorno ter características atmosféricas e climatológicas que permitem uma excelente modelagem através de inferências por satélite.

Por fim, pode-se perceber que, embora a MLP exija uma maior complexidade de implementação em linguagem de programação e demande maior tempo em ajustes de arquitetura e hiperparâmetros devido à sua complexidade, esse foi o modelo que, de maneira geral, apresentou melhor desempenho e resultou nos *ganhos* mais satisfatórios dentre todos os modelos utilizados no trabalho.

Agradecimentos

Agradecemos à UFPE-Universidade Federal de Pernambuco, ao DEN-Departamento de Energia Nuclear, ao CER-Centro de Energias Renováveis e ao Conselho Nacional de Pesquisa (CNPq) Bolsa nº302251-2017-0, por apoiarem os meios materiais e o ambiente científico para a execução desta pesquisa.

REFERÊNCIAS

- Aguiar, LM, J. Polo, J, J.M. Vindel, JM and Oliver, A., 2019. Analysis of satellite derived solar irradiance in islands with site adaptation techniques for improving the uncertainty, *Renewable Energy* 135 ,2019, 98e107.
- Gueymard, CA, Gustafson, WT, Bender, G, Etringer, A and Storck, P., 2012. Evaluation of procedures to improve solar resource assessments: optimum use of short-term data from a local weather station to correct bias in long-term satellite derived solar radiation time series, *World Renewable Energy Forum*, Denver, CO, USA.
- Hong, T. 2010. Short Term Electric Load Forecasting (a dissertation submitted to the Graduate Faculty of North Carolina State University). North Carolina State University. Raleigh, North Carolina.
- Martínez, A., Velázquez, E., 2018. Empirical models for estimation of diffuse solar radiation in a tropical, mountainous and humid place (Xalapa, México).
- Miranda, DR, Araújo, EV e Vilela, OC., 2020. Análise do recurso solar de longo prazo na cidade de São João do Rio do Peixe, sertão da Paraíba, VIII Congresso Brasileiro de Energia Solar –Fortaleza, 01 a 05 de junho de 2020.
- Narvaez, G., Giraldo, LF, Bressan, M and Pantoja, A., 2021. Machine learning for site-adaptation and solar radiation forecasting, *Renewable Energy* 167 333 – 342.
- Pereira, E. B., Martins, F. R., Gonçalves, A. R., Costa, R. S., Lima, F. L., Rütther, 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: INPE. 88 p.
- Peruchena, CMF., Polo, J., Luis Martín, L., Mazorra, L., 2020. Site-Adaptation of Modeled Solar Radiation Data: The Site Adapt Procedure, *Remote Sens.* 2020, 12, 2127; doi:10.3390/rs12132127.
- Polo, J., Peruchena, CF., Salamaliki, V., Aguiar LM., Turpin, M., Pomares, LM., Kazantzidis, A., Blanc, P., Remund, J., 2020. Benchmarking on improvement and site-adaptation techniques for modeled solar radiation datasets, *Solar Energy* 201, 469–479.
- Šuri, M., Cebecauer, T., Skoczek, A., Solar, G., 2011. SolarGIS: solar data and online applications for PV planning and performance assessment.
- Tiba, C. (Coord.). 2000. Atlas solarimétrico do Brasil: banco de dados terrestres. Recife: Editora Universitária, 111p.
- Tiba, C., Henriques, M., Souza L, 2019. Methods for Site-Adaptation of Satellite-Based DNI Time Series: Application to Brazilian Northeast, ISES Solar World Congress.

COMPUTATIONAL MODELING OF THE GLOBAL HORIZONTAL IRRADIATION TIME SERIES ESTIMATED FROM SATELLITE DATA

Abstract. *The item of greatest risk for a solar venture is imperfect knowledge of solar irradiation, due to the fact that there is no capillarity of measurement stations because it would be economically unfeasible to densify at the desired levels. Thus, it is standard in the market to seek or model data from satellite images because they have coverage of virtually the entire globe and long time series of over 20 years, however, they are subject to problems such as bias. The present work aims to develop models for estimating the long-term global horizontal irradiance levels GHI in the city of Patos in Paraíba (Northeast Region of Brazil) using satellite estimated data and fitting them with a time series measured at a high-performance solarimetric station during 01 year. The modeling was done through the methods: multivariate linear regression, Artificial Neural Networks (by use of Multilayer Perceptron approach or MLP) and machine learning (Support Vector Machine for Regression-SVR); implemented in Python language. To obtain better results, the datasets were preprocessed to identify and suppress gaps, remove outliers, synchronize the time series (satellite and terrestrial) and partition into 03 Kt bands. The following results were obtained: in the interval $0.0 < Kt < 0.35$ there was improvement in all statistical indicators for the 03 methods and considering all the 04 metrics, the best approach was the MLP; in the interval $0.3 < Kt < 0.70$ there was no improvement (the differences are rounding errors) and finally in the interval $0.7 < Kt < 1.1$ there was a slight improvement in the MLP method.*

Key words: Satellite GHI modeling, site adaptation, solar irradiance estimate