

# ESTUDO COMPARATIVO ENTRE OS MÉTODOS LASSO E MACHINE LEARNING PARA PREVISÃO DA IRRADIAÇÃO SOLAR DE CURTO PRAZO

**Samira Fontes Domingos** – samirafisica@yahoo.com.br

**Wallace do Couto Boaventura**

Programa de Pós-graduação em Engenharia Elétrica - Universidade Federal de Minas Gerais - UFMG

**Luís Guilherme Monteiro Oliveira**

Departamento de Engenharia Elétrica - IPUC - PUC Minas

1.1. Recurso solar e meteorologia da radiação solar

**Resumo.** A previsão da irradiação solar para curto prazo, especialmente a previsão para 5 minutos à frente, é adequada para a utilização no despacho de energia e sua precisão define a confiabilidade do sistema. Machine Learning - ML (Aprendizado de máquinas) é um método consagrado atualmente para previsões de diferentes naturezas e sua performance foi comparada ao método LASSO (least absolute shrinkage and selection operator), dentro do modelo de Regressão, para previsão da irradiação solar de curto prazo. O LASSO tem a capacidade de identificar preditores mais apropriados pois possui a propriedade de shrinkage, (encolhimento em português), que tem a capacidade de deletar completamente uma variável se esta não for tão relevante para o contexto de previsão. A simulação foi realizada no software R e, para tal, foram utilizados dados públicos do laboratório NREL, medidos em Oahu, no Havaí, com medidas de irradiação solar de nove estações solarimétricas, com taxa de amostragem com intervalos de 1 em 1 segundo. Foram utilizados quatro métricas notáveis (RMSE, nRMSE, MBE e MAE) para comparação dos métodos. Ambos os métodos se mostraram eficazes para a previsão de curto prazo. O LASSO alcançou RMSE de  $27.95 \text{ W/m}^2$ , nRMSE de  $0.06 \text{ W/m}^2$ , MAE de  $10.53 \text{ W/m}^2$  e MBE de  $-3.90$ . O ANN atingiu RMSE de  $28.42 \text{ W/m}^2$ , nRMSE de  $0.07 \text{ W/m}^2$ , MAE de  $10.56 \text{ W/m}^2$  e MBE de  $-3.89$ . Porém devido às interpretações que o método LASSO propicia, o método tem potencial para equiparar ao Machine Learning, para o horizonte de previsão em estudo, neste caso, previsão de curto prazo.

**Palavras-chave:** Previsão da irradiação solar, LASSO, Machine learning.

## 1. INTRODUÇÃO

A expansão da capacidade instalada de parques solares fotovoltaicos nos últimos anos, adicionada à projeção de alto crescimento desta fonte, impuseram ao Operador Nacional do Sistema Elétrico (ONS) a necessidade de aperfeiçoar os insumos dos estudos energéticos e elétricos. Face a essa grande inserção, faz-se necessário a alocação de um montante adequado de reserva de potência operativa. Nesse sentido, é de suma importância que a previsão de geração por fonte solar fotovoltaica seja feita com suficiente precisão para o planejamento de curtíssimo e curto prazo da Operação Eletroenergética, bem como para a Operação em Tempo Real do Sistema Interligado Nacional (SIN). (MME, ONS, 2023) A instabilidade elétrica em todo o país no dia 15/08/2023, reafirmou a importância primordial da previsão de curto prazo para o Sistema elétrico brasileiro.

As decisões para a operação do mercado de energia são baseadas em previsões. Para o mercado que utiliza previsão para os próximos 5 minutos, são dados de 65 minutos anteriores para construir a previsão. E essa previsão para intervalos de 5 minutos são produzidos até um horizonte de previsão de 5 a 75 minutos. (Kaur, Nonnenmacher, Pedro, e Coimbra, 2016).

Dessa maneira, o trabalho de previsão da irradiação e da geração de energia da fonte intermitente de energia solar, inclusive para o horizonte de previsão de 5 minutos, torna-se pertinente e com uma importância cada vez maior para integração dessa fonte no mercado de energia no Brasil e no mundo, e é para esse intervalo que o presente trabalho foi direcionado. A previsão da irradiação solar compõe um importante aliado contra a intermitência da irradiação solar, juntamente com banco de baterias que permitem armazenar a energia gerada.

Para iniciar uma previsão, é importante definir três parâmetros: Horizonte de previsão, Intervalo de previsão e Resolução da previsão (ou resolução espacial). Considerando que o mercado comece em  $t$ , então o Horizonte de previsão é definido como a quantidade de tempo antes de  $t$ , o intervalo de previsão é definido como o intervalo de tempo até o qual a previsão é emitida e a resolução de previsão indica as frequências com que as previsões são emitidas. (Kaur, Nonnenmacher, Pedro, e Coimbra, 2016).

O trabalho de (Das, et al., 2018) define as faixas para o Horizonte de previsão: Horizonte de previsão de muito curto prazo (*Very short-term forecast horizon*) é a previsão desde poucos segundos até muitos minutos (<1 h). Esse tipo de previsão é usado para suavização de energia, despacho de eletricidade em tempo real e reservas ideais. (Yang, Ye, Lim, e Dong, 2015) indica a previsão de muito curto prazo para aplicações em grandes instalações fotovoltaicas.

A Fig. 1 detalha a linha do tempo em tempo real do mercado energético na interconexão ocidental dos EUA, na Califórnia.

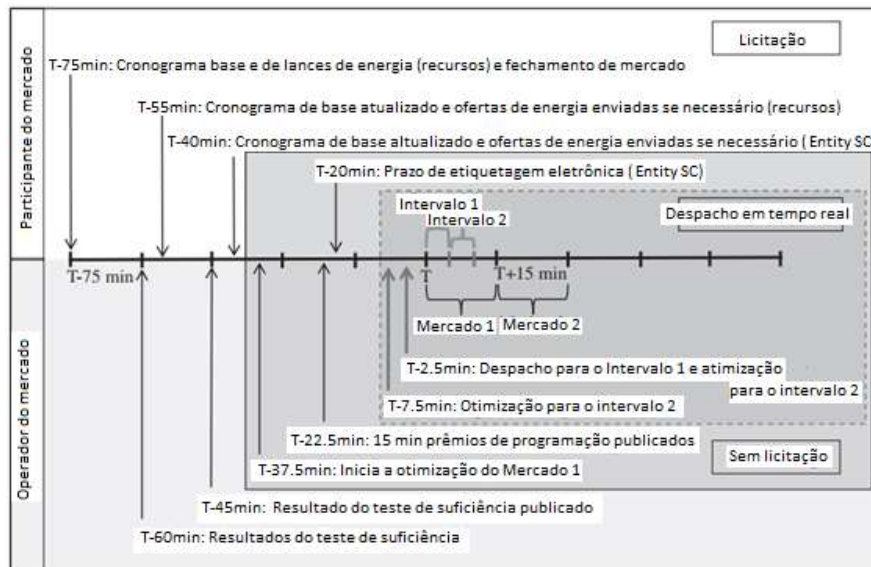


Figura 1- Linha do tempo do mercado energético na interconexão ocidental dos EUA.  
Fonte traduzida: (Kaur, Nonnenmacher, Pedro, e Coimbra, 2016).

É possível notar a importância da previsão intra-hora e de até 5 minutos para que o despacho de energia aconteça com segurança, tanto no mercado de energia quanto no mercado reserva.

Previsão de curto prazo (*Short-term forecasting*) é usado para previsão entre 1 h e 7 dias e garante a programação e despacho de energia elétrica sendo assim, útil para projetar a gestão de um sistema de energia. O trabalho de (Nobre, et al., 2016) aborda diferentes métodos de previsão para curto prazo. Previsão de Médio prazo (*Medium-term forecasting*) é realizada entre uma semana e um mês. Esse tipo de previsão facilita o planejamento do sistema de energia e o cronograma de manutenção, prevendo a disponibilidade da energia elétrica no futuro. Previsão de Longo prazo (*Long-term forecasting*) é feito de um mês até um ano sendo útil para planejar a geração, transmissão e distribuição de energia e sua operação de segurança.

## 2. MÉTODOS DE PREVISÃO

Na atualidade, existem diversas e distintas metodologias para previsão do recurso solar, sendo ampla a bibliografia disponível. A separação entre os métodos é dificultada pela enorme variedade de nomenclaturas utilizadas na comunidade científica (Yang, Kleissl, Gueymardc, Pedro, e Coimbra, 2018)

(Yang, Kleissl, Gueymardc, Pedro, e Coimbra, 2018), consideraram as publicações mais importantes das revistas científicas da área e conseguiram agrupar diferentes métodos em basicamente 5 grandes grupos: Séries temporais (*Times Series*), Regressão (*Regression*), Previsão numérica do tempo ou NWP (*Numerical weather prediction*), Aprendizado de máquina - ML (*Machine Learning*) e Previsão por base de imagem (*Image bases Forecasting*). O estudo (Domingos, Monteiro, e Boaventura, 2020) detalha os diferentes grupos de métodos.

Para o presente trabalho, foram estudados um método do grupo *Machine Learning* o ANN e outro do grupo de Regressão, o LASSO. Ambos os métodos já foram comparados por outros pesquisadores, porém com dados de naturezas distintas. O trabalho de (Lei, 2023) ao avaliar capacidade de suporte de membros de tubos retangulares de aço preenchidos com concreto comparou os mesmos métodos e verificou que o LASSO obteve um pior comportamento em relação ao método ANN. O trabalho de (Cahigas, 2023) utilizou dos métodos para examinar os fatores que influenciam as intenções dos turistas de visitar Siargao após o Supertufão Rai. Ao comparar apenas os modelos, LASSO teve o melhor desempenho e a combinação LASSO-ANN foi considerado o melhor algoritmo de aprendizado de máquina. O trabalho de (Zhu, 2022) usou desses métodos para prever o coeficiente de partição poldimetilsiloxano-ar (KPDMS-ar) na medida de concentrações de poluentes orgânicos persistentes (POPs) e LASSO e ANN tiveram resultados muito próximo, quase que idênticos. Inclusive as métricas para avaliação do comparativo foram incapazes de apontar diferenças entre eles. Dessa maneira, nota-se que o comportamento dos métodos depende da natureza dos dados.

### 2.0 Método de aprendizado de máquina ou ML (*Machine Learning*)

ML é importante devido à sua capacidade de executar mapeamentos não lineares arbitrários. Algoritmos ML baseiam-se no conceito de padrões de aprendizado e modelo dos dados, onde a aprendizagem implica classificação, regressão e previsão. Neste sentido, os algoritmos ML são adequados para a previsão da irradiação solar. Aplicações de previsão do recurso solar consistem em criar modelos preditivos para valores pontuais ou intervalos de previsão baseados em conjunto de dados de dados históricos.

Os métodos ML são técnicas estatísticas capazes de identificar tendências e reduzir, substancialmente, o viés em relação ao conjunto de validação. A robustez dos modelos ML depende da diversidade do conjunto de dados, o método e a habilidade das variáveis para capturar o espaço de soluções necessárias para reproduzir as saídas, e muito importante, a figura de mérito (como RMSE, NRMSE, MAE e MBE) usada para validar os resultados. Sobrepondo métodos ML com outros métodos quase sempre resulta em melhores previsões se um amplo conjunto de precisão de previsão métricas é utilizada.

**Método rede Neural Artificial ou ANN (Artificial Neural Network).** ANN é uma linha importante dentro do método ML e consiste em um processador paralelo maciço, composto de processamento simples (“neurônios”), que têm a capacidade de armazenar conhecimento (pesos sinápticos) e tornando-o disponível para uso posterior. Os principais benefícios das redes neurais podem ser ligados à sua capacidade de operar de forma não linear, o que os torna adequados para manipular fenômenos com modelos físicos complexos, que, na maioria dos casos, são difíceis ou impraticáveis para resolver através de funções lineares (Yang, Kleissl, Gueymardc, Pedro e Coimbra, 2018). A Eq. (1) define o modelo ML para 2 camadas ocultas.

$$h(x|w, b) = \sigma(w^4 \sigma(w^3 \sigma(w^2 x + b^2) + b^3) + b^4) \quad (1)$$

Onde:  $\sigma$  é uma função de ativação vetorizada,  $w$  é o peso sináptico da conexão com o neurônio e sua saída e  $b$  o viés do neurônio.

No trabalho de (Pedro e Coimbra, 2015) foi realizado previsões de 15 min a 2 h com a utilização dos métodos ANN e kNN (*k-Nearest Neighbor*) e se obteve resultados pertinentes, demonstrando melhorias em relação a métodos mais simples. Também mostra que os modelos são altamente dependentes para diferentes microclimas e que a alta densidade da rampa de irradiação e de séries temporais podem explicar com sucesso a boa performance da previsão tanto em diferentes localidades quanto em diferentes horizontes temporais.

No Brasil, o método ANN foi utilizado por (Martins, Pereira, e Guarnieri, 2012) para melhorar a previsão de curto prazo e obteve bons resultados.

O método ML e as técnicas estatísticas de previsão da irradiação solar são comumente usadas com uma resolução temporal de intra-hora e Previsão de curto prazo. (Yang, Quan, Disfani, e Liu, 2017) (Silva, 2019)

O método ML é uma das abordagens mais populares na previsão do recurso solar e o ANN foi selecionado para compor a simulação do presente estudo.

## 2.1 Método de Regressão (*Regression Method*)

O método da Regressão é um processo estatístico e o trabalho de (Das, et al., 2018) define Regressão como um método que é usado para estabelecer um relacionamento entre variáveis. Para o uso em modelo de previsão em sistemas fotovoltaicos, a potência prevista é considerada uma variável dependente e as variáveis meteorológicas são consideradas variáveis explicativas.

Nos métodos estatísticos espaço-temporais, o valor da irradiância futura para uma determinada estação é prevista baseada no passado e na irradiação solar atual no local determinado e nas estações vizinhas. O método LASSO pode ser usado dessa forma, como abordado no trabalho de (Aryaputera, Yang, Zhao, e Walsh, 2015).

A Regressão é um método utilizado para a previsão e modelagem de séries temporais. Os principais benefícios de se utilizar a regressão são: indicar a relação entre as variáveis, indicar a força do impacto de várias variáveis independentes em uma variável dependente e comparar os efeitos das variáveis medidas em diferentes escalas.

**Método de Regressão LASSO (*least absolute shrinkage and selection operator*).** O método LASSO tem sido usado na previsão solar ao lado de outros métodos de seleção de variáveis. O LASSO aos poucos tem sido usado para identificar preditores mais apropriados para um dado período (Yang, Ye, Lim e Dong, 2015), pois uma importante vantagem do LASSO em relação aos demais métodos de Regressão é a propriedade de *shrinkage* ou encolhimento, que tem a capacidade de zerar completamente uma variável se esta não for tão relevante para o contexto de previsão.

LASSO impõe uma restrição na soma dos valores absolutos dos parâmetros do modelo, inserindo como um limite superior, que faz com que os coeficientes de regressão para algumas variáveis sejam anulados. Este é o processo de encolhimento. O encolhimento permite uma melhor interpretação do modelo e identifica os preditores mais fortemente

associados com a variável dependente. Esse é o processo de seleção das variáveis. Ele vai obter o subconjunto de preditores que minimiza o erro de predição. (Prata, 2019).

De acordo com a publicação original de Tibshirani, LASSO minimiza a soma residual dos quadrados sujeito à soma do valor absoluto dos coeficientes sendo menos que uma constante. Devido à natureza dessa limitação tende-se a produzir alguns coeficientes que são exatamente zero e, conseqüentemente, resulta em modelos interpretáveis. (Tibshirani, 1996). A Eq. (2) é equivalente ao termo de regularização  $\ell_1$ :

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (2)$$

Onde  $\lambda$  é o parâmetro de sintonização com regulares forças de penalidade para o termo de regularização  $\ell_1$ ,  $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ , é o parâmetro de regressão,  $t \geq 0$  é o parâmetro de ajuste que controla quantidade de encolhimento ou shrinkage,  $(x^i, y_i)$  é o dado,  $i = 1, 2, \dots, n$ , onde  $x^i = (x_{i1}, \dots, x_{ip})^T$  são as variáveis preditoras,  $p$  (são variáveis independentes) e  $y_i$  são as respostas.

O método LASSO permite a inserção de diferentes preditores e o mesmo possui a capacidade de identificar qual preditor é relevante para a previsão e qual possui menor impacto. Uma importante vantagem da técnica LASSO é que ele permite modificar o preditor de forma instantânea e durante a aplicação do método. Isso é importante para a previsão de minutos, pois a mudança da direção do vento, por exemplo, pode acontecer a qualquer momento e não pode ser negligenciada.

LASSO apresenta um bom comportamento para previsões com resolução temporal de intra-hora como 5 minutos, funciona bem para microescala, permite a utilização de dados meteorológicos como preditores e o termo de regularização aliado ao encolhimento diminuem facilmente preditores possibilitando melhorar a precisão e simplificando o modelo. Dessa forma, LASSO foi o método escolhido para a realização desse estudo.

### 3. METODOLOGIA

Para a realização das simulações, foram utilizados dados coletados e disponibilizados gratuitamente pelo *National Renewable Energy Laboratory* (NREL) na *Measurement and Instrumentation Data Center* (MIDC) na estação de medição *Oahu Solar Measurement Grid* (Wilcox e Andreas, 2010), que fica localizada na ilha Oahu no arquipélago do Havaí, no Oceano Pacífico Norte.

Os dados de irradiação solar global de Oahu foram realizados em plano horizontal com uma taxa amostral de 1 em 1 segundo. Todos os piranômetros são do fabricante LI-COR, modelo LI-200, exceto o sensor denominado AP\_2 que é do fabricante *SOLRMAP Kalaeloa Oahu* RSR e o modelo R2R2. Em dados de alta frequência de amostragem, valores máximos e mínimos podem ocorrer causados por ruído e não por efeitos de nuvens, conforme explicado por (Yang, Ye, Lim, e Dong, 2015). Assim, os dados foram agregados em intervalos de 10 s. Para a previsão de 5 minutos, esta frequência de amostragem é suficiente e pode ser alterada para previsões com horizonte de tempo mais longo. Devido à alta qualidade dos dados fornecidos pelo NREL, não se fez necessário a aplicação de tratamento extra dos dados.

Para a aplicação desta simulação foram utilizados dados de um dia ensolarado com algumas nuvens e não foram fracionados os dados devido às condições de céu com e sem nebulosidade. Apesar de ser um intervalo pequeno para validar o método, para previsões de muito curto prazo, como 5 minutos, é suficiente, segundo com (Marquez e Coimbra, 2013).

Para realizar a previsão com 5 minutos de antecedência, empregou-se dados dos 75 minutos anteriores, que compõem o conjunto de treinamento (*training set*). Os 75 minutos são subdivididos em dados com intervalos de 10 segundos, totalizando 450 pontos. A Fig. 2 ilustra a utilização desses 450 dados (intervalo de 10 s), resultando em um somatório de 4500 segundos, equivalente a 75 minutos para cada sensor. A previsão é realizada em 30 pontos, formando um horizonte de 300 segundos, ou seja, 5 minutos. Estes dados de 300 segundos constituem o conjunto de teste (*test set*).

Em cada ciclo, são efetuadas seis previsões com 5 minutos de antecedência, acumulando um horizonte de previsão de 30 minutos. Dessa maneira, após treinar ambos os modelos com o mesmo conjunto de treinamento (75 minutos), a avaliação do desempenho ocorre no conjunto de teste (5 minutos).

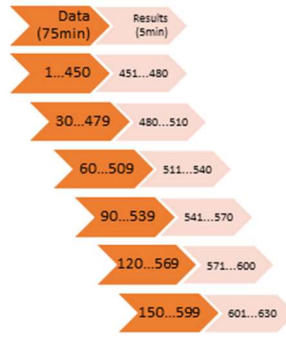


Figura 2- Exemplo de um ciclo de simulação de previsão.

A Fig. 3 ilustra a localização dos sensores, a direção preferencial do vento (seta vermelha) e a escala do mapa.

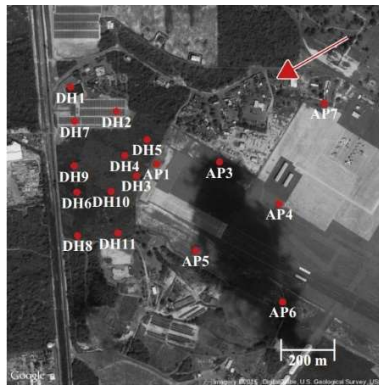


Figura 3- Localização dos sensores em Oahu. Fonte: (Yang, Ye, Lim, e Dong, 2015).

A previsão apresentada a seguir é baseada em dados de nove estações em relação aos dados do sensor DH10. (Os sensores utilizados são: DH9, DH6, DH8, AP5, AP6, AP4, AP1, DH3 e DH4).

#### 4. RESULTADOS

A Fig. 4 demonstra a propriedade do encolhimento do LASSO.

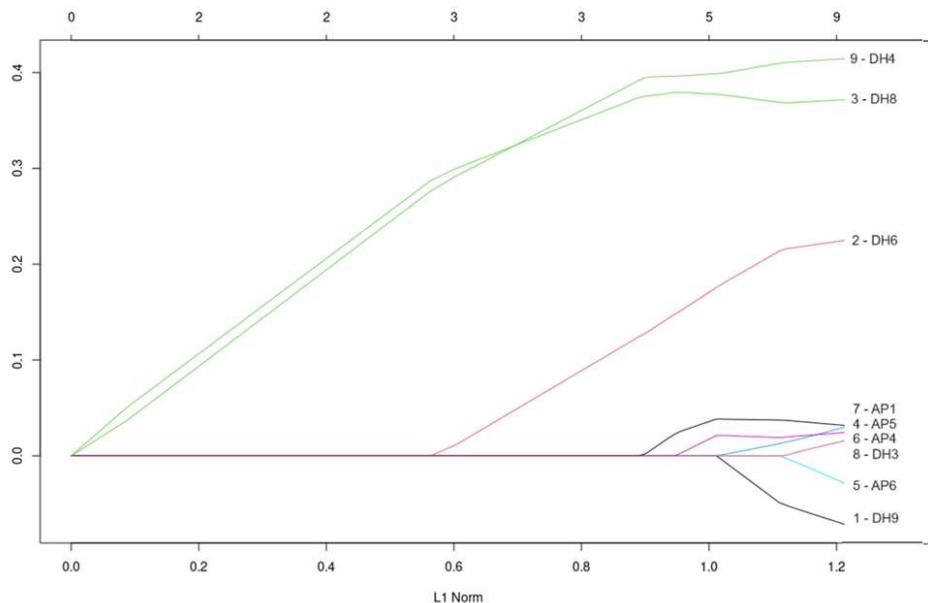


Figura 4 - Encolhimento para nove sensores.

Pela Fig. 4, é possível verificar que o sensor DH8 é o primeiro preditor a entrar no modelo e está contra o vento, e entra quase ao mesmo tempo que o DH4, que é um sensor mais próximo e está a favor do vento. O terceiro preditor é um sensor a favor do vento, também próximo de DH10, que é o DH6. Em sequência as estações AP1 e AP5 entram na sequência no modelo. A partir daí, as estações entram no modelo quase com o mesmo peso.

No caso de uma instalação futura de uma usina solar, na localização do sensor DH10, com a Fig. 4, é possível definir a melhor localização das estações para fazerem medições mais realísticas do que será a irradiação solar no local da DH10. Outra utilização possível para a Fig. 4 é definir quais os locais que mais influenciam o sensor DH10 e neles inserir medidores de irradiação solar. No caso seriam as estações DH8, DH4 e DH6. E um medidor no local onde o sensor AP6, por exemplo, pode ser dispensado.

Ainda para os mesmos preditores, foi aplicada a validação cruzada *k-fold* (onde é dividido os dados de entrada em subconjuntos de dados *k* ou *folds*. O teste acontece em todos, menos em um (*k-1*) dos conjuntos de dados e, em seguida, avalia-se o modelo no conjunto de dados que não foi usado para treinamento). A Fig. 5 mostra como é o erro para cada  $\log(\lambda)$ .

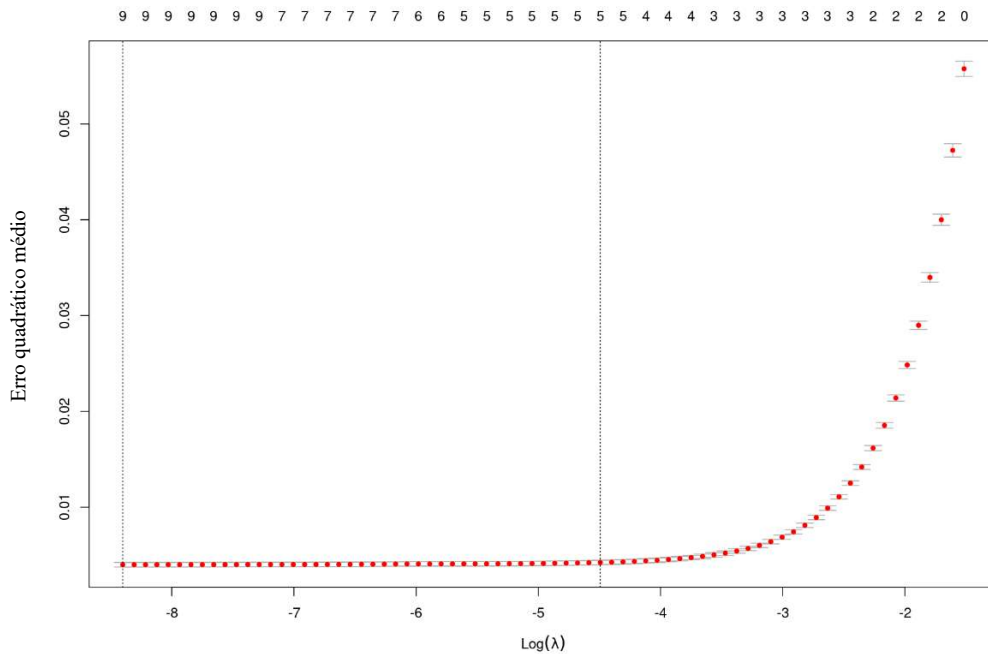


Figura 5 - Erros da Validação cruzada para a previsão do LASSO.

A Fig. 5 foi construída com dados dos nove sensores em relação à DH10, utilizando 10 dobras, ou  $n_{folds} = 10$  para a validação cruzada. Na Fig. 5 a curva de pontos vermelhos representa a validação cruzada e as barras representam o erro padrão durante a sequência de  $\log(\lambda)$ . A primeira linha tracejada localizada na esquerda do gráfico marca o valor de  $\lambda$  com o menor valor de erro médio, que para o exemplo corresponde ao valor de  $\lambda = 0.011184$ . Nesse ponto pode perceber-se que todos os 9 sensores estão no modelo (verificando a numeração na parte superior do gráfico) e que nesse ponto o modelo tem um menor erro (barras menores).

A outra linha tracejada (no centro do gráfico da Fig. 5) marca onde se encontra o modelo mais regularizado, de modo que o erro esteja dentro de um erro padrão do mínimo, que para o exemplo corresponde a  $\lambda = 0.00022472$ . É esse último valor de  $\lambda$  que será utilizado na previsão. Com a utilização desse  $\lambda$  apenas cinco sensores são utilizados para a previsão de 5 minutos à frente (são os sensores DH4, DH8, DH6, AP1 e AP5).

A Fig. 6 ilustra a comparação 5 min à frente, entre os dados previstos por ambos os métodos e com os dados do próprio sensor DH10, denominado I GHI.

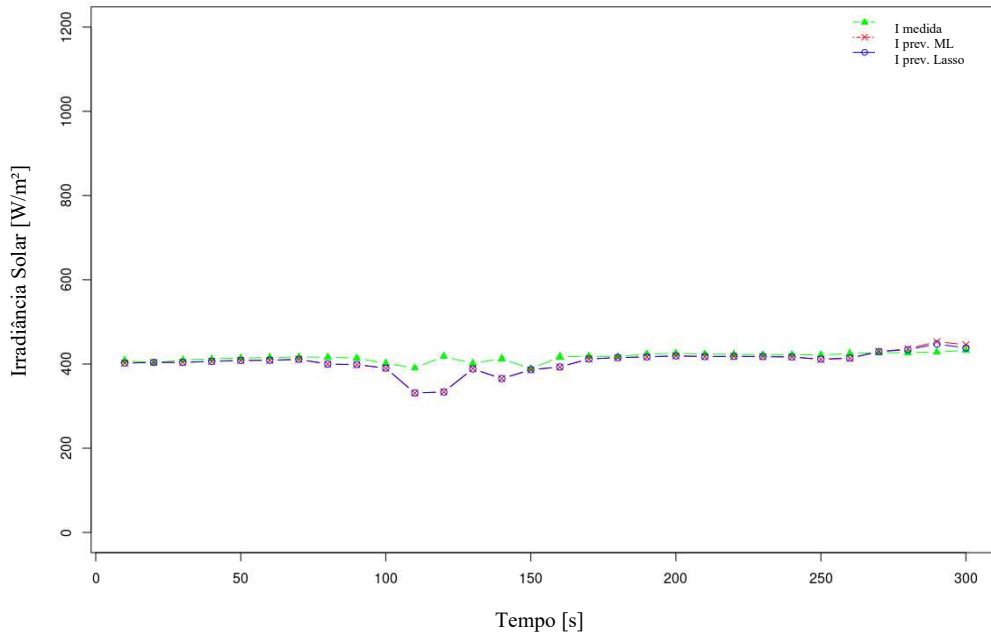


Figura 6 - Comparação de irradiação para previsão de 5 minutos para diferentes métodos em um intervalo de céu nublado.

Ambos os métodos selecionados para o presente trabalho apresentam um bom comportamento para previsões de curto prazo e acompanham bem a tendência da irradiação solar. Em momentos nublados, os métodos têm comportamentos ainda imprecisos, como ilustra a Fig. 7, ao mostrar a irradiação da previsão de 5 min em um intervalo acumulado de 30 min, para um ciclo de simulações de previsão.

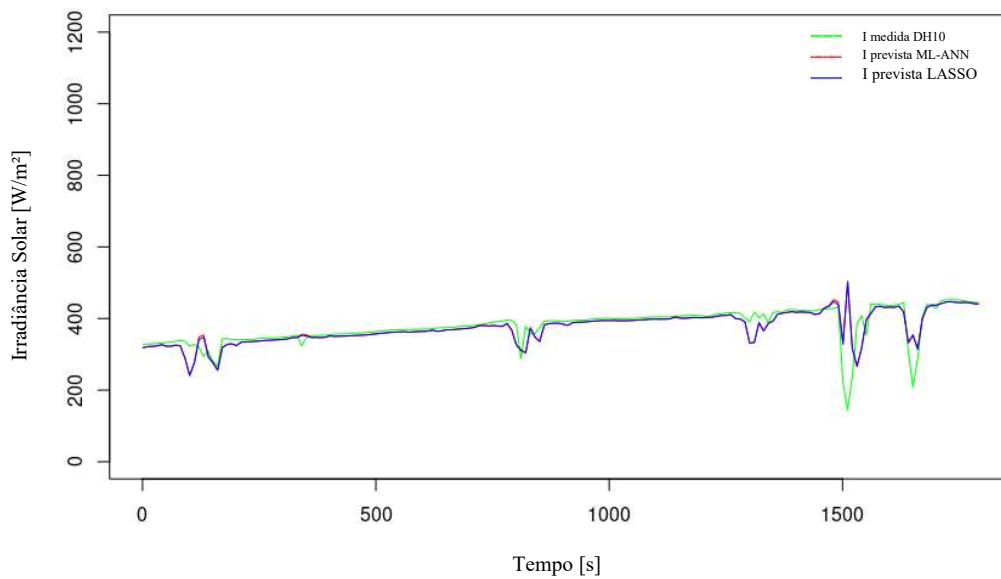


Figura 7 - Irradiação da previsão de 5 minutos em um intervalo acumulado de 30 minutos.

Em momentos sem nuvens, os métodos apresentam comportamento semelhantes gerando resultados consistentes, o que os torna capazes de trabalhar em locais com poucas influências de coberturas (nuvens).

A Fig. 8 ilustra a assertividade da previsão de 5 minutos em um intervalo acumulado de 30 minutos, mostrando o comportamento de cada método de previsão em relação aos dados originais de irradiação.

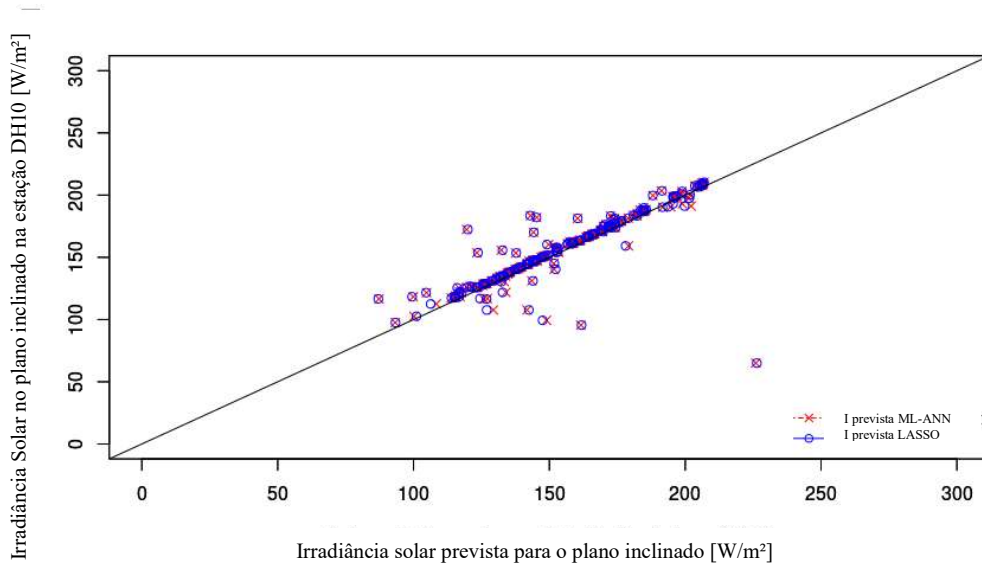


Figura 8 - Assertividade a partir da previsão de 5 minutos em intervalo acumulado de 30 minutos.

A Fig. 8 mostra que os pontos previstos sobre a linha têm uma assertividade de 100%. E quanto mais distantes da linha, menor proximidade entre os valores. Os pontos azuis são plotados com a relação entre a previsão do método LASSO e os dados do sensor DH10 e apresentam uma proximidade da linha. Os pontos vermelhos, são plotados com relação entre a previsão do método ANN e os dados do sensor DH10 e ambos têm comportamentos bem semelhantes. Ao analisar diferentes métricas estatísticas, a comparação entre métodos é mais clara, como mostra a Tab. 1.

Tabela 1: Métricas médias

MÉTODO	RMSE [W/m <sup>2</sup> ]	nRMSE [W/m <sup>2</sup> ]	MAE [W/m <sup>2</sup> ]	MBE
LASSO	27.95	0.06	10.53	-3.90
ML-ANN	28.42	0.07	10.56	-3.89

O RMSE tem uma escala de 0 a valores arbitrários, de acordo com (Eumetrain, 2020), e é a magnitude médias dos erros estimados e quanto mais próximo de zero, maior a qualidade dos valores estimados. Já o nRMSE, que é o RMSE normalizado também tem seu melhor valor próximo de zero. Dessa maneira ambos os métodos alcançaram admirável resultado. O MAE tem uma escala de 0 a valores arbitrários, de acordo com (Eumetrain, 2020), e quanto mais próximo de zero, maior a qualidade dos valores estimados. O MBE por ser um valor negativo, mostra que a previsão tem valores subestimados, e um valor tão próximo de 1 é um resultado muito bom.

Com as métricas estatísticas, é possível perceber com bastante clareza como o método LASSO se aproxima do ANN e ambos apresentam bons resultados para previsão de curto prazo com 5 minutos de antecedência para clima tropical. A Fig. 9 mostra o comportamento das métricas da Tab. 1.

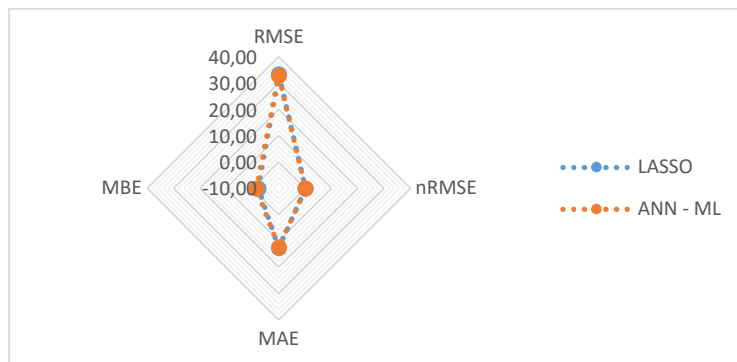


Figura 9 - Médias Métricas estatísticas.



A Fig. 9 mostra que o comportamento dos métodos é muito similar e a análise de diferentes métricas não se faz capaz de diferenciá-los.

## 5. CONCLUSÃO

O estudo comparativo entre o modelo consagrado ANN, dentro do *Machine Learning* e o modelo de regressão LASSO demonstrou que para previsão de curto prazo ambos os métodos são indicados, principalmente para regiões com poucas coberturas de nuvens.

As métricas utilizadas foram incapazes de demonstrar uma diferenciação relevante entre dos métodos de previsão estudados, mas indicam que ambos os métodos são efetivos para o cálculo da previsão de curto prazo. A métrica RMSE, que é muito sensível a pequenos desvios, idealmente se aproxima de zero, indicando um maior atributo da previsão, chegou a 27.95 W/m<sup>2</sup> para o LASSO e 28.42 W/m<sup>2</sup> para ANN-ML, que são valores muito satisfatórios para o RMSE. O seu valor normalizado, que é o nRMSE é capaz de suavizar pequenas falhas e tem seu melhor valor próximo de zero, mensurou 0.06 W/m<sup>2</sup> para o LASSO e 0.07 W/m<sup>2</sup> para o ANN-ML, que são resultados bastante interessantes. A métrica MAE tem também seu melhor valor próximo a zero e para o LASSO alcançou um valor de 10.53 W/m<sup>2</sup> e 10.56 W/m<sup>2</sup> para ANN-ML, que são valores baixos e muito similares, diferenciando apenas na segunda casa decimal. A métrica MBE, apresentou valores negativos para ambos os métodos, com -3,90 para o LASSO e -3,89 para ANN-ML, sinalizando que ambas as previsões têm valores subestimados, e pela proximidade do valor nulo, apresentam erros muito baixos.

Tanto ANN quanto o método Lasso podem fornecer ganhos de informação em relação a modelos simplórios, como persistência devido à sua capacidade de modelar relações mais complexas, adaptar-se aos dados e realizar seleção de características. As Redes Neurais Artificiais destacam-se por sua capacidade de aprendizado complexo, permitindo a identificação de relações não lineares e nuances nos dados. Sua adaptabilidade durante o treinamento é fundamental para ajustar pesos e otimizar o desempenho do modelo, tornando-as particularmente eficazes em cenários nos quais as relações entre variáveis são intrincadas e variáveis mudam ao longo do tempo. Porém, ao considerar a facilidade de implementação do método LASSO e a interpretação que os gráficos gerados detêm, ao indicar a relação entre as variáveis, indicar a força do impacto de várias variáveis independentes em uma variável dependente e comparar os efeitos das variáveis medidas em diferentes escalas, ele se torna uma opção viável e apreciável para a utilização para previsão de curto prazo. Ao introduzir um termo de regularização na função de custo, o Lasso promove paridade nos coeficientes do modelo, destacando a relevância das variáveis explicativas. Além disso, a penalização do Lasso desempenha um papel crucial na prevenção do *overfitting*, garantindo que o modelo não se torne excessivamente complexo e preservando sua capacidade de generalização para novos dados.

### *Agradecimentos*

Os autores agradecem ao PPGEE – Programa de Pós Graduação em Engenharia Elétrica da UFMG, à CAPES – Coordenadoria de Aperfeiçoamento de Pessoal de Ensino Superior, ao Projeto de P&D CEMIG-ANEEL D0727 – Análise de Arranjo Técnico e Comercial Baseado em uma Planta Piloto de Sistema Distribuído de Armazenamento de Energia em Alimentador Crítico da Rede de Distribuição de 13,8 kV – FASE II e à Pró-reitoria de Pesquisa da PUC Minas e Instituto Politécnico da PUC Minas pela viabilização e apoio financeiro para a realização deste trabalho.

## REFERÊNCIAS

- Aryaputera, A. W., Yang, D., Zhao, L., & Walsh, W. M. (november de 2015). Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging. *Solar Energy*, 122, pp. 1266-1278.
- Cahigas M. M. L., A. K. (23 de May de 2023). Super Typhoon Rai's Impacts on Siargao Tourism: Deciphering Tourists' Revisit Intentions through Machine-Learning Algorithms. *Sustainability*, 15 (11). doi:<https://doi.org/10.3390/su15118463>
- Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y., Deventer, W. V., . . . Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 81, pp. 912–928.
- Domingos, S. F., Monteiro, L. G., & Boaventura, W. d. (2020). ESTADO DA ARTE PARA PREVISÃO DA RADIAÇÃO SOLAR. *VIII Congresso Brasileiro de Energia Solar*. Fortaleza. Acesso em 29 de 11 de 2023, disponível em <https://anaiscbens.emnuvens.com.br/cbens/article/view/742/742>
- Eumetrain. (2020). Acesso em 13 de 03 de 2020, disponível em Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE): [http://www.eumetrain.org/data/4/451/english/msg/ver\\_cont\\_var/uos3/uos3\\_ko1.htm](http://www.eumetrain.org/data/4/451/english/msg/ver_cont_var/uos3/uos3_ko1.htm)

- Kaur, A., Nonnenmacher, L., Pedro, H. T., & Coimbra, C. F. (2016). Benefits of solar forecasting for energy imbalance markets. *Renewable Energy*(86), 819 - 830.
- Lei, J. X. (2023). Machine Learning-Based Prediction of Axial Load Bearing Capacity for CFRST Columns. *Latin American journal of Solids and Structures, solids struct* 20 (8). doi:<https://doi.org/10.1590/1679-78257807>
- Marquez, R., & Coimbra, C. F. (2013). Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*(91), 327–336.
- Martins, F. R., Pereira, E. B., & Guarnieri, R. A. (6 de december de 2012). Solar Radiation Forecast Using Artificial Neural Networks. *International Journal of Energy Science, 2Iss*.
- MME, ONS. (2023). *MODELO DE TERMO DE REFERÊNCIA - Projeto Meta - Subprojeto 24-1 – Previsão de Geração Solar Fotovoltaica de Curtíssimo e Curto*. Acesso em 29 de 11 de 2023, disponível em <https://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/Termo%20de%20Refer%c3%aaancia%20Projeto%20Meta%20-%20SP%2024%20Consultoria%201.pdf>
- Nobre, A., A.Severiano Jr., C., Karthik, S., Kubis, M., Zhao, L., R.Martins, F., . . . Reindl, T. (August de 2016). PV power conversion and short-term forecasting in a tropical, densely-built environment in Singapore. *Renewable Energy, 94*, pp. 196-509.
- Pedro, H. T., & Coimbra, C. F. (2015). Short-term irradiance forecastability for various solar micro-climates. *Solar Energy, 122*, 587-602.
- Portal Solar. (30 de outubro de 2017). Fonte: <https://www.portalsolar.com.br/o-que-e-geracao-distribuida.html>
- Prata, D. (2019). *Aprendizagem de máquina: Regressão Lasso*. (Universidade Federal de Tocantins) Acesso em 01 de 07 de 2019, disponível em <http://amuft.wikidot.com/aula10>
- Silva, C. Q. (2019). *Redes neurais aplicadas no reconhecimento de símbolos matemáticos manuscritos online*. Trabalho de conclusão de curso, universidade de São Paulo, Instituto de matemática e estatística, São Paulo. Acesso em 12 de 02 de 2020, disponível em <https://www.ime.usp.br/~map/tcc/2019/CaiqueQuaresmaV2.pdf>
- Tibshirani, R. (1996). Regression Shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, 58*(1), pp. 267-288. Acesso em 16 de 01 de 2020, disponível em <http://statweb.stanford.edu/~tibs/lasso/lasso.pdf>
- Wilcox, S., & Andreas, A. (2010). *Solar Resource & Meteorological Assessment Project (SOLRMAP): Rotating*. NREL Report No. DA-5500-56497. doi:<http://dx.doi.org/10.5439/1052228>
- Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T., & Coimbra, C. F. (2018). History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Solar Energy, 168*, 60-101.
- Yang, D., Quan, H., Disfani, V. R., & Liu, L. (April de 2017). Reconciling solar forecasts: Geographical hierarchy. *Solar Energy, 146*, pp. 276-286.
- Yang, D., Ye, Z., Lim, L. H., & Dong, Z. (2015). Very short term irradiance forecasting using the lasso. *Solar Energy*(114), 314–326.
- Zhu T., C. T. (5 de February de 2022). Prediction models with multiple machine learning algorithms for POPs: The calculation of PDMS-air partition coefficient from molecular descriptor. *Journal of Hazardous Materials, 423, Part B*. Acesso em 29 de 01 de 2024, disponível em <https://www-sciencedirect.ez27.periodicos.capes.gov.br/science/article/pii/S0304389421020057>

## COMPARATIVE STUDY BETWEEN LASSO AND MACHINE LEARNING METHODS FOR PREDICTING SHORT-TERM SOLAR IRRADIATION

**Abstract.** *The short-term solar irradiance forecast, especially the 5-minute ahead forecast, is suitable for use in power dispatch and its accuracy defines the reliability of the system. Machine Learning - ML is a currently established method for predictions of different natures and its performance was compared to the LASSO method (least absolute shrinkage and selection operator), within the Regression model, for predicting short-term solar irradiation. LASSO can ability to identify more appropriate predictors because it has the shrinkage property, which has the ability to completely reset a variable if it is not so relevant to the prediction context. The simulation was carried out in the R software and, for this purpose, public data from the NREL laboratory was used, measured in Oahu, Hawaii, with solar irradiation measurements from nine solarimetric stations, with a sampling rate at intervals of 1 second. Four notable metrics (RMSE, nRMSE, MBE and MAE) were used to compare the methods. Both methods have proven effective for short-term forecasting. LASSO achieved an RMSE of 27.95 W/m<sup>2</sup>, nRMSE of 0.06 W/m<sup>2</sup>, MAE of 10.53 W/m<sup>2</sup> and MBE of -3.90. The ANN achieved an RMSE of 28.42 W/m<sup>2</sup>, nRMSE of 0.07 W/m<sup>2</sup>, MAE of 10.56 W/m<sup>2</sup> and MBE of -3.89. However, due to the interpretations that the LASSO method provides, the method has the potential to be equivalent to Machine Learning, for the forecast horizon under study, in this case, short-term forecast.*

**Keywords:** Solar irradiation, Forecast, LASSO, Machine Learning.