

SITE ADAPTATION DA RADIAÇÃO GLOBAL HORIZONTAL PARA ESTAÇÕES SOLARIMÉTRICAS NA ARGENTINA E NO BRASIL

Diego Rodrigues de Miranda – diego.rmiranda@ufpe.br

Renan Soares Siqueira Costa – renan.soares@ufpe.br

Olga de Castro Vilela – olga.vilela@ufpe.br

Alexandre Carlos Araújo da Costa – alexandre.acosta@ufpe.br

Universidade Federal de Pernambuco, Centro de Energias Renováveis (CER-UFPE)

Germán Ariel Salazar – german.salazar@conicet.gov.ar

Universidad Nacional de Salta, Grupo de Estudio y Evaluación de la Radiación Solar (GEERS - INENCO – CONICET)

Alex Coutinho Pereira – alexcp@chesf.gov.br

Eduardo Boudoux Jatoba – ejatoba@chesf.gov.br

Alcides Codeceira Neto – alcidesc@chesf.gov.br

José Bione de Melo Filho – jbionef@chesf.gov.br

Companhia Hidro Elétrica do São Francisco (Chesf)

Resumo. Estimar a radiação solar incidente sobre certa localidade durante vários anos é necessário para o desenvolvimento de projetos fotovoltaicos e heliotérmicos de geração centralizada. Neste trabalho, técnicas estáticas de regressão linear múltipla, redes neurais do tipo multilayer perceptron e quantile mapping são empregadas para ajustar séries de irradiância global horizontal (GHI) de longo prazo, utilizando medições em terra, procedimento conhecido como site adaptation. São utilizadas duas estações solarimétricas, uma na cidade de Petrolina, no Nordeste do Brasil, e outra na província de Salta, no sítio de El Rosal, Argentina, como estudos de caso. Dados da Copernicus Atmosphere Monitoring Service (CAMS) e do European Centre for Medium-Range Weather Forecasts (ECMWF) são empregadas como variáveis regressoras nos modelos estatísticos. São testadas duas diferentes divisões dos dados em calibração (80% das amostras) e teste/operação (20% das amostras), uma delas realizando uma randomização temporal dos dados e a outra mantendo a componente temporal. Os resultados mostram que o site adaptation consegue melhores estatísticos para a estação de El Rosal (24,28% de nRMSD, -5,18% de nMBE e 1,00 de STDRatio), do que o modelo de referência da CAMS para GHI (34,67% de nRMSD, 19,28% de nMBE e 0,95 de STDRatio), enquanto que para a estação de Petrolina, os resultados obtidos (21,66% de nRMSD, 0,51% de nMBE e 0,92 de STDRatio) conseguiram resultados similares aos do modelo da CAMS (22,22% de nRMSD, 4,84% de nMBE e 0,98 de STDRatio); todos os estatísticos apresentados acima são para a divisão dos dados que mantém a componente temporal. Tanto para a estação de Petrolina quanto para El Rosal, os resultados obtidos com a divisão com randomização dos dados são mais acurados do que a que mantém a componente temporal.

Palavras-chave: site adaptation, irradiância global horizontal

1. INTRODUÇÃO

No desenvolvimento de projetos fotovoltaicos ou heliotérmicos de grande porte, uma boa estimativa do recurso solar da região de interesse é necessária para obter estimativas de produção de energia mais acuradas. Para tanto, campanhas de medições de radiação solar, temperatura, umidade e velocidade do vento, entre outras variáveis meteorológicas, são realizadas nos locais específicos com vistas à implementação de empreendimentos de energia. Atualmente no Brasil, as campanhas de medições para projetos fotovoltaicos de grande porte devem ter duração mínima de 1 ano, sendo a principal medição a da irradiância global horizontal (EPE, 2017). Já para projetos heliotérmicos ou projetos fotovoltaicos com concentração, ambos também de grande porte (mais que 5 MW), a principal medição é a da irradiância direta normal por, no mínimo, três anos. As instruções direcionadas aos empreendimentos de energia solar centralizada, o que inclui as instruções associadas às campanhas de medições, são publicadas pela Empresa de Pesquisa Energética para possibilitar às empresas a participação nos leilões de energia promovidos pela Agência Nacional de Energia Elétrica (ANEEL).

Contudo, para estimar o recurso solar em dada região, não basta somente realizar a campanha de medições no local específico, mas estimar o comportamento climatológico das variáveis de interesse. Como a fonte solar é intermitente, com variações estocásticas ao longo do tempo, a caracterização do recurso deve levar em consideração uma janela temporal de, no mínimo, 10 anos, período muitas vezes denominado como climatologicamente significativo (SENGUPTA *et al.*, 2017). Para isso, bases de dados históricas que fornecem estimativas de variáveis ambientais durante vários anos podem ser utilizadas. Como medir por tanto tempo é inviável, modelos estatísticos surgem como uma forma de adaptar as séries climatológicas fornecidas pelas bases de dados utilizando as medições realizadas em terra, procedimento conhecido na literatura como *site adaptation* (POLO *et al.*, 2016).

As séries históricas de radiação podem ser provenientes tanto de modelos de previsão numérica do tempo (*Numerical Weather Prediction* - NWP) ou reanálise, a exemplo da base de dados do *European Centre for Medium-Range Weather Forecasts* (ECMWF, e.g. modelos ERA-5 e ERA-5 Land) e da *National Aeronautics and Space Administration* (NASA, e.g. modelo MERRA-2), quanto a partir de modelos que utilizem imagens de satélite (e.g. PVGIS, CERES, NSDBR, CAMS). Atualmente, os modelos baseados em imagens de satélites são mais utilizados para avaliação do recurso solar, pois apresentam estimativas mais acuradas que modelos de NWP. Apesar disso, diferenças significativas ainda são encontradas entre as séries fornecidas por modelos baseados em imagens de satélites e as séries medidas por estações solarimétricas, principalmente para radiação direta normal. Fernández-Peruchena *et al.* (2020) citam quatro principais fontes de erros de modelos baseados em imagens de satélites: o efeito das nuvens na atenuação da radiação solar; o modelo de céu claro utilizado; a área do pixel do satélite que pode levar a erros principalmente durante dias nublados (nuvens intermitentes) e devido a mudanças na quantidade de aerossóis presentes na atmosfera; e os erros devido a terrenos elevados e altos albedos provenientes de desertos ou neve. Os efeitos desses potenciais erros podem resultar em séries com um *bias* e desvio (amplitude de variação) expressivos. Uma possível solução para superar esses potenciais erros é fazer o *site adaptation* das séries fornecidas pelas bases históricas.

No *site adaptation*, as séries de longo prazo devem ter uma cobertura temporal que inclua o período da campanha de medições realizada em terra. Assim, calibram-se e validam-se os modelos estatísticos no período concomitante entre as medições realizadas pela estação solarimétrica no local de interesse e as bases de dados históricas. Para gerar a série histórica adaptada ao local é necessário aplicar o modelo estatístico calibrado e validado à toda a série histórica. Essa série de longo prazo ajustada pode ser utilizada em diferentes aplicações. Em termos de empreendimentos solares, a série histórica é utilizada para o cálculo da incerteza da variabilidade interanual (KARIUKI E SATO, 2018) e para geração de um ano meteorológico típico da região de interesse (WILCOX E MARION, 2008), fatores cruciais para a viabilidade econômica dos projetos. Além disso, a série pode ser utilizada para o preenchimento das lacunas de campanhas de medições (SCHWANDT *et al.*, 2014), bem como em previsões da potência de saída de um empreendimento solar.

Nesse sentido, modelos regressivos ou modelos baseados em *quantile mapping* (QM) são largamente utilizados para adaptar as séries no período concomitante entre as medições em terra e as saídas das bases de dados históricas. As séries utilizadas nesses métodos estatísticos podem incluir saídas de modelos de satélite (irradiâncias global horizontal, difusa horizontal e direta normal), saídas de NWP (no geral, somente a irradiância global horizontal) e outras variáveis (ângulo zenital, hora solar aparente, entre outras). Neste trabalho, modelos regressivos baseados em regressões lineares e redes neurais do tipo *multilayer perceptron* são utilizados para realizar o *site adaptation* utilizando como *target* as medições das estações solarimétricas analisadas e, como variáveis regressoras, informações fornecidas pela CAMS e pelo ECMWF. Como a CAMS utiliza imagens do satélite METEOSAT, que têm uma resolução temporal de 15 min, a escala de tempo que será utilizada para realizar o *site adaptation* também será de 15 min.

2. BASES DE DADOS UTILIZADAS

Duas estações solarimétricas são utilizadas para aplicação do *site adaptation*, uma estação localizada no sítio de El Rosal na província de Salta, na Argentina, e outra estação localizada no município de Petrolina, no Brasil. As características de ambas são apresentadas na Tabela 1. Em relação às variáveis das bases de dados históricos, foram utilizadas variáveis provenientes do modelo da CAMS *Radiation Service* que fornece séries temporais de radiação de acordo com o campo de visão do satélite METEOSAT de -66° a 66° em ambas latitude e longitude ([CAMS radiation service - www.soda-pro.com](http://www.soda-pro.com)) e interpolações das saídas do ERA5-Land do ECMWF ([ERA5-Land hourly data from 1981 to present \(copernicus.eu\)](http://era5-climate.copernicus.eu)). A CAMS fornece as séries temporais já interpoladas para a região de interesse na escala temporal de minutos até meses; neste trabalho, foram utilizadas as saídas da CAMS na escala dos minutos e integradas para a escala de 15 min. Já para variáveis do ECMWF, foi utilizada uma interpolação bilinear simples para obter as séries temporais na coordenada de interesse e uma interpolação temporal simples para obter as séries de 15 min a partir das séries horárias. A Tabela 2 mostra todas as variáveis das bases de dados utilizadas neste trabalho.

Tabela 1 – Informações sobre as duas estações solarimétricas utilizadas.

Localização/Cidade	El Rosal	Petrolina
Província/Estado	Salta	Pernambuco
País	Argentina	Brasil
Latitude (°)	-24,393	-9,107
Longitude (°)	-65,7683	-40,442
Altitude (m)	3355	376
Piranômetro	Kipp&Zonen CMP3	EKO MS-80
Intervalo de Integração	1 min	1 min
Datalogger	Campbell CR1000	Campbell CR1000
Período de medição	2013 - 2016	2018 - 2021

Fonte: própria.

Tabela 2 – Variáveis utilizadas para realização do site adaptation.

1	TOA	Irradiação horizontal no topo da atmosfera (Wh/m ²)	CAMS
2	Clear sky GHI	Irradiação global horizontal de céu claro (Wh/m ²)	CAMS
3	Clear sky BHI	Irradiação direta horizontal de céu claro (Wh/m ²)	CAMS
4	Clear sky DHI	Irradiação difusa horizontal de céu claro (Wh/m ²)	CAMS
5	Clear sky BNI	Irradiação direta normal de céu claro (Wh/m ²)	CAMS
6	GHI	Irradiação global horizontal (Wh/m ²)	CAMS
7	BHI	Irradiação direta horizontal (Wh/m ²)	CAMS
8	DHI	Irradiação difusa horizontal (Wh/m ²)	CAMS
9	BNI	Irradiação direta normal (Wh/m ²)	CAMS
10	Sza	Ângulo zenital (°)	CAMS
11	summer/winter split	Divisão inverno/verão (1 - verão e 2 - inverno)	CAMS
12	tco3	Coluna total de ozônio (Dobson unit)	CAMS
13	Tcwv	Coluna total do vapor de água (kg/m ²)	CAMS
14	AOD BC	Profundidade óptica do aerossol a 550 nm para carbono negro	CAMS
15	AOD DU	Profundidade óptica do aerossol a 550 nm para poeira	CAMS
16	AOD SS	Profundidade óptica do aerossol a 550 nm para sal marinho	CAMS
17	AOD OR	Profundidade óptica do aerossol a 550 nm para matéria orgânica	CAMS
18	AOD SU	Profundidade óptica do aerossol a 550 nm para sulfato	CAMS
19	Fiso	MODIS BRDF parâmetro fiso	CAMS
20	Fvol	MODIS BRDF parâmetro fvol	CAMS
21	Fgeo	MODIS BRDF parâmetro fgeo	CAMS
22	Albedo	Albedo	CAMS
23	Cloud optical depth	Valor do tempo de aquisição mais próximo do pixel	CAMS
24	Cloud coverage of the pixel	Percentual do valor do tempo de aquisição mais próximo do pixel	CAMS
25	Vel	Velocidade do Vento (m/s) a 10 m	ERA5-Land
26	d2m	Temperatura do ponto de orvalho a 2 m (K)	ERA5-Land
27	t2m	Temperatura ambiente a 2 m (K)	ERA5-Land
28	Skt	Temperatura de uma fina camada da superfície em equilíbrio radiativo (K)	ERA5-Land
29	Sp	Pressão atmosférica na superfície (Pas)	ERA5-Land
30	Ssrd	Irradiância global horizontal (W/m ²)	ERA5-Land
31	Strd	Irradiância térmica da superfície (W/m ²)	ERA5-Land

Fonte: própria.

3. METODOLOGIA

Considerando o estado da arte no que se refere ao *site adaptation* das variáveis associadas à energia solar, a metodologia deste trabalho propõe a realização de um *site adaptation* em 4 etapas: i) divisão dos dados provenientes das bases históricas nos conjuntos de treinamento e operação/teste; ii) seleção de variáveis e/ou extração de variáveis das bases de dados históricas utilizando somente os dados pertencentes ao conjunto de treinamento; iii) aplicação de modelos regressivos e modelos baseados em *quantile mapping* para realização do *site adaptation*; vi) avaliação dos resultados no conjunto de operação/teste.

Neste trabalho, a base de dados histórica escolhida para radiação foi a da [Copernicus Atmosphere Monitoring Service \(CAMS Radiation Service\)](#). Salazar *et al.* (2020) valida a base de dados da CAMS para a cidade de Petrolina, Pernambuco, e afirma que, dentre as 11 bases de dados de radiação solar analisadas, a da CAMS é a mais consistente para utilização na região. Como os dados empregados neste trabalho englobam uma estação solarimétrica no município de Petrolina – PE, escolheu-se a base de dados da CAMS para realização do *site adaptation*. A CAMS fornece séries temporais de irradiância global horizontal, difusa horizontal, direta normal e as respectivas séries de irradiância em condições de céu claro no período de fevereiro de 2004 até dois dias anteriores à data atual, em escalas temporais de minutos, 15 minutos e horários e com resolução espacial interpolada para o ponto de interesse. Os dados são fornecidos pelo serviço atmosférico *Copernicus*, um programa europeu para desenvolver a capacidade de observação europeia do planeta com respeito ao monitoramento terrestre, marinho e atmosférico, planejamentos de emergência, segurança e mudanças climáticas. É utilizado um modelo físico que utiliza aproximações do modelo de transferência radiativa, o método do Heliosat-4 (QU *et al.*, 2017). O método é baseado no modelo *McClear* (LEFÈVRE *et al.*, 2013) para cálculo da radiação em condições de céu claro e no modelo *McCloud* (SCHROEDTER-HOMSCHEIDT *et al.*, 2019) para calcular a atenuação da radiação devido à presença de nuvens. As principais entradas do método Heliosat-4 são propriedades dos aerossóis,

coluna total de vapor d'água e coluna total de ozônio, que são fornecidos pelo serviço global da CAMS com uma resolução temporal de 3 horas e resolução espacial de 0,8° (até 20 de Junho de 2016) e 0,4° (desde 21 de Junho de 2016). Propriedades das nuvens são obtidas a partir da segunda geração dos satélites METEOSAT, que fornece imagens a cada 15 minutos em uma resolução espacial de 3 km a 10 km.

Os dados medidos de radiação utilizados passaram por um procedimento de garantia de qualidade, conforme descrito em Petribú *et al.* (2017), o que inclui os filtros de qualidade recomendados pela *Baseline Surface Radiation Network* (BSRN). Os dados são, então, integralizados para uma escala de 15 min, de acordo com os procedimentos descritos por Salazar *et al.* (2020) e Roesch *et al.* (2011). Em seguida, modelos estatísticos baseados em métodos regressivos e correções das distribuições de probabilidades das variáveis (*quantile mapping*) são aplicados para realização do *site adaptation*. Os modelos são validados por avaliação estatística.

Foi adotado neste trabalho duas diferentes divisões dos dados em calibração e teste/operação, uma que faz uma randomização de todos os conjuntos antes de dividi-los (Divisão 1) e outra que mantém a componente temporal no conjunto de teste/operação (Divisão 2), conforme Tabela 3. Os subconjuntos utilizam aproximadamente 80% dos dados para calibração e 20% para teste/operação. No conjunto de calibração, são feitas a seleção/extração de variáveis e o treinamento dos modelos regressivos utilizados no *site adaptation*, enquanto no conjunto de teste/operação é feita a aplicação da seleção/extração de variáveis definidas no conjunto de calibração e avaliação estatística dos resultados.

Tabela 3 – Exemplos da divisão nos conjuntos de calibração e teste/operação.

Divisão nos conjuntos de calibração e teste/operação*	Calibração (~80% da série)	Teste/Operação (~20% da série)
Divisão 1: <i>Time-steps</i> de 15 min escolhidos de forma aleatória	'26-Jul-2016 17:45'	'28-Feb-2013 10:15'
	'13-Jan-2015 10:00'	'10-Mar-2016 18:00'
	'31-Jul-2014 12:30'	'08-Abr-2013 18:30'
Divisão 2: Teste/operação corresponde aos 20% iniciais da série, mantendo a ordem temporal	'05-Jan-2014 12:30'	'01-Jan-2013 07:30'
	'09-Jul-2014 14:30'	'01-Jan-2013 07:45'
	'11-Feb-2014 18:30'	'01-Jan-2013 08:00'

*A série temporal adotada nesse exemplo compreende o período de '01-Jan-2013 07:30:00' até '28-Dez-2016 19:00:00'. Fonte: própria.

Após dividir os dados nos dois conjuntos citados acima, deve-se, primeiro, selecionar/extrair as variáveis mais relevantes do domínio estudado. Para tanto, podem ser utilizados tanto algoritmos de seleção de variáveis (*feature selection*), como filtros baseados em coeficientes estatísticos, quanto algoritmos de extração de variáveis, a exemplo de técnicas de redução de dimensionalidade como a análise de componentes principais. A técnica aqui utilizada para seleção de variáveis é um filtro baseado na correlação de *Pearson*. Diferente do que é proposto nos algoritmos de *feature selection* baseados na correlação (HALL, 1999; YU E LIU, 2004), neste trabalho é aplicado um filtro baseado na correlação de *Pearson* entre as *features* e o *target* individualmente, de forma a selecionar as variáveis com maior correlação com o *target*. No geral, os algoritmos de seleção buscam os grupos de variáveis que possuam maior correlação com o *target* e menor correlação entre si. Para eliminar a redundância dos grupos de variáveis selecionados, o que inclui as *features* que tenham uma alta correlação entre si, é adotado o algoritmo de redução de dimensionalidade de análise de componentes principais logo após a seleção de variáveis para alguns dos modelos de regressão utilizados no *site adaptation*. A Análise de Componentes Principais (*Principal Component Analysis* – PCA) é uma técnica de redução da dimensionalidade que pode ser aplicada a conjuntos de dados com muitas variáveis redundantes, ou seja, correlacionadas entre si. Para tanto, uma transformação linear é aplicada ao conjunto original dos dados, obtendo um novo conjunto, o das componentes principais, menor que o original e que explica boa parte da variância do conjunto original, tal como descrito em Jolliffe (1986). Neste trabalho, utilizam-se as primeiras componentes principais que explicam mais do que 95% do conjunto original dos dados.

Para o *site adaptation*, os modelos de regressão linear múltipla e redes neurais são utilizados, bem como a combinação proposta por Fernández-Peruchena *et al.* (2020) de utilizar sequencialmente um modelo regressivo seguido pelo *quantile mapping* (QM). Na regressão linear (WILKS, 2013), preditores (e.g., dados provenientes de modelos de satélites ou dados atmosféricos de reanálise) são utilizados em uma estimativa linear de um determinado preditando (e.g., dados observacionais de radiação solar ou temperatura). A estimativa dos coeficientes da regressão é feita empregando o método dos mínimos quadrados, que busca minimizar o erro quadrático médio (*Mean Square Error* – MSE) entre a variável observada e a variável estimada pelo modelo regressivo.

A técnica de QM aplica uma transformação inversa de forma a utilizar a curva de distribuição acumulada dos dados observados como referência para corrigir a curva dos dados modelados. O ajuste pode ser feito tanto utilizando funções paramétricas (*Cumulative Distribution Functions* – CDFs) quanto funções empíricas (*empirical cumulative density distribution, ecdfs*). Neste trabalho, a abordagem empírica é utilizada. Para gerar a *ecdf*, as séries das variáveis envolvidas devem ser ordenadas do menor valor para o maior e deve ser calculada a frequência de ocorrência de um certo valor na série, conforme Equação 1, onde n corresponde a quantidade de elementos da série temporal (FERNÁNDEZ-PERUCHENA *et al.*, 2020). Para corrigir a distribuição dos dados modelados, basta então aplicar a transformação inversa. Na Equação 2, y_c é a série da variável modelada corrigida.

$$ecfd = \frac{1}{n}p \quad p = 1,2,3, \dots, n \quad (1)$$

$$y_c = ecfd_{obs}^{-1}(ecfd_{mod}(x_{mod})) \quad (2)$$

Por fim, para as redes neurais do tipo *multilayer perceptron*, foi utilizada a função de ativação tangente hiperbólica com algoritmo de minimização de erro de Levenberg-Marquardt (GAVIN *et al.*, 2020). A arquitetura da rede *multilayer perceptron* considera uma camada de entrada (*input layer*), três camadas ocultas (com 5 neurônios na primeira camada, 4 na segunda e 3 na terceira – *hidden layers*) e uma camada de saída. A quantidade de épocas e inicializações utilizadas foram de 50 e 100, respectivamente.

Para comparar os modelos, será utilizado o diagrama proposto por Taylor (2001), pois ele é uma ferramenta útil para análise estatística gráfica. Os estatísticos associados ao diagrama são o desvio padrão, a correlação e o *skill score* (SS4). Nos diagramas apresentados neste trabalho, o ponto em magenta representa a observação e a linha que parte dele representa o desvio padrão da observação. Quanto mais próximo da linha em magenta, melhor os modelos representam o desvio padrão da série de dados observacionais. Quanto mais próximo da observação, em relação a posição azimutal, maior correlação tem o modelo com as medições. O *skill score* é representado pelas linhas tracejadas em vermelho, partindo-se do ponto da observação; ele indica o desempenho geral do modelo, já que leva em conta tanto a correlação (r) quanto a razão entre o desvio padrão do modelo (σ_m) pelo desvio padrão da observação (σ_o), estatísticos denominado *STD Ratio*, conforme Equação 3.

$$SS4 = \frac{(1 + r)^4}{4 (STD Ratio + 1/STD Ratio)^2} \quad (3)$$

Utiliza-se, também, o erro médio (*Mean Bias Error* – MBE) normalizado pela média (nMBE), a raiz do erro quadrático médio (*Root Mean Square Deviation* – RMSD), o RMSD normalizado pela média e o *STD Ratio* (σ_m/σ_o). Os estatísticos escolhidos são importantes para analisar a amplitude de variação da série modelada em relação a série observada (*STD Ratio*), a estrutura de fase e frequência entre as séries (r), o desempenho geral do modelo (SS4) e a dispersão dos dados (RMSD e MBE).

4. RESULTADOS

Para a seleção de variáveis, bem como para todos os modelos regressivos utilizados, o *target* é o índice de claridade, k_t (razão entre a irradiância global horizontal – GHI – e a irradiância extraterrestre horizontal). Todos os modelos aplicados consideram a seleção de variáveis seguida de extração ou somente a extração de variáveis, para diminuir a redundância existente no conjunto de 31 variáveis regressoras utilizado. A Tabela 4 apresenta as abreviações utilizadas para os diferentes modelos regressivos e suas respectivas entradas. Os modelos regressivos de regressão linear múltipla (MLR) e redes neurais (NN) do tipo *multilayer perceptron* (MLP) podem ser pós-processados com o *quantile mapping* (QM), de forma que as distribuições acumuladas dos dados nos conjuntos de validação e teste/operação se adequem à distribuição dos dados do conjunto de calibração. Vale salientar que o *quantile mapping* faz a correção dos modelos em relação à distribuição acumulada da irradiância global horizontal ao invés do k_t . A Tabela 5 apresenta os resultados para a seleção de variáveis feita. Pode-se notar que as maiores correlações das variáveis da CAMS e do ECMWF com k_t para a região de El Rosal estão abaixo de 0,6, enquanto para Petrolina estão abaixo de 0,8.

Tabela 4 – Abreviaturas utilizadas para os diferentes modelos avaliados.

Abreviatura	Descrição
GHI CAMS	GHI resultante do modelo da CAMS
GHI MLR	GHI obtida utilizando regressão linear múltipla com extração de variáveis
GHI MLRs	GHI obtida utilizando regressão linear múltipla com seleção de variáveis seguida de extração de variáveis
GHI MLRs+QM	GHI obtida utilizando regressão linear múltipla com seleção de variáveis seguida de extração de variáveis e pós-processamento feito com <i>quantile mapping</i>
GHI NN	GHI obtida utilizando redes neurais do tipo MLP com extração de variáveis
GHI NNs	GHI obtida utilizando redes neurais com seleção de variáveis global seguida de extração de variáveis
GHI NN+QM	GHI obtida utilizando redes neurais do tipo MLP com extração de variáveis e pós-processamento feito com <i>quantile mapping</i>
GHI NNs+QM	GHI obtida utilizando redes neurais do tipo MLP com seleção de variáveis seguida de extração de variáveis e pós-processamento feito com <i>quantile mapping</i>

Fonte: própria.

Tabela 5 – Resultados para seleção de variáveis considerando o k_t como *target*.

Estação El Rosal				Estação Petrolina			
Divisão 1		Divisão 2		Divisão 1		Divisão 2	
Variável	ρ	Variável	ρ	Variável	ρ	Variável	ρ
ClearSkyBNI	0,56	ClearSkyBNI	0,57	BNI	0,76	BNI	0,75
GHI	0,49	GHI	0,49	BHI	0,71	BHI	0,70
Ssrd	0,48	ssrd	0,49	GHI	0,64	GHI	0,62
BHI	0,43	ClearSkyBHI	0,44	ssrd	0,52	ssrd	0,51
BNI	0,43	ClearSkyGHI	0,42	skt	0,44	skt	0,43
ClearSkyBHI	0,42	BHI	0,42	ClearskyBNI	0,44	ClearskyBNI	0,42
ClearSkyGHI	0,40	BNI	0,41	ClearskyBHI	0,40	ClearskyBHI	0,38
TOA	0,38	TOA	0,40	ClearskyGHI	0,38	t2m	0,37
Skt	0,29	skt	0,31	t2m	0,38	ClearskyGHI	0,36
t2m	0,26	t2m	0,28	TOA	0,36	TOA	0,34
Fiso	0,13	DHI	0,13	tco3	0,16	tco3	0,14
Fgeo	0,12	tco3	0,11	fiso	0,11	fiso	0,11
tco3	0,12	fiso	0,10	fgeo	0,11	fgeo	0,10

Fonte: própria.

As Tabelas 6 e 7 apresentam os resultados dos modelos regressivos aplicados. Para El Rosal, considerando a divisão 1 dos dados adotada (dados randomizados), o modelo que apresenta o maior SS4 e razão entre os desvios unitária é o ‘GHI NN + QM’. Já, considerando a divisão 2, em que o conjunto de teste/operação manteve a sua componente temporal, o modelo que apresenta o melhor SS4 é o ‘GHI MLR’. Os diferentes resultados para cada divisão dos dados mostram que os modelos adotados não conseguem generalizar a sua aplicação em diferentes conjuntos de calibração e teste. O pós-processamento feito com *quantile mapping* consegue ajustar bem o BIAS e o desvio no caso da divisão 1 dos dados, já que as características estatísticas dos conjuntos de calibração e teste são similares para essa divisão. Já na divisão 2, o modelo de *quantile mapping* busca adequar os dados do conjunto de teste à distribuição dos dados do conjunto de calibração, sendo que ambos os conjuntos possuem características estatísticas distintas, pois a componente temporal foi mantida. Por isso, os modelos pós-processados com *quantile mapping* na divisão 2 dos dados apresentam uma superestimação da razão entre os desvios e um BIAS mais elevado, em valor absoluto, do que no caso da divisão 1. A Figura 1 apresenta o diagrama de Taylor para a estação de El Rosal considerando a divisão 1 dos dados, e a Figura 2 apresenta as dispersões dos melhores modelos apresentados na Tabela 6 em comparação com o modelo da CAMS para GHI.

Para Petrolina, o modelo com maior SS4 para a divisão 1 foi ‘GHI NN + QM’, enquanto que para divisão 2, os modelos regressivos aplicados não conseguiram melhores estatísticos que o próprio modelo da CAMS (Tabela 7). De fato, existem situações em que a acurácia do modelo de satélite é tão alta para GHI que é difícil conseguir melhorar ainda mais essa acurácia a partir do uso de medições em terra (POLO *et al.*, 2020). Vale salientar, também, que o modelo da CAMS é melhor ajustado à região de Petrolina do que na região de El Rosal, pois o ângulo de visão do satélite METEOSAT é melhor enquadrado em Petrolina do que em El Rosal, já que esta última está na borda do ângulo de visão desse satélite. Além disso, ajustes locais podem já ter sido feitos pela equipe da CAMS na região de Petrolina, pois há uma estação solarimétrica com dados públicos disponibilizados pela *Baseline Surface Radiation Network* (BSRN) na região (SCHROEDTER-HOMSCHEIDT *et al.*, 2019).

A Tabela 8 apresenta um comparativo entre os modelos que obtiveram os melhores resultados estatísticos para a região de El Rosal, o modelo da CAMS e um modelo da base de dados da SOLCAST (*Solar Forecast & Solar Irradiance Data*), que fornece, entre outros produtos, séries de longo prazo de GHI que utilizam imagens do satélite GOES. Pode-se notar que a GHI modelada pela SOLCAST possui os melhores estatísticos, com RMSDn de 17% e SS4 de 0,88. Como a família de satélite GOES engloba satélites que visualizam bem a região da América do Sul, os resultados dos produtos de radiação modelados utilizando as imagens do GOES são mais acurados que o *site adaptation* feito utilizando os dados da CAMS, que são gerados considerando as imagens do satélite METEOSAT.

Tabela 6 – Resultados para a Estação Solarimétrica de El Rosal.

DIVISÃO 1 – com randomização dos dados								
Estatísticos	BIAS	nMBE	Desvio	STDRatio	RMSD	RMSDN	Corr	SS4
Unidades	(W/m ²)	(%)	(W/m ²)	(.)	(W/m ²)	(%)	(.)	(.)
GHI Medido	0,00	0,00%	323,40	0,00	0,00	0,00%	1,00	0,00
GHI CAMS	159,04	24,26%	295,23	0,91	260,20	39,68%	0,78	0,63
GHI MLR	-2,32	-0,35%	306,41	0,95	145,48	22,19%	0,89	0,80
GHI MLR+QM	-0,03	-0,01%	323,25	1,00	146,59	22,36%	0,90	0,81
GHI MLRs	-1,60	-0,24%	294,80	0,91	148,59	22,66%	0,89	0,79
GHI MLRs+QM	-0,03	-0,01%	323,25	1,00	152,99	23,33%	0,89	0,79
GHI NN	-2,40	-0,37%	295,51	0,91	133,85	20,41%	0,91	0,83
GHI NNs	-1,41	-0,22%	292,35	0,90	137,99	21,05%	0,90	0,81
GHI NN+QM	-0,03	-0,01%	323,25	1,00	137,34	20,95%	0,91	0,83
GHI NNs+QM	-0,03	-0,01%	323,25	1,00	141,83	21,63%	0,90	0,82

DIVISÃO 2 – mantendo a componente temporal								
Estatísticos	BIAS	nMBE	Desvio	STDRatio	RMSD	RMSDN	Corr	SS4
Unidades	(W/m ²)	(%)	(W/m ²)	(.)	(W/m ²)	(%)	(.)	(.)
GHI Medido	0,00	0,00%	294,06	0,00	0,00	0,00%	1,00	0,00
GHI CAMS	120,81	19,28%	278,44	0,95	217,23	34,67%	0,80	0,66
GHI MLR	-32,47	-5,18%	295,22	1,00	152,17	24,28%	0,87	0,77
GHI MLR+QM	-36,37	-5,80%	329,78	1,12	163,39	26,07%	0,88	0,76
GHI MLRs	-13,75	-2,19%	281,70	0,96	146,05	23,31%	0,87	0,77
GHI MLRs+QM	-36,37	-5,80%	329,78	1,12	164,99	26,33%	0,87	0,76
GHI NN	-43,77	-6,99%	265,70	0,90	166,36	26,55%	0,84	0,71
GHI NNs	-17,15	-2,74%	276,70	0,94	139,24	22,22%	0,88	0,79
GHI NN+QM	-36,37	-5,80%	329,78	1,12	183,08	29,22%	0,84	0,71
GHI NNs+QM	-36,37	-5,80%	329,78	1,12	159,20	25,41%	0,88	0,77

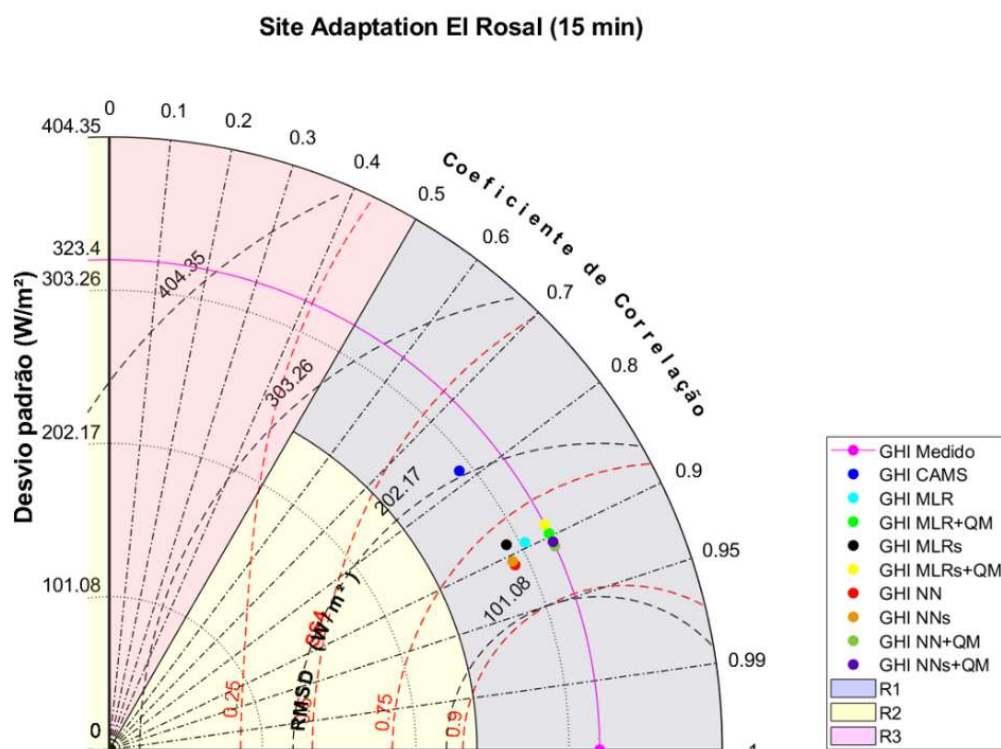


Figura 1 – Diagrama de Taylor para a divisão 1 dos dados (com randomização) da Estação Solarimétrica de El Rosal.

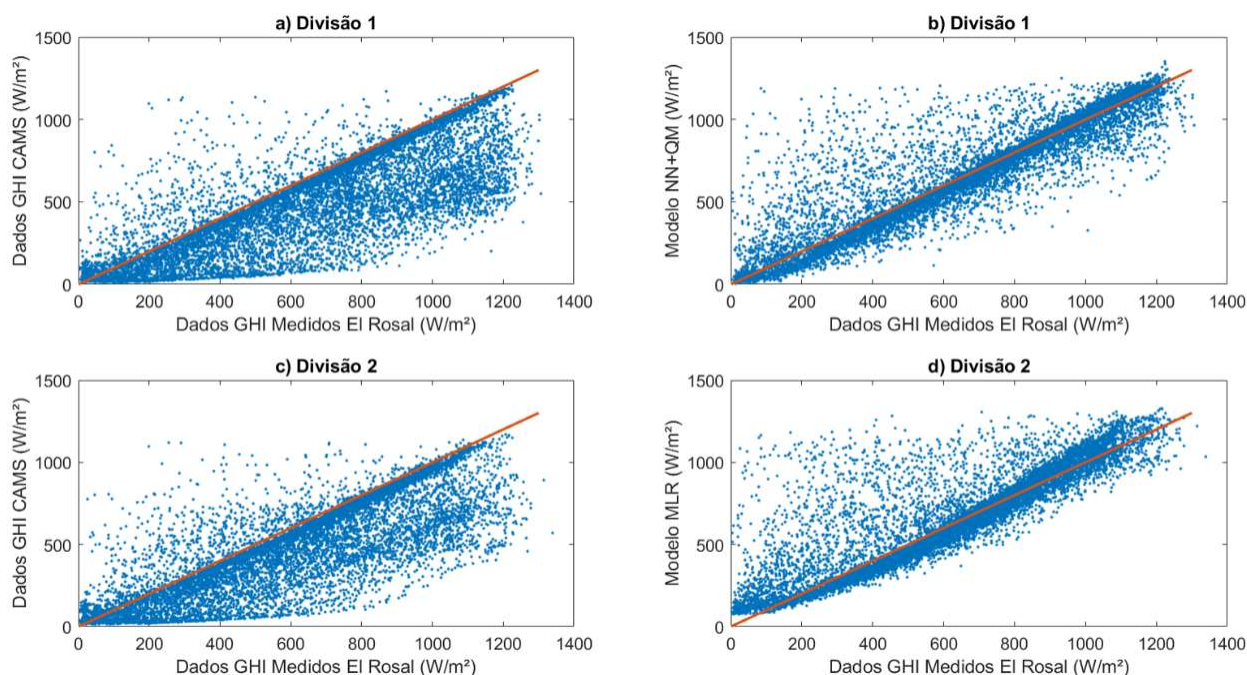


Figura 2 – Dispersões para a divisão 1 ('a' e 'b') e 2 ('c' e 'd') dos dados da Estação Solarimétrica de El Rosal.

Tabela 7 – Resultados para a Estação Solarimétrica de Petrolina.

DIVISÃO 1 – com randomização dos dados								
Estatísticos	BIAS	nMBE	Desvio	STDRatio	RMSD	RMSDN	Corr	SS4
Unidades	(W/m ²)	(%)	(W/m ²)	(.)	(W/m ²)	(%)	(.)	(.)
GHI Medido	0,00	0,00%	295,05	0,00	0,00	0,00%	1,00	0,00
GHI CAMS	-28,85	-5,75%	287,79	0,98	121,31	24,20%	0,92	0,85
GHI MLR	-0,04	-0,01%	276,54	0,94	118,40	23,62%	0,92	0,84
GHI MLR+QM	2,98	0,59%	299,30	1,01	122,01	24,34%	0,92	0,84
GHI MLRs	0,57	0,11%	274,66	0,93	120,88	24,11%	0,91	0,83
GHI MLRs+QM	2,98	0,59%	299,30	1,01	123,88	24,71%	0,91	0,84
GHI NN	-3,04	-0,61%	276,22	0,94	112,62	22,46%	0,92	0,85
GHI NNs	0,91	0,18%	272,15	0,92	115,18	22,97%	0,92	0,84
GHI NN+QM	2,98	0,59%	299,30	1,01	117,09	23,36%	0,92	0,85
GHI NNs+QM	2,98	0,59%	299,30	1,01	119,89	23,91%	0,92	0,85
DIVISÃO 2 – mantendo a componente temporal								
Estatísticos	BIAS	nMBE	Desvio	STDRatio	RMSD	RMSDN	Corr	SS4
Unidades	(W/m ²)	(%)	(W/m ²)	(.)	(W/m ²)	(%)	(.)	(.)
GHI Medido	0,00	0,00%	321,27	0,00	0,00	0,00%	1,00	0,00
GHI CAMS	-26,03	4,84%	313,48	0,98	119,56	22,22%	0,93	0,87
GHI MLR	3,06	0,57%	303,92	0,95	120,16	22,33%	0,93	0,86
GHI MLR+QM	48,99	9,10%	291,67	0,91	131,91	24,51%	0,92	0,85
GHI MLRs	-4,42	-0,82%	305,07	0,95	125,37	23,30%	0,92	0,85
GHI MLRs+QM	48,99	9,10%	291,67	0,91	135,44	25,17%	0,92	0,84
GHI NN	10,10	1,88%	293,46	0,91	116,69	21,68%	0,93	0,86
GHI NNs	2,73	0,51%	297,08	0,92	116,56	21,66%	0,93	0,87
GHI NN+QM	48,99	9,10%	291,67	0,91	129,84	24,13%	0,93	0,85
GHI NNs+QM	48,99	9,10%	291,67	0,91	130,34	24,22%	0,93	0,85

Tabela 8 – Comparativo entre *site adaptation*, modelo da CAMS e da SOLCAST para GHI da Estação de El Rosal*.

Estatísticos	BIAS (W/m ²)	nMBE	Desvio (W/m ²)	STDRatio	RMSD (W/m ²)	RMSDN	Corr	SS4
OBS	0,00	0,00%	320,53	0,00	0,00	0,00%	1,00	0,00
NN+QM (Divisão 1)	2,11	0,32%	320,19	1,00	135,68	20,59%	0,91	0,83
MLR (Divisão 2)	-9,89	-1,50%	310,64	0,97	149,42	22,68%	0,89	0,79
CAMS GHI	157,60	23,92%	295,62	0,92	260,17	39,48%	0,78	0,62
SOLCAST GHI	11,57	1,76%	297,69	0,93	111,51	16,92%	0,94	0,88

*Dados utilizados de 2013 até 2016.

5. CONCLUSÕES E PERSPECTIVAS FUTURAS

O objetivo do *site adaptation* é aumentar a acurácia da série fornecida pela base de dados históricas, no caso a série de GHI da CAMS, obtendo, assim, uma melhor estimativa do recurso solar de longo prazo em certa região. Os resultados apresentados mostram que a metodologia adotada obtém séries mais acuradas na estação de El Rosal, que está localizada na borda do campo de visão do satélite europeu METEOSAT, o que faz o modelo da CAMS apresentar resultados com alta variância para a região. Assim, é esperado que metodologias de *site adaptation* consigam melhorar a estimativa da radiação feita pela CAMS nessa região. Já quando se considera a estação de Petrolina, onde o modelo da CAMS reproduz de forma satisfatória a irradiância global horizontal medida para uma escala de 15 minutos, o *site adaptation* não conseguiu obter melhores resultados no conjunto de teste/operação. Além disso, a estratégia de randomização de dados (divisão 1) permitiu obter resultados mais acurados para GHI, considerando-se a aplicação em *site adaptation*.

Para tentar melhorar a acurácia do *site adaptation* na escala temporal de 15 minutos para irradiância global horizontal é necessário investigar melhor como aumentar a acurácia dos modelos de regressão. Fernández-Peruchena et al. (2020) mostram, nesse sentido, que a combinação de diferentes variáveis das bases de dados pode gerar melhores resultados em modelos regressivos. Assim, trabalhos futuros podem avaliar metodologias de combinação das variáveis disponibilizadas pelas bases de dados de forma a obter variáveis mais correlacionadas com o *target* para os modelos regressivos. Além disso, podem ser adicionadas variáveis provenientes de outros modelos, como, por exemplo, as variáveis disponibilizadas pela família de satélite GOES (*Geostationary Operational Environmental Satellite*) de forma a prover mais informação para a regressão. O satélite GOES-East possui, inclusive, um ângulo de visão mais favorável para a América do Sul do que os satélites europeus da série METEOSAT, o que pode ser um ponto favorável para utilização das variáveis fornecidas pela família de satélites GOES na solução de problemas de *site adaptation* na região da América do Sul. Os trabalhos futuros também devem incluir o *site adaptation* de outras componentes da radiação fornecidas pelas estações solarimétricas, como a irradiância direta normal, variável de maior interesse no desenvolvimento de projetos de geração heliotérmica.

Agradecimentos

Os autores agradecem o apoio financeiro da Agência Nacional de Energia Elétrica – ANEEL e da Companhia Hidro Elétrica do São Francisco – CHESF por meio do projeto 23076.009704/2020-56 no âmbito do Programa de P&D ANEEL, como também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pelo apoio aos bolsistas no âmbito do Programa de Pós-Graduação em Tecnologias Energéticas e Nucleares – PROTEN.

REFERÊNCIAS

- EPE, EMPRESA DE PESQUISA ENERGÉTICA. Instruções para Solicitação de Cadastramento e Habilitação Técnica de Empreendimentos Fotovoltaicos com Vistas à Participação nos Leilões de Energia Elétrica. 2017. Disponível em <http://www.epe.gov.br/sites-pt/leiloes-de-energia/Documents/EPE-DEE-065_2013_R5_2017_UFV.pdf> Acesso em 03 de dezembro de 2019.
- FERNÁNDEZ-PERUCHENA, CARLOS M., JESÚS POLO, LUIS MARTÍN, AND LUIS MAZORRA. Site-Adaptation of Modeled Solar Radiation Data: The SiteAdapt Procedure. *Remote Sensing* 12(13): 1–17, 2020.
- GAVIN, H. P. The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems. Department of Civil and Environmental Engineering Duke University, 2019.
- HALL, M. A. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. Conference: Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford University, Stanford, CA, USA, 2000.
- JOLLIFFE, I. T. Principal Component Analysis. 2ª. ed.: Elsevier, 1986.
- KARIUKI, B. W.; SATO, T. Interannual and Spatial Variability of Solar Radiation Energy Potential in Kenya Using Meteosat Satellite. *Renewable Energy* 116:88–96, 2018.

- LEFÈVRE, M.; OUMBE, A.; BLANC, P.; ESPINAR, B.; GSCHWIND, B.; QU, Z.; WALD, L.; SCHROEDTER-HOMSCHIEDT, M.; HOYER-KLICK, C.; AROLA, A.; BENEDETTI, A.; KAISER, J. W.; MORCRETTE, J. J. McClear: A New Model Estimating Downwelling Solar Radiation at Ground Level in Clear-Sky Conditions. *Atmospheric Measurement Techniques*, 6(9):2403–18, 2013.
- PETRIÚ, L. B.; SABINO, E.; BARROS, H.; COSTA, A.; BARBOSA, E.; VILELA, O.C. Procedimento objetivo para a garantia de qualidade de dados de radiação solar. Apresentado em XL Reunión de Trabajo de la Asociación Argentina de Energías Renovables y Ambiente (ASADES), San Juan, Argentina, 2017.
- POLO J., WILBERT S., RUIZ-ARIAS J. A., MEYER R., GUEYMARD C., SÚRI M., ET AL. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. *Solar Energy*, 132:25–37, 2016.
- POLO, J.; FERNÁNDEZ-PERUCHENA, C.; SALAMALIKIS, V.; AGUIAR, L. M.; TURPIN, M.; MARTÍN-POMARES, L.; KAZANTZIDIS, A.; BLANC, P.; REMUND, J. Benchmarking on Improvement and Site-Adaptation Techniques for Modeled Solar Radiation Datasets. *Solar Energy* 201(March):469–79, 2020.
- QU, ZHIPENG, ARMEL OUMBE, PHILIPPE BLANC, BELLA ESPINAR, GERHARD GESELL, BENOIT GSCHWIND, LARS KLÜSER, MIREILLE LEFÈVRE, LAURENT SABORET, MARION SCHROEDTER-HOMSCHIEDT, LUCIEN WALD. Fast Radiative Transfer Parameterisation for Assessing the Surface Solar Irradiance: The Heliosat-4 Method. *Meteorologische Zeitschrift* 26(1):33–57, 2017.
- ROESCH, A.; WILD, M.; OHMURA, A.; DUTTON, E.G.; LONG, C.N.; ZHANG T. Assessment of BSRN radiation records for the computation of monthly means. *Atmos. Meas. Tech.*, v. 4, p. 339–54, 2011.
- SENGUPTA, M., HABTE, A., GUEYMARD, C., WILBERT, S., & RENNE, D. Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications (No. NREL/TP-5D00-68886). National Renewable Energy Lab.(NREL), Golden, CO (United States), 2017.
- SALAZAR, G.; GUEYMARD, C.; GALDINO, J.; VILELA, O. C.; FRAIDENRAICH, N. Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil. *Renewable and Sustainable Energy Reviews*, 117, 109478, 2020.
- SCHROEDTER-HOMSCHIEDT, MARION, CARSTEN HOYER-KLICK, NIELS KILLIUS, JETHRO BETCKE, MIREILLE LEFÈVRE, LUCIEN WALD, ETIENNE WEY, LAURENT SABORET. User's Guide to the CAMS Radiation Service (CRS) Status December 2019. (December):1–74, 2019.
- SCHWANDT, MARKO, KAUSHAL CHHATBAR, RICHARD MEYER, KATHARINA FROSS, INDRADIP MITRA, RAMADHAN VASHISTHA, GODUGUNUR GIRIDHAR, S. GOMATHINAYAGAM, AND ASHVINI KUMAR. Development and Test of Gap Filling Procedures for Solar Radiation Data of the Indian SRRA Measurement Network. *Energy Procedia* 57:1100–1109, 2014.
- TAYLOR, K.E. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, v. 106, p. 7183-7192, 2011.
- WILCOX, S. E MARION, W. Users manual for TMY3 data sets. Technical Report NREL/TP-581-43156, 2008.
- WILKS, D. S. Statistical methods in the atmospheric sciences. 2ª. ed. [S.l.]: Elsevier, 2013.
- YU, L.; LIU, H. Efficient Feature Selection via Analysis of Relevance and Redundancy. *Journal of Machine Learning Research* 5, p. 1205–1224, 2004.

SITE ADAPTATION OF GLOBAL HORIZONTAL RADIATION FOR SOLARIMETRIC STATIONS IN ARGENTINA AND BRAZIL

Abstract. *Estimating the solar radiation incident on a certain location for several years is necessary to develop photovoltaic and heliothermic projects of solar power plants. In this work, multiple linear regression, multilayer perceptron neural networks and quantile mapping are used to adjust long-term global horizontal irradiance (GHI) series using ground measurements, a procedure known as site adaptation. Two solarimetric stations, one in the city of Petrolina, in northeastern Brazil, and another in the province of Salta, at the El Rosal site, Argentina, are used as case studies. Data from the Copernicus Atmosphere Monitoring Service (CAMS) and the European Center for Medium-Range Weather Forecasts (ECMWF) are used as regression variables in the statistical models. Two different divisions of the data in calibration (80% of the samples) and test/operation (20% of the samples) are tested. One of them performs the data's temporal randomization, and the other keeps the temporal component. The results show that the adaptation site achieves better statistics for the El Rosal station (24,28% de nRMSD, -5,18% de nMBE e 1,00 de STDRatio) than the CAMS reference model for GHI (34,67% de nRMSD, 19,28% de nMBE e 0,95 de STDRatio), while for the Petrolina station, the results obtained (21,66% de nRMSD, 0,51% de nMBE e 0,92 de STDRatio) did not achieve greater accuracy than the CAMS model (22,22% de nRMSD, 4,84% de nMBE e 0,98 de STDRatio). Both for the Petrolina station and El Rosal, the results obtained with the division with a randomization of the data are more accurate than the one that maintains the temporal component.*

Keywords: *site adaptation, global horizontal irradiance*