

Integração de dados e modelos de previsão de produção fotovoltaica do nordeste brasileiro

Data Integration and Forecast Models for Photovoltaic Production in the Brazilian Northeast

Hugo Abreu Mendes ^{1,2}  orcid.org/0000-0002-0474-2680

Henrique Ferreira Nunes ³  orcid.org/0000-0002-0474-2680

Paulo S.G. de Mattos Neto ⁴  orcid.org/0000-0002-2396-7973

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

² Especialização em Inteligência Artificial, Escola Politécnica de Pernambuco, Pernambuco, Brasil,

³ Fundação para Inovações Tecnológicas (FITec), Recife, Brasil,

⁴ Centro de Informática, Universidade Federal de Pernambuco, Pernambuco, Brasil.

E-mail do autor principal: Hugo Abreu Mendes ham@poli.br

Resumo

Todos os ramos produtivos da sociedade precisam de uma estimativa para poder controlar bem suas despesas. No ramo da energia, concessionárias de energia elétrica utilizam essa informação para controle de fluxo de potência na rede. Para que a estimativa de produção energética de sistemas fotovoltaicos seja bem realizada, há a necessidade da união de diversas variáveis geoespaciais e meteorológicas. Este trabalho propõe a criação de uma plataforma de integração de dados satelitais, modelos de estimativa de produção, estações base de medição e capacidade de produção real. São apresentados alguns modelos estatísticos, probabilísticos e inteligência artificial, geradores de estimativas de produção espaciais e temporais, que podem melhorar os ganhos de produção, bem como facilitar o acompanhamento e supervisão de novos empreendimentos.

Palavras-Chave Energia Fotovoltaica; Previsão; Otimização, Regressão;

Abstract

All productive branches of society need an estimate to be able to control their expenses well. In the energy business, electric utilities use this information to control the power flow in the grid. For the estimation of the energy production of photovoltaic systems to be well done, it is necessary to join several geospatial and meteorological variables. This work proposes the creation of a satellite data integration platform, production estimation models, measurement base stations and actual production capacity. Some statistical, probabilistic and artificial intelligence models that generate spatial and temporal production estimates that could improve production gains as well as facilitate the monitoring and supervision of new enterprises are presented.

Key-words: Photovoltaic Energy; Forecasting; Optimization; Regression.

1 INTRODUÇÃO

Estimativa de potencial de geração fotovoltaico é um tema que já obteve grande avanço [1]–[3]. Os trabalhos já realizados levam em consideração a tecnologia utilizada pela célula e modelos de satélite que visam definir os parâmetros físicos de entrada, tais como radiação, temperatura e velocidade do vento [4]–[7].

A utilização de técnicas de inteligência artificial aplicadas a este tema tem sido limitada a predição temporal de geração [8]–[10]. Este tipo de predição é muito útil levando-se em consideração o sistema elétrico completo de uma região ou país, para balanceamento de oferta e demanda, sendo possível uma previsibilidade maior para o operador do sistema elétrico, em relação a adequação do uso de outras fontes de energia. Entretanto, por necessitar de dados temporais sobre usinas instaladas, não é indicado para a recomendação dos melhores locais para a implantação de novas usinas.

As pesquisas utilizando técnicas de IA (Inteligência Artificial) visando a predição de geração fotovoltaica antes mesmo da existência do sistema ainda são incipientes, envolvendo apenas dados tipo GIS (Geographic Information System) [11]–[14].

Este trabalho busca unir o uso de técnicas de ciência de dados e algoritmos de IA para estimativa espacial do potencial de geração fotovoltaica, possibilitando uma melhor escolha do local para implantação de novas usinas fotovoltaicas. O trabalho consta também com a implementação de um modelo híbrido para previsão de series temporais, com dados de geração média diária de algumas usinas.

A escolha da região nordeste do Brasil para estudo de caso leva em consideração o fato de ser a região com a maior quantidade de dados de empreendimentos de geração fotovoltaicas disponíveis pela plataforma da ONS (Operador Nacional do Sistema) [15, 16].

A organização das seções deste trabalho se dá da seguinte forma. Na seção 2 será discutido dados espaciais obtidos de diferentes fontes, todos os dados desta seção são obtidos a partir de médias temporais durante todos os anos de coleta de cada base. Nas seções 3 e 4, são apresentados dois modelos de estimativa espacial de produção fotovoltaica. Posteriormente, na seção 5 é explicado o uso de series temporais para dados de geração

diária em algumas usinas do nordeste. Por fim, os resultados são apresentados na seção 6. Fechando o trabalho, há uma breve discussão na seção 7, em que são levantadas melhorias e ideias para a posterioridade.

2 DADOS ESPACIAIS NO NORDESTE

A distribuição espacial do recurso solar é um conhecimento necessário, obtido a partir de modelos de transferência radiativa e validação com dados observacionais. Unindo os dados das bases do INMET (Instituto Nacional de Meteorologia), PVGIS (Photovoltaic Geographic Information System), ONS e do Atlas Brasileiro de Energia Solar, é possível criar um sistema capaz de estimar o fator de capacidade médio de uma região, sendo a razão entre a produção gerada e a capacidade de produção nominal. O fator de capacidade é utilizado nestas análises, por ser independente do tamanho da usina e das tecnologias utilizadas nas células solares.

Pode-se integrar ao sistema inteligência de previsão temporal, podendo ser horária, diária ou mensal de geração para qualquer usina. O sistema, portanto, completo, seria capaz de possibilitar uma estimativa espacial de produção bem como uma estimativa específica temporal para um empreendimento existente. O mais importante é, com a união de bases de dados, facilitar a busca por novos empreendimentos, tanto para concessionárias quanto para microgeradores.

As subseções a seguir descrevem as bases utilizadas para geração de um cubo de dados, em que cada coluna corresponde a uma variável diferente. O que possibilita a manipulação destes dados é o fato, sendo a localização (latitude e longitude).

2.1 Atlas Brasileiro de Energia Solar

São dados provenientes do Atlas solarimétrico, de autoria do LABREN (Laboratório de Modelagem e Estudos de Recursos Renováveis de Energia). Dados de 17 anos (1999-2015) de coleta, que resulta por fim, ao realizar uma média no tempo, em um atlas de radiação no Brasil, com total anual e cada mês [17], [18]. Estes dados são obtidos por satélite, utilizando o modelo BRASIL - SR e validado com estações do INMET [19, 20]. São ao todo cinco colunas sobre diferentes formas de interação da irradiação no ambiente de medição.

- Global Horizontal - É a energia total por unidade de área incidente numa superfície horizontal, sendo a soma da componente direta e da componente difusa;
- Plano Inclinado - Energia total por unidade de área incidente sobre um plano inclinado na latitude do local em relação à superfície da Terra;
- Direta Normal - Energia por unidade de área proveniente diretamente do Sol que incide perpendicularmente à superfície.
- Difusa - Energia incidente sobre uma superfície horizontal por unidade de área, decorrente do espalhamento do feixe solar direto pelos constituintes atmosféricos (moléculas, material particulado, nuvens, etc.
- Fotossinteticamente Ativa - Energia com capacidade de produção fotossintética, associada à produção de biomassa, à morfologia vegetal e ao crescimento de plantas.

2.2 INMET

O Instituto Nacional de Meteorologia disponibiliza dados a partir de diversas estações espalhadas pelo país [21]. Nesta seção estes dados são obtidos pelo mesmo período que os dados do Atlas Brasileiro de Energia Solar, entre 1999 e 2015, é feita então uma média temporal nestes dados, resultado apenas em sua componente espacial. No Nordeste, pode ser visto a localização das estações, pela Figura 1. As seguintes variáveis foram utilizadas:

- Insolação Total (hs);
- Número de Dias com Precipitação (qtd)
- Pressão Atm nível Mar Média (mbar)
- Temperatura Máxima Média (°C)
- Umidade Relativa Média (%)
- Velocidade do Vento Média (mps)
- Nebulosidade Média (décimos)
- Precipitação Total (mm)
- Temperatura Compensada Média (°C)
- Visibilidade Média (%)
- Temperatura Mínima Média (°C)

Algumas destas variáveis estão compiladas para visualização na Figura 2.

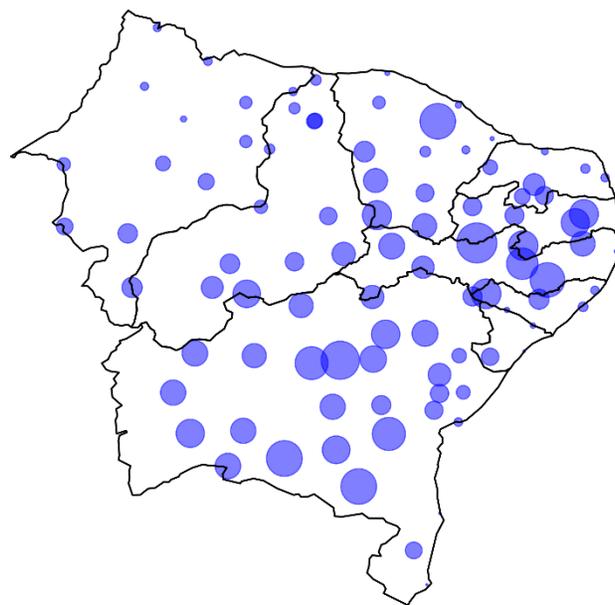


Figura 1: Localização das estações INMET. Quanto maior o círculo maior a altitude. Observa-se uma boa granularidade de estações.

2.3 PVGIS

É uma plataforma que tem sido desenvolvida há mais de 10 anos, pela comissão europeia JRC (Joint Research Centre). Pode-se usar o PVGIS para estimar a produção de diferentes tipos de sistemas fotovoltaicos em praticamente qualquer lugar do mundo [22]. Estes dados serão utilizados integralmente contendo as seguintes colunas: média mensal (para todos os meses do ano) de geração para o sistema fotovoltaico escolhido e desvio padrão da geração mensal devido a variações anuais.

Para o Brasil, o PVGIS disponibiliza dois modelos de dados, NSRDB [7], e SARA [23]. A produção média mensal estimada, pelo NSRDB e SARA, pode ser visualizada diretamente no site da plataforma. Mais sobre as metodologias e bases utilizadas pelo PVGIS podem ser encontradas em [24, 25].

Uma das razões para a não utilização direta dos dados do PVGIS para a estimativa da produção de energia é que não é baseado no modelo de radiação BRASIL - SR, que é o mais acurado por ser validado com dados locais [19, 20].

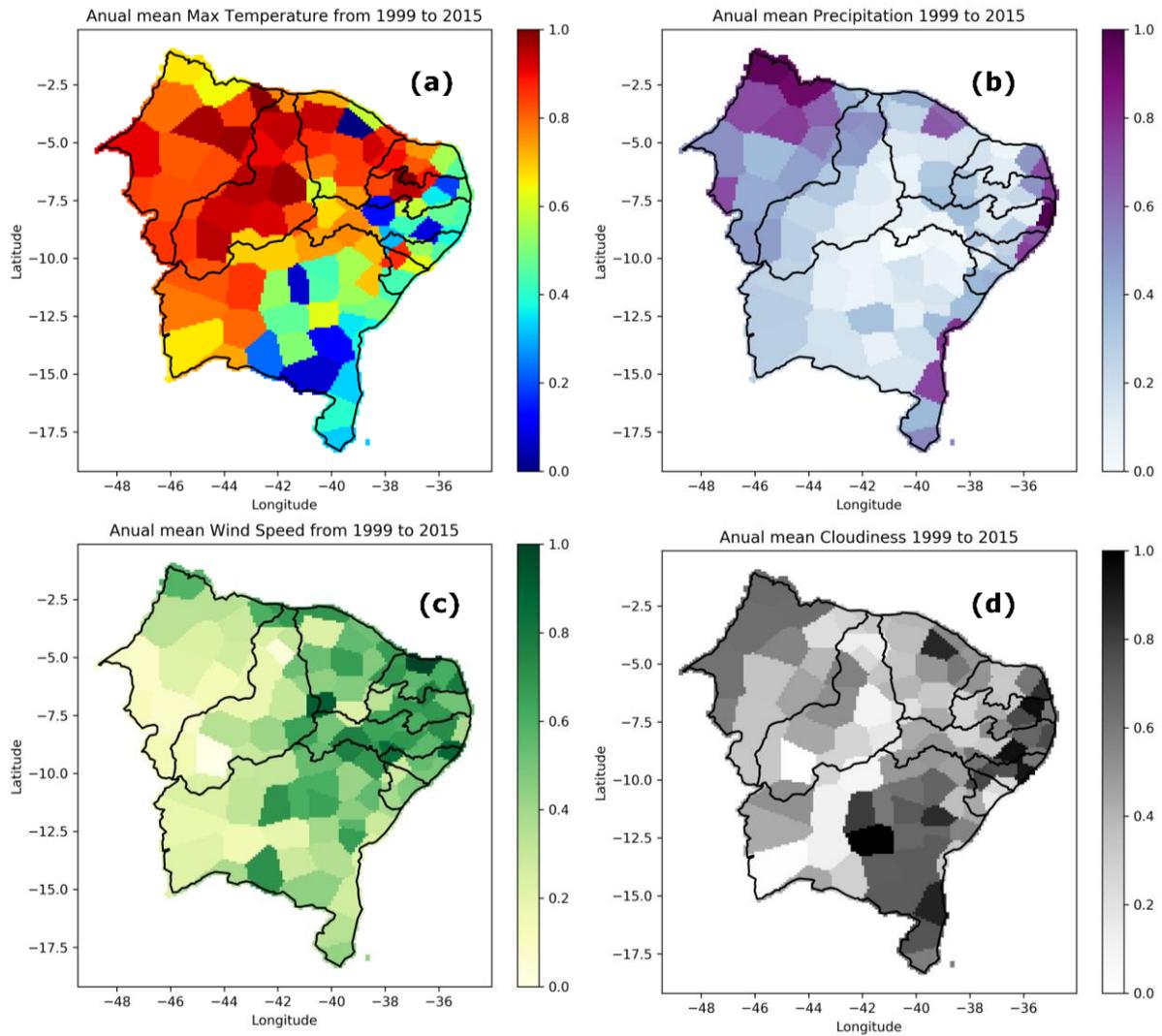


Figura 2: Compilação de variáveis, normalizadas, disponíveis pelo INMET: a) temperatura máxima média; b) precipitação; c) velocidade do vento média; d) nebulosidade.

Para obtenção dos dados do PVGIS foi utilizado o serviço por requisição, utilizando por padrão potência de pico de 1kW. Também foi especificada a inclinação ótima do módulo [26].

2.4 ONS

Entre todas as bases já discutidas, nenhuma delas é referente a informações de geração real de usinas fotovoltaicas. A ONS disponibiliza em seu balanço de geração dados de algumas usinas do Brasil [15, 16], em sua maioria no Nordeste.

Foram utilizadas as séries temporais de fator de capacidade das usinas fotovoltaicas do Nordeste.

Para a extração da componente espacial foi feita uma média e utilizado um interpolador. A metodologia de interpolação utilizada é a Kriging [27, 28], com resultado apresentado na Figura 3. Foi utilizado um modelo de variograma calculado com uma função de potência [29].

3 OTIMIZAÇÃO DE COMITÊ DE REGRESSORES

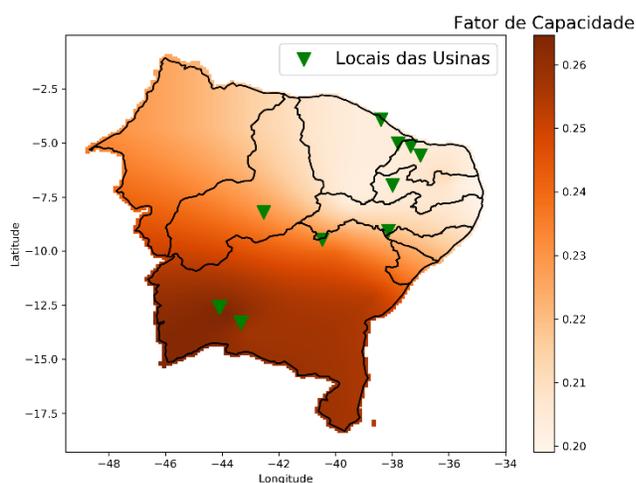


Figura 3: Mapa de fator de capacidade de geração obtido por interpolação e localização das usinas com dados disponíveis pela ONS no Nordeste.

A partir do cubo de dados gerado pela concatenação de todas as bases de espaciais apresentadas, é possível criar um modelo de regressão utilizando inteligência de votação por comitê. A intenção é visualizar como IA aprenderia sobre o fator de capacidade de produção espacialmente, visto na Figura 3.

A votação utilizada é pela média homogênea do obtido por diversos modelos de regressão: Regressão linear; Random Forest Regressor [30]; Support Vector Machine Regressor [31]; Adaboost [32]; Bagging [33]; GradientBoosting [34]; RANSAC (Random Sample Consensus) [35]; Passive Aggressive Regressor [36]; SGD (Stochastic Gradient Descendent) Regressor. Todos os regressores foram aplicados utilizando o pacote scikit-learn do Python [37].

Cada um dos regressores possui diversos parâmetros diferentes, e não necessariamente o uso de todos é a melhor solução. Para isto foi feito um algoritmo de busca, visando fazer uma combinação ótima de regressores e seus parâmetros. Esta combinação é homogênea, cada regressor possui o mesmo peso no comitê.

O algoritmo, pode ser acessado em: https://github.com/hugoabreu1002/Optm_ensembl_e_data_cube/blob/master/Ensemble_on_capacidade.ipynb. Gera uma grande população de comitês, formada pela combinação desses regressores com parâmetros aleatórios. Formada esta primeira população aleatória, é feita uma avaliação sobre os

comitês e a partir disto, o algoritmo faz recombinações e cruzamentos entre os regressores para cada um dos comitês da população, depois disso todos os comitês da população são reavaliados e o ciclo de cruzamento e recombinações continua após um determinado número de ciclos ou condição de parada tenha sido atingida.

4 ESTIMATIVA DE PRODUÇÃO BASEADA NAS MATRIZES COVARIÂNCIA DE CORRELAÇÃO

Um estudo de caso, que pode ser acessado em: https://github.com/hugoabreu1002/PVIA-PE-CovCor/blob/master/Analise_CovCorr_KNN_e_Integrando_Bases.ipynb foi feito para o estado de Pernambuco, a partir do cubo de dados oriundo da metodologia da seção 2. Buscando gerar uma estimativa de produção baseada apenas nas matrizes covariância K e correlação R entre diversas variáveis, dos dados INMET e do LABREN.

A motivação para esta estimativa é levar em consideração todas as variáveis possíveis, além da radiação direta incidente, em uma determinada região. O desafio então é definir, quais e como outras variáveis vão influenciar a produção fotovoltaica.

Primeiramente, todas as colunas do cubo de dados devem estar normalizadas em escala de zero a unidade. Isto é necessário para que a ordem de grandeza de cada variável não influencie. Após isto, são necessárias as matrizes de covariância de correlação do cubo.

A estimativa é dada por E , na equação 1. Sendo X um vetor coluna de variáveis para cada ponto no espaço discreto (mapa de Pernambuco); a é um vetor linha referente a covariância entre a radiação direta normal e todas as outras variáveis (INMET e LABREN), enquanto b é, analogamente, o vetor linha referente a correlação.

$$E = a \cdot X^2 + b \cdot X \quad (1)$$

A estimativa consiste de uma combinação linear de dois termos, um quadrático e um linear. No termo quadrático, as variáveis do vetor coluna são potencializadas e multiplicadas pela covariância, isto indica que, como todas as variáveis são menores ou iguais à unidade, aquela que se distanciar mais influenciará menos. Nas regiões em que as variáveis

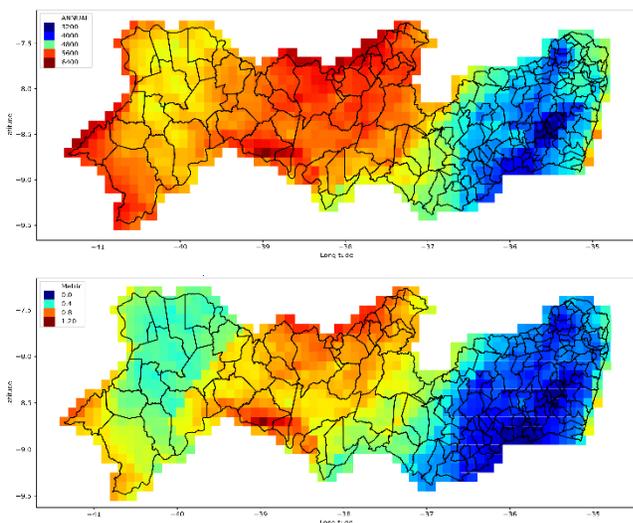


Figura 4: Visualização da radiação direta normal acima no estado de Pernambuco e avaliação da métrica proposta.

se distanciarem do seu máximo, o efeito da informação da covariância será reduzido.

O termo linear, da correlação serve para inserir a informação de quanto cada variável influencia a produção, a partir da sua relação com a radiação direta.

Os sinais dos valores de correlação das variáveis temperatura máxima média, umidade e precipitação, foram invertidos, por estas influenciarem negativamente o potencial de geração [38]–[40]. Uma visualização da métrica de estimativa proposta é possível com a Figura 4, juntamente com a radiação direta normal, percebe-se que em algumas regiões de máximo são atenuadas, levando a apenas um máximo global.

5 Dados e modelos temporais

A metodologia utilizada consiste em gerar um sistema Híbrido composto por um modelo ARIMA (*Autoregressive Integrated Moving Average*) e duas RNAs (Redes Neurais Artificiais) do tipo MLP (*Multi Layer Perceptron*). Primeiramente, o modelo ARIMA gera a previsão para a série temporal. No segundo passo, a primeira RNA estima o erro de previsão obtido pela diferença entre a série temporal e a estimação do ARIMA. Finalmente, a segunda RNA busca a função mais adequada para combinar a previsão do ARIMA com a previsão da primeira MLP. Este tipo de modelagem já foi muito utilizada e discutida na literatura [41]–[47].

As séries temporais utilizadas são obtidas a partir do balanço disponibilizado pela ONS [15]. Se tratam da média diária de geração de energia de quatro usinas fotovoltaicas. O modelo ARIMA utilizado neste trabalho faz uso de variáveis exógenas (a literatura referência este uso como modelo ARIMAX). Para uma das usinas, a informação de sazonalidade também é aproveitada, modelo SARIMAX (*Seasonal AutoRegressive Integrated Moving Average with exogenous regressors model*).

As séries temporais que atuam como variáveis exógenas são obtidas pelo INMET [21], pelo mesmo período que as séries temporais de geração de energia. Foram utilizadas como variáveis exógenas: Precipitação, temperatura máxima, temperatura mínima, insolação, temperatura compensada média, umidade relativa média e velocidade do vento média. Todas as séries são manipuladas após escalonamento entre 0 e 1.

Para obter as duas RNAs bem treinadas e parametrizadas, tanto para a função de associação não linear quanto para a previsão do erro, todo o modelo é envolto por busca baseado em um algoritmo genético. O algoritmo procura os melhores parâmetros para o MLP que estima o erro e o MLP que gera uma função não linear que associa o ARIMA puro e erro modelado.

Nesta busca, o algoritmo gera uma população de MLPs, com parâmetros aleatórios, avalia estes e ranqueia os melhores parâmetros. Depois disso, uma nova população é gerada, após um cruzamento entre melhores e piores MLPs. A melhor MLP é repetida na próxima geração. Depois do cruzamento, os parâmetros numéricos sofrem mutação. Esta nova população é reavaliada e o ciclo continua.

Para todas as séries foi estipulado 80% de dados para treino, que são utilizados para realizar o treinamento do algoritmo de otimização e hibridização, bem como todas as MLPs e 20% para teste, que utilizado para avaliação do modelo híbrido final.

Encapsulando, outra busca evolucionária, é realizada para achar, as variáveis de *Lag*, quantidade de dados do passado que serão utilizados para prever o próximo dado futuro e *Forecast*, quantidade de previsões a partir de um modelo. As variáveis em questão são descritas:

- *Lag Error*. Quantidade de amostras do passado utilizadas para prever a próxima amostra futura da série temporal do erro

entre a serie real e a do modelo ARIMA. Obtida sobre a serie do erro, é usado para modelar o erro entre a série original e a obtida pelo ARIMA;

- *Forecast Association Error*. Quantidade de amostras futuras geradas a partir do erro modelado; *Lag Association Error* Quantidades de amostras do passado do erro modelo: são usadas para modelar função associativa entre erro modelado e serie ARIMA.
- *Lag Association ARIMA*. Quantidade de amostras do passado da série do modelo ARIMA: obtida sobre a série do ARIMA, é usada para modelar função associativa entre erro modelado e serie ARIMA.

O algoritmo evolucionário escolhido não leva em consideração a variabilidade implícita de treinamento de MLPs, entretanto este foi programado de forma a salvar a cada geração os modelos com os pesos treinados, não apenas as topologias. Cada MLP utilizada no modelo híbrido proposto pode ter as seguintes características modificadas. Função de ativação (*activation*): Identidade; logística; tangente hiperbólica; relu. Atualização da taxa de aprendizagem (*learning rate*): constate, *invscaling*, adaptativa. Otimizador (*solver*): adam, lbfgs. Quantidade de neurônios nas três camadas escondidas.

6 RESULTADOS

6.1 Estimativa de produção baseada nas matrizes de covariância e correlação

Pelo que foi explicado na seção 3, é possível avaliar se a estimativa proposta está mais próxima da potência fotovoltaico real do que a radiação direta normal. Para tanto, é preciso que todos os dados estejam na mesma escala e fazer uma comparação com métricas MSE (Mean Squared Error) e MAE (Mean Absolute Error), com os dados de modelo de satélite das bases PVGIS NSRDB e SARAH, descritos na seção 2.3. Esta avaliação pode ser vista na Tabela 1. Percebe-se que a métrica proposta de fato se aproxima mais da estimava gerada pelos modelos PVGIS, do que a radiação direta normal.

Tabela 1: Avaliação do erro entre modelos de satélite, radiação direta normal e métrica proposta.

	NSRDB		SARAH	
	MSE	MAE	MSE	MAE
Radiação Direta Normal	0,0203	0,1247	0,0515	0,1993
Métrica Proposta	0,0194	0,1057	0,0263	0,1310

6.2 Otimização de comitês de regressores

A partir da metodologia descrita na seção 2.5, o resultado da otimização de comitês resultou em uma redução no erro entre a capacidade de produção real interpolada, da Figura 5 e a capacidade obtida pelo comitê de regressores. É visto a partir destas que a IA tem certa dificuldade em suavizar as fronteiras das estações INMET

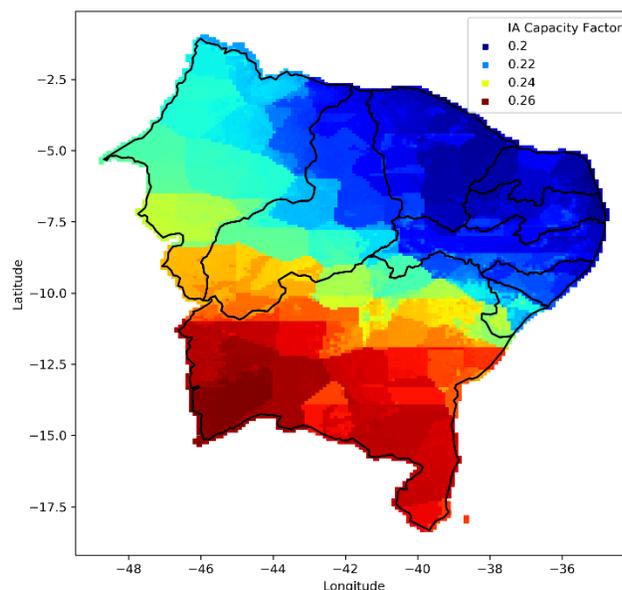


Figura 5: Fator de capacidade obtido a partir do comitê de regressores otimizado.

Comparando-se um comitê não otimizado, em que todos os objetos são inicializados na sua forma padrão pelo pacote do Scikit-Learn do Python [37], há uma redução de 10,43% (de 0,00488 para 0,00377) do erro médio absoluto, o erro quadrático médio foi reduzido em 23% (de 0,05544 para 0,04967).

6.3 Otimização de modelo híbrido de predição de series temporais

Pelo que foi apresentado na seção 4, é apresentado o resultado para algumas usinas, com os dados de teste, visualmente pode ser as séries não escalonadas, na Figura 6. Numericamente a Tabela 2 utiliza algumas métricas para avaliar o mesmo resultado apenas sobre os dados de teste. Cada uma das series de geração média diária resultou em uma topologia otimizada para o modelo híbrido. Estas topologias são mostradas a seguir. Todos os resultados podem ser acessados em: https://github.com/hugoabreu1002/series_temp_hibrid_ons.

Tabela 2: Avaliação do resultado do algoritmo de predição de series temporais utilizando métricas MAE, MSE e (Mean Absolute Percentage Error).

Rio Alto	MAE	MSE	MAPE
ARIMA	0,0722	0,0089	0,1084
Híbrido	0,0640	0,0067	0,0924
Assu 5	MAE	MSE	MAPE
ARIMA	0,0888	0,0132	0,1370
Híbrido	0,0839	0,0126	0,1250
Fontes Solar 1	MAE	MSE	MAPE
ARIMA	0,1181	0,0285	0,2647
Híbrido	0,1049	0,0184	0,2511
BJL solar	MAE	MSE	MAPE
ARIMA	0,0466	0,0042	0,0728
Híbrido	0,0626	0,0072	0,0975

Sobre as series temporais das usinas, a seguir está a compilação de suas características e as topologias dos modelos híbridos encontrado.

Usina Rio Alto: Nesta série são utilizados os últimos 86 pontos para teste. Sendo os primeiros 340 para treino. ARIMA (P=1, D=1, Q=1); MLP para modelagem do erro [activation: 'identity'; learning rate: 'adaptive'; solver: 'lbfgs'; neurônios nas camadas escondidas: (1ª=114, 2ª=30, 3ª=10)]; Lag Error: 10]; MLP para função de associação [activation : 'identity'; learning rate: 'invscaling';

solver: 'lbfgs'; neurônios nas camadas escondidas: (35, 44, 3)]; Lag Association ARIMA: 13; Lag Association Error: 13; Forecast Association Error: 15].

Assu 5: Nesta série são utilizados os últimos 31 pontos para teste. Sendo os primeiros 122 para treino. ARIMA (P=0, D=1, Q=1); MLP para modelagem do erro [activation: 'identity'; learning rate: 'invscaling'; solver: 'adam'; neurônios nas camadas escondidas: (1ª=78, 2ª=51, 3ª=8)]; Lag Error: 6]. MLP para função de associação [activation: 'tanh'; learning rate: 'constant'; solver: 'lbfgs'; neurônios nas camadas escondidas: (1ª=65, 2ª=30, 3ª=3)]; Lag Association ARIMA: 7; Lag Association Error: 2; Forecast Association Error: 15].

Fontes Solar 1: Nesta série são utilizados os últimos 107 pontos para teste. Sendo os primeiros 428 para treino. ARIMA (P=1, D=1, Q=1); MLP para modelagem do erro [activation: 'tanh'; learning rate: 'adaptive'; solver: 'adam'; neurônios nas camadas escondidas: (9, 30, 7)]; Lag Error: 4]. MLP para função de associação [activation: 'relu'; learning rate: 'invscaling'; solver: 'lbfgs'; neurônios nas camadas escondidas: (1ª=39, 2ª=33, 3ª=3)]; Lag Association ARIMA: 7; Lag Association Error: 18; Forecast Association Error: 8].

BJL Solar: Nesta série são utilizados os últimos 64 pontos para teste. Sendo os primeiros 256 para treino. Para esta série foi utilizada informação sazonal, resultando em um modelo SARIMAX. ARIMA (P=1, D=1, Q=1); Seasonal ARIMA (P=1, D=0, Q=1, s = Mensal); MLP para modelagem do erro [activation: 'relu'; learning rate: 'adaptive'; solver: 'adam'; neurônios nas camadas escondidas: (1ª=111, 2ª=24, 3ª=5)]; Lag Error: 4]; MLP para função de associação [activation: 'identity'; learning rate: 'invscaling'; solver: 'lbfgs'; neurônios nas camadas escondidas: (45, 43, 7)]; Lag Association ARIMA: 8; Lag Association Error: 5; Forecast Association Error: 18].

7 DISCUSSÃO

Na metodologia explicada na seção 3 é possível acrescentar o uso de um algoritmo de busca, no intuito de retornar pesos a cada variável e suas covariância e correlação, como foi feito, cada variável tem o mesmo peso na estimativa.

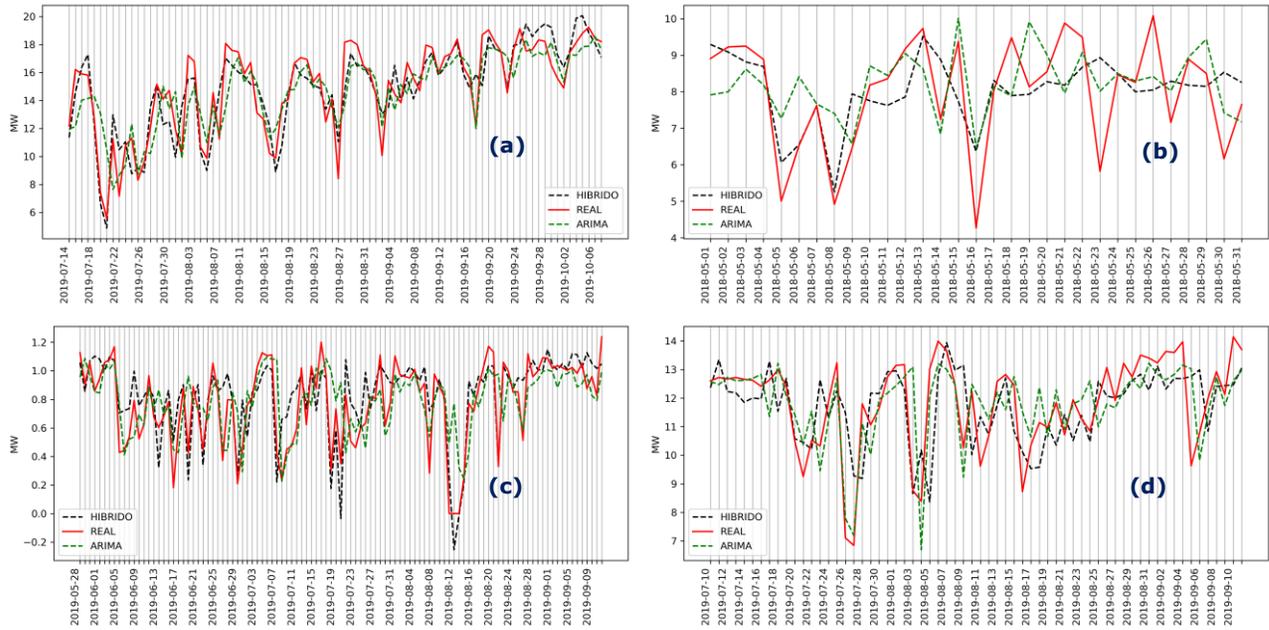


Figura 6: Resultado das análises de séries temporais para geração média diária de algumas usinas do nordeste. (a) Série temporal de geração média diária, modelo híbrido e modelo ARIMA para a usina fotovoltaica Rio Alto, localizada em Coremas-PB. (b) Série temporal de geração média diária, modelo híbrido e modelo ARIMA para a usina fotovoltaica Assu 5, localizada em Assu-RN. (c) Série temporal de geração média diária, modelo híbrido e modelo ARIMA para a usina fotovoltaica Fontes Solar 1, localizada em Tacaratu-PE. (d) Série temporal de geração média diária, modelo híbrido e modelo ARIMA para a usina fotovoltaica BJJ Solar, localizada em Bom Jesus da Lapa-BA.

A adição de um algoritmo de busca também pode ser usada para adicionar pesos a cada regressor integrante dos comitês da seção 2.5, ainda sobre a seção 2.5, uma melhor forma de avaliar os resultados seria fazer uma comparação estatística, a partir de várias execuções tanto do algoritmo de otimização de comitês utilizado, quanto das inicializações padrão dos regressores, preferencialmente definindo parâmetros aleatórios para estes.

Na seção 2.2 pode-se sugerir uma interpolação mais elaborada para os dados do INMET, neste trabalho optou-se por uma interpolação de vizinho mais próximo, basicamente um KNN (*k nearest neighbor*) com $K = 1$.

Outras melhorias e sugestões são bem vindas ao trabalho, cujo intuito é gerar um ambiente amigável para a implementação de inteligência artificial em projetos envolvendo usinas fotovoltaicas.

AGRADECIMENTOS

À todas as parcerias deste trabalho FITec/SECTI/CMParqtel/UPE/FACEPE que é fruto do

Edital RTIA, e a todos os funcionários e colegas da FITec e do Parqtel.

Referências

- [1] CHIN, V. J.; SALAM, Z.; ISHAQUE, K. **Cell modelling and model parameters estimation techniques for photovoltaic simulator application: A review**. Applied Energy, v. 154, p. 500-519, 2015.
- [2] JORDEHI, A. R. **Parameter estimation of solar photovoltaic (PV) cells: A review**. Renewable and Sustainable Energy Reviews, v. 61, p. 354-371, 2016.
- [3] DE AZEVEDO DIAS, C. L. et al. **Performance estimation of photovoltaic technologies in Brazil**. Renewable Energy, v. 114, p. 367-375, 2017.
- [4] MUELLER, R. W. et al. **The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance—A LUT based eigenvector hybrid approach**. Remote Sensing of Environment, v. 113, n. 5, p. 1012-1024, 2009.
- [5] HULD, T.; MÜLLER, R.; GAMBARELLA, A. **A new solar radiation database for estimating**

PV performance in Europe and Africa. Solar Energy, v. 86, n. 6, p. 1803-1815, 2012.

[6] AMILLO, A.; HULD, T.; MÜLLER, Richard. **A new database of global and direct solar radiation using the eastern meteosat satellite, models and validation.** Remote sensing, v. 6, n. 9, p. 8165-8189, 2014.

[7] HABTE, A.; SENGUPTA, M.; LOPEZ, A. **Evaluation of the national solar radiation database (NSRDB): 1998-2015.** National Renewable Energy Lab. (NREL), Golden, CO (United States), 2017.

[8] VOYANT, C. et al. **Machine learning methods for solar radiation forecasting: A review.** Renewable Energy, v. 105, p. 569-582, 2017.

[9] WOLFF, B.; LORENZ, E.; KRAMER, O. **Statistical learning for short-term photovoltaic power predictions.** Computational sustainability. Springer, Cham, 2016. p. 31-45.

[10] LI, Z. et al. **A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting.** Energies, v. 9, n. 1, p. 55, 2016.

[11] KHAN, G.; RATHI, S. **Optimal site selection for solar PV power plant in an Indian state using geographical information system (GIS).** International Journal of Emerging Engineering Research and Technology, v. 2, n. 7, p. 260-266, 2014.

[12] FERNANDEZ-JIMENEZ, L. A. et al. **Site selection for new PV power plants based on their observability.** Renewable energy, v. 78, p. 7-15, 2015.

[13] CARRIÓN, J. A. et al. **The electricity production capacity of photovoltaic power plants and the selection of solar energy sites in Andalusia (Spain).** Renewable Energy, v. 33, n. 4, p. 545-552, 2008.

[14] BORAN, F. E.; MENLIK, T.; BORAN, K. **Multi-criteria axiomatic design approach to evaluate sites for grid-connected photovoltaic power plants: A case study in Turkey.** Energy Sources, Part B: Economics, Planning, and Policy, v. 5, n. 3, p. 290-300, 2010.

[15] **Dados da geração solar fotovoltaica no sin.** <http://www.ons.org.br/Paginas/resultados-da-operacao/boletim-geracao-solar.aspx>. Acessado: 2019-09-03.

[16] **Matriz de energia elétrica.** <http://www2.aneel.gov.br/aplicacoes/capacidade-brasil/OperacaoCapacidadeBrasil.cfm>. Acessado: 2019-09-03.

[17] BUENO, E. P. et al. **Atlas Brasileiro de Energia Solar.** 2a. Edição Revisada e Ampliada São José dos Campos.

[18] **Atlas brasileiro de energia solar.** http://labren.ccst.inpe.br/atlas_2017.html. Acessado: 2019-09-04.

[19] MARTINS, F. R. et al. **Base de dados climático-ambientais aplicados ao setor energético-Projeto SONDA.** Proc. of XII Simpósio Brasileiro de Sensoriamento Remoto, INPE, Sao José dos Campos, Brazil, 2005.

[20] MARTINS, F. R. et al. **Mapeamento dos recursos de energia solar no brasil utilizando modelo de transferência radiativa brasil-sr.** Anais do I Congresso Brasileiro de Energia Solar. 2007. p. 8-10.

[21] **Instituto nacional de meteorologia.** <http://www.inmet.gov.br/portal/>. Acessado: 2019-10-02.

[22] **Photovoltaic geographical information system.** Disponível em: <https://re.jrc.ec.europa.eu/pvgis.html>, acessado: 2019-09-04.

[23] MÜLLER, R. et al. **Sarah-a new homogeneous climate data record of surface radiation.** EGU General Assembly Conference Abstracts. 2015.

[24] **Overview of pvgis data sources and calculation methods.** Disponível em: https://re.jrc.ec.europa.eu/pvg_static/methods.html. Acessado: 2019-10-11.

[25] **Other free solar radiation data and pv tools.** Disponível em: https://re.jrc.ec.europa.eu/pvg_static/data_sources.html. Acessado: 2019-10-11.

[26] **Interface for accessing pvgis data and calculations.** Disponível em: https://re.jrc.ec.europa.eu/pvg_static/web_service.html. Acessado: 2019-10-02.

[27] STEIN, M. L. **Interpolation of spatial data: some theory for kriging.** Springer Science & Business Media, 2012.

[28] P. developers, **PyKrige Documentation**, 2019, release 1.4.1. [Online]. Disponível em: <https://pykrige.readthedocs.io/en/latest/index.htm>

[29] OLIVER, M. A.; WEBSTER, R. **A tutorial guide to geostatistics: Computing and**

modelling variograms and kriging. Catena, v. 113, p. 56-69, 2014.

[30] SVETNIK, V. et al. **Random forest: a classification and regression tool for compound classification and QSAR modeling.** Journal of chemical information and computer sciences, v. 43, n. 6, p. 1947-1958, 2003.

[31] DRUCKER, H. et al. **Support vector regression machines.** Advances in neural information processing systems. 1997. p. 155-161.

[32] SOLOMATINE, D. P.; SHRESTHA, D. L. **AdaBoost. RT: a boosting algorithm for regression problems.** 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541). IEEE, 2004. p. 1163-1168.

[33] BREIMAN, L. **Bagging predictors.** Machine learning, v. 24, n. 2, p. 123-140, 1996.

[34] FRIEDMAN, J. H. **Stochastic gradient boosting.** Computational statistics & data analysis, v. 38, n. 4, p. 367-378, 2002.

[35] FISCHLER, M. A.; BOLLES, R. C. **Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.** Communications of the ACM, v. 24, n. 6, p. 381-395, 1981.

[36] CRAMMER, K. et al. **Online passive-aggressive algorithms.** Journal of Machine Learning Research, v. 7, n. Mar, p. 551-585, 2006.

[37] PEDREGOSA, F. et al. **Scikit-learn: Machine learning in Python.** Journal of machine learning research, v. 12, n. Oct, p. 2825-2830, 2011.

[38] SUN, Y. et al. **Correlation feature selection and mutual information theory based quantitative research on meteorological impact factors of module temperature for solar photovoltaic systems.** Energies, v. 10, n. 1, p. 7, 2017.

[39] KAZEM, H. A.; CHAICHAN, M. T. **Effect of humidity on photovoltaic performance based on experimental study.** International Journal of Applied Engineering Research (IJAER), v. 10, n. 23, p. 43572-43577, 2015.

[40] HAILEGNAW, B. et al. **Rain on methylammonium lead iodide based perovskites: possible environmental effects of perovskite solar cells.** The journal of physical chemistry letters, v. 6, n. 9, p. 1543-1547, 2015.

[41] ZHANG, G. P. **Time series forecasting using a hybrid ARIMA and neural network model.** Neurocomputing, v. 50, p. 159-175, 2003.

[42] KHASHEI, M.; BIJARI, M. **An artificial neural network (p, d, q) model for timeseries forecasting.** Expert Systems with applications, v. 37, n. 1, p. 479-489, 2010.

[43] BABU, C. N.; REDDY, B. E. **A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data.** Applied Soft Computing, v. 23, p. 27-38, 2014.

[44] DE OLIVEIRA, J. F. L.; LUDERMIR, T. B. **A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting.** 2014 Brazilian Conference on Intelligent Systems. IEEE, 2014. p. 73-78.

[45] DE OLIVEIRA, J. F. L.; LUDERMIR, T. B. **A hybrid evolutionary decomposition system for time series forecasting.** Neurocomputing, v. 180, p. 27-34, 2016.

[46] DOMINGOS, S. de O.; DE OLIVEIRA, J. F. L.; DE MATTOS NETO, P. S. G. **An intelligent hybridization of ARIMA with machine learning models for time series forecasting.** Knowledge-Based Systems, v. 175, p. 72-86, 2019.

[47] DE MATTOS NETO, P. S. G.; CAVALCANTI, G. D. C.; MADEIRO, F. **Nonlinear combination method of forecasters applied to PM time series.** Pattern Recognition Letters, v. 95, p. 65-72, 2017.