



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

## **USO DE APRENDIZADO DE MÁQUINA PARA GERAÇÃO DE SÉRIES TEMPORAIS DE IRRADIAÇÃO SOLAR NA SUPERFÍCIE**

Mariane Souza Guarachi

Relatório de Iniciação Científica do programa PIBIC, orientada pelo Dr. André Rodrigues Gonçalves e Dr. Fernando Ramos Martins.





MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

## **USO DE APRENDIZADO DE MÁQUINA PARA GERAÇÃO DE SÉRIES TEMPORAIS DE IRRADIAÇÃO SOLAR NA SUPERFÍCIE**

Mariane Souza Guarachi

Relatório de Iniciação Científica do  
programa PIBIC, orientada pelo Dr.  
André Rodrigues Gonçalves e Dr.  
Fernando Ramos Martins.

URL do documento original:

INPE  
São José dos Campos  
2022



## **AGRADECIMENTOS**

A Deus, por ter me dado saúde e proporcionado condições para que eu conquistasse meus sonhos e vencesse os obstáculos encontrados ao longo da realização deste estudo.

Aos meus pais e irmãos por todo apoio, amor e carinho, além de sempre me incentivarem a ir atrás do que eu realmente desejo.

Aos meus amigos que acreditaram em mim e que contribuíram de alguma forma para que eu desenvolvesse este projeto.

Ao auxílio financeiro do Instituto Nacional de Pesquisas Espaciais (INPE) e o Centro de Pesquisa e Desenvolvimento Científico e Tecnológico (CNPq), contemplada pelo PIBIC.

Em especial gostaria de agradecer meu orientador André Rodrigues Gonçalves e o meu coorientador Fernando Ramos Martins pelos ensinamentos passados, pela paciência no meu processo de aprendizagem, permitindo que eu desempenhasse o meu melhor e por todos os conselhos compartilhados.



## RESUMO

É de suma importância ter disponível séries temporais de irradiância solar de um determinado local na superfície da Terra para aplicação em áreas científicas e tecnológicas. Em um caso mais específico, como a geração de eletricidade pela energia solar são necessários dados detalhados de irradiância solar para modelar o desempenho de uma planta fotovoltaica. Frequentemente, estas séries temporais são obtidas pelas medições realizadas no solo ou são derivadas de imagens de satélites. No entanto, as observações em solo podem apresentar descontinuidades, que duram de minutos a dias, no registro dos dados causadas pela manutenção preventiva e corretiva deficientes e falta de energia nos equipamentos. Esta falta de dados é prejudicial para o estudo de viabilidade de uma planta fotovoltaica, pois isso acarreta no crescimento das incertezas sobre o potencial técnico disponível no local. Portanto, é necessário a utilização de séries temporais geradas sinteticamente ou matematicamente para preencher as lacunas de dados das medições. Com isso, o presente projeto de pesquisa tem como objetivo desenvolver metodologias de aprendizado de máquina para obter estimativas de irradiação solar de São Martinho da Serra e Petrolina por meio da geração de séries temporais. A metodologia de aprendizado de máquina com a utilização de redes neurais artificiais vai ser aplicada para a geração de séries temporais, utilizando a linguagem de programação Python. Por fim, a série artificial possibilitou investigar o desempenho no preenchimento de lacunas de dados em séries temporais.

Palavra-chave: Energia solar; modelagem numérica; séries temporais.



# **USE OF MACHINE LEARNING FOR GENERATING TIME SERIES OF SOLAR SURFACE IRRADIATION**

## **ABSTRACT**

It is imperative to have available time series of solar irradiance data from a given location on the Earth's surface for application in scientific and technological areas. In a more specific case, such as solar power generation, detailed solar irradiance data are needed to model the performance of a photovoltaic plant. Often, these time series are obtained from ground measurements or derived from satellite imagery. However, time series of ground observations can present discontinuities, which last from minutes to days, caused by poor preventive and corrective maintenance or power outages. The data gaps hinder the feasibility study of a photovoltaic plant, as it leads to growing uncertainties about the technical potential available at the site. Several studies discuss computational methods to fill data gaps based on the time series attributes. The current study aims to develop a machine learning methodology to fill solar irradiation data gaps for ground data acquired at São Martinho da Serra and Petrolina. INPE will provide access to the data acquired in both SONDA Network sites.

Keywords: Solar energy. Numerical modeling. Time series.



## LISTA DE FIGURAS

	<u>Pág.</u>
Figura 1 – Representação do Ângulo de Zênite, de Azimute e do Ângulo Solar.	5
Figura 2 – Representação da declinação solar. ....	6
Figura 3 – Neurônio artificial. A saída do neurônio é representada pela letra $y$ , as entradas são representadas por $X_1, X_2, \dots, X_n$ , o potencial interno do neurônio ( $\xi$ ), a função de ativação não constante ( $\sigma$ ) que determina a saída do neurônio. ....	11
Figura 4 – Fluxograma dos próximos passos a serem realizados no projeto...	13
Figura 5 – Radiação Máxima Mensal. ....	16
Figura 6 – Radiação Média Mensal. ....	17



## SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO.....	1
2 OBJETIVOS.....	3
2.1 Objetivo Geral.....	3
2.2 Objetivos Específicos.....	3
3 REVISÃO BIBLIOGRÁFICA .....	4
3.1 Radiação Solar .....	4
3.2 Modelagem e Séries Temporais .....	7
3.3 Redes Neurais Artificiais.....	10
4 METODOLOGIA .....	13
4.1 Modelagem Numérica.....	13
4.2 Dados Observados .....	14
5 RESULTADOS .....	16
6 CONCLUSÃO .....	18
REFERÊNCIAS BIBLIOGRÁFICAS .....	19



## 1 INTRODUÇÃO

Atualmente tem-se o conhecimento que o consumo e a geração de eletricidade através de fontes não renováveis ocasiona danos ao meio ambiente, como o aquecimento global e as mudanças climáticas, provocados pelo aumento da temperatura do planeta e pela emissão de gases do efeito estufa (FIORIN et al., 2011). Por isso, a transição para uma matriz energética baseada em fontes renováveis se faz cada vez mais necessária (FIORIN et al., 2011). Além da questão ambiental é importante diversificar a matriz para garantir maior desempenho e segurança na distribuição de eletricidade para o país (FIORIN et al., 2011).

Uma alternativa para as questões levantadas anteriormente é a utilização da energia solar para geração de eletricidade. No Brasil, a utilização do recurso solar para geração de energia enfrenta vários desafios, mesmo que seja um país tropical, como a demanda por capacitação de profissionais, desenvolvimento tecnológico na avaliação/previsão e na conversão do recurso solar no território nacional, elevado investimento em infraestrutura no sistema elétrico nacional, lacunas de dados de campo e grande influência do clima (PEREIRA et al, 2017, FIORIN et al., 2011). Com isso, baseado nas questões ambientais e nos desafios encontrados surge a necessidade de se desenvolver pesquisas para avanço nos conhecimentos e desenvolvimento tecnológico baseado em dados sobre a irradiação solar na superfície.

O recurso solar apresenta intermitências, tornando necessário o uso de séries de alta resolução temporal para simular com mais confiabilidade o desempenho dos sistemas de conversão de energia. As séries temporais de irradiância solar disponíveis influenciam diretamente na tomada de decisão sobre a implementação e operação de uma planta fotovoltaica para geração de eletricidade em um determinado local (POLO et al., 2011).

A obtenção destas séries temporais ocorre a partir de medições realizadas no solo ou são derivadas de imagens de satélites (POLO et al., 2011). No entanto, as medições de radiação solar podem apresentar descontinuidades no registro dos dados, que duram de minutos a dias, o que geralmente é causado pelo

mau funcionamento dos instrumentos de medições, manutenção preventiva e corretiva deficientes e ausência de energia nos equipamentos (SCHWANDT et al., 2014).

A ausência de dados provoca um aumento das incertezas envolvidas na modelagem do potencial energético disponível e com isso compromete economicamente a projeção de instalação de uma planta ou sistema fotovoltaico para um determinado local, pois os investimentos necessários são altos para serem realizados em meio a grandes incertezas.

Por isso, de acordo com a natureza dos dados ausentes se faz necessária a utilização de séries temporais sintéticas geradas matematicamente para preencher as lacunas de dados das medições (POLO et al., 2011). A geração de séries temporais pode ser realizada a partir de modelos estatísticos utilizando várias metodologias diferentes, inclusive o aprendizado de máquina com a utilização de redes neurais artificiais capazes de acumular conhecimentos a partir de exemplos e repetições. Este trabalho propõe o desenvolvimento e emprego de metodologias de aprendizado de máquina para o preenchimento de lacuna de dados de irradiação solar por meio de séries temporais.

Os métodos estatísticos preveem e estimam os dados de irradiação solar que incidem na atmosfera em um determinado local de interesse por meio de equações empíricas. Estes métodos podem combinar dados obtidos pelas imagens de satélites com medições feitas no solo (FIORIN et al., 2011).

Os modelos de Redes Neurais Artificiais contemplam um modelo estatístico baseado em unidades simples de processamento. Logo, as Redes Neurais Artificiais possuem o potencial de auto-organização, generalização e processamento temporal das informações (FIORIN et al., 2011).

## **2 OBJETIVOS**

### **2.1 Objetivo Geral**

Esse trabalho tem como objetivo geral o desenvolvimento de metodologia de aprendizado de máquina para obtenção de estimativas de irradiação solar por meio da geração de séries temporais de dados.

### **2.2 Objetivos Específicos**

- i. Qualificar e organizar uma série temporal de dados das localidades de estudo;
- ii. Treinar e aplicar uma metodologia de aprendizado de máquina para preenchimento de lacunas de observações realizadas em estações de coleta de dados em superfície;
- iii. Fazer uma análise comparativa dos resultados alcançados em cada região.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Radiação Solar

A Terra possui o Sol como sua principal fonte de energia (GUARNIERI, 2006). A radiação solar representa a principal fonte de energia para que ocorram na Terra e na atmosfera os processos químicos, dinâmicos e térmicos (MARTINS et al., 2004). A energia solar incide na Terra por meio de ondas eletromagnéticas em uma faixa contínua de comprimentos de onda que compreende o intervalo entre 200 nm e 4000 nm, de acordo com a lei de Planck (GUARNIERI, 2006).

Da energia emitida pelo Sol, cerca de 49% representa a radiação infravermelha, 43% a radiação de luz visível, 7% a radiação ultravioleta e menos de 1% representa os raios gama, raios-X e ondas de rádio (FIORIN et al., 2011).

Ao incidir na atmosfera a radiação pode ser espalhada ou absorvida, e aproximadamente 25% dessa radiação entra em contato com a superfície sem sofrer influência da atmosfera e os outros 75% são absorvidos ou espalhados de volta ao espaço ou em direção a Terra (FIORIN et al., 2011).

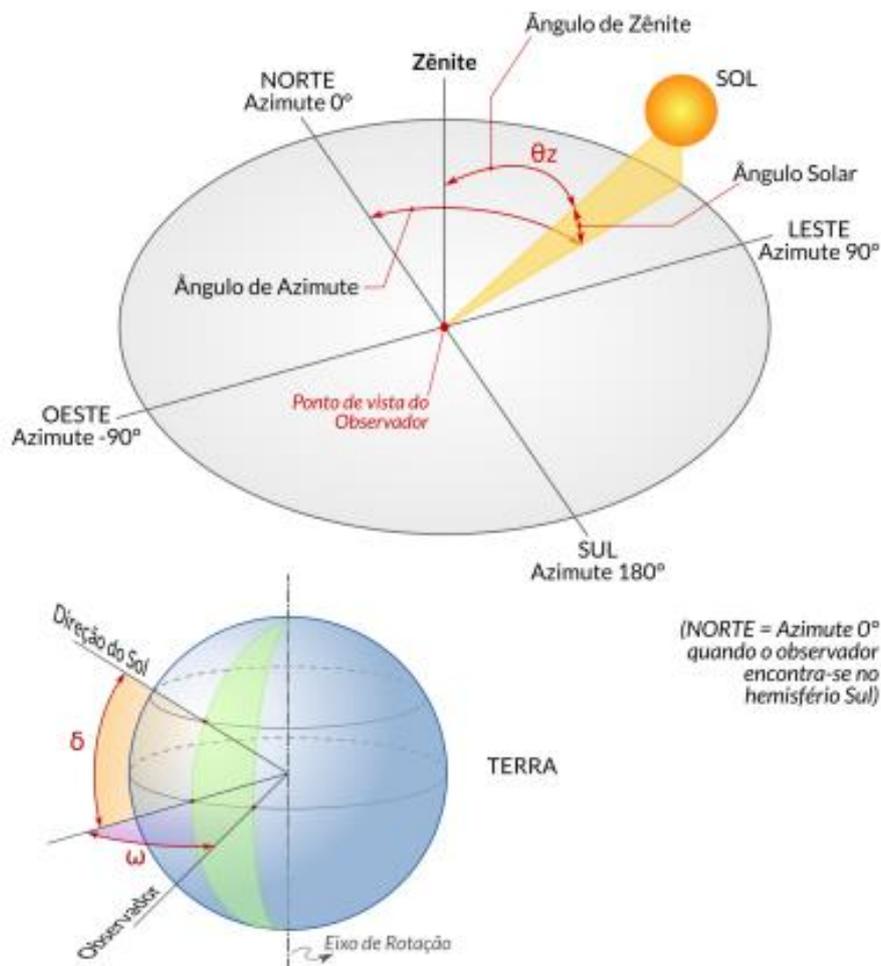
O fluxo radiante é a taxa com que ocorre a transferência de energia por radiação eletromagnética, sendo definido em unidades de energia por unidade de tempo (Joules por segundo -  $J.s^{-1}$ ) ou Watts - W) (GUARNIERI, 2006). Já o fluxo radiante que passa por uma unidade de área é caracterizado pela densidade de fluxo radiante, também conhecida como irradiância, expressa em Watts por metro quadrado ( $W.m^{-2}$ ) (FIORIN et al., 2011; MARTINS et al., 2004).

A radiação solar incide na superfície plana com diferentes ângulos de inclinação (GUARNIERI, 2006). Definimos o zênite local como linha que forma um ângulo de  $90^\circ$  com a superfície terrestre. O ângulo zenital solar ( $\theta_z$ ) é definido como o ângulo entre a vertical da superfície (normal) e a direção da radiação no mesmo local da Terra ao Sol, podendo variar de  $0^\circ$  a  $90^\circ$ . O ângulo

azimutal solar é definido como o ângulo formado entre o meridiano da localização do observador e a projeção da linha imaginária entre observador e o Sol no plano horizontal, podendo variar de  $-180^\circ$  a  $+180^\circ$  (MARTINS e PEREIRA, 2019).

O ângulo azimutal solar é definido como o ângulo formado entre o meridiano da localização do observador e a projeção da linha imaginária entre observador e o Sol no plano horizontal, podendo variar de  $-180^\circ$  a  $+180^\circ$  (PEREIRA et al., 2017).

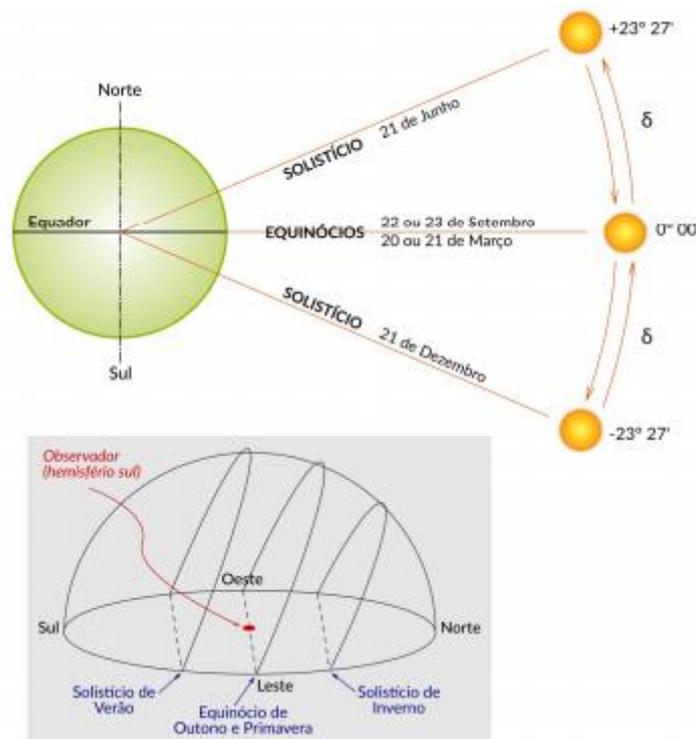
Figura 1 – Representação do Ângulo de Zênite, de Azimute e do Ângulo Solar.



Fonte: Pereira et al. (2017).

O ângulo horário solar ( $\omega$ ) é definido como o deslocamento angular do movimento aparente do Sol devido à rotação da Terra, podendo variar entre  $-180^\circ$  e  $+180^\circ$  (PEREIRA et al., 2017). A cada uma hora o ângulo horário solar varia  $15^\circ$ , e, por convenção utilizamos o ângulo positivo na parte da manhã e negativo a tarde, e ao meio dia é considerado zero. A declinação solar ( $\delta$ ) é definida como o ângulo formado pela linha de radiação que liga o Sol a Terra e a inclinação do plano equatorial da Terra, podendo variar entre  $-23^\circ 27'$  e  $+23^\circ 27'$  no período de um ano (PEREIRA et al., 2017).

Figura 2 – Representação da declinação solar.



Fonte: Pereira et al. (2017).

A separação das componentes da radiação solar incidente depende de todos os processos radiativos que ocorreram até que ela atinja a superfície da Terra. Com isso, a irradiância difusa é aquela que incide na superfície horizontal resultante do espalhamento ocasionado pelas nuvens, material particulado e gases presentes na atmosfera (FIORIN et al., 2011).

A irradiância direta é a que incide na superfície horizontal sem sofrer nenhum processo de espalhamento. Já a irradiância direta normal é a irradiância que possui a sua incidência em um ângulo de  $90^\circ$  com a superfície e a irradiância global é a soma da irradiância direta com a irradiância difusa, incidentes no plano horizontal (FIORIN et al., 2011).

### **3.2 Modelagem e Séries Temporais**

A disponibilidade de energia solar incidente na superfície terrestre pode ser avaliada através de uma rede de radiômetros e de interpolações das medidas de radiação (GUARNIERI, 2006). Porém, para que esta avaliação seja confiável é necessário que se tenha uma quantidade elevada de radiômetros com pouco distanciamento entre si, cerca de 40-50 km (MARTINS e PEREIRA, 2011). Logo, uma proposta mais interessante para a obtenção de estimativas de radiação em áreas de grande extensão é a utilização de modelos computacionais (PEREIRA et al., 2017).

Os modelos computacionais utilizados para estimar a radiação solar incidente são divididos em modelos estatísticos e modelos físicos (MARTINS e PEREIRA, 2019). Os modelos estatísticos são definidos pelo emprego de expressões empíricas que estimam a radiação incidente em um determinado local na Terra em detrimento das propriedades do sistema Terra-Atmosfera obtidas pelos instrumentos no solo ou pelos satélites (FIORIN et al., 2011).

A tecnologia de sensoriamento remoto por meio de satélites artificiais proporcionou a obtenção de imagens de satélites, as quais são importantes para a definição da cobertura de nuvens e outros parâmetros atmosféricos (MARTINS et al., 2004). Já os modelos físicos são definidos pelo emprego de métodos matemáticos que simulam as condições atmosféricas e resolvem a equação de transferência radioativa. Além disso, a credibilidade do modelo físico está relacionada com a confiabilidade das informações disponíveis sobre as condições atmosféricas (MARTINS e PEREIRA, 2019).

Independente do modelo utilizado, possuir séries temporais de irradiância solar de um local em específico com dados confiáveis é imprescindível para modelar

e prever o desempenho de uma planta fotovoltaica (POLO et al., 2011). Portanto, quando não se é possível ter todos os dados de irradiação disponíveis, devido a alguma falha nas medições ou outros motivos descritos anteriormente, uma possível solução é empregar o modelo estatístico de aprendizado de máquina para geração de séries temporais sintéticas de irradiação solar na superfície.

O trabalho desenvolvido por Moravčik (2016) verificou a possibilidade de corrigir a medição automática utilizando redes neurais artificiais e obteve como resultado a ferramenta PANN de linha de comando flexível para treinamento de redes neurais. De acordo com Moravčik (2016), a rede neural GRU (Gated Recurrent Unit) possui o potencial de contemplar as dependências temporais dos dados, o que apresenta uma vantagem pois o percentual de erros encontrado foi baixo. A utilização desta rede neural proporcionou bons resultados, confirmando ser a melhor escolha para o preenchimento de lacunas de dados (MORAVČIK, 2016).

Moravčik (2016) utilizou uma rede definida como um codificador automático empilhado, com uma camada de saída de neurônios com tamanho igual ao da entrada, a camada de saída foi determinada de uma forma para não limitar os valores a um intervalo. As variáveis obtidas por diferentes sensores independentes são as que apresentaram um preenchimento melhor e os sensores que estão próximos ao solo são os mais significativos na formação de dependências, uma vez que são estáveis e com isso apresentam menos ruídos (MORAVČIK, 2016).

Spokas e Forcella (2006) desenvolveram um modelo empírico para estimar a radiação solar incidente com intervalo de tempo de uma em uma hora, os parâmetros de entrada utilizados no modelo empírico foram latitude, longitude, elevação do local de campo, precipitação diária com temperaturas do ar mínimas e máximas diárias. Os resultados apresentaram vantagens, pois pode-se prever a radiação solar durante o ano todo a partir de dados climáticos limitados e sem necessidade de calibração específica do local (SPOKAS; FORCELLA, 2006).

O modelo apresentado por Spokas e Forcella (2006) superestimou os valores de radiação solar na primavera e no verão e subestimou no outono e inverno, a superestimação pode acontecer em razão do aumento da reflexão da radiação solar na cobertura de neve ou pela dispersão atmosférica devido aos ângulos de declinação solar mais baixos. De acordo com Spokas e Forcella (2006) a precisão dessas previsões poderia ser mais alta se a precipitação e a temperatura fossem registradas durante as lacunas nas leituras do radiômetro. Este modelo pode ser aplicado no preenchimento de lacunas de radiação solar nas medições obtidas em estações meteorológicas (SPOKAS; FORCELLA, 2006).

O estudo realizado por Karim e Singh (2014) aplica o ajuste polinomial para modelar a radiação solar global, os dados foram filtrados usando métodos de ajuste e suavização com diferentes graus de ajustes de curvas polinomiais. O modelo de ajuste polinomial pode ser utilizado para dimensionar o sistema de geração de energia solar (KARIM; SINGH, 2014). Karim e Singh (2014) concluíram que pelos resultados numéricos, o modelo de ajuste com ordem de segundo grau fornece melhores resultados sem qualquer oscilação nos pontos finais do gráfico.

Gerber et al. (2018) propuseram um método de estimativa espaço-temporal para preencher lacunas em conjuntos de dados de sensoriamento remoto, este método observa a coerência espacial e temporal que são característicos aos conjuntos de dados. De acordo com Gerber et al. (2018), a regressão quantílica proporciona uma recuperação de parâmetros mais precisa na análise de conjuntos de dados de sensoriamento remoto, a validação mostrou que o método proposto tem um bom desempenho em termos de erro de predição quadrático médio.

O método de previsão desenvolvido por Gerber et al. (2018), indicou ótimos resultados para valores ausentes em conjuntos de dados de sensoriamento remoto. O método conseguiu prever todos as lacunas em conjuntos de dados com grandes proporções de valores ausentes, os dados previstos reconstruíram os padrões espaciais e temporais dos conjuntos de dados com

muitos detalhes e forneceu intervalos de previsão, que quantificam as incertezas de previsão com base em considerações estatísticas (GERBER et al., 2018).

Pelos resultados obtidos por Schwandt et al. (2014), pode-se perceber que a maioria das lacunas nos dados de irradiância horizontal global e horizontal difusa ocorrem quando valores de radiação solar são altos, provavelmente durante o meio-dia. Para lacunas superiores a três horas na irradiância horizontal global, irradiância horizontal difusa e irradiância direta normal, os dados dos dias vizinhos são usados para preencher as lacunas, com isso, as mudanças na cobertura das nuvens levam a alterações nos valores da radiação solar (SCHWANDT et al., 2014).

A metodologia aplicada depende da disponibilidade da irradiância horizontal global, irradiância horizontal difusa e irradiância normal direta e da duração da lacuna (SCHWANDT et al., 2014). O método de preenchimento de lacunas diferencia se apenas um, dois ou todos os três componentes de radiação solar não estiverem disponíveis (SCHWANDT et al., 2014). Uma melhoria adicional no procedimento de preenchimento de lacuna pode ser alcançada quando dois componentes da irradiância solar estiverem faltando por interpolação linear dos índices de clareza ( $k_t$ ) definido pela razão entre a irradiância solar incidente na superfície e no topo da atmosfera. (SCHWANDT et al., 2014).

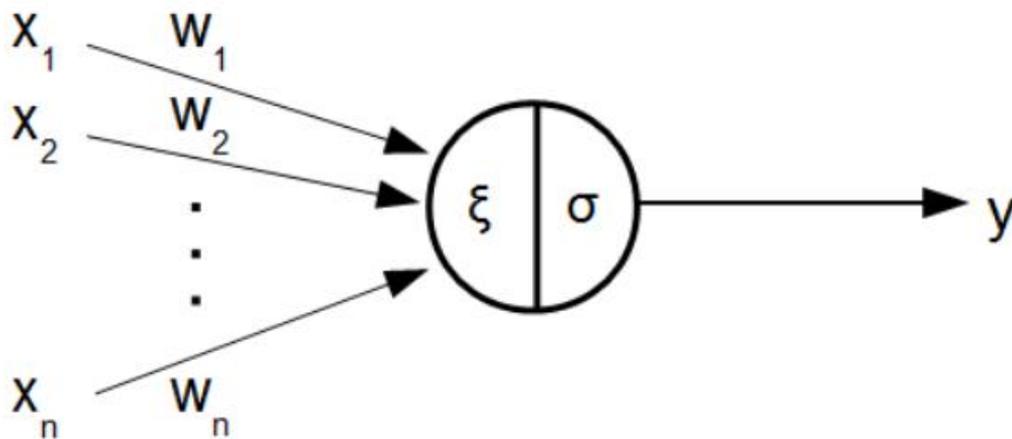
### **3.3 Redes Neurais Artificiais**

As Redes Neurais Artificiais são consideradas uma ferramenta estatística e possuem em sua composição neurônios, que a partir dos dados de entrada fazem a computação de funções matemáticas não lineares. Os neurônios, também chamados de unidades de processamento devido a sua principal função, são colocados em paralelo formando camadas que são interligadas entre si a partir de conexões (sinapses) associadas a pesos sinápticos (SRIKRISHNAN et al., 2015).

Os conhecimentos adquiridos pela RNA são armazenados pelos pesos sinápticos, por meio da ponderação das entradas de cada um dos neurônios. A

princípio, para que se utilize a RNA é necessário que ela seja treinada para aprender os processos requeridos na resolução do problema, e para isso os pesos sinápticos são modificados até que a resolução seja aceitável. Logo, após o treinamento os pesos sinápticos são fixados, e o modelo de rede neural está pronto para, a partir de dados de entrada disponíveis estimar os dados de saída que seriam observados no local de estudo (DONGARE et al., 2012).

Figura 3 – Neurônio artificial. A saída do neurônio é representada pela letra  $y$ , as entradas são representadas por  $X_1, X_2, \dots, X_n$ , o potencial interno do neurônio ( $\xi$ ), a função de ativação não constante ( $\sigma$ ) que determina a saída do neurônio.



Fonte: Moravčik (2016).

Os treinamentos são divididos em supervisionados e não-supervisionados. O treinamento supervisionado consiste na utilização de um vetor de entrada para calcular o vetor de saída e comparar o resultado obtido com o vetor alvo, logo, os pesos sinápticos são controlados para que os vetores de saída coincidam com o alvo. O ajuste dos pesos é realizado após a apresentação de todos os exemplos de dados de treinamento para a RNA (treinamento por lote) ou a cada iteração (treinamento sequencial) (GUARNIERI, 2006).

As RNAs possuem diferentes arquiteturas, e a escolha de qual utilizar está diretamente ligado com o problema a ser resolvido. O que varia de uma arquitetura para outra é a quantidade de neurônios e de camadas, a topologia da RNA e o tipo de conexão entre os neurônios.

As RNAs podem possuir apenas uma camada, ou seja, apenas um neurônio entre a camada de entrada e a de saída, ou podem possuir multicamadas com um conjunto de neurônios, uma camada de entrada e saída e uma ou mais camadas intermediárias (FIORIN et al., 2011).

Os neurônios possuem conexões que são classificadas em feedback (cíclica), onde a saída de um neurônio é a entrada de outro pertencentes a mesma ou a distintas camadas, e em feedforward (acíclica), onde a saída de um neurônio condiz com a entrada de outro neurônio pertencente a uma camada posterior.

As conexões do tipo cíclica ou recorrente são aquelas que possuem realimentação da saída na entrada. Logo, as RNAs podem ser divididas em Perceptrons de Múltiplas Camadas (MLP), Processamento Temporal, Self-Organizing, Perceptron e Adaline, Memórias Matriciais, entre outras (FIORIN et al., 2011).

De acordo com o trabalho de Fiorin et al. (2011), pode-se perceber que a utilização de um modelo empírico como a RNA Perceptrons de Múltiplas Camadas (MLP) se mostrou eficaz para estimar os parâmetros relacionados com a incidência de irradiação na superfície e modelar parâmetros instantâneos de radiação derivados dos dados de cobertura de nuvens.

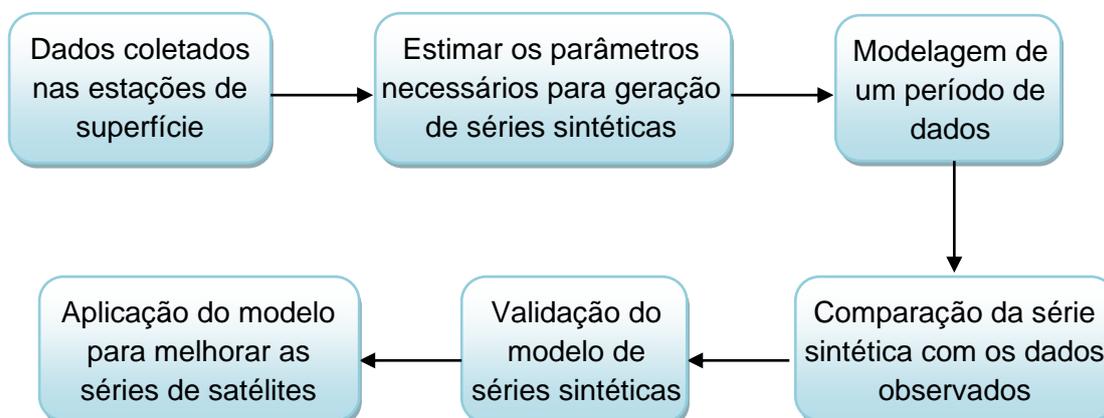
Em geral, as RNAs demonstraram melhor desempenho nas previsões do que os modelos que utilizam regressão linear múltipla, em razão da RNA ser eficiente na exploração das não-linearidades das análises realizadas e das informações dos diferentes parâmetros de entrada. As redes MLP se destacam por serem o tipo mais simples capaz de representar qualquer função contínua, como processos físicos, por isso são muito utilizadas em geociências (processos naturais são tipicamente contínuos).

## 4 METODOLOGIA

O trabalho realizado compreendeu uma revisão bibliográfica, por meio de pesquisas em periódicos (Google Scholar, Science Direct e CAPES) e pesquisas em busca de artigos científicos, teses ou dissertações. Esta revisão bibliográfica possuiu o objetivo de familiarização com o tema e com alguns conceitos fundamentais e essenciais para se iniciar o trabalho.

Com isso, foi realizado um levantamento bibliográfico do conhecimento publicado sobre a geração de séries temporais e capacitação no uso da linguagem Python. Em continuidade foram utilizados diferentes conjuntos de preditores e topologias de redes neurais com o intuito de investigar o desempenho no preenchimento de dados em séries temporais. O fluxograma abaixo demonstra os passos de desenvolvimento do trabalho.

Figura 4 – Fluxograma dos próximos passos a serem realizados no projeto.



Fonte: Produção da autora.

### 4.1 Modelagem Numérica

O presente estudo utiliza redes neurais artificiais alimentadas por estimativas de irradiação solar na superfície produzidos pelo modelo de transferência radiativa BRASIL-SR e dados meteorológicos observados em estações da rede SONDA. O modelo BRASIL-SR simula os processos físicos de atenuação da radiação solar na atmosfera utilizando dados observados por satélite.

As estimativas produzidas pelo modelo possuem resolução temporal igual às imagens de satélite (entre 30 minutos e 1 hora). As redes neurais artificiais serão utilizadas para refinar as estimativas produzidas pelos modelos numéricos mencionados anteriormente produzindo estimativas de irradiação solar incidente na superfície com resolução de 10 minutos ou menos. Considerando essa uma resolução temporal adequada para a simulação de operação de plantas de geração fotovoltaica.

A proposta tem como ponto de partida a aplicação de Redes Neurais Artificiais (RNA) amplamente adotadas para simulação de processos não-lineares. A utilização desta metodologia de aprendizagem de máquina é resultado da capacidade de auto-organização, generalização e processamento temporal que permite solucionar problemas de alta complexidade (FIORIN et al., 2011). A organização de uma RNA é dependente da complexidade do problema a ser resolvido, variando desde problemas mais simples (linearmente separáveis) até problemas mais complexos (não-lineares).

## **4.2 Dados Observados**

O Sistema de Organização Nacional de Dados Ambientais (SONDA) é um projeto do Instituto Nacional de Pesquisas Espaciais (INPE) que visa a implementação de infraestrutura, tanto física quanto de recursos humanos, para aprimoramento na base de dados coletados em superfície indispensáveis para o levantamento do potencial de recurso eólico e solar, este que é obtido através do modelo de satélite, BRASIL-SR (SONDA, 2021).

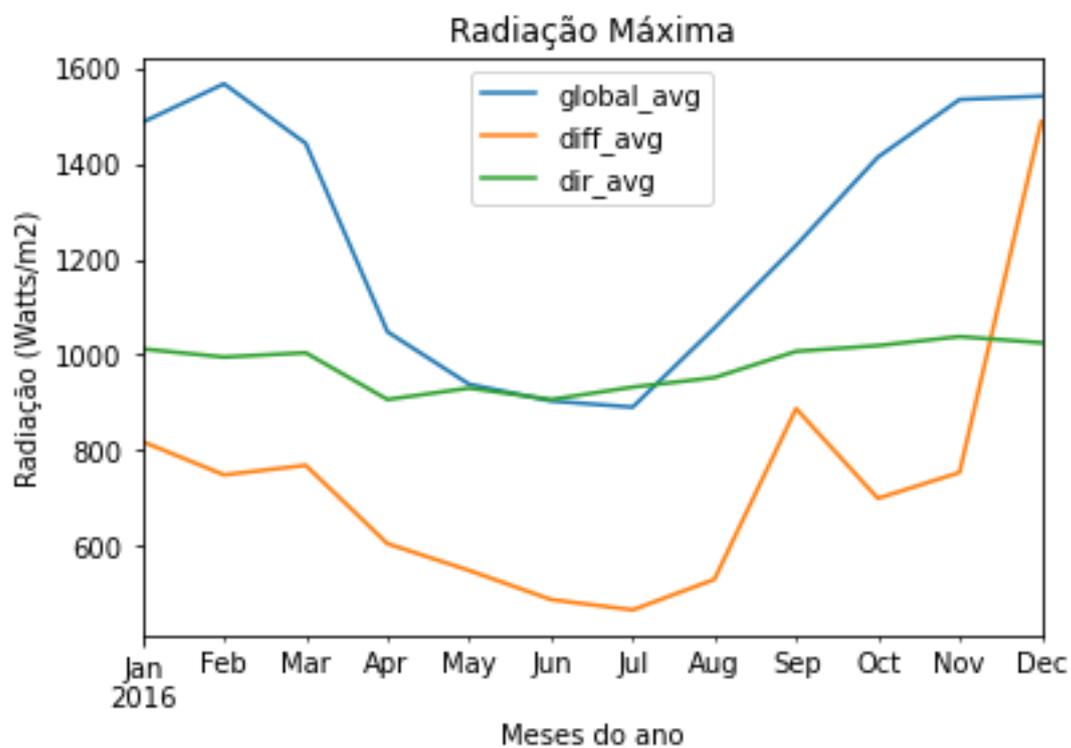
A base de dados da Rede SONDA será utilizada no treinamento da RNA para duas localidades com climatologias distintas: Petrolina (PE) e São Martinho da Serra (RS). O principal motivo da escolha dessas duas localidades é devido as particularidades climáticas que apresentam.

Petrolina se localiza na região Nordeste do Brasil, região quente e seca, com elevada incidência de radiação solar na superfície. A precipitação é baixa ao longo de todo o ano assim como a variabilidade sazonal da irradiação solar na superfície. São Martinho da Serra está localizada no Sul do país, bem na

região central do Rio Grande do Sul. Nesta região, há uma variabilidade sazonal bastante elevada na incidência de radiação solar na superfície conforme indicado no Atlas Brasileiro de Energia Solar (PEREIRA et al., 2017).

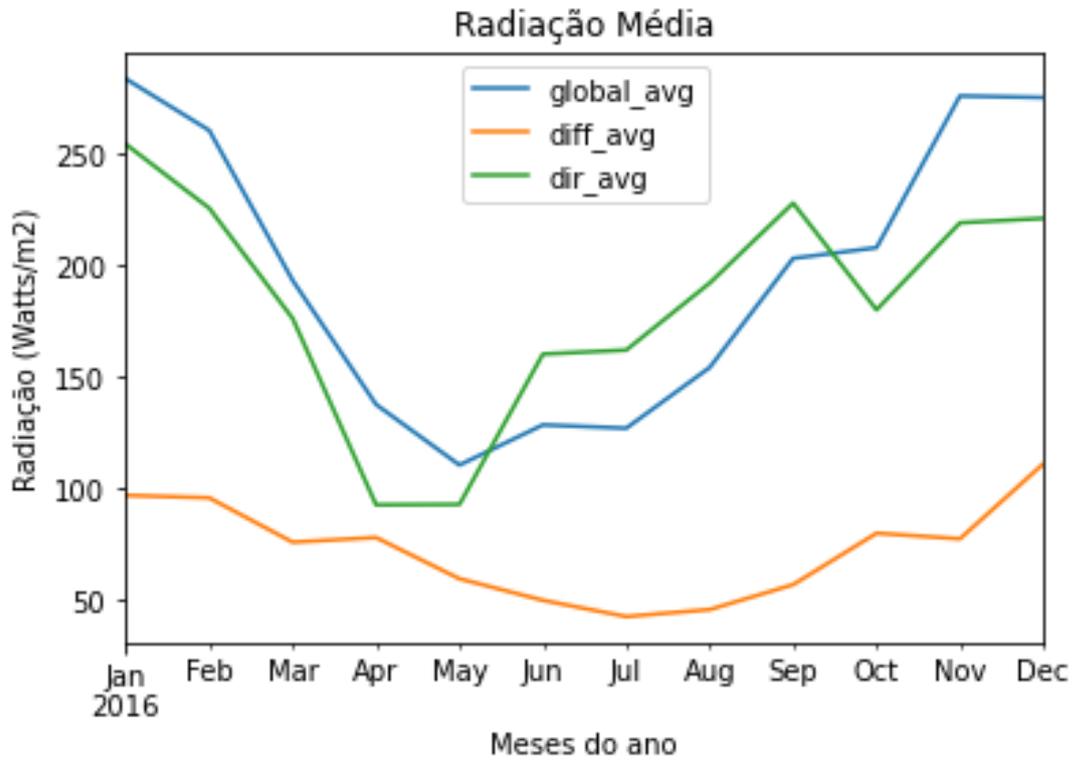
## 5 RESULTADOS

Figura 5 – Radiação Máxima Mensal.



Fonte: Produção da autora.

Figura 6 – Radiação Média Mensal.



Fonte: Produção da autora.

## **6 CONCLUSÃO**

A partir da revisão bibliográfica pude compreender melhor os conceitos envolvidos na geração de energia a partir de fontes renováveis, como a energia solar e sobre os processos de incidência de radiação solar na superfície, além da interação da radiação com a atmosfera.

Na literatura, o aprendizado de máquina para geração de séries temporais de radiação solar sintéticas se mostrou eficiente e capaz de fazer simulações e previsões dos dados de radiação solar. Além disso, ficou evidente a importância desses dados gerados sinteticamente para a modelagem do desempenho de uma planta fotovoltaica.

## REFERÊNCIAS BIBLIOGRÁFICAS

DONGARE, A. D. et al. Introduction to artificial neural network. **International Journal of Engineering and Innovative Technology (IJEIT)**, v. 2, n. 1, p. 189-194, 2012.

FIORIN, Daniel V. et al. Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. **Revista Brasileira de Ensino de Física**, v. 33, n. 1, p. 1309-2-1309-20, mar.2011. Disponível em: <https://www.scielo.br/j/rbef/i/2011.v33n1/>. Acesso em: 6 out. 2021.

GERBER, Florian et al. Predicting Missing Values in Spatio-Temporal Remote Sensing Data. **Ieee Transactions On Geoscience And Remote Sensing**. p. 2841-2853. Maio 2018. Disponível em: 10.1109/TGRS.2017.2785240. Acesso em: 22 jun. 2022.

GUARNIERI, Ricardo André. **Emprego de redes neurais artificiais e regressão linear múltipla no refinamento das previsões de radiação solar do modelo ETA**. 2006. 171 f. Dissertação (Mestrado em Meteorologia) - INPE, São José dos Campos, 2006. Disponível em: <http://mtc-m16b.sid.inpe.br/col/cptec.inpe.br/walmeida/2003/04.25.17.13/doc/mirrorget.cgi?languagebutton=pt-BR&metadatarpository=sid.inpe.br/MTC-m13@80/2006/05.30.18.50.03&index=0&choice=full>. Acesso em: 18 out. 2021.

KARIM, Samsul Ariffin Abdul; SINGH, Balbir Singh Mahinder. Global Solar Radiation Modeling Using Polynomial Fitting. **Applied Mathematical Sciences**. p. 367-378. 2014. Disponível em: [www.m-hikari.com](http://www.m-hikari.com) <http://dx.doi.org/10.12988/ams.2014.39516>. Acesso em: 21 jun. 2022.

MARTINS, F. R.; PEREIRA, Enio Bueno. Estudo comparativo da confiabilidade de estimativas de irradiação solar para o sudeste brasileiro obtidas a partir de dados de satélite e por interpolação/extrapolação de dados de superfície. **Revista Brasileira de Geofísica (Impresso)**, v. 29, p. 265-276, 2011.

MARTINS, F. R. e PEREIRA, E. B. **Energia Solar**. Estimativa e Previsão de Potencial Solar. Curitiba: Editora Apris, 2019. Disponível em: <http://www.ccst.inpe.br/publicacao/energia-solar-estimativa-e-previsão-de-potencial-solar/>. Acesso em: 15 mai. 2022.

MORAVČIK, Matúš. **Gap-filling Meteorology data using Neural Networks**. 2016. 60 f. Tese (Doutorado) - Fakulta Informatiky, Masarykova Univerzita, Brno, 2016. Disponível em: <https://is.muni.cz/th/d09hs/>. Acesso em: 20 jun. 2022.

PEREIRA, E. B.; MARTINS, F. R.; GONÇALVES, A. R.; COSTA, R. S.; LIMA, F. L.; RÜTHER, R.; ABREU, S. L.; TIEPOLO, G. M.; PEREIRA, S. V.; SOUZA, J. G. **Atlas brasileiro de energia solar**. 2.ed. São José dos Campos: INPE, 2017. 80p. Disponível em: <http://doi.org/10.34024/978851700089>. Acesso em: 10 dez. 2021.

POLO, J. et al. A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. **Solar Energy**, v. 85, n. 5, p. 1164-1170, mai.2011. Disponível em: <https://doi.org/10.1016/j.solener.2011.03.011>. Acesso em: 6 out. 2021.

SONDA. **SISTEMA DE ORGANIZAÇÃO NACIONAL DE DADOS AMBIENTAIS**. Disponível em: <http://sonda.ccst.inpe.br>. Acesso em: 15 out. 2021.

SCHWANDT, Marko et al. Development and test of gap filling procedures for solar radiation data of the Indian SRRA measurement network. **Energy Procedia**, v. 57, p. 1100-1109, 2014. Disponível em: <https://doi.org/10.1016/j.egypro.2014.10.096>. Acesso em: 6 dez. 2021.

SPOKAS, Kurt; FORCELLA, Frank. Estimating hourly incoming solar radiation from limited meteorological data. **Weed Science**. S.L., p. 182-189. 2006. Disponível em: <https://doi.org/10.1614/WS-05-098R.1>. Acesso em: 20 jun. 2022.

SRIKRISHNAN, V. et al. Using multi-pyranometer arrays and neural networks to estimate direct normal irradiance. **Solar Energy**, v. 119, p. 531–542, 2015.

```

# -- coding: utf-8 --
"""
@author: Mariane Souza Guarachi
"""
#global_avg -> Radiação Global Horizontal
#dir_avg -> Radiação Direta Normal
#diff_avg -> Radiação Difusa

import calendar
import datetime
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt

def JulianDate_to_date(y, jd):
    month = 1
    while (jd - calendar.monthrange(y,month)[1] > 0) & (month <= 12):
        jd = jd - calendar.monthrange(y,month)[1]
        month += 1
    date = datetime.date(y,month,jd).strftime("%d-%m-%Y")
    return date

#Importando os dados
path = r'C:\Users\maria\Downloads\codigomor\codigomor\SMS_SD_2016\SMS16'
month = ['01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12']

#Necessário ajustar a lista de datas quando mudar ano ou período de meses
date_list = pd.date_range(start='2016-01-01 00:00', end='2017-01-01 00:00', freq
= 'min', closed= 'left')
df_SMS = pd.DataFrame()

for j in month:
    filename = path + j + 'ED_prequalificado.csv'
#Print do nome do arquivo
    print (filename)
    df_temp = pd.read_csv (filename , delimiter=',', names =
        ['id','year', 'day', 'minuto', 'global_avg', 'global_std',
'global_max', 'global_min',
        'diff_avg', 'diff_std', 'diff_max', 'diff_min', 'par_avg',
'par_std', 'par_max', 'par_min',
        'lux_avg', 'lux_std', 'lux_max', 'lux_min', 'tp_sfc',
'humid', 'press', 'rain', 'ws_10m', 'wd_10m', 'wd_10m_std',
        'CosAngZen', 'dir_avg', 'dir_std', 'dir_max', 'dir_min',
'lw_avg', 'lw_std', 'lw_max', 'lw_min',
        'Temp_glo', 'Temp_dir', 'Temp_diff',
'Temp_dome', 'Temp_case'])
    #Print (df_temp.shape)
    df_SMS = df_SMS.append(df_temp, ignore_index=True)

#Print da quantidade de linhas e colunas
print (df_SMS.shape)
df_SMS.index = date_list

#Substituindo valores negativo por zero

```

```

df_SMS.loc[df_SMS.global_avg<0,'global_avg']=0
df_SMS.loc[df_SMS.diff_avg<0,'diff_avg']=0
df_SMS.loc[df_SMS.dir_avg<0,'dir_avg']=0

#Excluindo as colunas que não serão utilizadas
df_SMS = df_SMS.drop(['id','year', 'day', 'minuto', 'global_std', 'global_max',
'global_min','diff_std', 'diff_max', 'diff_min',
'par_avg', 'par_std', 'par_max', 'par_min','lux_avg', 'lux_std',
'lux_max', 'lux_min', 'wd_10m_std','CosAngZen',
'dir_std', 'dir_max', 'dir_min', 'lw_avg', 'lw_std', 'lw_max',
'lw_min','Temp_glo', 'Temp_dir', 'Temp_diff',
'Temp_dome','Temp_case'], axis=1)
print(df_SMS)

df_SMS2 = df_SMS.loc[df_SMS.global_avg != 3333]
df_SMS2 = df_SMS2.loc[df_SMS.diff_avg != 3333]
df_SMS2 = df_SMS2.loc[df_SMS.dir_avg != 3333]

#df_SMS3 = df_SMS2.loc[df_SMS.index >= '01-01-2016' & <= '31-11-2016']

#Plotando média mensal
df_SMS_monthly = df_SMS2.resample('M').mean()
df_SMS_monthly['global_avg'].plot(legend=True,title="Radiação
Média",ylabel="Radiação (Watts/m2)",xlabel="Meses do ano")
df_SMS_monthly['diff_avg'].plot(legend=True)
df_SMS_monthly['dir_avg'].plot(legend=True)
plt.show()

plt.savefig('C:/Users/maria/Downloads/Plot Graficos/media_mensal.png')

#Função para salvar a figura como imagem (ver qual é)

#plotando valor máximo mensal
df_SMS_maximum = df_SMS2.resample('M').max()
df_SMS_maximum['global_avg'].plot(legend=True,title="Radiação
Máxima",ylabel="Radiação (Watts/m2)",xlabel="Meses do ano")
df_SMS_maximum['diff_avg'].plot(legend=True)
df_SMS_maximum['dir_avg'].plot(legend=True)
plt.show()

```