ENG1467/ELE2742 – Otimização

Alexandre Street e Pedro Ferraz

Data de entrega: 31 de março de 2022

Previsão de série de geração eólica

A geração de energia eólica tem crescido expressivamente nas últimas décadas, trazendo diversos benefícios ao meio ambiente quando comparado a suas alternativas. Uma tarefa crucial para a operação do Sistema Interligado Nacional é a previsão da geração de energia a partir de usinas eólicas.

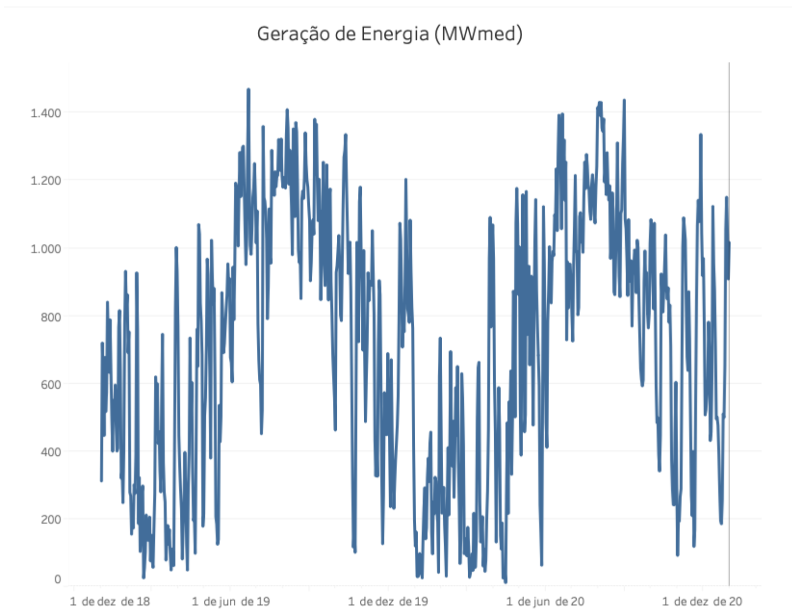

Figure 1: Série diária de geração eólica agregada no estado do Piauí

Metodologia e exercícios

Questão 1 - Análise exploratória e modelagem

Em anexo, o arquivo “eolica.csv” contém a série de geração eólica horária na região NE, obtido do site da ONS. Utilizando os dados desse arquivo, realizaremos a análise exploratória da série e estimaremos modelos para explicar e prever a série.

Para prever a série temporal y_t, devemos investigar como a série se comporta no tempo e tentar identificar padrões e comportamentos. Uma forma de verificar se há uma relação entre instantes de tempo diferentes é realizar um gráfico de dispersão. Ao plotar o gráfico de dispersão de y_t em função de y_{t-k}, podemos averiguar se o a série em y_{t-k} ajuda a explicar e/ou prever a série em y_t.

Item i) Faça um gráfico de dispersão de y_t em função de y_{t-1}. Em seguida, faça também gráficos de dispersão de y_t em função de y_{t-k} para diferentes valores de k. Indique quais termos apresentam alta correlação e podem potencialmente ser usados para prever a série.

Utilizando os termos relevantes encontrados no **item i)**, podemos desenvolver um modelo para prever y_t. Chamaremos de $\hat{y}_t$ a estimativa dada pelo modelo para a série no instante t.

O modelo que utiliza apenas a informação sobre a própria série em instantes anteriores para prever é chamado de modelo auto-regressivo. Em particular, se o modelo utiliza apenas o instante de tempo anterior y_{t-1} para prever y_t é chamado de modelo auto-regressivo de ordem 1, AR(1). A estimativa desse modelo é dada por

$$\hat{y}_t(\beta) = \beta_0 + \beta_1 y_{t-1}$$

Podemos, ainda, acrescentar outros termos que acharmos pertinentes. Se a série em questão é uma série horária, podemos tentar utilizar a mesma hora do dia anterior para prever a próxima hora, por exemplo. Assim, podemos utilizar o modelo

$$\hat{y}_t(\beta) = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-24}$$

Uma vez que temos formulado o modelo $\hat{y}$, podemos estimar os coeficientes β através de regressão quantílica, cujo problema de otimização está mostrado a seguir:

$$\min_{\beta} \sum_{t=1}^{T} |y_t - \hat{y}_t(\beta)|$$

Item ii) Utilizando os termos relevantes identificados no **item i)**, estime diferentes modelos auto-regressivos através de regressão quantílica.

Além dos termos auto-regressivos, podemos também levar em conta efeitos sazonais, como por exemplo o efeito das estações do ano. Para investigar se a série possui efeitos sazonais, podemos realizar agregações da série. Se quisermos investigar se a série tem sazonalidade anual (isto é, se o mês do ano passado ajuda a prever o mesmo mês deste ano), podemos calcular a média da série ao longo de cada um dos meses e visualizá-la mês a mês, o que facilitará encontrar padrões.

Item iii) Realize agregações da série de diferentes formas (exemplo: dia, semana, mês, ano), faça gráficos da série em cada agregação e verifique que tipos de sazonalidade estão presentes.

Com base nas componentes sazonais encontradas no **item iii)**, podemos levar em conta efeitos sazonais diários, semanais, mensais e/ou anuais. Para isso, podemos introduzir termos trigonométricos, que são adequados para modelar fenômenos sazonais devido a suas propriedades cíclicas e pela capacidade de aproximar funções periódicas. Para incluir um efeito sazonal que cujo ciclo é de s observações, devemos incluir os seguintes termos em nosso modelo:

$$\hat{y}_t(\phi, \psi) = \sum_{k=1}^{M} \{\phi_k \cos(\omega_k t) + \psi_k \sin(\omega_k t)\}$$

onde $M \leq \frac{s}{2}$ é o número de termos a serem utilizados e

$$\omega_k = \frac{2\pi}{s} k$$

Assim, para considerar o efeito de sazonalidade diária em uma série horária, deve-se utilizar $s = 24$, já que cada 24 observações corresponderão a 1 ciclo, por exemplo. Já para considerar o efeito de sazonalidade anual, deve-se utilizar $s = 24 \cdot 365 = 8760$. Como exemplo, se escolhermos representar sazonalidade diária e anual cada um com $M = 1$, teremos o modelo

$$\hat{y}_t(\phi, \psi) = \phi_1 \cos\left(\frac{2\pi}{24} t\right) + \psi_1 \sin\left(\frac{2\pi}{24} t\right) + \phi_2 \cos\left(\frac{2\pi}{8760} t\right) + \psi_2 \sin\left(\frac{2\pi}{8760} t\right)$$

Os coeficientes ϕ e ψ de tais modelos podem também ser estimados via regressão quantílica.

Item iv) A partir da análise de sazonalidade realizada no **item iii)**, adote um modelo de sazonalidade (simples ou múltipla) utilizando a abordagem descrita acima e estime-o via regressão quantílica. Discuta o impacto de diferentes números de termos M.

Podemos também combinar os termos auto-regressivos e os termos sazonais para formular um modelo que leva em conta esses diferentes fatores. O modelo AR(1) incluindo termos de sazonalidade, por exemplo, pode ser expresso como:

$$\hat{y}_t(\beta, \phi, \psi) = \beta_0 + \beta_1 y_{t-1} + \sum_{k=1}^{M} \{\phi_k \cos(\omega_k t) + \psi_k \sin(\omega_k t)\}$$

Questão 2 - Avaliação dos modelos

Para avaliar o desempenho do modelo, devemos verificar a capacidade do modelo de estimar corretamente valores anteriores da série e também a capacidade de prever corretamente novos dados nunca antes vistos para o modelo. As previsões feitas para instantes de tempo que já foram utilizados para estimar o modelo são chamadas de previsões *in-sample*, enquanto previsões feitas para instantes de tempo além dos utilizados são chamadas de previsões *out-of-sample*.

Para realizar os itens a seguir, separe uma parte da série para ser usada para a estimação dos modelos (que será o conjunto *in-sample*) e uma outra parte para testar os modelos em observações não vistas (que será o conjunto *out-of-sample*).

Uma forma de fazer isso é escolher um instante T tal que o conjunto *in-sample* será de y_1 até y_T, e o conjunto *out-of-sample* será de y_{T+1} até o fim da série.

Para avaliar a capacidade do modelo de se adequar aos dados e prever corretamente instantes passados, podemos utilizar a métrica R^2 no conjunto de estimação (*in-sample*):

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

onde y_t é o valor da série no instante t, $\hat{y}_t$ é o valor estimado no instante t utilizando toda informação disponível até $t-1$, e $\bar{y}$ é a média da série no conjunto de estimação.

Além disso, para avaliar o desempenho do modelo em prever dados no conjunto de teste (*out-of-sample*), podemos utilizar a métrica MAE:

$$MAE = \frac{1}{K}\sum_{t=T+1}^{T+K}|y_t - \hat{y}_t|$$

onde K é o tamanho do conjunto de avaliação/teste, e $\hat{y}_t$ é o valor previsto para o instante de tempo t utilizando toda a informação disponível até $t-1$.

Item i) Separe o conjunto de dados em conjuntos *in-sample* e *out-of-sample*.

Item ii) Estime os modelos propostos na **questão 1** no conjunto *in-sample* e verifique o valor de R^2 para cada modelo. Compare os valores de R^2 de modelos diferentes e discuta o significado.

Item iii) Para cada modelo estimado no **item ii)**, realize a previsão 24 horas à frente e calcule a métrica de erro MAE. Faça também os gráficos da série e da previsão. Avalie a qualidade das previsões qualitativamente e quantitativamente.

Item iv) Com base nos itens anteriores, discuta qual o melhor modelo para prever 24 horas à frente.