

6.

Estratégia de Especificação Metodológica

No presente Capítulo é apresentada uma proposta de especificação não paramétrica para a obtenção de previsões probabilísticas da geração de energia eólica de curto prazo. Esta metodologia se fundamenta em técnicas não paramétricas tais como: Análise Espectral Singular (SSA), Agrupamento Hierárquico Aglomerativo e a Densidade Condicional por *Kernel* calculada através do estimador de *Nadaya-Watson*.

A maioria das abordagens para a previsão da velocidade do vento limita-se ao uso de modelos estatísticos ou de Inteligência Computacional como foi explicitado em Ramirez-Rosado et al. (2009) e Methaprayoon et al. (2007) (para mais detalhes ver a Seção 3.4). Estes enfoques são caracterizados por utilizar modelos paramétricos para prever a velocidade do vento, embora outras metodologias como Redes Neurais, Lógica Fuzzy ou SSA possam ser usadas. Apesar disso, SSA é a única metodologia de interesse para o desenvolvimento desta tese devido à limitada quantidade de trabalhos realizados para previsão de séries de energia eólica ou da velocidade do vento.

SSA é uma técnica não paramétrica para a previsão de séries temporais. Embora ela tenha sido usada para a previsão nesta área, até agora sua aplicação se restringe a séries cuja frequência é mensal ou diária. Por conseguinte, SSA torna-se uma proposta alternativa interessante que abrange a previsão de séries que possuem uma frequência horária (no contexto da geração eólica é conhecida como alta frequência), sem a necessidade de uma especificação funcional do modelo estocástico que governe a série. Além disso, devido a que os registros da velocidade do vento costumam possuir valores atípicos, conhecidos como outliers, SSA conta com a vantagem de que a metodologia faz a filtragem destes valores.

Por sua vez, o estimador de *Nadaraya-Watson* possui um papel fundamental no processo de previsão da geração de energia eólica, já que ele permite estimar a densidade condicional da geração eólica, isto é, a estimação estocástica da curva de potência, a qual é fundamental para a conversão da velocidade do vento em potência gerada. Por conseguinte, é densidade condicional quem faz a ponte entre a previsão da velocidade do vento via SSA e a previsão probabilística da produção eólica. A estratégia metodológica geral de previsões probabilísticas da geração eólica que este trabalho seguirá é uma adaptação do enfoque em dois estágios para previsões pontuais da geração de energia eólica (ver Seção 3.1), cuja estrutura é a seguinte:

1. **Primeiro estágio:** utilizar *SSA Sequencial* para efetuar as previsões horárias da velocidade do vento h passos à frente.
2. **Segundo estágio:** com base nos dados históricos da velocidade do vento e da potência gerada, estimar de forma estocástica a curva de potência, ou seja, a densidade condicional $P(\text{potência}|\text{velocidade})$ por médio do estimador *kernel* de *Nadaya-Watson*, que não é mais do que a densidade de probabilidade da potência condicionada a um determinado valor da velocidade do vento.
3. **Terceiro estágio:** usar os valores previstos da velocidade do vento em combinação com a curva de potência estocástica para gerar a previsão da energia eólica produzida.

A Figura 6.1 ilustra os três estágios que devem ser executados para obter as previsões probabilísticas da geração eólica. No entanto, nas próximas seções cada um destes estágios será apresentado detalhadamente indicando o processo de especificação para a obtenção tanto das previsões da velocidade quanto a estimação da densidade condicional via *kernel*.

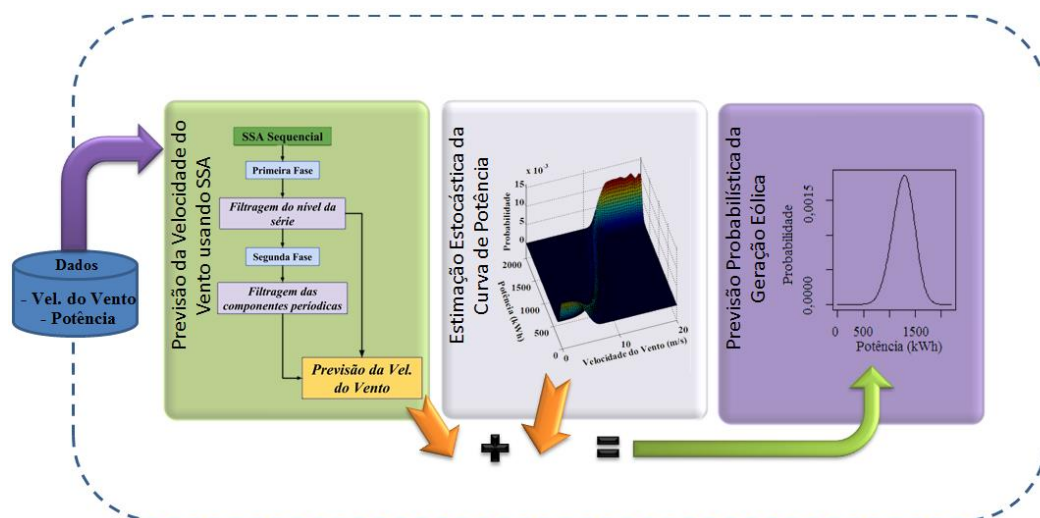


Figura 6.1 – Diagrama da metodologia proposta. (Fonte: Elaboração do Autor)

6.1.

Primeiro Estágio: Previsão Horária da Velocidade do Vento

A previsão da série temporal da velocidade do vento constitui o primeiro estágio do arcabouço metodológico. Para este fim, emprega-se *SSA sequencial*, a qual está

composta de duas fases. Uma primeira fase, onde o nível da série de velocidade do vento é extraído e; em seguida, na segunda fase, a partir dos residuais obtidos na primeira fase, as componentes periódicas são detectadas e extraídas. A diferença entre a metodologia SSA para *previsão* e *SSA sequencial* está baseada em que o método básico de SSA (unicamente compreende dois estágios: *decomposição* e *reconstrução*) é executado duas vezes, uma em cada fase do método *sequencial* e, na segunda fase além de ser aplicado o método básico é acrescentado um terceiro estágio que é o estágio de *previsão*. A Figura 6.2 exibe com maior clareza o processo de *SSA sequencial*.

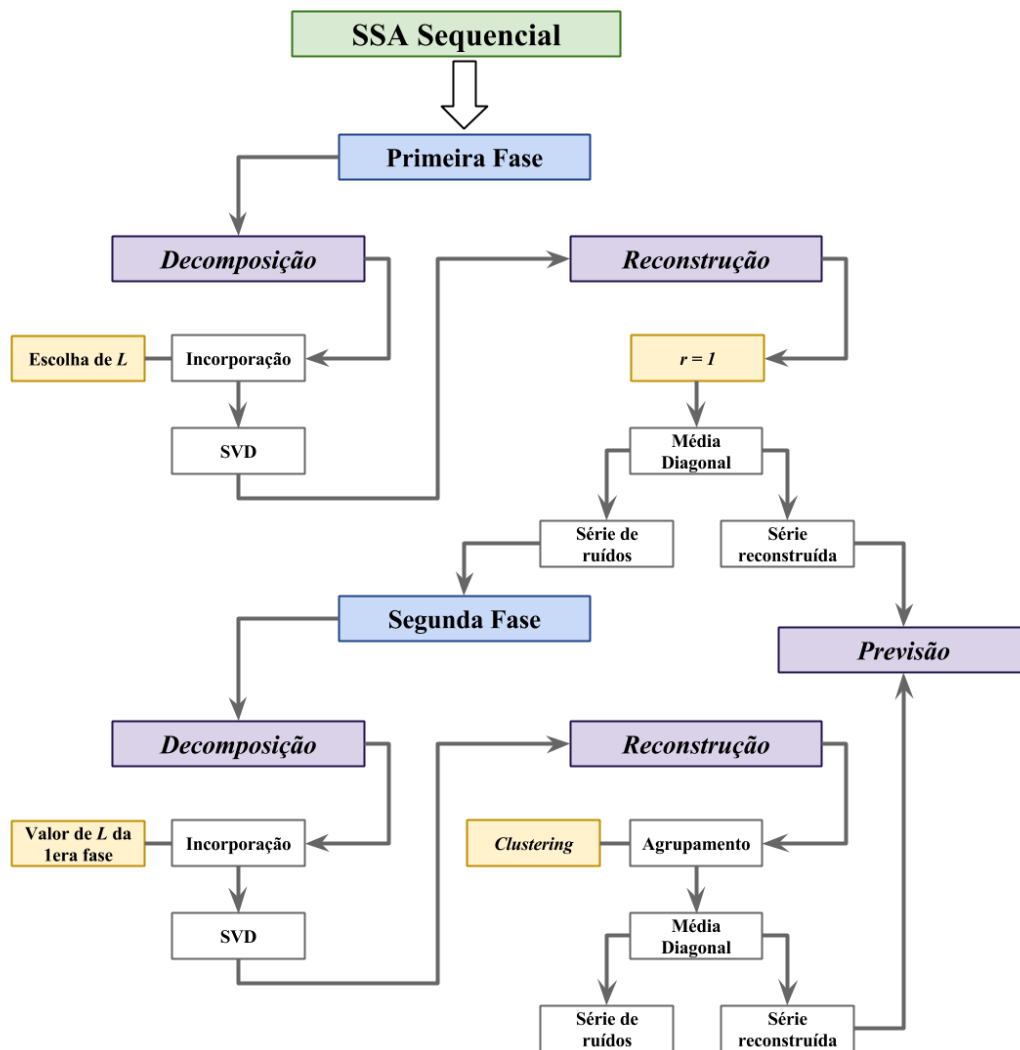


Figura 6.2 – Diagrama do processo de *SSA Sequencial*. (Fonte: Elaboração do Autor)

Os dados da velocidade do vento possuem uma frequência horária, esta série temporal de alta frequência contém um grande volume de dados para serem analisados e está caracterizada por ser altamente sazonal e por apresentar uma estrutura muito

complexa em suas componentes periódicas; demandando conseqüentemente, uma abordagem que as permita identificar e extrair. Por esta razão, usar SSA *sequencial* facilita o processo de identificação dos diferentes tipos de componentes, toda vez que em cada sequencia é executado SSA para a extração de uma tipologia específica de componente.

A determinação do parâmetro L de defasagem é um elemento fundamental em SSA e, portanto, constitui o ponto de partida da metodologia. Segundo a literatura técnica de SSA, o valor de L deve ser um suficientemente grande, mas não superior a $T/2$. Assim que a escolha de L será efetuada na primeira fase de SSA *sequencial* para previsão de séries temporais, levando em consideração o problema e as indicações da escolha do valor de L cujo valor máximo pode ser igual a $T/2$ caso T seja par ou de L igual à $(T + 1)/2$ caso T seja impar.

Na primeira fase de SSA *sequencial* o nível da série é extraído. Para inicializar o método é preciso determinar o comprimento da janela L . Lembrando que o tamanho de L será determinado pelo problema, é importante levar em consideração que a série da velocidade do vento está caracterizada por possuir um padrão sazonal diário, válido para todas as épocas do ano e, dependendo da região, podem existir padrões semanais, mensais ou trimestrais (Yu et al., 2013). Assim que, sabendo que a série temporal possui uma componente periódica diária (24 horas), com o propósito de garantir uma melhor separabilidade das componentes, será escolhido L como sendo proporcional a este período.

Uma vez obtido o comprimento de L , prossegue-se com a implementação dos dois estágios de *decomposição* e *reconstrução* de SSA na série temporal da velocidade do vento. O estágio de *decomposição* é aplicado diretamente, sem restrição alguma. No entanto, o estágio de *reconstrução* contém a fase de *agrupamento* (Cassiano, 2014) para a qual é necessário determinar o número r de autotriplas a serem agrupadas. Para a série da velocidade do vento, isto é muito mais simples, toda vez que a primeira componente da SVD corresponde ao nível da série, o que conseqüentemente faz com que $r = 1$, e não seja preciso usar nenhuma ferramenta de agrupamento para a determinação deste valor.

Posteriormente, é aplicada a fase da *média diagonal* para obter a **série da velocidade do vento reconstruída**. Os ruídos resultantes são guardados para serem usados como sendo a série temporal da segunda fase de SSA *sequencial*, pois até agora unicamente foi descrito o procedimento para a obtenção da componente interpretável do nível. Nos resíduos encontra-se informação das componentes periódicas da série temporal sem ser extraída, a qual será efetuada na segunda fase de SSA *sequencial*.

Na segunda fase de SSA *sequencial* é realizada a extração das componentes periódicas da série da velocidade do vento, mantendo o mesmo tamanho de L que foi

obtido na primeira fase e usando a série dos resíduos da primeira fase de SSA *sequencial*, como a nova série temporal sobre qual será aplicado SSA. A diferença da primeira fase onde o valor de r sempre toma o valor de 1, nesta segunda fase é preciso calcular o tamanho de r . Por conseguinte, será necessário fazer uso da metodologia Hierárquica Aglomerativa para o agrupamento das autotriplas, pois tanto a matriz de correlações ponderadas quanto à análise dos valores singulares e a análise gráfica dos vetores singulares, apresentam limitações na visualização e identificação das componentes periódicas a serem agrupadas, isto, como consequência da grande quantidade de componentes que constituem a matriz de trajetória.

Para o processo de agrupamento será utilizada a metodologia de clusterização Hierárquica Aglomerativa, a qual usa a matriz de *correlações ponderadas* que foi obtida na segunda fase de SSA *sequencial* e, representa um conjunto de medidas de similaridade entre cada par de indivíduos. Propõe-se usar esta matriz como medida de similaridade no processo de agrupamento, já que a o cálculo da matriz de *correlações ponderadas* envolve o termo de ponderação ω_k (Equação 4.17), o qual ao assumir o valor de 1 ($\omega_k = 1 \forall k$) faz que a matriz de *correlações ponderadas* seja idêntica à medida de similaridade denominada “*Coefficiente de separação angular*” (Ver Seção 10.1.5 para um maior detalhamento).

Com a matriz de *correlações ponderadas* inicializa-se o processo de agrupamento, e para isto, é transformada cada entrada da matriz, em uma medida de dissimilaridade através da operação:

$$d_{ij} = 1 - s_{ij} \quad (6.1)$$

com s_{ij} elemento genérico da matriz de similaridades. Posteriormente, para realizar o agrupamento dos indivíduos, é escolhido o critério que medi a dissimilaridade entre os grupos, o qual usualmente é conhecido como método de agregação de classes. Assim que, pelos seus bons resultados e simplicidade de cálculo, o método selecionado é o método do vizinho mais próximo (ou *single-link*).

Seguidamente, constrói-se o dendrograma com base no método aglomerativo que foi exposto na Seção 10.1.3.1. Uma vez é obtido o dendrograma, o agrupamento das autotriplas é realizado fazendo cortes horizontais na figura o mais perto possível da raiz. Embora a metodologia hierárquica permita fazer cortes em diferentes níveis do dendrograma de acordo com o número de *clusters* desejados, o procedimento proposto nesta tese sugere que estes cortes sejam realizados o mais próximo possível da parte superior da raiz, para poder obter um melhor agrupamento das autotriplas.

Tendo efetuado o agrupamento das autotriplas, a série da velocidade do vento é reconstruída, dando por finalizado o segundo estágio (*reconstrução*) de SSA, dando passo ao início do terceiro e último estágio: a previsão da série temporal. Neste terceiro estágio é executado o algoritmo de previsão recorrente para gerar os valores previstos h passo à frente, descrito na Seção 4.1.3.

6.2.

Segundo Estágio: Estimação Estocástica da Curva de Potência

O cálculo da densidade condicional através do estimador *kernel* constitui o segundo estágio deste arcabouço metodológico. Esta densidade faz uso da técnica não paramétrica denominada: *Estimador de Nadaraya-Watson*, que basicamente é um estimador *kernel* para calcular a densidade de probabilidade de um par de variáveis aleatórias quaisquer, como foi apresentado na Seção 5.5, mas que para o caso específico desta tese, é a densidade da produção eólica condicionada a um determinado valor da velocidade do vento. Assim, para as variáveis aleatórias da potência (p) e da velocidade do vento (v) deseja-se obter:

$$f_P(p|v) = \frac{f_{VP}(v,p)}{f_V(v)}. \quad (6.2)$$

O procedimento para o cálculo estocástico da curva de potência, ou seja, da estimação da densidade condicional da potência, a qual consiste em calcular a densidade da variável aleatória p (potência), condicionada à variável aleatória $V = v$ (Velocidade do vento), que pode ser formulada como segue:

$$\hat{f}_P(p|V = v) = \frac{\hat{f}_{VP}(v,p)}{\hat{f}_V(v)}. \quad (6.3)$$

A estimação da Equação anterior, que representa a função de densidade de probabilidade condicional da geração de energia eólica, foi obtida através do estimador *Nadaya-Watson* (Equações 5.34 ou 5.35 e 5.36):

$$\hat{f}_P(p|V = v) = \frac{\frac{1}{n} \sum_{j=1}^n K_{h_v}(v - V_j) K_{h_p}(p - P_j)}{\frac{1}{n} \sum_{j=1}^n K_{h_v}(v - V_j)}. \quad (6.4)$$

Rescrevendo os termos tem-se:

$$\hat{f}_P(p|v) = \sum_{j=1}^n \mathcal{W}_j(v) K_{h_p}(p - P_j), \quad (6.5)$$

sendo

$$\mathcal{W}_j(v) = \frac{K_{h_v}(v - V_j)}{\sum_{i=1}^n K_{h_v}(v - V_i)}. \quad (6.6)$$

Onde K_{h_v} e K_{h_p} são respectivamente, as funções *kernel* da velocidade do vento e da potência. Apesar de que a escolha do *kernel* pode afetar a estimativa da densidade, estudos sugerem que tal efeito é bastante pequeno. Por conseguinte, será escolhido o *kernel* Gaussiano, pois a construção do estimador apresenta no numerador (Equação 6.4) o produto de duas funções *kernel*, que como é sabido, o produto de dois *kernels* gaussianos univariados produz os mesmos resultados que o uso do *kernel* Gaussiano multivariado.

Por outra parte, para o cálculo da largura de h_v e h_p , será usada à abordagem de validação cruzada para a estimação dos parâmetros de suavização, pois apesar da existência de dois enfoques para sua estimação (um baseado em regras e outro orientado à dados), não existe uma regra geral para o tamanho ótimo do comprimento; no entanto, a validação cruzada proporciona tamanhos de largura dos parâmetros de suavização mais apropriados, especialmente quando várias variáveis explicativas estão envolvidas.

Para a construção estocástica da curva de potência, será necessária a repetição do estimador de *Nadaraya-Watson* dado pelas Equações 6.5 e 6.6. Desta forma, os valores da potência (p) podem variar no intervalo de zero até a capacidade do parque eólico, ou como nesta pesquisa, p pode variar com a capacidade da turbina. As variações de p são obtidas através de incrementos pequenos até atingir a capacidade estipulada. O procedimento para o cálculo da densidade condicional está baseada nos passos seguintes:

1. Criar uma grade com os valores da potência no intervalo que vai de zero até a capacidade do parque eólico/turbina, constituindo os valores da abcissa. De forma similar, constrói-se uma grade para a ordenada com os valores da velocidade, no intervalo de zero até uns metros acima da velocidade máxima registrada na base de dados utilizada. Neste trabalho foi escolhida a velocidade de 20 m/s como sendo a velocidade máxima possível, pois dentro da base de dados a máxima velocidade registrada é um valor muito próximo

ao selecionado. Cabe destacar que a discretização tanto da potência quanto da velocidade do vento deve ser realizada com incrementos regulares.

2. Estimar a densidade condicional da geração eólica através do estimador de *Nadaraya-Watson* dado pelas Equações 6.5 e 6.6, em conjunto com as grades obtidas no passo anterior. Para produzir uma estimativa da densidade condicional da produção eólica, em cada valor discretizado, a potência deve estar condicionada a uma determinada velocidade do vento. Estas estimações são armazenadas para serem usadas no seguinte estágio do arcabouço metodológico.

Ao executar estes dois passos, obtém-se uma estimativa estocástica da curva de potência como a apresentada na Figura 6.3, a qual é fundamental para poder obter através das previsões da velocidade do vento as previsões de energia eólica produzida. Pode também se observar como esta Figura descreve a forma, comumente conhecida, da curva de potência real fornecida pelos fabricantes das turbinas.

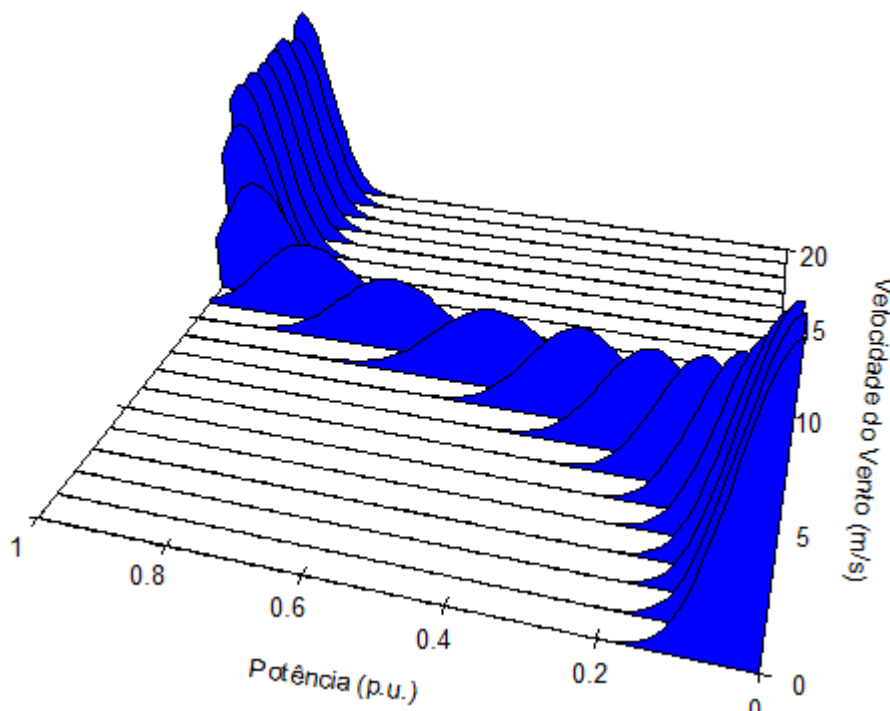


Figura 6.3 – Estimação da densidade probabilidade condicional da geração de energia eólica dada a velocidade do vento. (Fonte: Elaboração do autor)

6.3.

Terceiro Estágio: Previsão Probabilística da Geração Eólica

A previsão da densidade de probabilidade da geração de energia eólica consiste em determinar a *pdf* em um instante t (sendo t a origem da previsão), para $t + h$ passos à frente de um determinado horizonte de tempo:

$$f_P(p_{t+h}|V = v_{t+h|t}) = \frac{f_{VP}(v_{t+h|t}, p_{t+h})}{f_V(v_{t+h|t})}. \quad (6.7)$$

Sendo p_{t+h} a potência prevista h passos à frente, $v_{t+h|t}$ é a previsão da velocidade do vento em t para h passos à frente. f_{VP} é a função de densidade conjunta da velocidade do vento e a produção eólica e, f_V é densidade marginal prevista da velocidade do vento.

Com a estimativa da densidade de probabilidade condicional da geração eólica por *kernel* obtida no estágio anterior (Seção 6.2), em conjunto com as previsões da velocidade do vento geradas com *SSA sequencial* no primeiro estágio (Seção 6.1) deste arcabouço metodológico, é prevista a produção eólica probabilística. Assim por exemplo, supondo previsões da velocidade do vento com valores de 7 m/s e de 12 m/s, e utilizando a curva de potência estocástica estimada, obtém-se as respectivas previsões da densidade de probabilidade da produção de energia eólica condicionadas a estas velocidades do vento, as quais podem ser visualizadas nas Figuras 6.4 e 6.5.

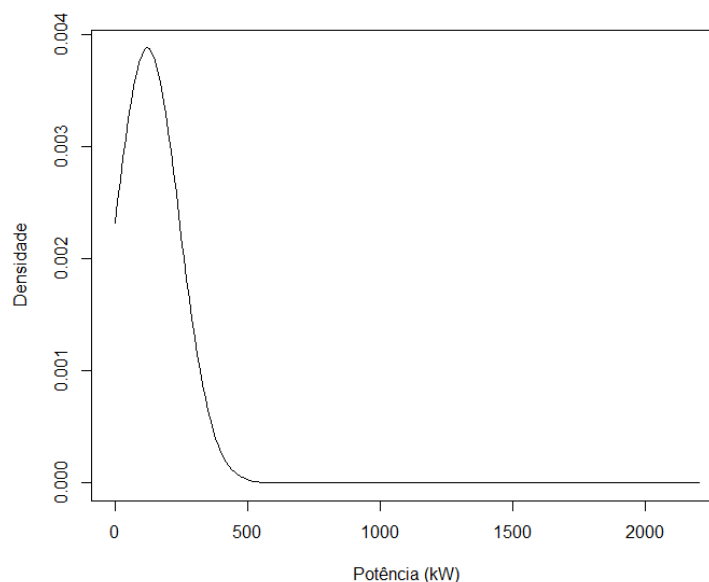


Figura 6.4 – Densidade de probabilidade para a geração de energia eólica dada uma velocidade do vento de 7 m/s. (Fonte: Elaboração do autor)

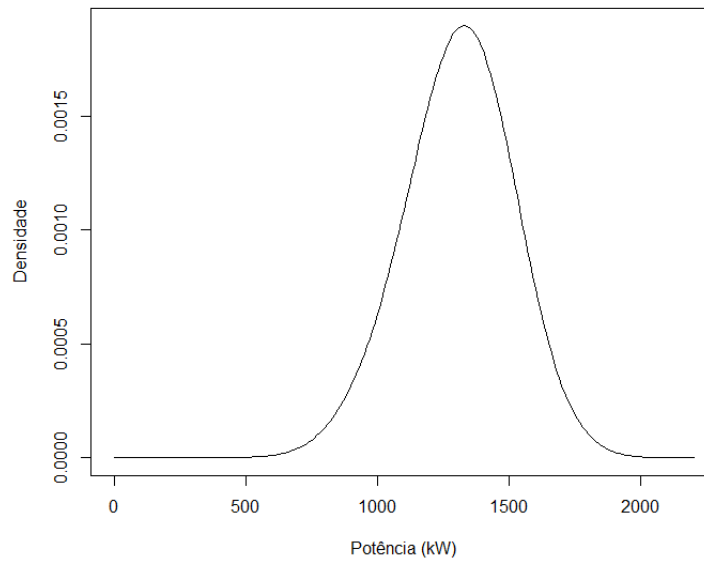


Figura 6.5 – Densidade de probabilidade para a geração de energia eólica dada uma velocidade do vento de 12 m/s. (Fonte: Elaboração do autor)

Nestas Figuras pode-se observar a dependência da potência com a velocidade. Por exemplo, ao comparar a $Prob(Potência > 500)$ é baixa na Figura 6.4, mas muito alta na Figura 6.5, isso mostra de forma muito clara a dependência (a condicionalidade) da velocidade do vento.

Este constitui, portanto, o arcabouço metodológico para a determinação das previsões probabilísticas da geração de energia eólica através do enfoque tradicional de dois estágios, fazendo uso de técnicas não paramétricas. Desta forma, no próximo capítulo serão apresentados resultados da implementação desta estratégia de especificação.