# Resultados

7.

Neste capítulo são apresentados os resultados da metodologia proposta relacionada à previsão probabilística de curto prazo da geração de energia eólica.

## 7.1. Descrição da Base de Dados

A base de dados utilizada no desenvolvimento deste trabalho pertence a um parque eólico brasileiro. O conjunto de dados contém registros horários para um período compreendido entre o dia 1 de janeiro de 2007 e o dia 7 de Janeiro de 2008; i.e., um total de 8.928 médias horárias da velocidade do vento (m/s) e da potência (kWh) associadas a uma turbina eólica com potência nominal de 2.200 kWh. Desta amostra, 8760 observações são utilizadas na modelagem (*in-sample*) e as 168 restantes na validação do modelo (*out-of-sample*) para diferentes horizontes de previsão. As séries temporais da velocidade do vento média horária e da potência média horária gerada são exibidas na Figura 7.1. De forma análoga, na Figura 7.2 apresenta-se o diagrama de dispersão da velocidade do vento média horária contra a potência média horária para o período referido acima, o qual descreve a curva de potência da base de dados históricos. Esta Figura mostra a relação de dependência não linear entre a energia eólica produzida e a velocidade do vento, pois por definição a potência varia com o cubo da velocidade do vento.



Figura 7.1 – Séries temporais da velocidade do vento (m/s) e da potência gerada (kWh).

A geração eólica guarda uma relação não linear com a velocidade do vento e ambas são variáveis aleatórias distribuídas segundo uma densidade conjunta, a qual permite derivar a densidade condicional da geração de energia eólica para uma dada velocidade do vento não conhecida.



Figura 7.2 – Curva de potência (Diagrama de dispersão da Velocidade do Vento (m/s) e a Potência (kWh)).

O histograma da velocidade do vento é apresentado na Figura 7.3. As maiores frequências da velocidade se concentram na faixa dos 7, 8 e 9 m/s, além disso, observa-se no histograma a forte presença de velocidades menores e algumas poucas frequências para velocidades maiores, validando que a velocidade do vento possui uma distribuição viesada.



Figura 7.3 – Histograma da velocidade do Vento (m/s).

Antes de aplicar o procedimento metodológico proposto para a previsão probabilística da geração de energia eólica usando as técnicas de SSA e Densidade Condicional por *kernel*, são apresentadas algumas informações descritivas que permitem caracterizar esta série temporal. As informações armazenadas nas Tabelas 7.1 e 7.2, contem algumas estatísticas descritivas do conjunto de dados tanto da velocidade do vento quanto da potência gerada.

Tabela 7.1 – Estatísticas descritivas da							
velocidade do vento (m/s).							
ESTATÍSTICA VALOR (m/s)							
Tamanho	8928						
Média	7,6609						
Mediana	7,5942						
Mínimo	0,0023						
Máximo	20,5850						
Desvio Padrão	4,1952						
Assimetria	0,1413						
Curtose	2,3000						

Tabela 7.2	2 – Estatística	s descritivas	da
no	ntência gerada	(kWh)	

1 8	(
ESTATÍSTICA	VALOR (kWh)
Tamanho	8928
Média	801,1282
Mediana	515,1211
Mínimo	1,3600e-08
Máximo	2100,3062
Desvio Padrão	771,6718
Assimetria	0,5188
Curtose	1,6790

O valor da média, da série temporal da velocidade do vento, é maior do que o valor de sua mediana, isto indica que possui uma distribuição assimétrica positiva, como foi evidenciado no histograma da Figura 7.3. Esta característica da série é reforçada com o valor do estatístico de assimetria, o qual é superior a zero. Os valores mínimo e máximo são coerentes com o comportamento da variável física, pois existem horas do dia em que a velocidade do vento é muito baixa, ou seja, o vento é quase nulo; mas tem outros momentos em que o vento possui uma maior velocidade. Por sua vez, o coeficiente da curtose indica uma alta intensidade das frequências na vizinhança dos valores centrais (média e mediana), o que traduz que a distribuição é também leptocúrtica.

Como a geração eólica tem uma relação de dependência com a velocidade, os valores para a produção eólica conduzem a interpretações semelhantes às da velocidade do vento, isto é, a série temporal da potência gerada possui uma distribuição assimétrica positiva, além de ser leptocúrtica. Os valores mínimo e máximo correspondem, respectivamente, a zero potência gerada, como consequência da ausência de vento e, ao valor (muito próximo) da capacidade da turbina.

#### Resultados do Primeiro Estágio: Previsão da Velocidade do Vento

Com o intuito de avaliar a metodologia proposta foram implementados modelos que envolvem um enfoque diferente de previsão da velocidade do vento que são: os modelos multiplicativos ARIMA (*AutoRegressive Integrate Moving Average*) também conhecidos como SARIMA, pertencentes à família de modelos proposto por Box & Jenkins (1970). Os modelos Double Seasonal Holt-Winters (Taylor, 2003), além do tradicional método Naive, o qual se baseia no conceito de que a melhor previsão um passo à frente será o valor observado um passo anterior. Para um maior detalhe dos modelos SARIMA e Double Seasonal Holt-Winters consultar os Apêndices C e D, respectivamente. Os resultados destas modelagens são apresentados posteriormente para serem comparados conjuntamente com os resultados obtidos na modelagem de SSA *sequencial* para a previsão da velocidade do vento.

#### 7.2.1.

#### Resultados da Primeira Fase da Metodologia de SSA Sequencial

A metodologia de especificação descrita no capítulo 6 no seu primeiro estágio requer o uso de L, o qual foi escolhido como sendo igual a 720 devido a que, como foi explicado na Seção 4.2, a velocidade do vento é uma série sazonal com períodos de 12, 24 e 720, assim que o tamanho do parâmetro L deve ser proporcional ao maior período da série. Portanto, seguindo o indicado por Golyandina et al. (2001) L é igual a 720. Este valor foi avaliado através da seleção de diferentes tamanhos de L que levassem em consideração a sazonalidade da série temporal, fazendo com que os valores testados fossem proporcionais a um dos períodos sazonais.

O procedimento de seleção dos tamanhos de *L* utiliza SSA *sequencial*, isto é, na primeira fase se extrai a componente do nível aplicando os dois estágios de SSA (*Decomposição e Reconstrução*), aqui não existe a fase de agrupamento já que a componente do nível corresponde à primeira autotripla na decomposição em valores singulares (SVD), razão pela qual o parâmetro de agrupamento toma o valor de r = 1, como pode ser verificado na Figura 7.4, onde o gráfico para a primeira componente possui um comportamento suave e sem crescimentos, próprio das séries de velocidade do vento, as quais apresentam características de estacionariedade.



Figura 7.4 – Autovetores da SVD para um L = 720.

Deste processo obtém-se a série do nível reconstruída e a série de ruídos. Posteriormente, considerando esta série dos ruídos como a nova série da segunda fase, novamente aplicam-se sobre ela os dois estágios de SSA (*Decomposição e Reconstrução*) para extrair as componentes sazonais. Esta segunda fase difere da primeira em que as componentes periódicas podem ser qualquer conjunto de autotriplas resultantes da SVD. Portanto, o parâmetro de agrupamento *r* varia de acordo com o número de componentes agrupadas, determinadas pelo método de *Análise dos Valores Singulares*. É assim como nesta segunda fase, é gerada a série das componentes periódicas reconstruídas e uma nova série de ruídos.

Posteriormente, as séries reconstruídas com base na componente do nível e nas componentes periódicas são usadas no terceiro estágio de SSA, no qual são calculadas as previsões das séries da velocidade do vento reconstruídas, para cada um dos cumprimentos de *L* avaliados apresentados na primeira coluna da Tabela 7.3. Para efetuar a comparação entre os distintos tamanhos de *L* foi escolhido o RMSE (*Root Mean Square Error*) como medida de erro, e cujos resultados estão contidos nesta mesma Tabela na

terceira coluna. O critério para determinar o parâmetro L estabelece que este deve ser escolhido com sendo aquele com o menor valor de RMSE. Por conseguinte, o menor valor do RMSE é 1,4822, e corresponde ao cumprimento de janela L = 720. Este resultado reforça a escolha inicialmente feita com base nos padrões sazonais presentes na série da velocidade do vento.

Parâmetro L	Parâmetro r	RMSE
12	1 até 6	3,7554
24	1 até 10	1,9557
36	1 até 16	2,0143
168	1 até 25	1,9217
720	1 até 18	<u>1,4822</u>
2920	1 até 28	2,6318
4380	1 até 29	2,4642

Tabela 7.3 – Parâmetro de defasagem *L*.

Fonte: Elaboração do autor.

É importante destacar que na segunda fase da metodologia SSA *sequencial* para determinar o tamanho de *L* o critério usado para usado para estabelecer o número de autotriplas agrupadas (parâmetro *r*) foi à *Análise dos Valores Singulares* das componentes obtidas da SVD. A técnica consiste em considerar cada valor singular  $\lambda_i$  obtido na SVD, através da razão dos autovalores  $\sum_{j=1}^{p_i} (\lambda_{l_{ij}})^{1/2} / \sum_{l=1}^{d} (\lambda_l)^{1/2}$ , a qual é ordenada de forma decrescente em magnitude e expressa a contribuição de cada um dos autovalores na SVD, que é o principio da análise de componentes principais (ACP) (Vasconcelos, 2011). Desta forma, são consideradas apenas as *r* primeiras autotriplas.

Devido à resolução temporal da serie em estudo, o tamanho de *L* é muito grande, o que consequentemente gera uma matriz trajetória de similares proporções. A SVD desta matriz trajetória fornece um número muito grande de autotriplas igual a 720, valor que corresponde ao tamanho de *L*. Este vasto número de autotriplas torna o processo de remoção das componentes sazonais todo um desafio, pois a maioria das técnicas sugeridas na literatura de SSA para o agrupamento das autotriplas faz uso da análise gráfica. Assim, ao considerar na *Análise dos Valores Singulares* todas as autotriplas na mesma figura, o gráfico resultante não é visualmente claro para a interpretação e extração das componentes como pode ser observado na Figura 7.5.



Figura 7.5 – Valores singulares para L = 720.

Outra dificuldade a ser superada é o fato de que na análise dos autovetores, os polígonos regulares conformados que indicam a presença de componentes sazonais nem sempre coincidem com as planícies explícitas no espectro dos valores singulares, os quais também indicam números ordinais das autotriplas emparelhadas que estão relacionados com períodos harmônicos específicos e, em muitos casos, quando usada com o gráfico da matriz de correlação ponderada não há evidência de correlação entre as componentes. As Figuras 7.6, 7.7 e 7.8 obtidas através do pacote "Rssa" do software "R Project for **Statistical Computing**", permitem evidenciar dita dificuldade para o caso em que L =36. A Figura 7.6 apresenta os valores singulares obtidos na SVD estabelecendo os postos das componentes harmônicas pertencentes às planícies explícitas entre autotriplas. Desta forma os círculos na Figura 7.6 assinalam este fato. No entanto, existem regiões nas quais é possível que não seja claramente identificável esta planície, o que faz necessário que o processo de identificação seja auxiliado com a Análise Gráfica dos Autovetores Pareados e da matriz de correlação ponderadas correspondentes às Figuras 7.7 e 7.8. Estas análises são complementares ao procedimento de agrupamento das componentes harmônicas devido a que acrescenta ou confirma as informações contidas nas Figuras. Pode existir uma situação inversa, acreditar que a região apresenta características de planície, isto é, postos consecutivos cujas conexões descrevem linhas horizontais (círculos da Figura 7.6). A região que se encontra no círculo vermelho, cujos postos correspondem às componentes emparelhadas 20-21 e 21-22 indica a existência de componentes periódicas. Ao inspecionar visualmente a Figura 7.7, os gráficos esboçados para as componentes emparelhadas 20-21 e 21-22 não descrevem polígonos regulares para estas componentes que fundamentem este fato. Além disso, ao analisar a Figura 7.8 constata-se também que a partir da componente 4 as componentes não estão fortemente correlacionadas (veja-se

Seção 4.4.2 para um melhor detalhe). As componentes pareadas que forma um polígono regular estão sublinhadas em amarelo na Figura 7.7.



Figura 7.6 – Valores singulares para L = 36.

Por conseguinte, é possível perceber como o problema de identificação pode crescer exponencialmente em função primeiro da complexidade da série temporal, mas também do valor assignado ao parâmetro L, o qual está relacionado ao tamanho da série de dados e que para o caso particular desta tese, este valor é muito grande, o que impossibilita realizar de forma eficiente a identificação e agrupamento das autotriplas através da análise gráfica; demandando métodos mais adequados que façam o agrupamento das componentes, como será analisado mais adiante no processo de implementação da metodologia para o cumprimento da janela L = 720.







Figura 7.8 – Matriz de correlações ponderadas.

#### Resultados da Segunda Fase da Metodologia de SSA Sequencial

Como foi indicado na Seção anterior, para identificar as componentes harmônicas que formam o sinal através da execução dos estágios de *decomposição* e de *reconstrução* de SSA básico, é preciso da série de ruídos obtida na fase anterior e o valor do parâmetro r, o qual é determinado no estágio de *reconstrução* SSA na fase de *agrupamento*. O valor do parâmetro L de defasagem para a segunda fase de SSA *sequencial* é o mesmo valor de L da primeira fase, que para este trabalho equivale a L = 720. Com esta informação é construída a matriz trajetória da fase de *incorporação* e, posteriormente, são obtidos as autotriplas (valores singulares e autovetores direitos e esquerdos) na fase da SVD. Logo, na primeira fase do estágio de *reconstrução* são identificadas as autotriplas que constituem o sinal referido às componentes harmônicas; para seguidamente, estabelecer o respectivo agrupamento das componentes detectadas através dos Métodos Hierárquicos Aglomerativos, os quais permitem contornar o problema de agrupamento empregando análise gráfica discutido na Seção anterior.

Para poder utilizar o Método Hierárquico Aglomerativo inicialmente é preciso da escolha de uma medida de similaridade. No escopo desta tese foi selecionada a matriz de *correlações ponderadas* como medida de similaridade pela sua semelhança com o "*Coeficiente de separação angular*" da área de clusterização. Uma vez que a matriz foi calculada se procede a transformar esta matriz em uma matriz de medidas de dissimilaridade usando a Equação (6.1). Adicionalmente é preciso estabelecer o critério de agregação de *clusters*, o qual foi escolhido como sendo o Método do vizinho mais próximo conhecido também como *single-link*. Tanto a medida de similaridade na implementação computacional.

Com o fornecimento de todas as informações, o seguinte passo no processo de agrupamento das autotriplas é a construção do dendrograma com base no Método Hierárquico Aglomerativo definido na Seção 10.1.3.1. Para isto, usou-se o pacote "**Rssa**" e outras funções disponibilizadas no software "**R Project for Statistical Computing**", além das informações anteriormente indicadas, as quais permitem definir os parâmetros de entrada da função de agrupamento. Os dendrogramas foram elaborados com a matriz de *correlação ponderada* (ver Seção 4.4.2), considerando 50 autotriplas para cada dendrograma, pois dado que o total destas autotriplas ascende a 720 a inspeção visualização se impossibilita por conta do volumem de informação. Esta problemática que está presente nas outras metodologias de análise gráfica para identificação e agrupamento das autotriplas vem sendo discutida com antecedência; não entanto, com o

uso do dendrograma se torna mais simples o processo de inspeção visual, devido a que o Método Hierárquico Aglomerativo já identifica os pares de autotriplas candidatos a serem componentes harmônicas. Para estabelecer quais autotriplas constituem o sinal é preciso efetuar cortes horizontais (linhas vermelhas das Figuras 7.9 e 10.2 até 10.14) que permitam determinar, a partir deles, os *clusters* constituídos, *clusters* que nesta segunda fase de SSA *sequencial* podem ser classificados unicamente em duas tipologias: componentes harmônicas ou ruído.

Para identificar a cardinalidade (entenda-se por ordem matricial por coluna de uma autotripla qualquer na matriz resultante da SVD) de uma autotripla basta só somar o número que a folha respectiva possui com o *número inicial* do intervalo (ao qual esta folha pertence), indicado *no título* de cada dendrograma, menos um. Assim por exemplo, para identificar a cardinalidade de uma autotripla que se encontra na Figura do dendrograma 10.2 e cuja folha corresponde ao número 40, ela é identificada como tendo uma cardinalidade igual a 90, como resulta de fazer a operação algébrica: 40 + 51 - 1 = 90. É importante ressaltar que a grande maioria dos cortes efetuados encontra-se no intervalo (0,4, 0,6), isto é, o eixo vertical das Figuras dos dendrogramas que contem as distâncias cartesianas das medidas de dissimilaridade.

Os valores da matriz de dissimilaridade são valores de distância que variam de 0 até 1. Assim que para começar com o procedimento de agrupamento, efetuam-se cortes horizontais tais que eles estejam relativamente próximos do nível mais alto da jerarquia, mas que não muito afastados das distâncias horizontais no intervalo (0,4, 0,6); de fato, se possível, tentar inicialmente a distância de 0,5 como corte horizontal para fazer os agrupamentos respectivos. Como foi mencionado anteriormente, devido à dimensão muito grande da matriz de dissimilaridades, foi preciso considerar grupos de 50 componentes para a construção de cada dendrograma; razão pela qual se obteve um total de 14 dendrogramas para serem analisados. O agrupamento das autotriplas foi realizado usando um dendrograma por vez.

Um aspecto muito importante que deve ser sublinhado é que o dendrograma gerado através Métodos Hierárquicos Aglomerativos comprovou ser uma ferramenta muito mais eficiente e precisa no processo de agrupamento das componentes harmónicas porque não se requere considerar nenhuma outra informação além do dendrograma e os cortes a serem efetuados, fato que reveste o processo de eficiência, se comparado com a metodologia tradicional adotada em SSA para a qual devem ser analisadas 4 tipologias de gráficos (espectro dos valores singulares, autovetores emparelhados conformado polígonos regulares, matriz de *correlação ponderada* e o periodograma) para problemas de porte menor. A continuação na Figura 7.10, apresenta-se o primeiro dendrograma do Método Hierárquico Aglomerativo, fazendo uso da estratégia indicada acima para a realização dos cortes. Os dendrogramas restantes podem ser visualizados no Apêndice B.



Figura 7.9 – Dendrograma das autotriplas 1 até 50.

As linhas transversais de cor vermelha presentes em todos os dendrogramas são denominadas: cortes. Estes cortes permitem identificar e escolher os *clusters* formados, isto é, os pares de autotriplas que são componentes harmônicas que constituem o sinal. Assim, por exemplo, do dendrograma da Figura 7.9 pode-se identificar a partir do corte efetuado, que as autotriplas escolhidas como sendo componentes harmônicas em ordem hierárquica são: 5-6, 29-30, 3-4, 39-40, 1-2, 17-18, 11-12, 20-21, 23-24 e 36-37 (ordenando ascendentemente as autotriplas tem-se: 1-2, 3-4, 5-6, 11-12, 17-18, 20-21, 23-24, 29-30 e 36-37, 39-40). Posteriormente, continua-se operando da mesma forma com os dendrogramas restantes até obter todos os pares de autotriplas que conformam o sinal. O total de autotriplas que são componentes harmônicas do sinal da série é igual a r = 259 autotriplas.

Um exercício que poderia ser interessante é validar se efetivamente estas autotriplas são componentes harmônicas. Isto pode ser constatado por médio do gráfico dos autovetores emparelhados apresentado na Figura 7.10. Nesta Figura se verifica que efetivamente estão contidas as informações dos pares de autotriplas que constituem o sinal referido às componentes harmônicas, formando polígonos aproximadamente regulares com *p*-vértices, como os sublinhados em cor amarela, correspondendo exatamente com as componentes indicadas pelo dendrograma e, cujos períodos são 24, 12, 8, 24, 22, 28, 26, 6, 27 e12 respectivamente. No entanto, se estivéssemos trabalhando

com a metodologia tradicional de identificação e agrupamento de SSA, observe-se que existem componentes harmônicas que deveriam ser descartadas como as componentes 17-18, 20-21, 23-24 e 36-37, devido a que os períodos delas (22, 28, 26 e 27) não são interpretáveis para dados horários desde o ponto de vista desta técnica, que analisa 4 tipos de figuras, mas que para a Metodologia de Agrupamento Hierárquico não são consideradas. Aqui é importante destacar que dependendo do problema, a abordagem com SSA poder diferir na fase de agrupamento se comparada com outros problemas, pois no caso de problemas nos quais seja requerida a filtragem, é importante reconstruir a série considerando somente componentes que tenham uma interpretação desde o ponto de vista do fenômeno que imerso na série, situação oposta que não é requerida em previsão. Embora isto não seja preciso para o caso da previsão das séries temporais, o que é fundamental é que as medidas de erro atinjam valores muito baixos, pois estas são indicadoras da bondade do ajuste dos dados. Por esta razão, apesar da existência de períodos não interpretáveis para as autotriplas identificadas como componentes harmônicas neste trabalho, elas serão consideradas efetivamente como componentes harmônicas que constituem o sinal. Por conseguinte, uma pergunta natural que lhe pode surgir a um analista é saber se ao efetuar cortes em distintos níveis do dendrograma, como identificar se as componentes que está considerando como sendo sinal sim melhoram a previsão?

A estratégia adota nesta pesquisa consiste em ir incorporando um par de autotriplas por vez ao estágio de *reconstrução* da série temporal, além de efetuar previsões no seguinte estágio da metodologia com estas autotriplas. Com a previsão podem ser calculadas as medidas de erro que permitiram decidir se deve ou não ser considerada a autotriplas em estudo como sendo uma componente harmônica, porém se recomenda somente fazer isto para as autotriplas que ficam muito mais afastadas do alto da hierarquia, pois poderia se tornar, dependendo da quantidade de autotriplas, um processo muito mais árduo e exaustivo do que ele já é atualmente.



Figura 7.10 – Autovetores emparelhados.

Uma vez as todas as autotriplas correspondentes às componentes harmônicas foram identificadas, prossegue-se com a segunda fase do estágio de *reconstrução* de SSA: o cálculo das médias diagonais; para posteriormente obter a série de componentes harmônicas reconstruída. Na Tabela 7.4 são apresentadas as autotriplas resultantes da SVD que foram identificadas e selecionadas através Método de Agrupamento Hierárquico Aglomerativo. Cada coluna desta Tabela contem as autotriplas identificadas e escolhidas de cada um dos catorze dendrogramas analisados. Assim por exemplo, a primeira célula da primeira coluna da Tabela 7.4, tem o título "D1", o qual representa a abreviação de "Dendrograma 1" (esta mesma analogia aplica para as colunas seguintes). A informação restante desta primeira coluna corresponde às autotriplas de dito dendrograma. Similarmente acontece para as colunas remanentes desta Tabela.

Com a componente do nível reconstruída na primeira fase de SSA *sequencial*, além das componentes harmônicas reconstruídas na segunda fase de SSA *sequencial*, prossegue-se com o terceiro e último estágio de SSA, relacionado à previsão da série temporal para obter as previsões horárias da velocidade do vento para *h* passos à frente, dando por finalizado com o cálculo destas previsões a segunda fase da metodologia SSA

*sequencial*. Com base nas informações resultantes do primeiro e do segundo estágio de SSA foram estimadas as previsões do terceiro estágio de SSA, para h = 8, 24, 72 e 168 passos à frente, as quais são apresentadas nas Figuras 7.11, 7.12, 7.13 e 7.14 respectivamente.

Tabela 7.4 – Autotriplas que conformam o sinal referente às componentes harmônicas escolhidas pelo Método Hierárquico.

D1	D2	D3	D4	D5	D6	<b>D</b> 7	<b>D8</b>	D9	<b>D10</b>	D11	D12	D13		D14		
1	54	101	151	201	251	308	351	404	456	501	564	604	651	677	700	
2	55	102	152	202	252	309	352	405	457	502	565	605	652	678	701	ĺ
3	63	103	153	203	253	312	357	426	459	506	582	649	653	679	702	Î
4	64	104	154	204	266	313	358	427	460	507	583	650	654	680	707	
5	67	113	155	205	267	314	359	445	461	520	598		655	681	708	
6	68	114	156	206	268	315	360	446	462	521	599		656	682	709	
11	71	117	157	207	269	316	361		464	524			657	683	710	
12	72	118	158	208	270	317	362		465	525			658	684	711	
17	78	136	168	209	271	318	363		466	528			659	685	712	
18	79	137	169	218	272	319	364		467	529			660	686	713	
20	88	138	174	219	273	320	365		468	547			664	687	714	
21	89	139	175	220	286	321	366		469	548			665	694	715	
23	90		176	221	287	322	367		475	549			668	695	716	
24	91		177	222	297	324	368		476	550			669	696	717	
25	97		178	223	298	325	380		477				672	697	718	
26	98		179	224	299	326	381		478				673	699	719	
29			196	229	300	327	383		491				674		720	
30			197	230		328	384		492							
39				231		329	386		493							
40				232		330	387		494							
				233			388		495							
				234			389		496							
				235			390		497							
				236			391		498							
				241					499							
				242					500							



Figura 7.11 – Previsão 8 horas à frente da Velocidade do Vento usando SSA sequencial.



Figura 7.12 - Previsão 24 horas à frente da Velocidade do Vento usando SSA sequencial.



Figura 7.13 - Previsão 72 horas à frente da Velocidade do Vento usando SSA sequencial



Figura 7.14 – Previsão 168 horas à frente da Velocidade do Vento usando SSA *sequencial* 

Tabela 7.5 – Medidas de erro das previsões da Velocidade do Vento para diferentes horizontes de tempo.

	Para $h = 8$	Para $h = 24$	Para $h = 72$	Para $h = 168$
RMSE (m/s)	1,2011	1,2050	1,7773	2,3928
MAD (m/s)	0,8939	0,9585	1,3699	1,8605
<b>U-THEIL</b>	0,5435	0,5345	0,7402	1,1396

É importante destacar que as previsões para o planejamento da operação para 72 e 168 horas à frente, segundo as classificações do horizonte de previsão da geração eólica apresentadas na Seção 3.1, são consideradas previsões de longo prazo. As previsões 24 horas à frente são classificadas como previsões de médio prazo. As previsões de curto prazo, por sua vez, estão em um intervalo que oscila entre os 30 minutos e no máximo 8 horas à frente. No entanto, nesta tese é considerada previsão de curto prazo como sendo aquela que pode ir de 1 até 24 horas à frente; médio prazo, de mais de um dia até 3 dias (72 horas) e no longo prazo, de mais 3 dias até 7 dias (168 horas). Em análises posteriores não será considerada a previsão de 8 horas à frente.

A Tabela 7.5 contem os resultados das medidas de erro RMSE, MAD e U-THEIL calculadas para os distintos horizontes de previsão. Apesar de que a previsão 168 horas à frente possui valores das métricas bastante razoáveis, o estatístico U-THEIL evidencia

que este modelo não é adequado para fazer previsões nesse horizonte de tempo, já que este estatístico avalia o desempenho das previsões com SSA contra os valores da previsão Naive (cuja estimativa futura é igual à estimativa do valor atual), indicando que quanto mais próximo de zero o U-THEIL estiver, melhor é a previsão com o Modelo SSA. Como o valor do estatístico é superior a 1, ou seja 1,1396, isto significa que o erro de SSA é maior do que da previsão Naive. Portanto, a previsão Naive se torna uma melhor opção de previsão do que empregar SSA. Os resultados das previsões para 8, 24 e 72 horas à frente a obtiveram bons resultados, mas como é sabido com o aumento do horizonte de previsão aumenta a incerteza na previsão e, por conseguinte, os valores das medidas de erro. Torna-se importante comparar os resultados do modelo e os horizontes de previsão de curto e médio prazo (24 e 72 horas) com outros modelos, já que a comparação permite ter uma melhor ideia do desempenho do modelo. Desta forma, a seguir nas Figuras 7.15 e 7.16 se ilustrará a previsão para 24 e 72 horas à frente dos modelos SSA *sequencial*, Double Seasonal Holt-Winters e SARIMA. Os modelos SARIMA e Double Seasonal Holt-Winters são detalhados no Apêndice C e D respectivamente.



Figura 7.15 – Previsão 24 horas à frente da Velocidade do Vento para diferentes modelos.

		*		
	SSA	SARIMA	D.S. Holt- Winters	Naive
RMSE (m/s)	<u>1,2050</u>	1,5630	1,8467	2,3121
MAD (m/s)	<u>0,9585</u>	1,2692	1,4500	1,8911
<b>U-THEIL</b>	<u>0,5345</u>	0,6932	0,8191	1,0255

Tabela 7.6 – Medidas de erro das previsões 24 horas à frente da Velocidade do Vento para diferentes modelos.



Figura 7.16 – Previsão 72 horas à frente da Velocidade do Vento para diferentes modelos.

	SSA	SARIMA	D.S. Holt- Winters	Naive
RMSE (m/s)	<u>1,7773</u>	1,9223	2,1662	2,4192
MAD (m/s)	<u>1,3699</u>	1,5475	1,7079	1,9209
<b>U-THEIL</b>	<u>0,7402</u>	0,8006	0,9022	1,0075

Tabela 7.7 – Medidas de erro das previsões 72 horas à frente da Velocidade do Vento para diferentes modelos.

Pode-se observar que o modelo de SSA *sequencial* tanto para as previsões 24 horas à frente quanto para 72 horas à frente obtiveram as melhores medidas de erro, proporcionando assim previsões mais exatas da velocidade do vento o que terá futuramente uma repercussão importante na previsão da geração da energia eólica, pois a potência depende diretamente do cubo da velocidade do vento o que implica que em essa proporção também são amplificadas as incertezas da previsão produção eólica. Apesar de SSA *sequencial* ter obtido melhores previsões da velocidade, os modelos SARIMA e D.S. Holt-Winters também possuem resultados bastante similares aos obtidos com SSA, além de ser consistentes com o horizonte de previsão.

Na seguinte Seção são apresentados os resultados da previsão probabilística da geração de energia eólica com base nas previsões obtidas em SSA *sequencial*, além de usar as previsões de outros modelos para efeitos comparativos.

#### 7.3.

#### Segundo Estágio: Estimação Estocástica da Curva de Potência

A estimação da densidade condicional por kernel constitui o segundo estágio do arcabouço metodológico. Para a estimação da densidade condicional da geração da potência, ou seja, a curva de potência empregou-se o estimador de kernel de Nadaya-Watson (Equações 6.5 e 6.6), o qual fez uso da função kernel Gaussiana. Inicialmente para estimar o kernel criou-se uma grade com valores da potência média horária que variam no intervalo de 0 até 2.200 (kWh) (capacidade da turbina), com incrementos de 10 kWh, conformando assim os valores da abcissa no plano cartesiano. Posteriormente, para construir a ordenada, cria-se uma grade com os valores da velocidade do vento que oscilam no intervalo fechado de 0 a 20 m/s com incrementos de 0,1 m/s. Ao estabelecer esta grade os resultados obtidos são formados por 221 valores diferentes para a potência gerada e 211 valores da velocidade do vento, constituindo assim uma matriz de 211×221, os quais foram usados em conjunto com estimador kernel Nadaya-Watson para obter consequentemente, a estimação da densidade condicional para cada um destes valores. Estes resultados podem ser observados graficamente nas Figuras 7.17 e 7.18, sendo os mesmos resultados, mas só que com uma rotação das Figuras. Pode-se observar na Figura 7.18 como esta densidade condicional descreve claramente o padrão da curva de potência para os distintos valores da grade. Estimações adicionais discretizando a potência média em intervalos menores foram realizadas, porém a maior resolução resultou apenas no aumento na carga computacional sem uma apreciável melhoria.



Figura 7.17 – Densidade Condicional da Potência calculada com estimador *Nadaya-Watson*.



Figura 7.18 – Densidade Condicional da Potência calculada com estimador *Nadaya-Watson*.

Desta forma, a matriz das densidades condicionais obtida com o cálculo do estimador de *Nadaya-Watson* foi armazenada para seguidamente, calcular a geração eólica.

#### 7.4.

### Terceiro Estágio: Previsão Probabilística da Geração Eólica

Uma vez geradas as densidades de probabilidade da produção eólica, condicionadas às velocidades do vento previstas com o modelo SSA *sequencial*, calculou-se a previsão pontual como sendo a média de cada densidade de probabilidade

condicional. A primeira coluna da Tabela 7.8 apresenta as medidas de erro segundo cada previsão pontual da geração de energia eólica. A Figura 7.19 apresenta as previsões pontuais 24 horas à frente da produção eólica, definidas pelas médias das distribuições condicionais.



Figura 7.19 – Previsão 24 horas à frente da geração de energia eólica usando SSA *sequencial*.

Para uma análise mais completa da densidade condicional por *kernel*, também foram estimadas as previsões dos modelos SARIMA multiplicativos e os métodos D.S.Holt-Winters, que como foi indicado anteriormente os detalhes de cada um destes modelos podem ser obtidos no Apêndice C e D. A Figura 7.20 esboça as previsões obtidas da produção eólica para cada dos modelos. Além disso, na Tabela 7.8 são apresentadas as medidas de erro dos modelos, incluindo o modelo SSA que foi detalhado na Figura anterior.

Tabela 7.8 – Medidas de erro das previsões pontuais 24 horas à frente da Geração de Energia Eólica para diferentes modelos.

Pevisão Out- of-sample	SSA	SARIMA	D.S. Holt- Winters	Naive
RMSE (m/s)	362,81	359,01	464,73	507,57
MAD (m/s)	248,77	259,01	330,44	347,27
<b>U-THEIL</b>	0,7303	0,7226	0,9354	1,0021



Figura 7.20 – Previsão da potência média horária (kWh) gerada pela turbina eólica 24 horas à frente.

Como pode ser observado na Tabela 7.8, o modelo de previsão da velocidade do vento que gera a melhor previsão horária da produção eólica 24 passos à frente é o modelo SARIMA para duas das mediadas de erro (RMSE e U-THEIL). Caso a escolha do melhor modelo de previsão esteja baseada na métrica MAD esta indicaria o modelo SSA. Segundo os resultados da Tabela, desde uma ótima muito rigorosa, pode-se dizer que não vale a pena usar a técnica SSA, pois duas das três medidas de erro favorecem outra tipologia de modelo. No entanto, ao comparar as medidas de erro da previsão da produção eólica correspondente aos modelos SSA e SARIMA, percebe-se que as diferencias entre estes valores não são realmente significativas e, portanto, as velocidades previstas com SSA poderiam ser consideradas uma boa opção para efetuar a previsão da produção eólica.

Para validar se efetivamente o modelo SSA não é apto para a previsão da velocidade do vento, foram calculadas as previsão para um horizonte de 72 horas à frente. Os resultados das novas previsões da geração eólica, empregando os distintos modelos de previsão da velocidade, encontram-se na Figura 7.21 e na Tabela 7.9. A melhor previsão de energia eólica é fornecida pelas previsões da velocidade do vento obtidas com SSA.

Intuitivamente se poderia pensar que o modelo com melhor previsão para o horizonte de previsão de 72 horas deveria continuar sendo o modelo SARIMA, como o foi para o horizonte de 24 horas. Porém, os resultados das medidas de erro são completamente opostos a isso, pois para todas as medidas a melhor previsão foi obtida com o modelo de SSA. Este resultado permite inferir que a previsão com SSA para o

horizonte de 24 horas não atingiu os melhores resultados para todas as medidas de erro, possivelmente porque o período de previsão possui características particulares que podem afetar a precisão da previsão, como por exemplo, a mudança do padrão sazonal da série ou valores atípicos presentes unicamente nessa faixa de tempo. Portanto, SSA é o modelo que possui melhores qualidades no ajuste da previsão.



Figura 7.21 – Previsão da potência média horária (kWh) gerada pela turbina eólica 72 horas à frente.

Tabela 7.9 – Medidas de erro das previsões pontuais 72 horas à frente da Geração de Energia Eólica para diferentes modelos.

Pevisão Out- of-sample	SSA	SARIMA	D.S. Holt- Winters	Naive
RMSE (m/s)	375,73	380,49	458,96	498,24
MAD (m/s)	252,66	264,47	324,84	329,52
<b>U-THEIL</b>	0,7541	0,7637	0,9212	1,0000

Por outro lado, os valores resultantes da medida de erro U-THEIL indicam que as previsões pontuais da produção eólica determinadas pela média das densidades condicionais são melhores que a previsão Naive. Com relação às medidas de erro RMSE e MAD, estas apresentam valores altos como consequência de duas situações: a primeira é o erro inerente à previsão da velocidade do vento, os quais são amplificados na previsão da geração de energia eólica, dado que a potência gerada é proporcional ao cubo da velocidade do vento e, a segunda, reside na incapacidade do estimador *kernel* de

conseguir reproduzir fielmente a densidade condicional, conforme evidenciado na Figura 7.22, onde se relaciona a curva de potência real com a média condicional calculada acima para obter as previsões pontuais, na qual se observa claramente que a média condicional da potência média para pequenos valores de velocidade do vento assume valores superiores aos valores observados. Isto se deve à pouca concentração de dados na faixa inferior a 3,5 m/s e 0 da velocidade do vento, a qual corresponde ao intervalo no qual teoricamente não deveria existir aproveitamento energético, ou seja, a geração eólica deveria ser zero. A outra razão obedece à discrepância de alguns valores da velocidade do vento, por exemplo, 2,64 m/s a potência gerada registrou um valor de 16,15 kWh, em quanto que para 2,77 m/s a potência gerada foi de 13,05 kWh, o que resulta inconsistente com a teoria de que a potência depende do cubo da velocidade acrescentada de um fator que depende da densidade do ar, da área que cobre o aerogerador entre outros. Como o estimador *kernel* utilizado para o cálculo da densidade de probabilidade condicional não é um estimador adaptativo no tempo, não consegui capturar estes valores baixos.



Figura 7.22 – Potência média condicional calculada com estimador *Nadaya-Watson* (vermelho) vs Curva de potência do histórico (azul).

Em resumo, a metodologia proposta atende as expectativas da geração de previsões probabilísticas, no contexto da energia eólica, de forma satisfatória.