

8. Aplicação a uma Série Real de Velocidade do Vento

A energia de origem hidrelétrica predomina na matriz elétrica brasileira. No entanto, alguns fatos têm gerado preocupações quanto ao atendimento da demanda de energia elétrica: o crescimento da população; a confiabilidade da rede; o custo da energia e a possibilidade de escassez de recursos hídricos. O advento da energia eólica e a tendência de crescimento do consumo de energia elétrica no Brasil apontam para um importante recurso complementar à geração hidrelétrica. Para este tipo de geração de energia, a principal variável a ser estudada é a velocidade do vento, pois a partir da velocidade do vento se avalia o potencial de energia eólico de certo local. Com a crescente cultura de geração eólica de energia, gerar previsões específicas para este tipo de energia, e diretamente para séries de velocidade de vento, torna-se uma tarefa necessária para o bom aproveitamento deste recurso e tomada de decisões do setor elétrico. Previsões de velocidade do vento são essenciais para reduzir custos de reserva de energia, e muitos trabalhos têm sido desenvolvidos aplicando às séries de velocidade do vento diferentes ferramentas de filtragem e modelagem de séries temporais, com o objetivo de melhorar a qualidade da previsão dos modelos para estas séries.

A potência gerada guarda uma relação cúbica com a velocidade do vento (PESSANHA et al., 2010). Esta relação pode ser observada na curva de potência representada na Figura 6.3 e na equação (34).

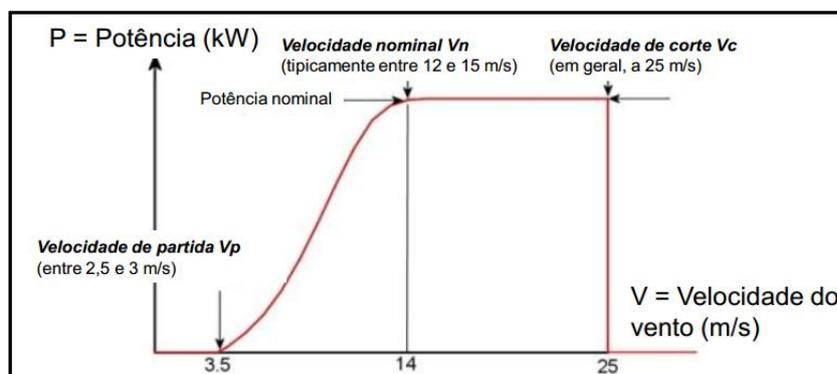


Figura 8.1. Curva de Potência Elétrica Produzida em função da velocidade do vento.

A Figura 8.1 mostra a relação entre a potência gerada e a velocidade do vento através da Curva de Potência. Segundo esta relação, a potência gerada pela turbina é nula se a velocidade do vento for inferior a 3,5 m/s. Neste ponto, a velocidade do vento recebe o nome de “velocidade de partida” (V_p). Normalmente, a utilização em sistemas pequenos exige uma velocidade média de 3,5 m/s a 4,5 m/s e aplicações em larga escala requerem uma velocidade média de, no mínimo, 6,5 m/s a 7,5 m/s.

A partir de 3,5 m/s até o ponto cuja velocidade do vento está entre 12 e 15 m/s, conhecida como “velocidade nominal” (V_n), a potência gerada é dada pela equação (49):

$$P = \frac{1}{2} C_p \cdot \eta_m \cdot \eta_g \cdot \rho \cdot A \cdot v^3, \quad (51)$$

onde:

P é a potência gerada;

ρ é a densidade do ar (kg/m^3);

A é a área de giro das lâminas da turbina eólica (m^2);

C_p, η_m, η_g são as constantes que expressam os rendimentos nos estágios da conversão de energia;

v é a velocidade do vento.

No intervalo em que a velocidade do vento está entre 15 m/s e 25 m/s, a potência gerada se estabiliza e se a velocidade do vento superar os 25 m/s, a turbina é desligada para evitar esforços mecânicos que possam danificá-la. Neste ponto, a velocidade do vento é chamada de “velocidade de corte” (V_c) e a partir destas informações, pode-se definir a potência como em (52):

$$P = \begin{cases} 0, & \text{se } v < V_p \\ P = \frac{1}{2} C_p \cdot \eta_m \cdot \eta_g \cdot \rho \cdot A \cdot v^3, & \text{se } V_p \leq v \leq V_n \\ P = \frac{1}{2} C_p \cdot \eta_m \cdot \eta_g \cdot \rho \cdot A \cdot v^3, & \text{se } V_n \leq v \leq V_c \\ 0, & \text{se } v > V_c \end{cases} \quad (52)$$

Para a aplicação do método proposto nesta tese foi utilizada uma série de 192 observações da média mensal de velocidade do vento medida a cada 10 minutos em uma estação anemométrica localizada em Petrolina-PE a 50 metros de altura, entre os meses de janeiro de 1996 a dezembro de 2011. A série é do Instituto Nacional de Meteorologia (INMET), e está disponível em

<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>.

8.1. Modelagem e Previsão Tradicional (sem SSA)

A Figura 8.2 a seguir mostra o comportamento da série de médias mensais de Velocidade do Vento medidas na Estação de Petrolina, na Bahia. Na Figura 8.3 tem-se o histograma da distribuição da velocidade do vento nesta amostra e o Box plot desta distribuição. A Tabela 8.1 traz as principais estatísticas amostrais dos dados.

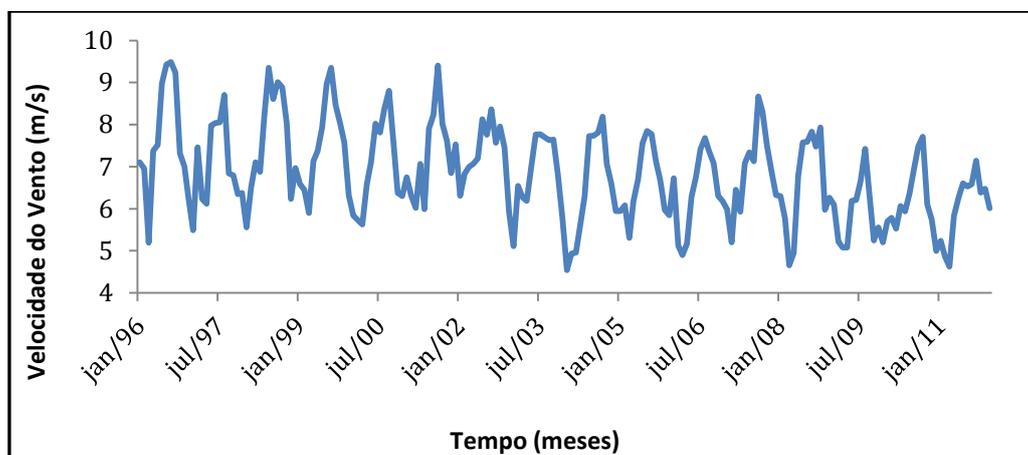


Figura 8.2: Médias mensais de velocidade do vento na Estação Petrolina.

Tabela 8.1. Principais Estatísticas da Série Original de Petrolina

Estatística	Valor
Tamanho	192
Média	6,832
Mediana	6,768
Mínimo	4,543
Máximo	9,491
Desvio Padrão	1,110
Coeficiente de Variação	0,162
Assimetria	0,182

Estatística	Valor
Curtose	-0,425
p-valor do teste de Normalidade de Jarque Bera	0.206
p-valor teste de Normalidade de Kolmogorov-Smirnov	0.323

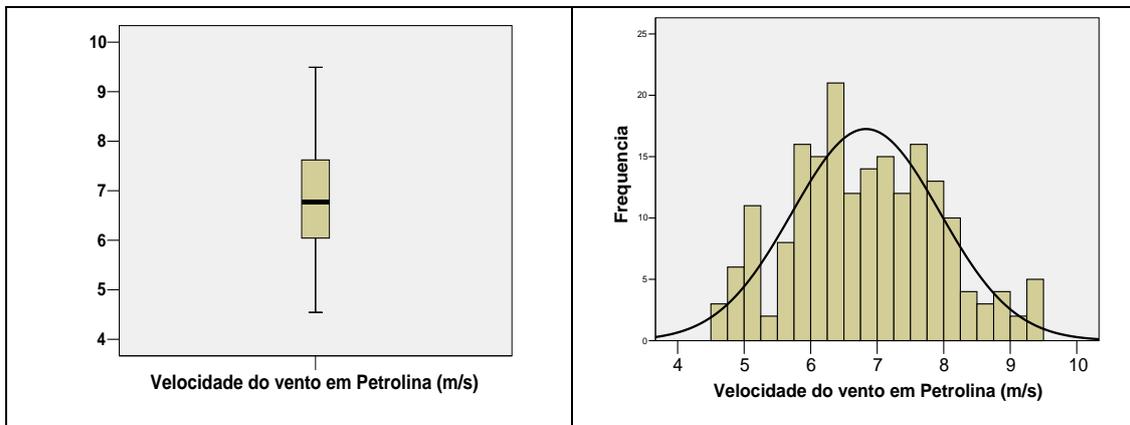


Figura 8.3: Box Plot e Histograma da Distribuição de velocidade do vento na Estação Petrolina - PE.

Os valores da série de velocidade do vento são compreendidos entre 4.543 m/s a 9.4941 m/s, com média de 6.832 m/s, sendo a classe modal de velocidades compreendidas entre 6.25 e 6.5 m/s. Este ponto de Petrolina onde se localiza esta estação pode então ser considerado um ponto eólico de potencial razoável, cuja média está sempre acima da média da velocidade mínima requisitada para produção das turbinas, 3,5 m/s. A velocidade mediana é 6.768 m/s e há baixa variabilidade entre os valores da série: a amplitude amostral é de 4,9511 m/s, desvio padrão de 1,110 m/s, que ocasiona baixo coeficiente de variação, igual a 0.162, caracterizando baixo grau de heterogeneidade nos dados. Como pode ser visto no Box Plot exibido na Figura 8.3 não há presença de outliers na série, que tem distribuição com baixos graus de assimetria e curtose (-0,425 e 0.206). De fato, como mostra o histograma da Figura 8.3, a distribuição é bem aproximada à Normal, o que é comprovado significativamente pelos testes de Normalidade Jarque Bera e Kolmogorov-Smirnov. De acordo com os p-valores de ambos os testes, exibidos na Tabela 8.1, não há evidência que levem à rejeição da hipótese nula de Distribuição Normal para a série de Velocidade do Vento de Petrolina.

Ao submeter a série ao teste da Raiz Unitária de Dickey-Fuller verificou-se que a hipótese de raiz unitária é rejeitada com significância, p-valor=0.000. O resultado é confirmado pelo teste KPSS que resultou em um p-valor= 0.7390, levando a concluir que não há evidência para rejeição da hipótese nula, que no teste

KPSS é H_0 : o processo não tem raiz unitária. O que se conclui a partir dos dois testes é o mesmo: a série dada não tem raiz unitária e é estacionária. Sendo assim não precisa ser diferenciada, e para uma modelagem ARIMA (p,d,q) de Box-Jenkins, $d = 0$. Tal conclusão é condizente com a dinâmica da série exibida na Figura 8.2, onde não são verificadas alterações bruscas de nível e reflete alguma forma de equilíbrio.

O gráfico exibido na Figura 8.4 exibe o Box-plot das distribuições de velocidade do vento de cada mês. A Tabela 8.2 traz as principais estatísticas da velocidade do vento, por mês.

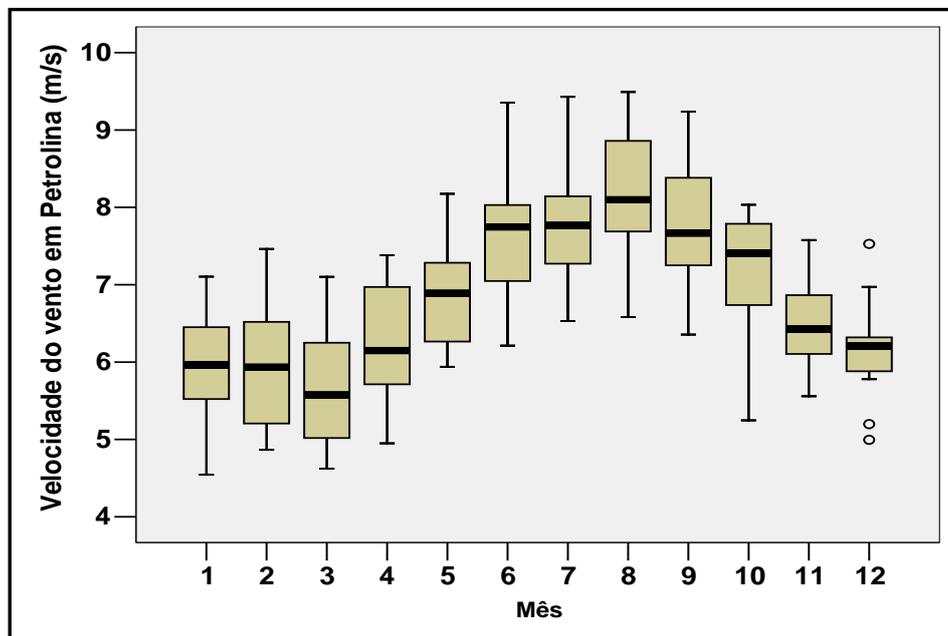


Figura 8.4: Boxplots das médias de velocidade de vento na Estação de Petrolina, agrupadas por mês.

Tabela 8.2: Principais estatísticas, por mês, das medidas de velocidade do vento em Petrolina/BA.

Mês	Média	Mediana	Desvio Padrão	Mínimo	Máximo	Coefficiente de Variação	Amplitude Amostral
1	5,954	5,964	0,678	4,543	7,104	0,114	2,561
2	5,984	5,934	0,783	4,865	7,463	0,131	2,598
3	5,677	5,574	0,785	4,622	7,103	0,138	2,481
4	6,202	6,150	0,769	4,950	7,382	0,124	2,432
5	6,864	6,889	0,685	5,939	8,175	0,100	2,236
6	7,641	7,748	0,859	6,213	9,352	0,112	3,139
7	7,781	7,767	0,782	6,530	9,429	0,101	2,899

Mês	Média	Mediana	Desvio Padrão	Mínimo	Máximo	Coefficiente de Variação	Amplitude Amostral
8	8,230	8,099	0,831	6,581	9,491	0,101	2,910
9	7,818	7,669	0,794	6,356	9,236	0,102	2,880
10	7,184	7,408	0,785	5,247	8,033	0,109	2,786
11	6,510	6,426	0,569	5,558	7,578	0,087	2,020
12	6,136	6,207	0,595	4,995	7,528	0,097	2,533

Pela Figura 8.4 e Tabela 8.2 é fácil perceber quão distintas dos demais meses são as distribuições da velocidade do vento nos meses de junho a outubro. Somente nestes meses a velocidade do vento ultrapassa 8 m/s e as médias e medianas ultrapassam 7.1 m/s. O mês de agosto é o mês com parâmetros de maior velocidade do vento (média, mediana e máximo), ao passo que no mês de março são registradas a menor média e menor mediana de velocidade do vento. Estes distintos padrões em meses distintos sugerem a presença de sazonalidade na série temporal da velocidade do vento.

Com o intuito de identificar os componentes presentes no padrão básico da série histórica de dados (tendência, ciclo, sazonalidade e aleatoriedade), foi feita análise de decomposição clássica pelo software FPW e foram obtidos os seguintes percentuais relativos a cada componente: Tendência/ciclo: 22,93%, Sazonal: 57,00% e Irregular: 20,07%, com formação multiplicativa.

As estimativas dos coeficientes de autocorrelação e autocorrelação parcial para a série foram calculadas e os primeiros 48 valores são exibidos nas Figuras 8.5 e 8.6 a seguir. Os valores para os lags 1, 12, 24, 36 e 48 são destacados em cor magenta. Além de calcular as autocorrelações, foram aplicados os testes de Ljung-Box. Para todas as autocorrelações até lag 48 os p-valores resultantes do teste foram iguais a 0,0000. Sendo assim os coeficientes de autocorrelação são significativamente distintos de zero.

O gráfico do correlograma exibido na Figura 8.5 apresenta oscilações ao longo das defasagens. Inicialmente as autocorrelações estimadas decrescem lentamente e em seguida é observado um crescimento gradual. A FAC é infinita em extensão com comportamento senoidal indicando um processo autoregressivo e sazonal. O gráfico da PACF tem uma autocorrelação parcial significativa, de valor 0,740, e

depois o valor cai bruscamente. Embora não seja observada a presença de todas as autocorrelações parciais a partir do lag 2 dentro do intervalo assintótico $\left[-1,96 \frac{1}{\sqrt{T}}, +1,96 \frac{1}{\sqrt{T}}\right]$, somente quatro correlações parciais não estão dentro deste intervalo, isso ainda é um indício da existência de um componente AR(1), logo temos $p = 1$.

Não há um corte em algum ponto q da função de Autocorrelação. Conclui-se então que não há componentes de médias móveis na série, logo temos $q=0$. Em geral séries físicas não tem mesmo componente de médias móveis.

Os altos valores das autocorrelações nos lags 1, 12, 24, 36 e 48 confirmam a existência de uma componente periódica de período $s=12$. Como o decrescimento da autocorrelação nestes lags é lento, conclui-se que há não estacionariedade sazonal na série e assim ela deve ser submetida a um filtro linear instável $(1 - B^{12}) = \nabla_{12}^1$, logo temos $D=1$.

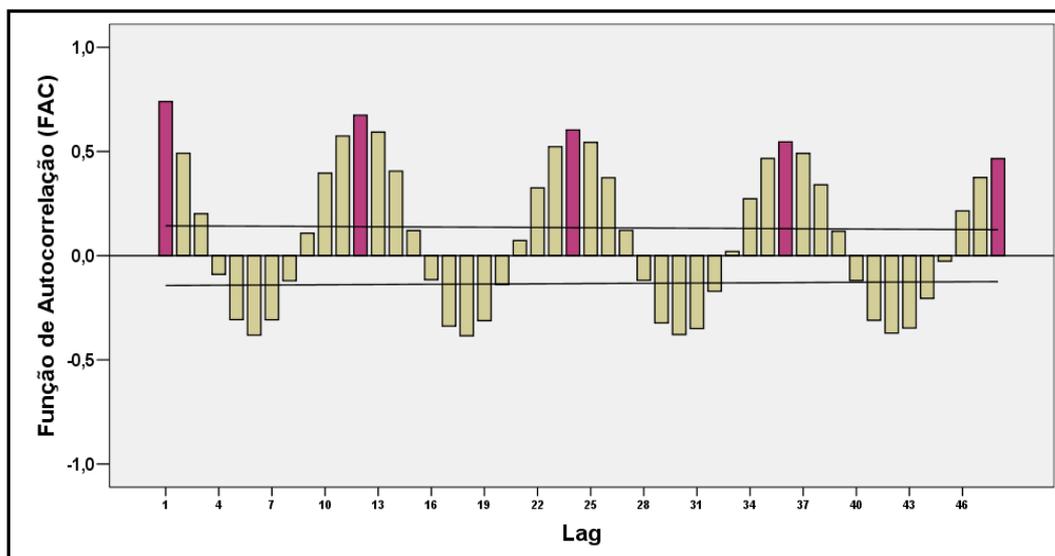


Figura 8.5. Função da Autocorrelação da série de Velocidade do Vento de Petrolina.

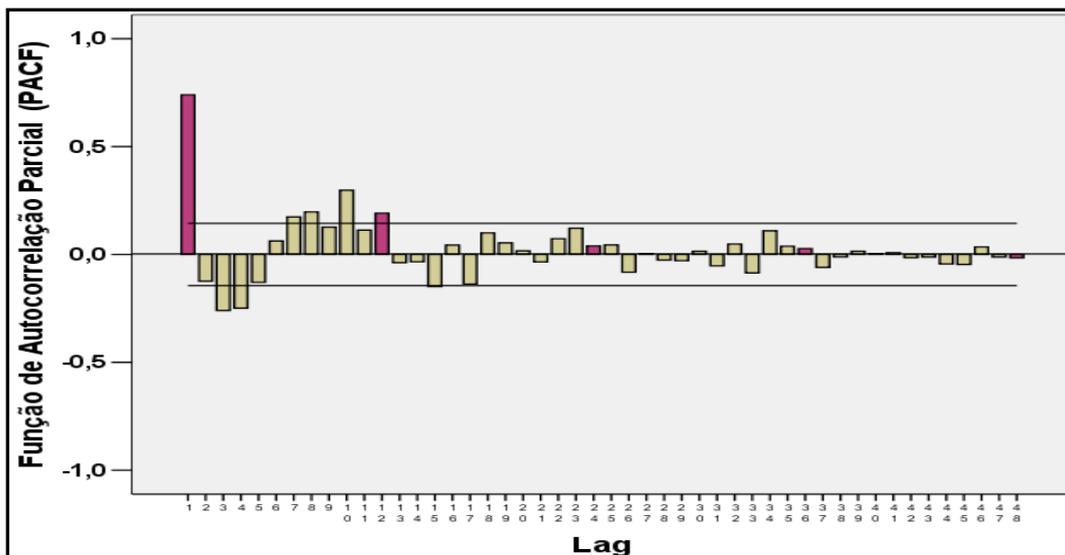


Figura 8.6. Função de Autocorrelação Parcial da série de Velocidade do Vento de Petrolina.

Aplicando então a diferença ∇_{12}^1 à série original foram estimadas sua FAC e FACP, exibidas nas Figuras 8.7 e 8.8. Como pode ser visto na Figura 8.7, existe um corte em $Q=2$ na FAC de $\nabla_{12}^1 Y_t$. Logo, isso sugere que $Q=2$.

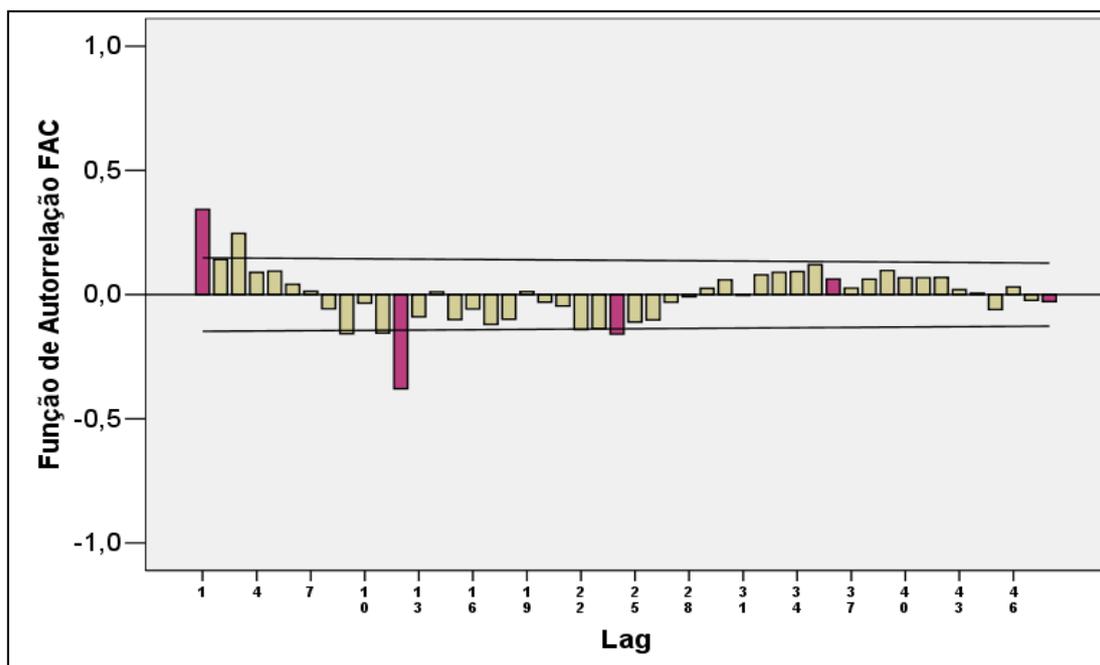


Figura 8.7. Função da Autocorrelação da série $\nabla_{12}^1 Y_t$.

E como pode ser visto na Figura 8.8, existe um corte em $P=3$ na FAC de $\nabla_{12}^1 Y_t$ que sugere que $P=3$.

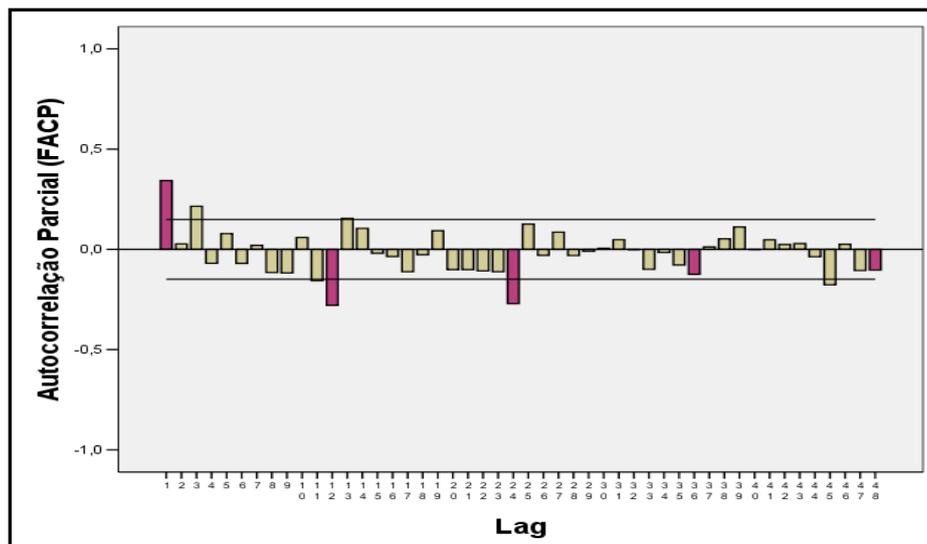


Figura 8.8. Função da Autocorrelação Parcial da série $\nabla_{12}^1 Y_t$.

Enfim tem-se uma sugestão de um SARIMA $(1,0,0) \times (3,1,2)_{12}$ para modelar a dinâmica da série de velocidade do vento de Petrolina-BA, cujos parâmetros estimados e medidas de qualidade de ajuste são exibidos na Figura 8.9 abaixo.

Forecast Model for Petrolina				
ARIMA(1,0,0) * (3,1,2)				
Term	Coefficient	Std.Error	t-Statistic	Signif.
a[1]	0.4572	0.0675	6.7686	1.0000
A[12]	0.6675	0.1132	5.8990	1.0000
A[24]	-0.2627	0.0827	-3.1773	0.9985
A[36]	-0.3170	0.0821	-2.3869	0.8012
B[12]	1.4678	0.0929	15.7920	1.0000
B[24]	-0.6838	0.0697	-9.8136	1.0000
Within-Sample Statistics				
Sample size	192		Number of parameters	6
Mean	6.832		Standard deviation	1.11
R-square	0.7582		Adjusted R-square	0.7517
Durbin-Watson	2.06	**	Ljung-Box(18)	=42.46 P=0.999
Forecast error	0.5533		BIC	0.5912
MAPE	0.07051		RMSE	0.5445
MAD	0.4233			

Figura 8.9. Estimação do modelo SARIMA $(1,0,0) \times (3,1,2)_{12}$ para Y_t pelo FPW.

Esta identificação do modelo é então comparada com a identificação automática expert oferecida pelo software FPW. O software sugere que o melhor modelo para explicar a dinâmica da série de velocidade do vento de Petrolina é o

modelo $ARIMA(1,0,0)*(1,0,1)_{12}$ com transformação logarítmica, cujas estimativas dos parâmetros e medidas de qualidade de ajuste são exibidos na Figura 8.9 a seguir. Então dois modelos SARIMA são candidatos a representar a série de velocidade do vento de Petrolina: o modelo 1, SARIMA $(1,0,0) \times (3,1,2)_{12}$ e o modelo 2 SARIMA $(1,0,0) \times (1,0,1)_{12}$. Comparando as medidas de qualidade de ajuste, *Forecast Error*, BIC, MAPE, RMSE e MAD conclui-se que o Modelo 1 ajusta melhor a série. A vantagem do modelo 2 é que ele é um modelo mais parcimonioso. Pra este trabalho, o critério a escolher o melhor modelo será melhoramento na qualidade de ajuste, assim, elege-se pelo método tradicional, o modelo 1, SARIMA $(1,0,0) \times (3,1,2)_{12}$ para modelar e gerar previsões para a série de Velocidade do Vento de Petrolina.

Forecast Model for Petrolina				
ARIMA(1,0,0)*(1,0,1) with log transform				
Term	Coefficient	Std.Error	t-Statistic	Signif.
a[1]	0.4946	0.0633	7.8091	1.0000
A[12]	0.9843	0.0139	70.9102	1.0000
B[12]	0.7700	0.0554	13.9041	1.0000
_CONST	0.0152			
Within-Sample Statistics				

-				
Sample size	192	Number of parameters	3	
Mean	1.908	Standard deviation	0.1639	
R-square	0.711	Adjusted R-square	0.7079	
Durbin-Watson	2.136	* Ljung-Box(18)	=32.81	P=0.9824
Forecast error	0.08861	BIC	0.6175	
MAPE	0.07489	RMSE	0.5751	
MAD	0.4553			

Figura 8.10. Estimação do modelo SARIMA $(1,0,0) \times (1,0,1)_{12}$ para Y_t pelo FPW.

Ao analisar a série dos erros ou resíduos das previsões geradas pelo modelo SARIMA $(1,0,0) \times (3,1,2)$ escolhido, verificou-se que a média dos erros é igual a $-0,020$ (aproximadamente nula) e variância $0,03$. O gráfico da série dos erros exibido na Figura 8.10 mostra comportamento de média e variância constante ao longo do tempo, sugerindo que os erros seguem processo estacionário. Quando submetida ao teste ADF, verificou-se que há evidências para afirmar que a série de

resíduos é estacionária, com p-valor do teste ADF= 0,000, para as três configurações (nenhuma, constante, constante e tendência), rejeitando a hipótese nula de raiz unitária. O gráfico da FAC dos erros exibido na Figura 8.11 abaixo leva a concluir que os erros podem ser considerados não correlacionados uma vez que eles apresentam, em geral, valores não significativos. Adicionalmente, a série dos resíduos foi submetida ao teste de Ljung Box que com p-valor=0,9171 leve nos a concluir que não há evidência para rejeitar a hipótese de que as autocorrelações nos resíduos até lag k=48 são nulas. As características observadas para os erros ou resíduos do modelo são compatíveis com as características de um processo ruído branco e assim pode-se concluir que o modelo proposto para modelar a velocidade de vento em Petrolina é eficiente.

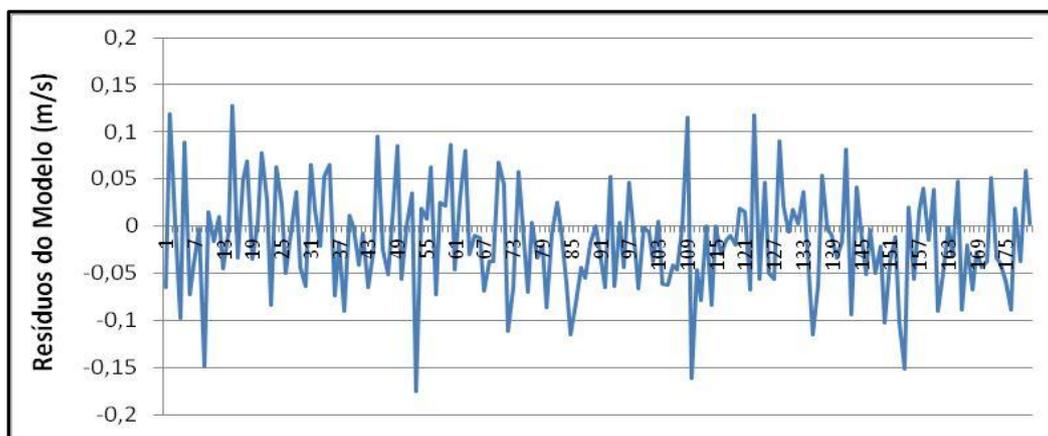


Figura 8.11. Série dos resíduos do modelo SARIMA (1,0,0) \times (3,1,2)₁₂.

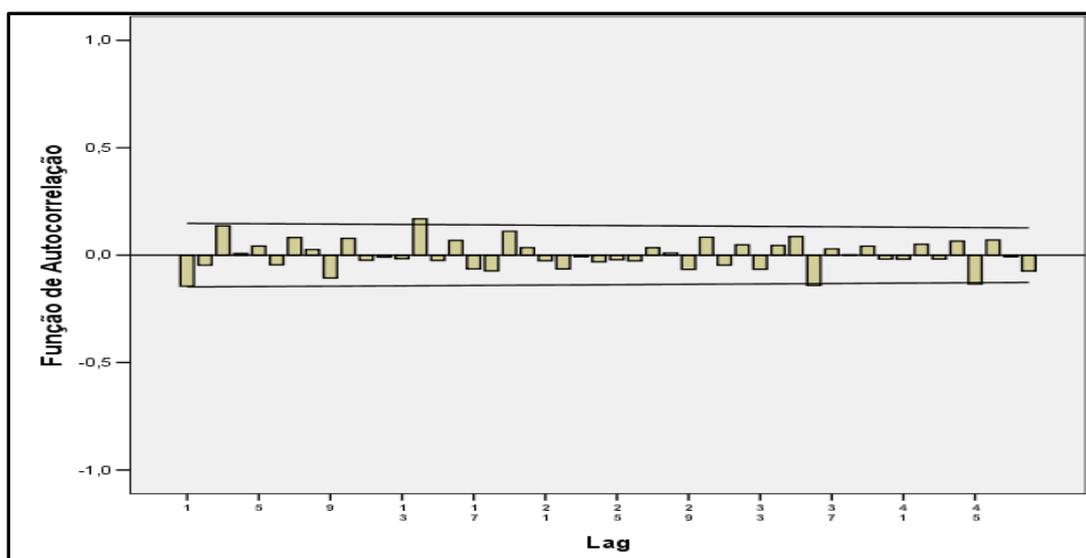


Figura 8.12. Autocorrelação dos resíduos do modelo SARIMA (1,0,0) \times (3,1,2)₁₂

Conforme proposto por SOUZA & CAMARGO (2004), após a escolha de um modelo via metodologia Box Jenkins é aconselhável elaborar novos modelos com um número de parâmetro superior ao do modelo fixado, que cubra as supostas direções de discrepâncias. Neste trabalho foram propostos os seguintes modelos para sobrefixação ao modelo $ARIMA(1,0,0) \times (3,1,2)_{12}$: $ARIMA(2,0,0) \times (3,1,2)_{12}$, $ARIMA(1,0,0) \times (4,1,2)_{12}$, $ARIMA(1,0,0) \times (3,1,3)_{12}$, $ARIMA(0,0,0) \times (3,1,2)_{12}$, $ARIMA(1,0,0) \times (2,1,2)$ e $ARIMA(1,0,0) \times (3,1,3)_{12}$. Além de não reduzirem as estatísticas de interesse dos erros, todos os modelos de sobrefixação propostos apresentaram não significância nos parâmetros adicionais, reafirmando a significância do modelo $SARIMA(1,0,0) \times (3,1,2)_{12}$.

8.2. Modelagem e Previsão usando SSA

Para a abordagem SSA, foi utilizado o comprimento de janela $L = 96$. Com isso a matriz XX^T , onde X é a matriz trajetória, tem dimensão (96×96) e são obtidos 96 autovalores. A Figura 8.13 apresenta o comportamento dos logaritmos destes autovalores.

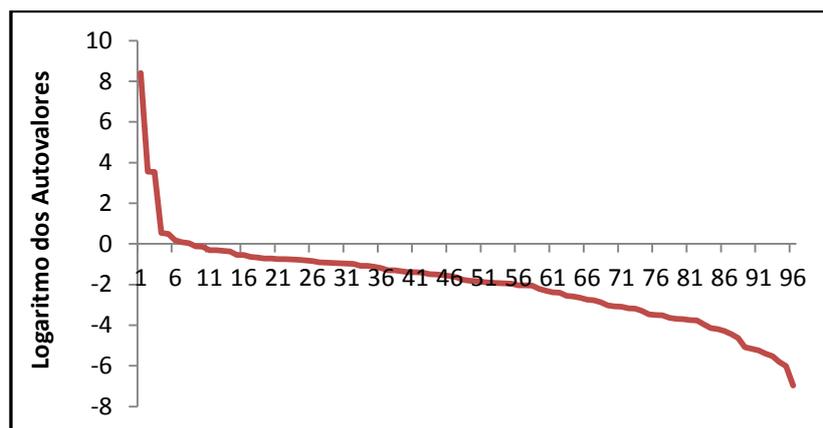


Figura 8.13. Logaritmo dos autovalores da matriz XX^T , decomposição SSA.

8.2.1. Identificação do Ruído por Verificação Visual dos Vetores Singulares

Na determinação das componentes de sinal e ruído, a análise do comportamento dos gráficos dos vetores singulares permite identificar a estrutura de cada componente. A Figura 8.13 mostra os seis primeiros vetores singulares. Pode-se perceber na Figura 8.14, através do comportamento exibido nos gráficos que, os vetores 1, 4, 5 e 6 pertencem à componente de tendência por conta de seus comportamentos mais suaves. Já os vetores 2 e 3 sugerem pertencer à componente harmônica. Porém, estudos mais específicos são necessários para chegar a esta conclusão. Um destes estudos está baseado no comportamento do gráfico de dispersão entre dois vetores singulares consecutivos no diagrama de dispersão. Caso ele se comporte como um polígono regular, os vetores envolvidos pertencem à componente harmônica e o seu período é notado pelo número de lados deste polígono. A Figura 8.14 mostra alguns gráficos de dispersão.

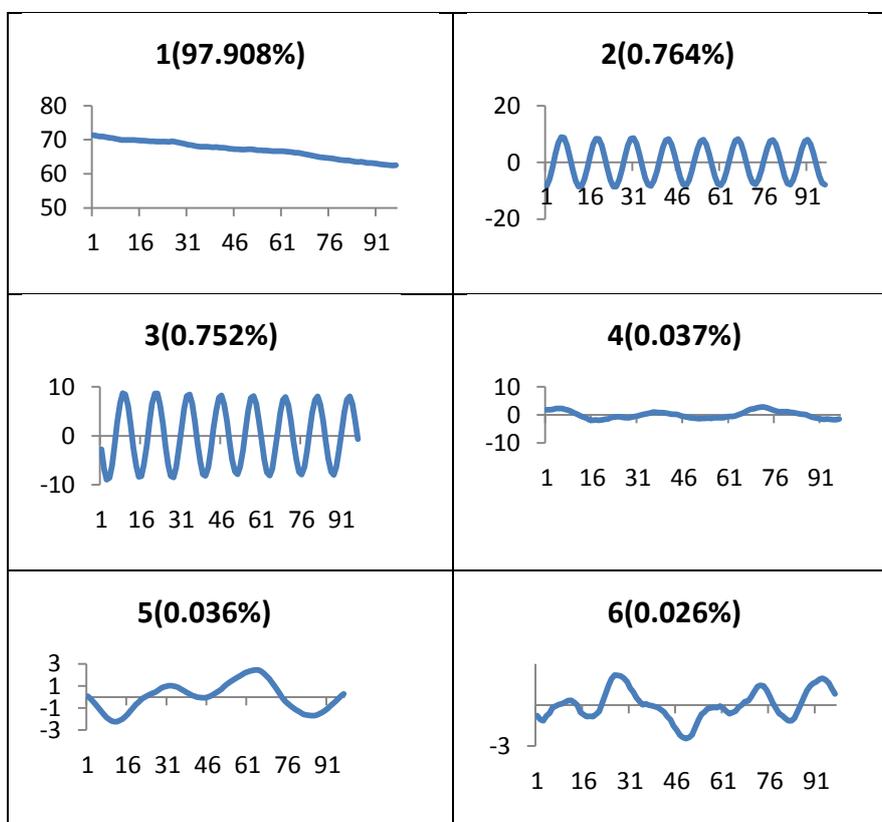


Figura 8.14. Os seis primeiros vetores singulares em ordem de significância.

Conforme pode ser observado na Figura 8.14, os diagramas de dispersão têm comportamento diferenciado de acordo com a componente à qual o vetor pertence. Pode-se observar que o diagrama de dispersão do par (2 e 3) tem uma forma poligonal com 12 lados e sugere que estes dois vetores pertencem à componente harmônica com período igual à 12. O diagrama de dispersão do par de componentes 5 e 6 não apresenta um comportamento definido, mas como foi visto na Figura 8,15, são vetores pertencentes à componente de tendência. Já o diagrama de dispersão do par de componentes 55 e 56 apresenta um comportamento caótico, indicando serem ruidosos os componentes 55 e 56.

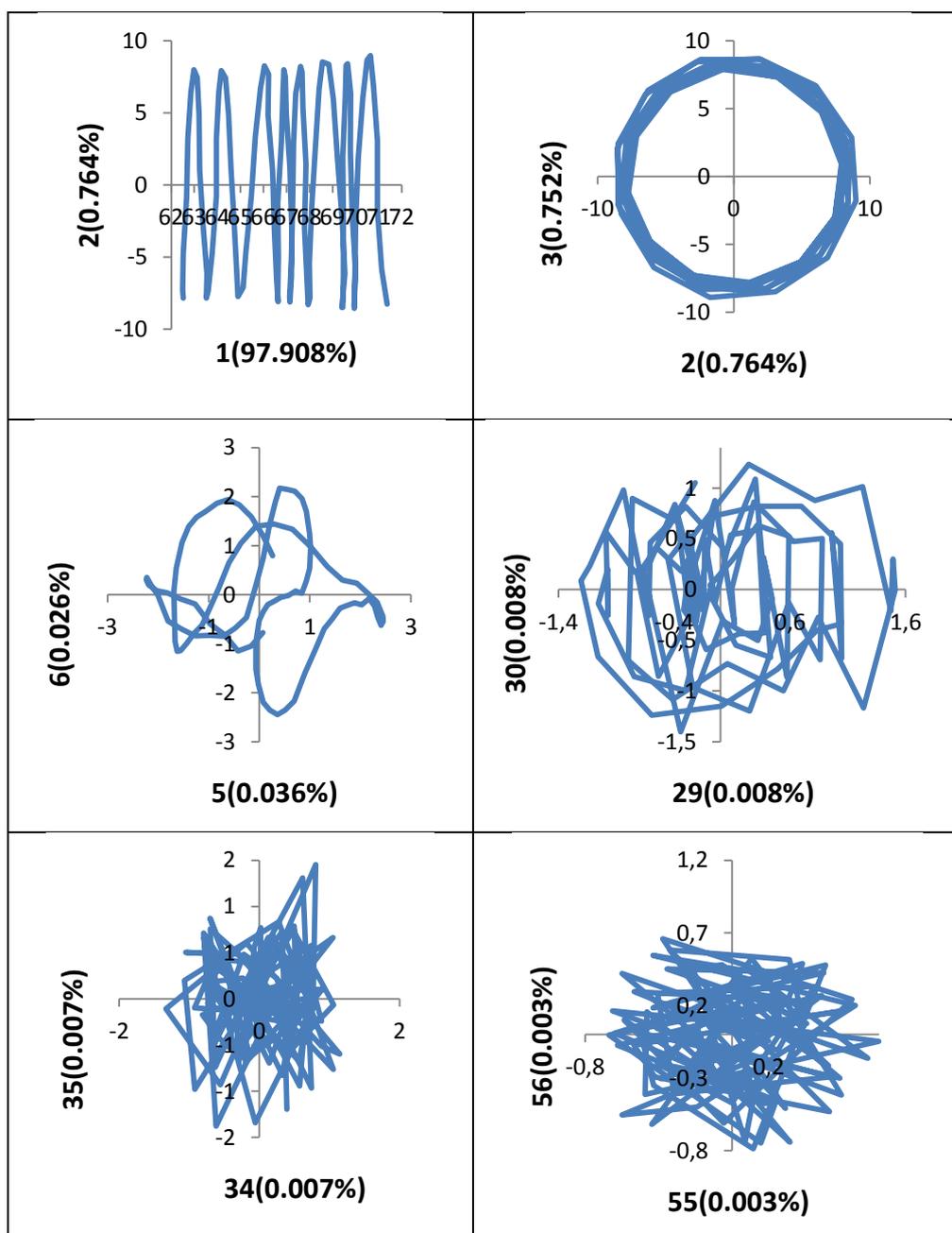


Figura 8.15. Alguns diagramas de dispersão

Outra forma de se definirem os vetores que compõem a componente harmônica pode ser feita através da análise de periodograma. A Figura 8.16 apresenta os periodogramas dos componentes 2 e 3 pertencentes à componente harmônica. Pode-se ver que há um pico no período 12, confirmando os indícios verificados na Figura 8.14.

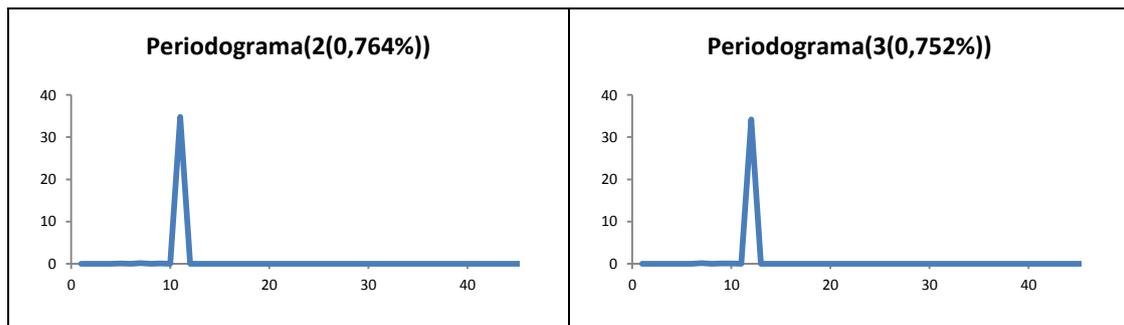


Figura 8.16. Periodograma dos vetores singulares 2 e 3.

Esta análise gráfica é feita para todos os 96 vetores singulares e então são definidas as componentes de tendência (com comportamento mais suave), harmônica (com comportamento senoidal, polígonos regulares nos diagrama de dispersão e picos nos periodogramas) e ruído (os demais). Ao final, as componentes são definidas pelos vetores apresentados na Tabela 8.3. As componentes definidas por estes vetores estão representadas graficamente na Figura 8.17.

Tabela 8.3. Definições das componentes pelo método da Verificação Visual.

Componente	Vetores
Tendência	1, 4 – 8, 10, 29, 41, 42
Harmônica	2, 3, 11 – 20, 33, 34
Ruído	9, 21 – 28, 30 – 32, 35 – 40, 43 – 96

O teste BDS foi utilizado para verificar a independência entre os dados na série de ruídos identificada na abordagem SSA com análise gráfica dos vetores singulares. A Tabela 8.4 mostra os resultados obtidos no teste BDS.

Tabela 8.4. Teste BDS aplicado à série de ruídos identificados por Verificação Visual.

Dim.	Estatística BDS	Erro Padrão	Estat. Z	<i>p</i> -valor
2	0,006858	0,002848	2,407909	0,1601
3	0,006505	0,002477	2,626266	0,0865
4	0,003256	0,001615	2,015849	0,4380
5	0,001361	0,000922	1,476115	0,1399
6	0,000420	0,000487	0,861956	0,3887

A partir do teste BDS pode-se observar que a hipótese nula de independência dos dados não é rejeitada ao nível de 5% de significância em todas as dimensões medidas. Isso confirma que a série identificada realmente possui estrutura de ruído. Estes resultados também confirmam que o comprimento de janela foi bem escolhido.

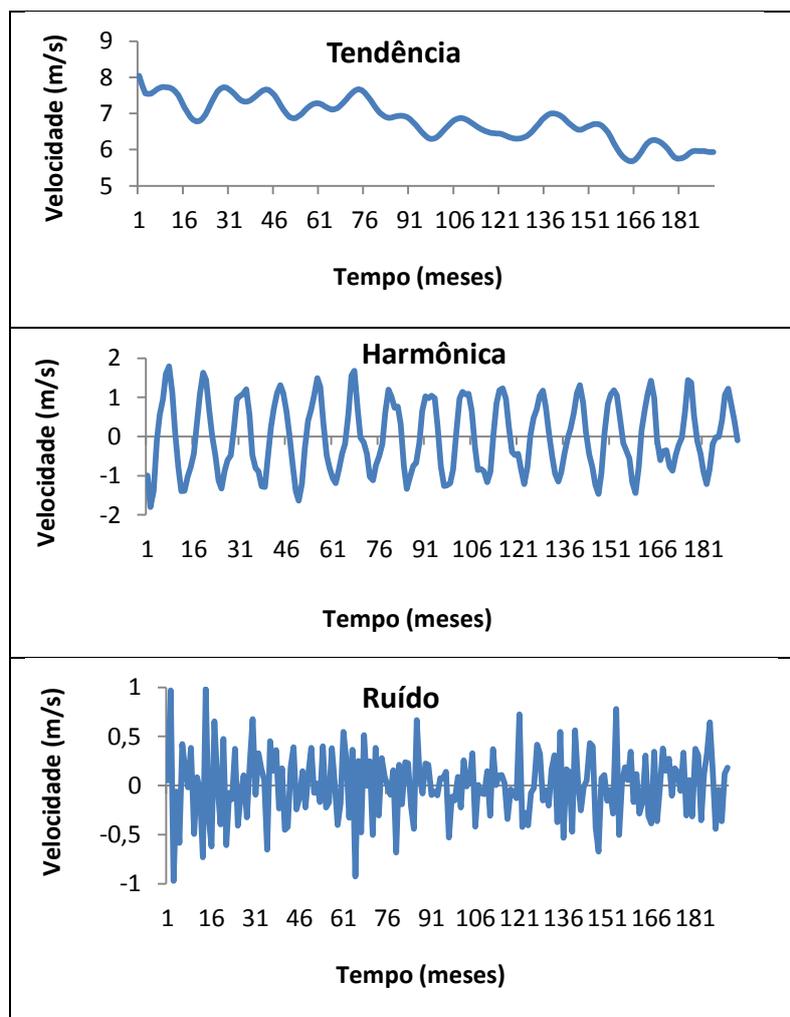


Figura 8.17. Componentes da metodologia SSA identificados por Verificação Visual do Comportamento dos Autovetores.

Uma avaliação de robustez do método SSA é a separabilidade. Para se verificar se as componentes estão bem separadas, uma boa medida é a correlação ponderada. A Tabela 8.5 mostra a correlação ponderada entre as componentes. Conforme pode ser verificado, as componentes são bem separadas. Então pode-se separar o sinal do ruído através da série aproximada menos ruidosa somando-se as componentes de tendência e harmônica.

Tabela 8.5. Correlação ponderada entre as componentes identificadas por Verificação Visual.

	Tendência	Harmônica	Ruído
Tendência	1	0,001	0,003
Harmônica	0,001	1	0,005
Ruído	0,003	0,005	1

8.2.2. Identificação do Ruído por Análise de Componentes Principais

Para a análise de componentes principais (ACP), utiliza-se o log dos autovetores da matriz XX^T . A contribuição dos autovetores a partir do autovetor 26 é menor que 0,01%. Assim, na SVD, as 25 primeiras matrizes passam a fazer parte da componente de sinal enquanto as demais fazem parte da componente ruidosa. A SVD na ACP passa a ser:

$$X = (E_1 + E_2 + \dots + E_{25}) + (E_{26} + E_{27} + \dots + E_{96}) = \text{Sinal} + \text{Ruido}$$

As Figuras 8.18 e 8.19 mostram as duas séries de sinal e ruído, respectivamente determinadas pelo agrupamento baseado em SVD.

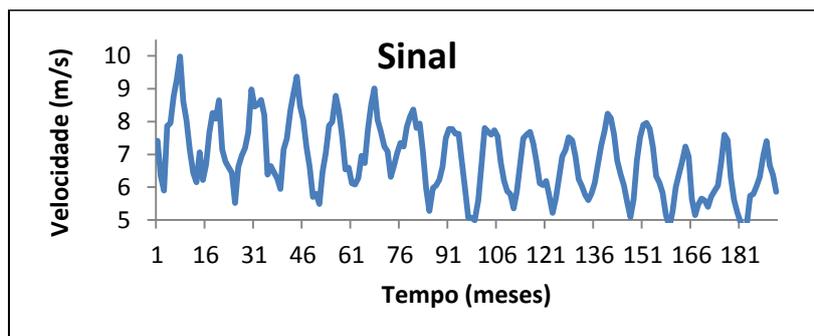


Figura 8.18. Série de Sinal identificado por SVD.

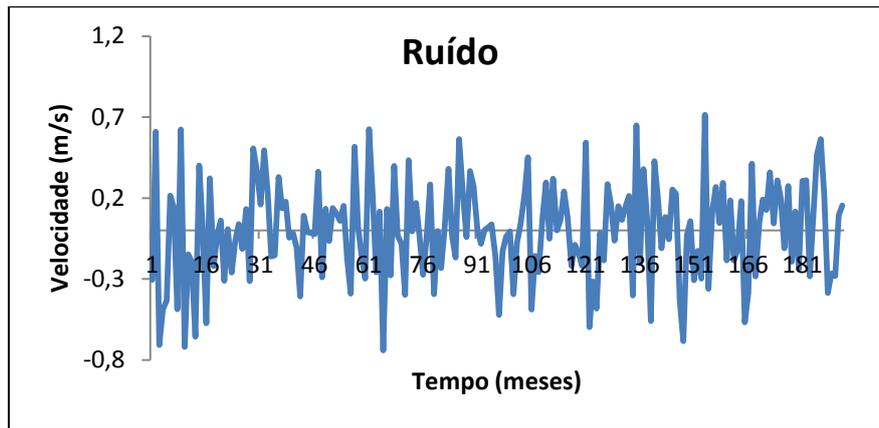


Figura 8.19. Série de ruído identificado por SVD.

Com o objetivo de verificar a separabilidade, a correlação ponderada foi calculada e encontrou-se correlação entre sinal e ruído igual a 0.011, sendo portanto descorrelacionados e bem separados. Isso significa que a probabilidade de haver sinal na série de ruído ou de haver ruído na série de sinal é muito pequena. Para verificar se os dados da série ruidosa são independentes, o teste BDS foi aplicado e os resultados mostrados na Tabela 8.6. A partir do teste BDS pode-se observar que a hipótese nula de independência dos dados não é rejeitada ao nível de 5% de significância em todas as dimensões medidas. Isso confirma que a série separada tem de fato característica de ruído.

Tabela 8.6. Teste BDS aplicado a série de ruídos identificados por SVD.

Dim.	Estatística BDS	Erro Padrão	Estat. Z	<i>p</i> -valor
2	0,007965	0,002785	2,860412	0,0524
3	0,005937	0,002385	2,489065	0,1280
4	0,003321	0,001532	2,167397	0,3021
5	0,001957	0,000862	2,270226	0,2323
6	0,000787	0,000449	1,752917	0,7968

8.2.3. Identificação do Ruído pelo Método de Clusterização Hierárquica

Usando a Clusterização Hierárquica para 3 clusters, teve-se a distribuição dos clusters como segue: as séries de 1 a 4 são sinais e as séries de 5 a 96 são ruídos. As Figuras 8.20 e 8.21 mostram as duas séries de sinal e ruído, respectivamente determinadas pelo agrupamento baseado em Clusterização Hierárquica.

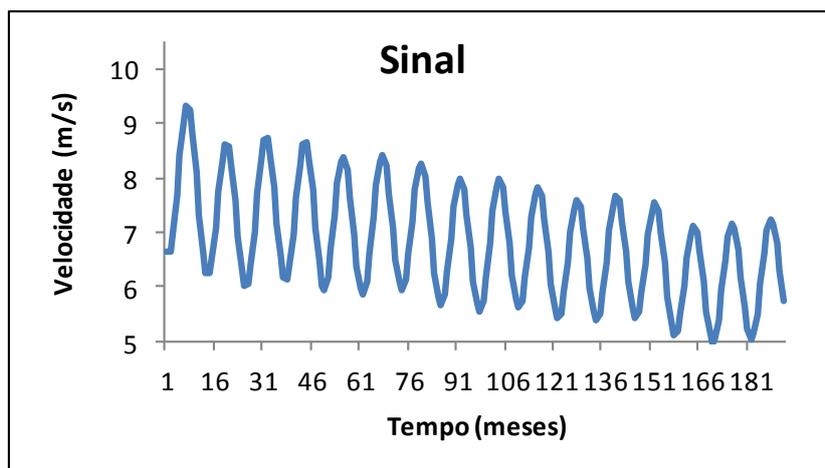


Figura 8.20. Série de Sinal Determinado por Clusterização Hierárquica.

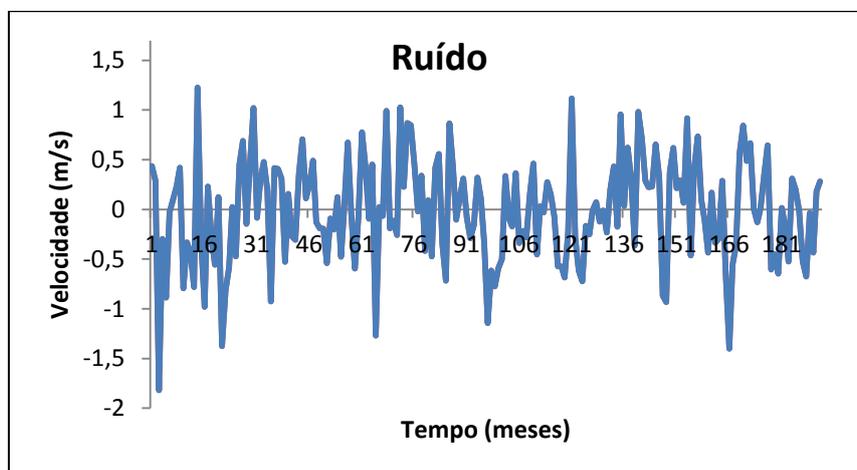


Figura 8.21. Série de Ruído Determinada por Clusterização Hierárquica.

Com o objetivo de verificar a separabilidade, a correlação ponderada foi calculada e encontrou-se correlação entre sinal e ruído igual a $-0,02784$ e as séries são classificadas como bem separadas. Para verificar se os dados da série de ruído são independentes, o teste BDS foi aplicado e os resultados mostrados na Tabela 8.7.

Tabela 8.7. Teste BDS aplicado à série de ruídos identificados pela Clusterização Hierárquica.

Dim.	Estatística BDS	Erro Padrão	Estat. Z	<i>p</i> -valor
2	0,002721	0,004739	0,574172	0,5659
3	0,006604	0,007536	0,876243	0,3809
4	0,010374	0,008978	1,155496	0,2479
5	0,013238	0,009359	1,414400	0,1572
6	0,012054	0,009027	1,335231	0,1818

A partir do teste BDS pode-se observar que a hipótese nula de independência dos dados não é rejeitada ao nível de 5% de significância em todas as dimensões medidas. Isso confirma que a série separada tem de fato característica de ruído.

8.2.4. Identificação do Ruído pelo Método de Clusterização DBSCAN

A Figura 8.23 abaixo traz o gráfico da Dist4 das L componentes de SVD da série de Petrolina, necessário para a estimação do parâmetro ε do algoritmo DBSCAN. ESTER et al. (2007) definem que o ponto ideal para inferir um valor de ε , é ordenada da curva no ponto onde é formado um “joelho” ou a ordenada de algum percentil, ou o ponto onde há a primeira mudança de tendência no Gráfico de Dist4. Aqui o ε é a ordenada do ponto de abscissa 80, igual a 0,5447. O algoritmo então busca formar clusters de no mínimo 4 de componentes conectadas por densidade, com $\varepsilon = 0,5447$.

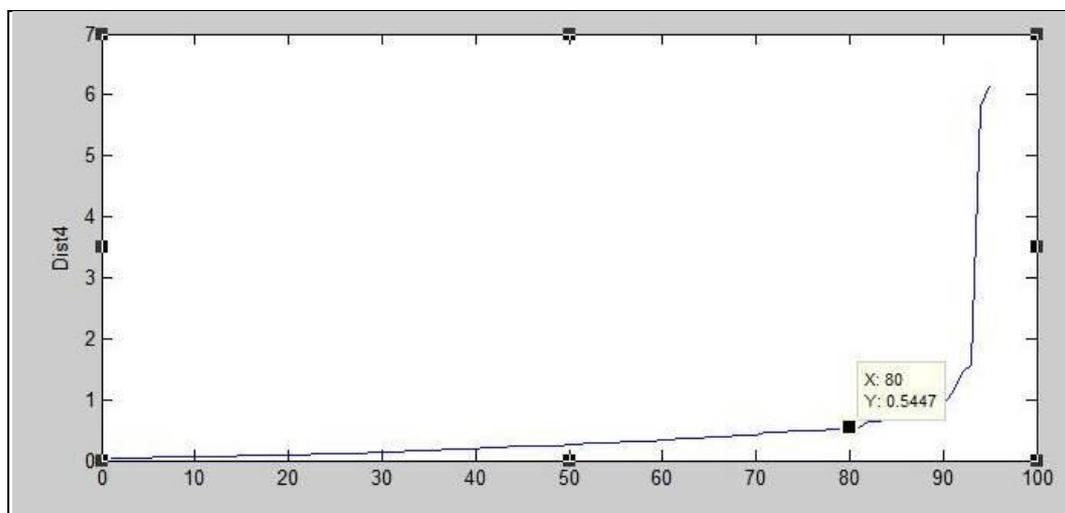


Figura 8.22: Dist4 das componentes de SVD da série de Petrolina para a estimação do parâmetro ε do algoritmo DBSCAN.

O DBSCAN identificou então 2 clusters e 1 cluster de ruído. A distribuição dos componentes de acordo como método DBSCAN é mostrada na Tabela 8.8.

Tabela 8.8. Definições das componentes pelo método DBSCAN.

Componente	Vetores
Sinal	1-9, 11-13,15-17,
Ruído	10, 14, 18-96

As Figuras 8.23 e 8.24 mostram as duas séries de sinal e ruído, respectivamente determinadas por DBSCAN.

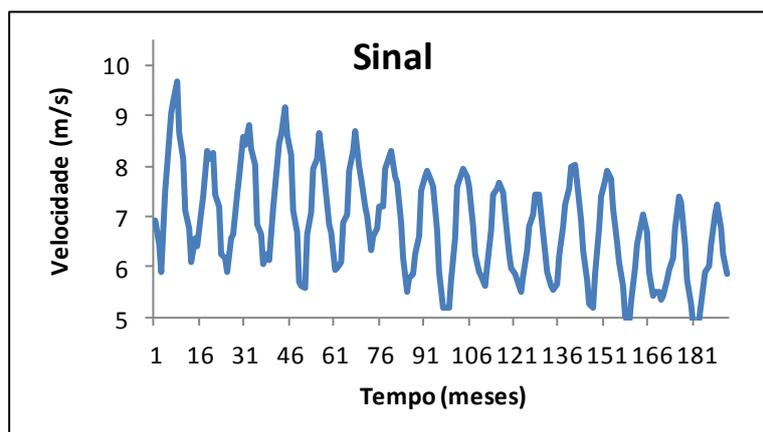


Figura 8.23 . Série de Sinal Identificada por DBSCAN.

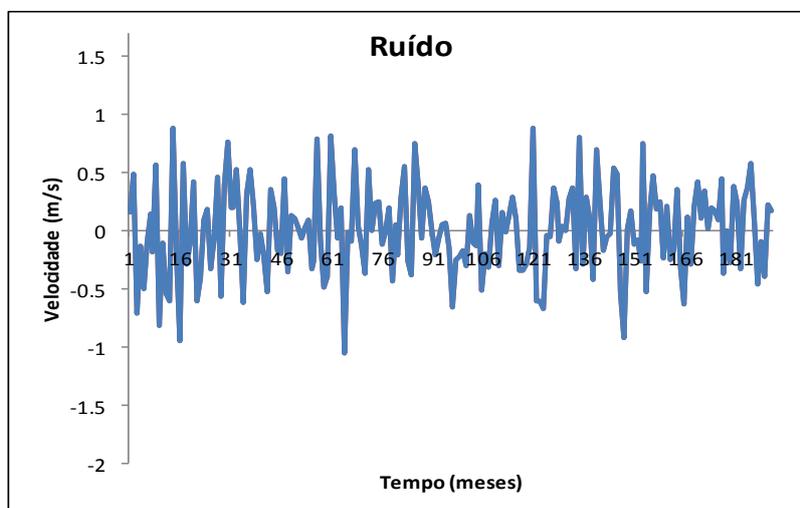


Figura 8.24. Série de Ruído Identificada por DBSCAN.

Com o objetivo de verificar a separabilidade, a correlação ponderada foi calculada e encontrou-se correlação entre sinal e ruído igual a 0.0105. O teste BDS foi aplicado à série de ruído e os resultados são mostrados na Tabela 8.9. A partir do teste BDS pode-se observar que a hipótese nula de independência dos dados não é rejeitada ao nível de 5% de significância em todas as dimensões medidas. Isso confirma que a série separada tem de fato característica de ruído.

Tabela 8.9. Teste BDS aplicado a série de ruídos identificados por DBSCAN.

Dim.	Estatística BDS	Erro Padrão	Estat. Z	<i>p</i> -valor
2	0,004015	0,004626	0,867841	0,3855
3	0,010032	0,007337	1,367328	0,1715
4	0,012137	0,008716	1,392483	0,1638
5	0,013730	0,009062	1,515178	0,1297
6	0,011371	0,008716	1,304594	0,1920

8.3. Previsão Mista: Filtro SSA + Previsão Clássica

Na abordagem de previsão mista considera-se a série aproximada, obtida por alguma abordagem de SSA, mas é feita a modelagem e previsão por Métodos Clássicas de previsão. Logo, para cada uma das séries aproximadas nas seções 8.2.1 a 8.2.4 é feita uma metodologia de identificação do modelo tal como feito para a série original na seção 8.1.

Para todas as séries aproximadas obtidas pelo filtro SSA, nas 4 distintas abordagens, o modelo identificado foi o mesmo modelo para a série original, um SARIMA $(1,0,0) \times (3,1,2)_{12}$. Este fato infere que a filtragem SSA não altera a estrutura de formação da série temporal original. Entretanto como as séries são distintas, os coeficientes, previsões e medidas de ajustes do modelo são também distintos. As seções seguintes trazem para cada abordagem, características do modelo ajustado à série aproximada após agrupamento SSA. As medidas de ajuste são omitidas porque elas não fazem sentido, são medidas de ajuste em relação à própria

série aproximada e deseja-se saber sobre quão bem estes modelos ajustam a série original. Por isso, as medidas de ajuste que interessam destes modelos, calculados com erros medidos com relação a série original, são calculados na seção 8.4.

8.3.1. Modelagem para a série aproximada obtida por verificação visual dos autovetores

Forecast Model for PetrolinaVisual				
ARIMA (1, 0, 0) * (3, 1, 2)				
Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.8653	0.0372	23.2429	1.0000
A[12]	-0.5352	0.1089	-4.9140	1.0000
A[24]	-0.7757	0.0596	-13.0085	1.0000
A[36]	-0.6407	0.0731	-8.7662	1.0000
B[12]	-0.1516	0.1173	-0.4396	0.9398
B[24]	-0.1561	0.0993	-1.5713	0.9839

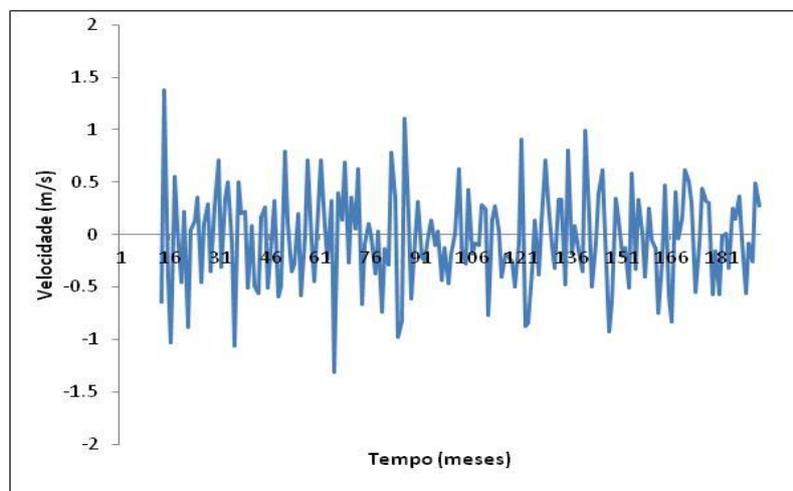


Figura 8.25: Resíduos do Modelo ARIMA para séries filtradas em SSA através da Verificação Visual dos Autovetores, calculados em relação à série original.

8.3.2. Modelagem para a série aproximada obtida por ACP

Forecast Model for PetrolinaACP				
ARIMA (1, 0, 0) * (3, 1, 2)				
Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.7284	0.0497	14.6635	1.0000

A[12]	-0.1165	0.1261	-2.9232	0.9441
A[24]	-0.7472	0.0508	-14.7171	1.0000
A[36]	-0.4119	0.0845	-4.8758	1.0000
B[12]	0.4950	0.1169	4.2330	1.0000
B[24]	-0.5152	0.0937	-5.4992	1.0000

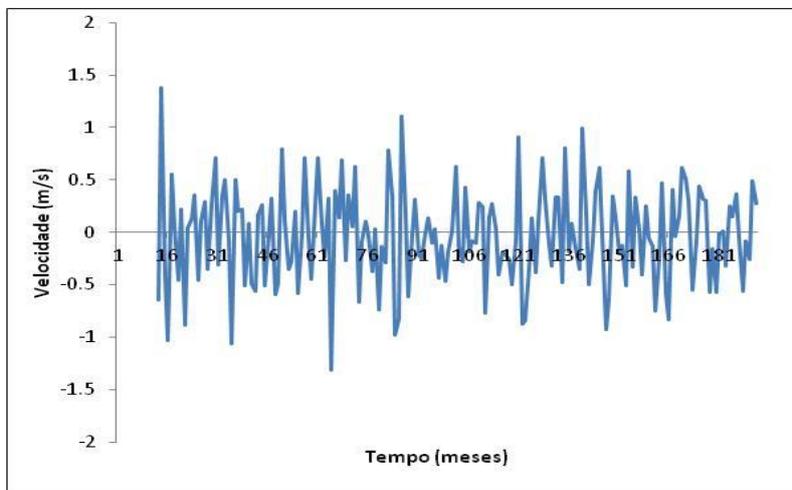


Figura 8.26: Resíduos do Modelo ARIMA para séries filtradas em SSA através da Análise de Componentes Principais, calculado em relação à série original.

8.3.3. Modelagem para a série aproximada obtida por Clusterização Hierárquica

Forecast Model for PetrolinaHierarquico				
ARIMA(1,0,0) * (3,1,2)				
Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.9997	0.0001	7946.1306	1.0000
A[12]	0.0731	0.0402	2.8174	0.9308
A[24]	-0.0626	0.0395	-3.5851	0.9871
A[36]	0.9249	0.0177	52.2046	1.0000
B[12]	-0.1291	0.0642	-2.0117	0.9558
B[24]	0.4882	0.0627	7.7803	1.0000

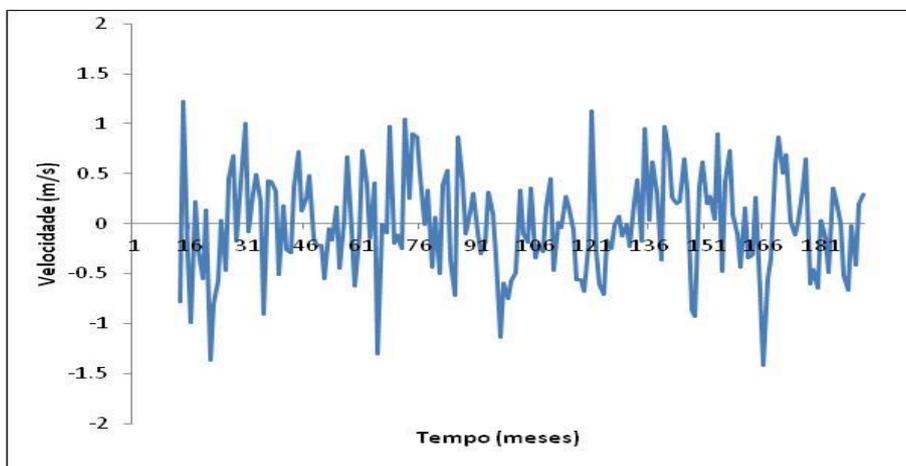


Figura 8.27: Resíduos do Modelo ARIMA para séries filtradas em SSA através da Clusterização Hierárquica, calculados em relação à série original.

8.3.4. Modelagem para a série aproximada obtida por Clusterização DBSCAN

Forecast Model for PetrolinaDBSCAN				
ARIMA (1,0,0) * (3,1,2)				
Term	Coefficient	Std. Error	t-Statistic	Significance
a[1]	0.9325	0.0270	34.5042	1.0000
A[12]	0.2550	0.2635	2.9679	0.8669
A[24]	-0.9510	0.0365	-26.0372	1.0000
A[36]	0.1433	0.2384	2.6012	0.9523
B[12]	0.2159	0.2544	2.8487	0.9040
B[24]	-0.3071	0.0768	-3.9981	0.9999

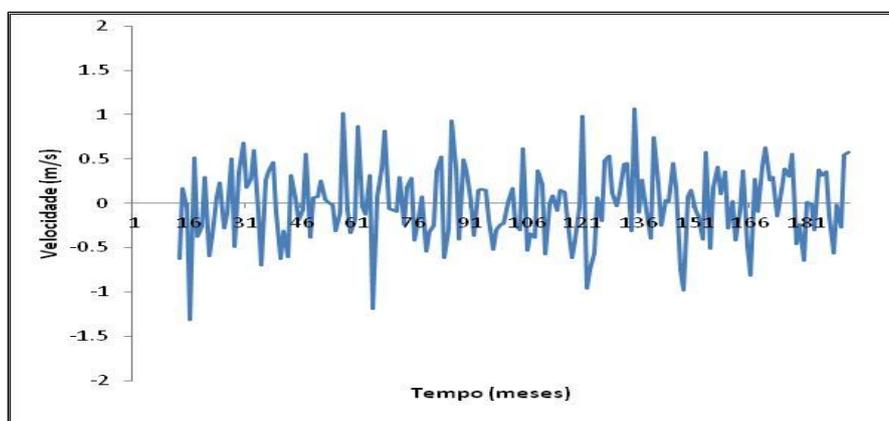


Figura 8.28: Resíduos do Modelo ARIMA para séries filtradas em SSA através do DBSCAN, calculados em relação à série original.

8.4. Previsão SSA: Filtro SSA + Previsão SSA

Na abordagem de previsão SSA considera-se a série aproximada, obtida por alguma abordagem de SSA e faz-se a modelagem também por SSA, descrita na seção 3.6. Os gráficos exibidos nas Figuras de 8.29 a 8.32 mostram as séries de resíduos das previsões SSA em cada uma das abordagens. As principais medidas de erros que avaliam a acurácia de previsão de cada uma das abordagens podem ser vistas na Tabela 8.5.

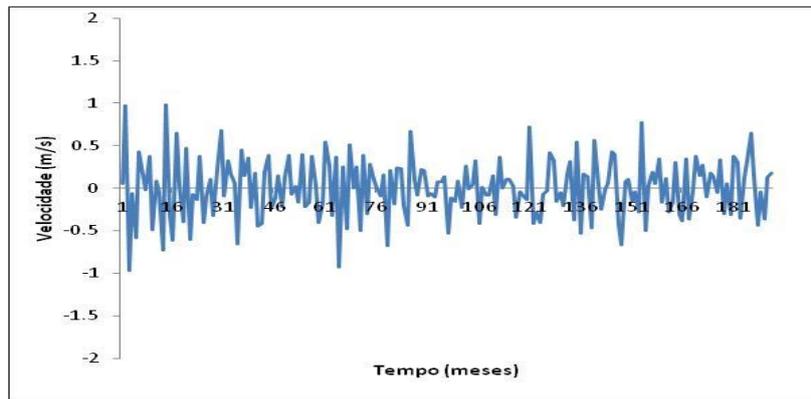


Figura 8.29: Resíduos da Previsão SSA para séries filtradas através da Verificação Visual do Comportamento dos Autovetores;

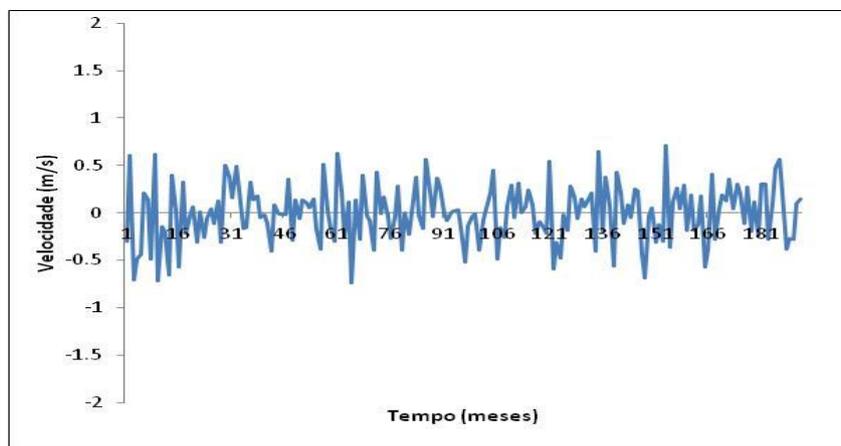


Figura 8.30: Resíduos da Previsão SSA para séries filtradas através da Análise de Componentes Principais.

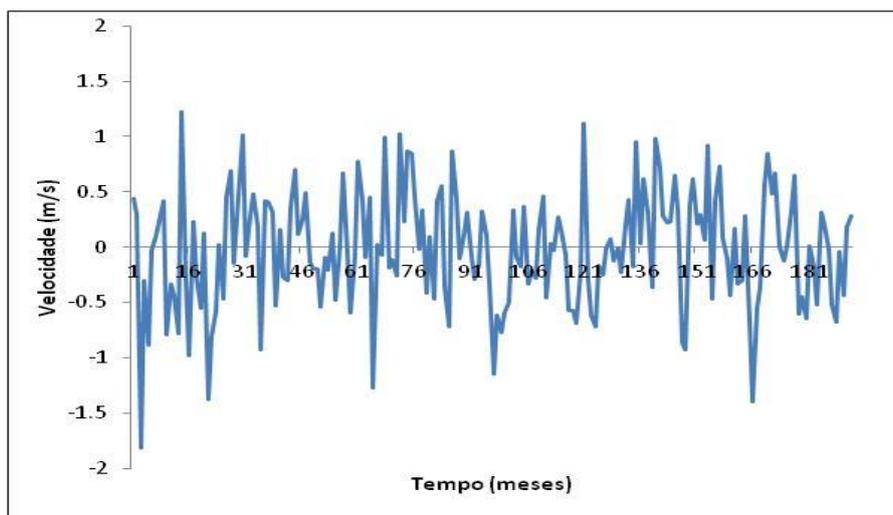


Figura 8.31: Resíduos da Previsão SSA para séries filtradas através da Clusterização Hierárquica.

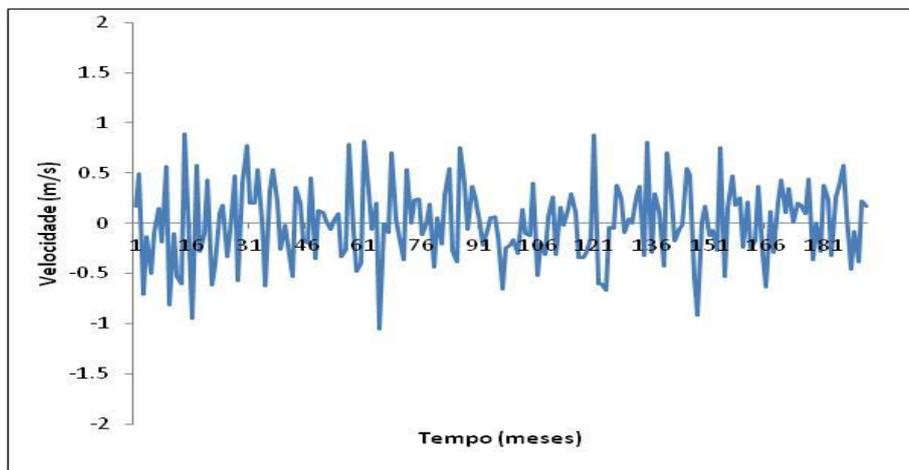


Figura 8.32: Resíduos da Previsão SSA para séries filtradas através do DBSCAN.

8.5. Comparação dos Resultados da Previsão Clássica, Previsão Mista e Previsão SSA

Cada abordagem usada na separação de ruído na fase 3 de SSA reconheceu seus particulares componentes de ruído. Todos os métodos concordam que as componentes de ordem 26-28; 30-32; 35-40 e 43-96 são componentes ruidosas, e as componentes de 1 a 3 referem-se a Sinal. O diferencial entre os métodos está na forma com que eles qualificam as componentes de ordem 4 a 26.

A Tabela 8.10 a seguir traz os valores das estatísticas de ajuste das previsões (*in sample*) de cada modelo proposto à série original de velocidade do vento de Petrolina.

Tabela 8.10: Medidas de qualidade de ajuste das previsões *in sample* da série de Petrolina por todos os métodos experimentados.

Método de Filtragem SSA	Método de Previsão	MAPE	RMSE	RMSE Normalizado	R^2 (%)
Nenhum	Clássica	0,07051	0,5764	0,5375	71,11
Análise Visual dos Autovetores	Clássica	0,0403	0,3395	0,3300	87,02
	SSA	0,0400	0,3313	0,3082	90,5
Análise de Componentes Principais	Clássica	0,0602	0,4980	0,4350	78,46
	SSA	0,0461	0,2996	0,2705	84,64

Clusterização Hierárquica	Clássica	0,0523	0,5209	0,473	77,87
	SSA	0,0522	0,5077	0,470	78,61
Clusterização DBSCAN	Clássica	0,0413	0,4205	0,392	90,37
	SSA	0,0301	0,2766	0,341	92,67

De acordo com os valores exibidos na Tabela 8.10, pode-se verificar que:

- a) O pior ajuste, de acordo com todas as medidas, ocorreu na situação em que o tratamento de séries temporais é feito por abordagem clássica, sem SSA, tal como foi verificado nas Simulações com modelos clássicos, na Seção 7.
- b) O melhor ajuste ocorreu com a combinação que é a proposta da tese: uso do DBSCAN+ SSA usando também SSA na previsão.
- c) O segundo melhor ajuste advém do método de Análise Visual dos Autovetores com Previsão SSA. E o terceiro ajuste advém do método que usa DBSCAN na Separação SSA só para filtrar a série e faz Modelagem com um Método de Previsão Clássica. Entretanto o método que usa Análise Visual dos Autovetores tem uma grande desvantagem: a necessidade de um conhecimento especialista do pesquisador, que por natureza não será exato; pesquisadores diferentes podem chegar a diferentes classificações de ruídos e sinal por este método.
- d) Para as quatro abordagens SSA, a previsão com SSA supera o método parcial onde é feita somente a filtragem SSA e a previsão é feita por modelos de Box e Jenkins. O uso da previsão SSA melhora principalmente o método que usa o agrupamento das autotriplas por Análise de Componentes Principais.
- e) Entre as quatro abordagens propostas pra filtrar ruído em SSA, a menos recomendada é a que usa a Clusterização Hierárquica, pois apresenta as menores medidas de ajuste. De fato, o método hierárquico fez uma atribuição muito generalizada ao cluster de ruído, uma vez que considerou somente 4 séries de sinal e todas as outras 92 séries da decomposição SSA

foram consideradas como ruído. Possivelmente foram perdidas informações importantes para modelagem da série equivocadamente classificadas como ruído pelo método, como prevê a literatura, já que o método não é especialista em classificar ruído, como é o DBSCAN.

As Tabelas 8.11 a seguir traz os valores das estatísticas de ajuste das previsões *out of sample* 24 passos a frente, de cada modelo proposto à série original de velocidade do vento de Petrolina. Para previsões *out of sample*, o modelo que utiliza DBSCAN+ SSA e SSA na previsão também apresenta melhor desempenho em relação aos demais modelos propostos.

Tabela 8.11: Medidas de qualidade de ajuste das previsões 24 passos a frente *out of sample* da série de Petrolina por todos os métodos experimentados.

Método de Filtragem SSA	Método de Previsão	MAPE	MADE	GMRAE
Nenhum	Clássica	0,069	0,414	0,895
Análise Visual dos Autovetores	Clássica	0,039	0,248	0,637
	SSA	0,039	0,224	0,650
Análise de Componentes Principais	Clássica	0,059	0,335	0,780
	SSA	0,045	0,235	0,624
Clusterização Hierárquica	Clássica	0,049	0,399	0,936
	SSA	0,047	0,359	0,739
Clusterização DBSCAN	Clássica	0,035	0,323	0,665
	SSA	0,027	0,180	0,373

Conclui-se, portanto que o uso do DBSCAN para separar as componentes de Ruído na fase de Agrupamento SSA e Previsão SSA é o melhor método para gerar previsões de velocidade do vento na estação de Petrolina. Sendo assim, tem-se uma proposta especialista em identificação de ruído e um método de previsão que pouco depende de palpite subjetivo do pesquisador (como é necessário na identificação de modelos ARIMA e na identificação das séries de ruído por análise visual do

comportamento dos autovetores). As previsões do método escolhido DBSCAN em SSA com previsão SSA são listadas na Tabela 8.12 a seguir, estimadas pelo software Cartepillar segundo expressão (38) da seção 3.6.

Tabela 8.12. Previsões de velocidade do vento da estação de Petrolina (em m/s) 24 passos a frente pelo método de Previsão SSA, usando DBSCAN na identificação de ruído em SSA.

Mês	Previsão de Velocidade do Vento	Mês	Previsão de Velocidade do Vento
jan/12	4,831	jan/13	5,174
fev/12	4,128	fev/13	4,728
mar/12	4,215	mar/13	4,863
abr/12	4,850	abr/13	5,350
mai/12	5,418	mai/13	5,811
jun/12	5,873	jun/13	6,372
jul/12	6,434	jul/13	7,084
ago/12	6,773	ago/13	7,383
set/12	6,538	set/13	6,959
out/12	6,092	out/13	6,330
nov/12	5,896	nov/13	6,006
dez/12	5,695	dez/13	5,773