

## 5 Modelagem

O modelo a ser proposto deve levar em consideração as características mais marcantes da série, de maneira a conseguir captar o fenômeno em estudo. Esse ainda deve ser coerente e parcimonioso.

Recentemente, diversas técnicas vem sendo desenvolvidas para prever a velocidade de vento no curto prazo. Essas técnicas tem sido baseadas em métodos estatísticos e computacionais. A título de ilustração podemos citar: Haslett e Raftery (1989) que utilizaram modelos ARFIMA e a correlação espacial entre estações meteorológicas para fazer previsões sobre velocidade de vento. Alexiadis et al (1999) que propuseram um modelo de redes neurais artificiais que considera além das velocidades de vento a correlação espacial entre diferentes estações de modo a melhorar suas previsões. Genton et al (2006) que apresentaram modelos que levam em consideração diferentes estações meteorológicas, a direção que o vento sopra, além do comportamento diário do vento. Taylor et al (2009) que introduz conjuntos de modelos atmosféricos (ensemble) para ajudar a fazer previsões. Diversos outros métodos tem sido empregados dado a relevância do assunto.

O primeiro modelo proposto nessa dissertação levaria em conta mais de uma estação anemométrica (São João do Cariri - PB, Triunfo - PE, Petrolina - PE, Belo Jardim - PE) para tentar utilizar a correlação espacial entre as estações de maneira que se obtivesse melhores previsões. Tal modelo se tornou inviável dado a baixa correlação e a falta de dados confiáveis entre as estações disponíveis no projeto SONDA.

A alternativa proposta e escolhida foi utilizar um segundo modelo. Este utilizaria dados apenas da velocidade de vento dentro de uma única estação anemométrica. A estação escolhida foi a de São João do Cariri - PB por apresentar mais regularidade de dados, além da existência de trabalhos prévios com a série temporal de ventos na estação, permitindo assim a comparação de resultados.

O modelo proposto aplicado a série temporal escolhida leva em consideração todos os fatores explicitados na análise descritiva de dados. Além disso, considera uma característica importante em séries de vento que é o comportamento de memória longa. Recentemente, Beblo e Schmid (2010) obtiveram bons resultados na previsão de velocidade de vento na estação anemométrica de Manschow no leste da Alemanha. Nessa dissertação será utilizado o mesmo tipo de modelagem aplicado pelos autores acima citados.

Logo, o modelo que representará a velocidade de vento na estação anemométrica de São João do Cariri-PB é o seguinte:

$$Z_t = \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{24}\right) + \beta_2 \sin\left(\frac{2\pi t}{24}\right) + \beta_3 \cos\left(\frac{4\pi t}{24}\right) + \beta_4 \sin\left(\frac{4\pi t}{24}\right) + \beta_5 I_{\{jan\}}(t) + \dots + B_{15} I_{\{nov\}}(t) + \epsilon_t \quad (5-1)$$

Onde:

$Z_t$  é a velocidade de vento horária em metros por segundo no instante  $t$

$I_{\{mês\}}(t)$  é a variável indicadora do mês no instante  $t$

$\epsilon_t$  é um processo ARFIMA  $(p, d, q)$

Portanto, temos que  $\epsilon_t$  pode ser definido de maneira resumida como:

$$\phi(B) \nabla^d \epsilon_t = \theta(B) a_t \quad (5-2)$$

Onde:

$a_t \sim iid(0, \sigma^2)$ ,

$d \in (0; 0,5)$  e

$B$  é o operador retardo definido como  $B\epsilon_t = \epsilon_{t-1} \dots$

Temos ainda que

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \end{aligned} \quad (5-3)$$

e  $\nabla^d$  pode ser definido como:

$$(1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j \quad (5-4)$$

Onde:

$$\pi_j = 1, \quad \text{se } j = 0$$

$$\pi_j = \prod_{k=1}^{k=j} \frac{k-1-d}{k}, \quad \text{se } j = 1, 2, \dots$$
(5-5)

O modelo é basicamente uma abordagem de série Fourier onde o padrão sazonal é modelado usando termos de Fourier (pares de harmônicas) e a parte de curto e longo prazo é explicada estocasticamente dentro da parte de erro,  $\epsilon_t$ , do modelo. Logo, o mesmo contém uma parte determinística e outra estocástica. O modelo pode ser reescrito de forma geral como:

$$Z_t = f(t) + y_t$$
(5-6)

Onde:

$Z_t$  é a série observada

$f(t)$  é a componente determinística, neste caso, expressa por harmônicas e variáveis indicadoras

$y_t$  é a componente estocástica, neste caso, expressa por um modelo de memória longa

A equação de previsão  $k$  passos à frente pode ser escrita da seguinte maneira:

$$Z_t(k) = f(t+k) + y_t(k)$$
(5-7)

Utilizar modelos dessa maneira traz uma série de vantagens práticas, dentre elas cabe destacar:

- Permite incluir diferentes frequências de Fourier, sendo ideal para séries com mais de um período sazonal
- Permite modelar longos períodos sazonais, mais de um ano por exemplo, sem utilizar um número excessivo de parâmetros
- Os padrões de curto e longo prazo presentes na série podem ser tratados no termo de erro,  $\epsilon_t$ , por algum tipo de modelagem autoregressiva (ARMA, ARFIMA etc)

Assim como em Beblo e Schmid (2010), a escolha de modelar a parte de erro usando um modelo ARFIMA se deu pela forte autocorrelação positiva e pelo fato de séries de vento apresentarem comportamento compatível com memória longa, ver Kavasseri e Seetharaman (2009) e Haslett e Raftery (1989).

O modelo proposto, excetuando a parte residual  $\{\epsilon_t\}$  (ARFIMA), é estimado via mínimos quadrados. Essa estimação permite que sejam coletados os resíduos, diferença entre o estimado e o observado, até o tempo  $t$ .

A próxima etapa do processo é a estimação de todos os parâmetros do modelo ARFIMA. Após a captação dos resíduos, estes são utilizados de maneira separada pesquisando entre diversos modelos ARFIMA aquele que apresenta o menor AIC (Akaike's Information Criterion) e BIC (Bayesian Information Criterion). Para tal, foram considerados 50 dias de dados de resíduos. O motivo da escolha de se utilizar 50 dias advém do fato de Gneiting et al (2006) terem advertido que períodos entre 30 e 60 dias de treino apresentam melhores previsões em modelos autoregressivos.

Tendo todos os parâmetros estimados, podemos fazer as previsões da parte residual. Esses valores previstos são incorporados aos componentes diários e mensais já previstos no modelo via mínimos quadrados, resultando então na previsão final.

Yajima (1991) demonstra importantes propriedades assintóticas (consistência e eficiência) dos estimadores de mínimos quadrados em regressões com erros estacionários e de memória longa.

## 5.1

### Aplicação do modelo

O modelo foi testado com diferentes configurações, variando o número de pares de harmônicas e verificando a significância de cada parâmetro incluído. Nessa etapa foi concluído que os parâmetros estimados referentes as variáveis indicadoras dos meses de Janeiro e Setembro são não significativos. Portanto o modelo de regressão harmônica final encontrado foi o seguinte:

$$\begin{aligned} \hat{Z}_t = & 6,156 + 0,909\cos\left(\frac{2\pi t}{24}\right) - 1,227\sen\left(\frac{2\pi t}{24}\right) + 0,789\cos\left(\frac{4\pi t}{24}\right) + \\ & + 0,123\sen\left(\frac{4\pi t}{24}\right) - 1,340I_{fev}(t) - 1,676I_{mar}(t) - \\ & - 2,386I_{abr}(t) - 1,752I_{mai}(t) - 1,502I_{jun}(t) - 1,361I_{jul}(t) - \\ & - 0,601I_{ago}(t) + 0,230I_{out}(t) + 0,180I_{nov}(t) \end{aligned} \tag{5-8}$$

Coefficientes	Estimativa	Erro padrão	Estatística $t$	$\Pr(>  t )$
Intercepto	6,15595	0,03273	188,075	$< 2^{-16}$
$\cos(2\pi t/24)$	0,90867	0,02324	39,1000	$< 2^{-16}$
$\sen(2\pi t/24)$	-1,22712	0,02324	-52,8030	$< 2^{-16}$
$\cos(4\pi t/24)$	0,78893	0,02324	33,9480	$< 2^{-16}$
$\sen(4\pi t/24)$	0,12311	0,02324	5,29800	$1, 2^{-7}$
fevereiro	-1,34041	0,06776	-19,7820	$< 2^{-16}$
março	-1,67621	0,06520	-25,7090	$< 2^{-16}$
abril	-2,38564	0,06601	-36,1430	$< 2^{-16}$
maio	-1,75254	0,06520	-26,8800	$< 2^{-16}$
junho	-1,50160	0,06601	-22,7490	$< 2^{-16}$
julho	-1,36137	0,06520	-20,8800	$< 2^{-16}$
agosto	-0,60119	0,06520	-9,22100	$< 2^{-16}$
outubro	0,23027	0,06520	3,53200	$< 2^{-16}$
novembro	0,18049	0,06601	2,73400	$< 2^{-16}$

Tabela 5.1: Coeficientes estimados pelo modelo

A tabela 5.2 abaixo apresenta as estatísticas de resumo da série de resíduos gerado pelo modelo harmônico e a figura 5.1 abaixo apresenta os resíduos.

Medidas de Resumo	
Mínimo	-5,55
Máximo	6,581
Primeiro Quartil	-1,004
Terceiro Quartil	1,016
Média	0
Mediana	0,031
Variância	2,362
Desvio Padrão	1,537

Tabela 5.2: Resumo dos Resíduos

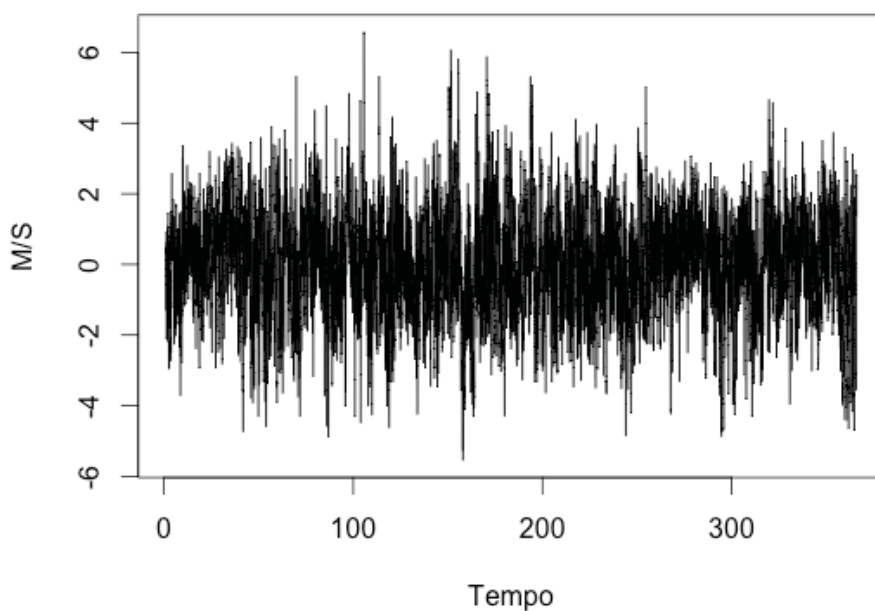


Figura 5.1: Resíduos

Os gráficos da função de autocorrelação e autocorrelação parcial dos resíduos mostram que esses possuem uma forte estrutura de correlação. Além disso, o teste de Ljung-Box rejeitou a hipótese  $H_0$  ( $p\text{valor} < 2 \cdot 2^{-16}$ ) de que os dados são independentes, ver figuras 5.2 e 5.3.

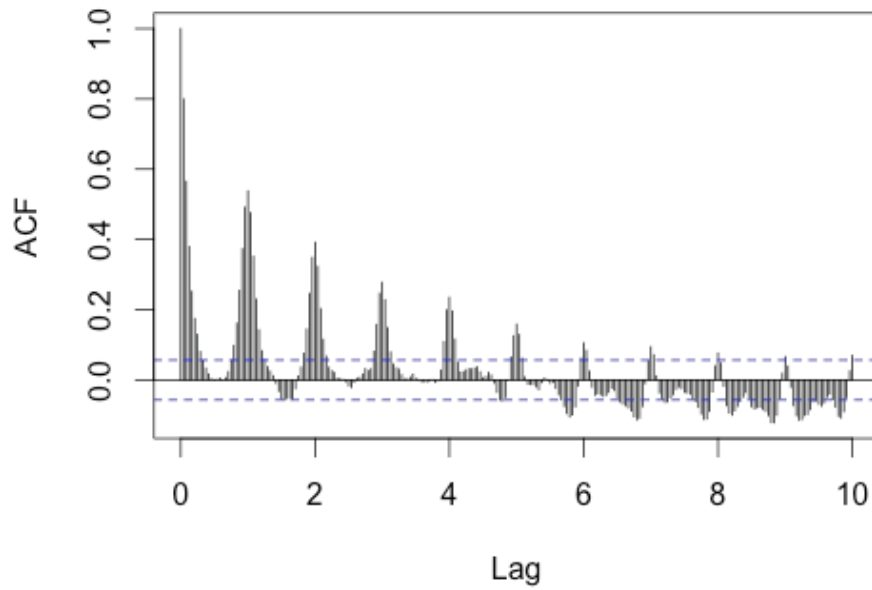


Figura 5.2: Função de autocorrelação dos resíduos

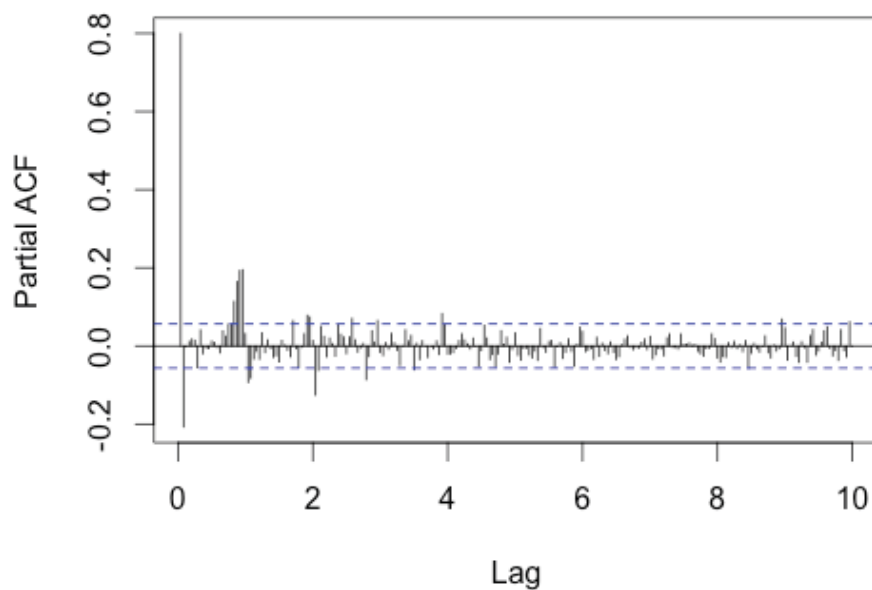


Figura 5.3: Função de autocorrelação parcial dos resíduos

Os gráficos, figuras 5.4 e 5.5, mostram que os dados não se afastam muito da distribuição normal. O gráfico do histograma apresenta formato semelhante ao de sino, característico da distribuição gaussiana, e no gráfico dos quantis teóricos da distribuição normal contra os quantis da série de resíduos (QQ-Plot) é possível perceber que temos quase uma reta bissetriz.

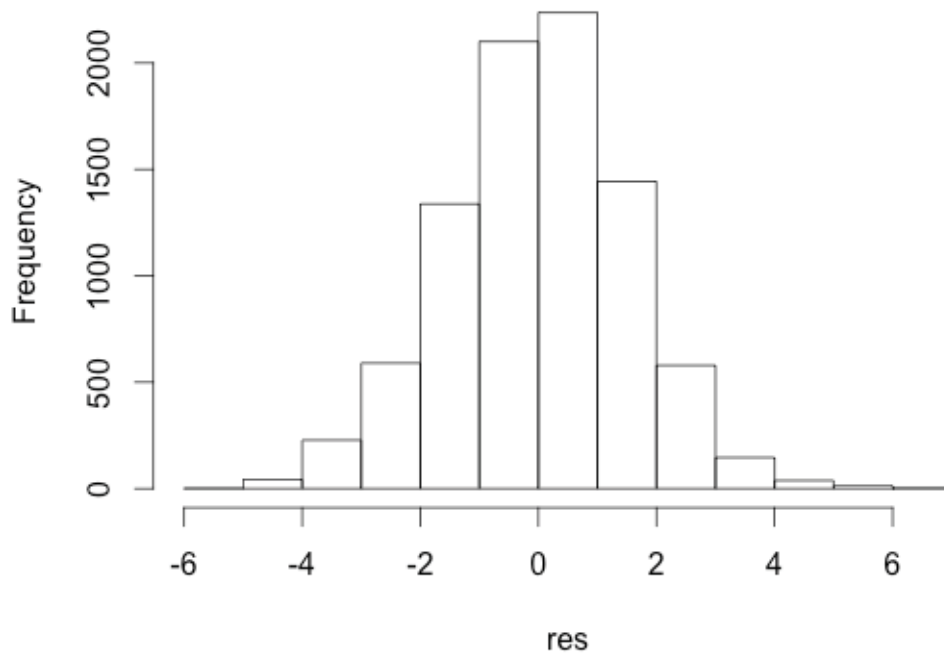


Figura 5.4: Histograma dos resíduos

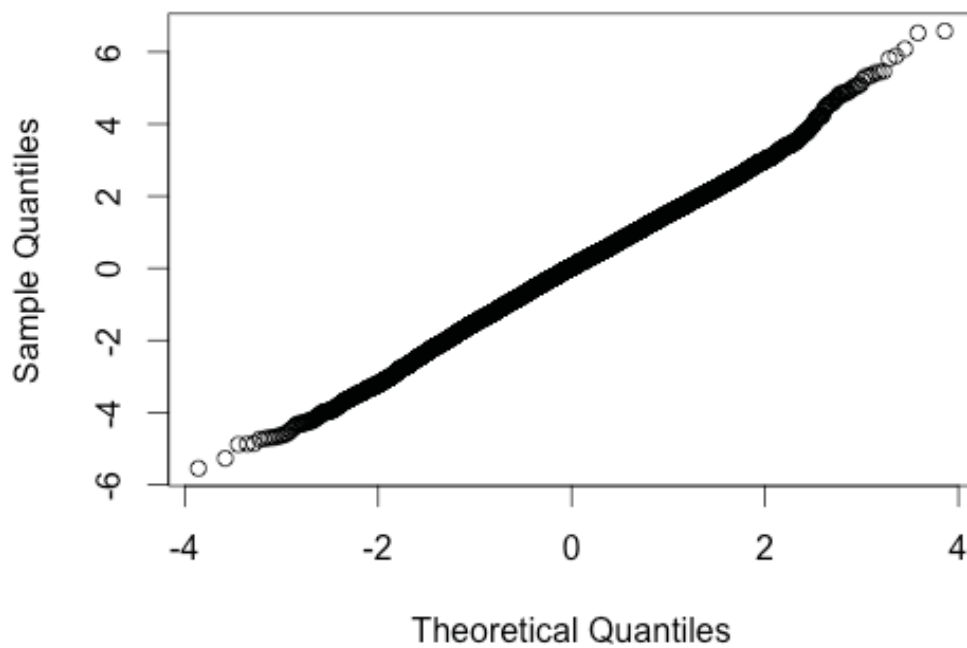


Figura 5.5: Q-Q Plot dos resíduos

Logo, pelos motivos já citados acima, os resíduos foram modelados via ARFIMA. A seleção do modelo  $ARFIMA(p,d,q)$  foi feita utilizando o algoritmo de Hyndman e Khandakar (2008) para selecionar os parâmetros  $p$  e  $q$  e o



algoritmo de Haslett e Raftery (1989) para estimar todos os parâmetros  $(p, d, q)$ .

O modelo ARFIMA, neste estudo, foi reestimado a cada nova informação trazida ao modelo. Isto significa que se, por exemplo, estivermos interessados em fazer previsões um passo à frente, a cada nova observação o resíduo obtido pelo modelo harmônico é incorporado a série. É importante ressaltar que o fato de reestimar o modelo ARFIMA a cada nova informação não traz nenhuma perda de interpretação já que é a parte residual que está sendo estimada.

A previsão final é obtida com a incorporação da previsão da parte residual, no horizonte de previsão de interesse, aos componentes diários e mensais representados pela regressão harmônica.

Considerando a previsão um passo à frente, a figura 5.6 apresenta os resultados para o modelo final (modelo harmônico incorporado com os termos de erro previstos pelo modelo ARFIMA), onde as observações previstas são as 168 horas da primeira semana do ano de 2007:

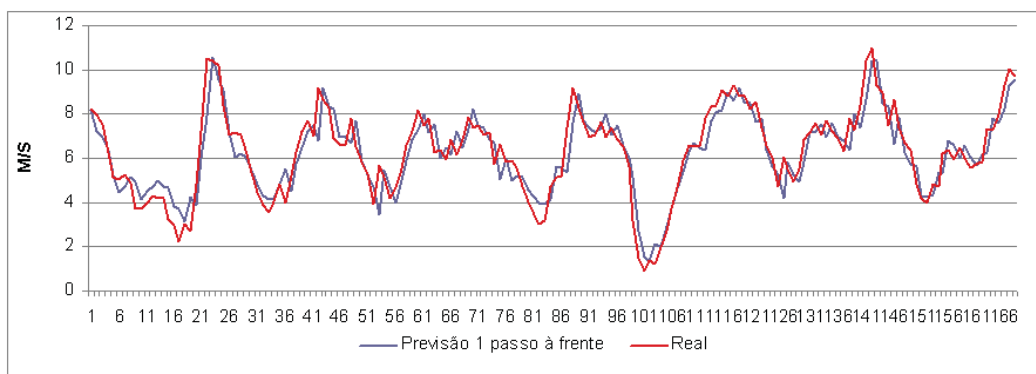


Figura 5.6: Previsão 1 passo à frente e valores observados

A análise do gráfico da previsão gerada pelo modelo e das observações no período em estudo, mostra que os valores previstos estão bem próximos dos valores verdadeiros, evidenciando que para a primeira semana do ano de 2007 o modelo foi capaz de prever com relativa habilidade utilizando previsões um passo à frente. Além disso, a tabela 5.3 abaixo mostra que os valores previstos não apresentam grandes desvios dos valores verdadeiros.

Métricas	Resultados
RMSE	0,807
MAE	0,630
MAPE	12,05%

Tabela 5.3: Métricas de comparação

O modelo proposto foi ainda testado utilizando previsões dois, três, quatro, cinco e seis passos (horas) à frente, também na primeira semana do ano de 2007. Os resultados podem ser vistos nas figuras 5.7, 5.8, 5.9, 5.10e 5.11 abaixo.

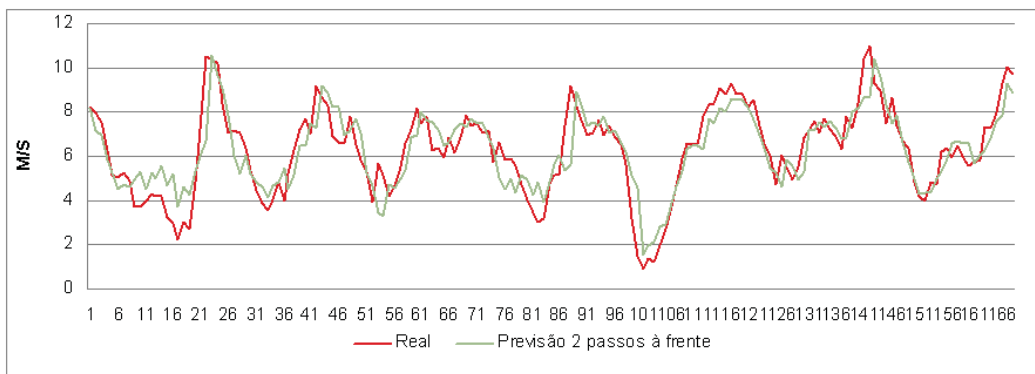


Figura 5.7: Previsão 2 passos à frente e valores observados

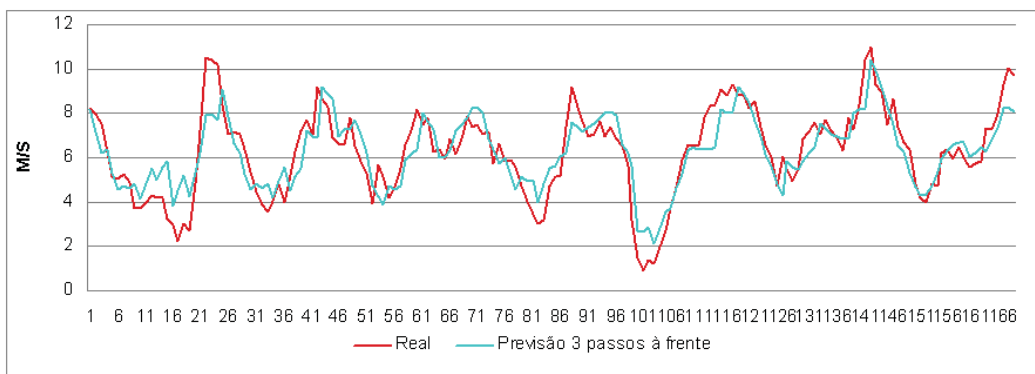


Figura 5.8: Previsão 3 passos à frente e valores observados

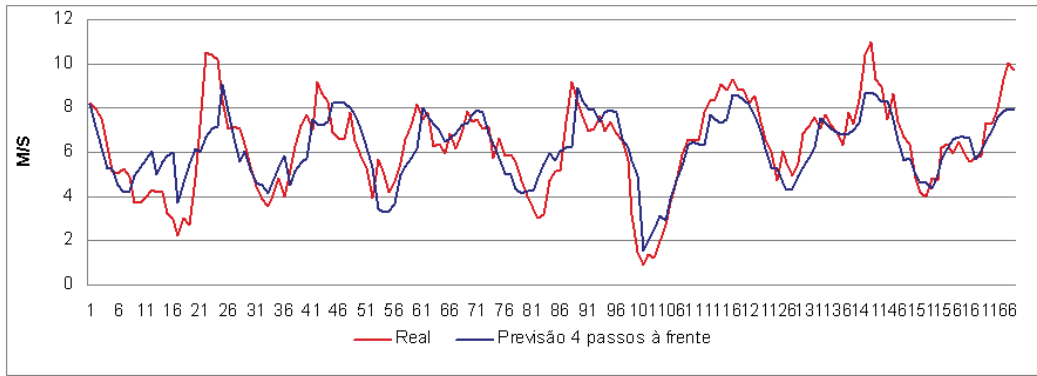


Figura 5.9: Previsão 4 passos à frente e valores observados

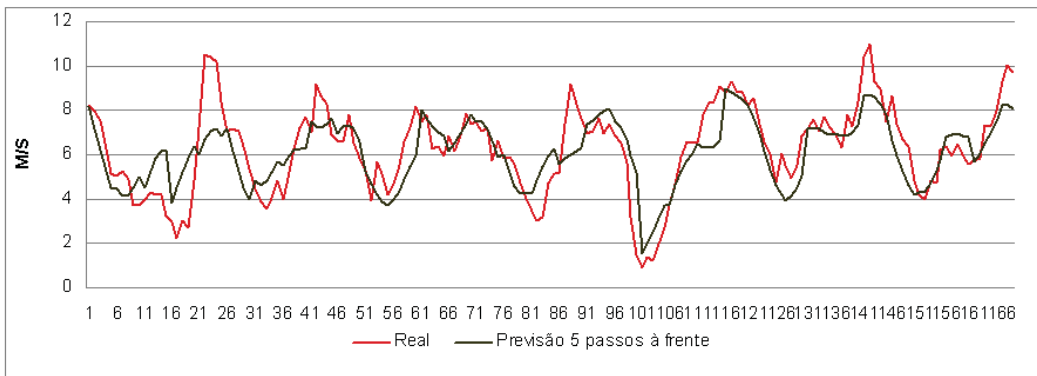


Figura 5.10: Previsão 5 passos à frente e valores observados

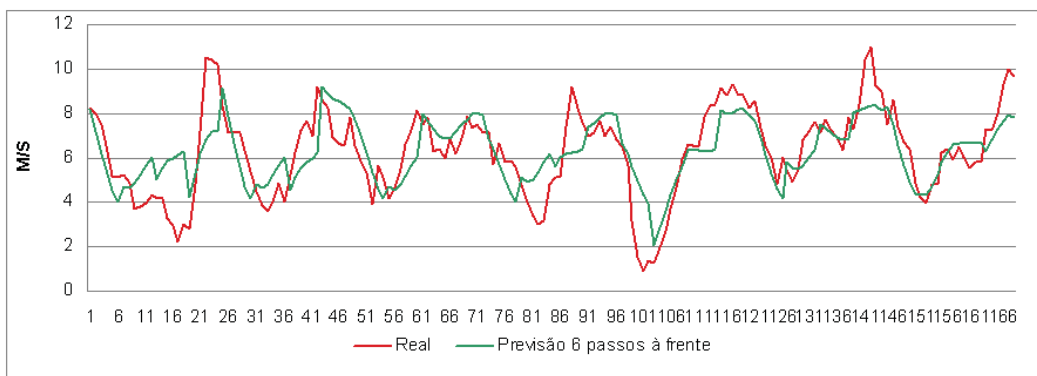


Figura 5.11: Previsão 6 passos à frente e valores observados

Os resultados mostram que o modelo proposto consegue acompanhar as observações de uma maneira geral. Além disso, é possível verificar que o modelo apresenta melhores resultados que o modelo ingênuo. Entretanto, fica evidente que o aumento do horizonte de previsão, faz com que ocorra uma significativa piora na habilidade de previsão do modelo. A tabela 5.4, apresenta os resultados das métricas de comparação.

Modelo Proposto	MAPE	MAPE (Ingênuo)	RMSE	RMSE (Ingênuo)
1 passo à frente	12,05%	13,74%	0,8066 <i>m/s</i>	0,9333 <i>m/s</i>
2 passos à frente	15,64%	21,80%	0,9984 <i>m/s</i>	1,4629 <i>m/s</i>
3 passos à frente	16,83%	30,76%	1,0101 <i>m/s</i>	1,9299 <i>m/s</i>
4 passos à frente	19,37%	38,11%	1,2097 <i>m/s</i>	2,2930 <i>m/s</i>
5 passos à frente	19,87%	44,79%	1,2608 <i>m/s</i>	2,6214 <i>m/s</i>
6 passos à frente	24,01%	49,54%	1,3613 <i>m/s</i>	2,8849 <i>m/s</i>

Tabela 5.4: Métricas de comparação

A seguir, na figura 5.12 abaixo, temos a evolução da métrica MAPE a medida que é aumentado o horizonte de previsão do modelo no período de teste (primeira semana de 2007), para o modelo até seis passos à frente.

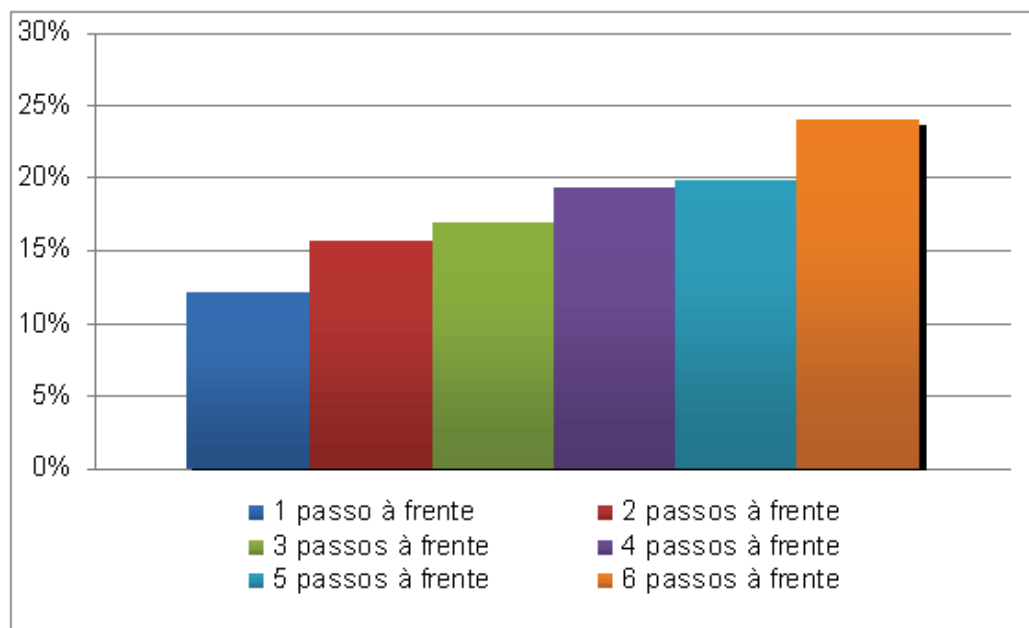


Figura 5.12: Evolução do MAPE na primeira semana de 2007

Foram feitas ainda previsões vinte quatro passos à frente utilizando o modelo proposto para a primeira semana de 2007. A figura 5.13 abaixo mostra os resultados.

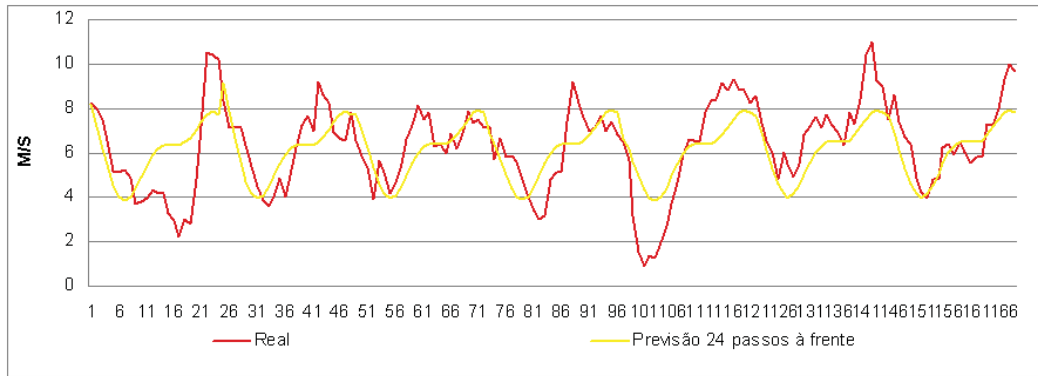


Figura 5.13: Previsão 24 passos à frente e valores observados

Como era de se esperar, a previsão vinte quatro passos à frente mostra que o modelo sofre uma piora ainda maior na sua capacidade preditiva. Entretanto, o mesmo ainda consegue acompanhar o comportamento da série. A tabela 5.5 abaixo apresenta as métricas de comparação para o modelo vinte quatro passos à frente.

Métricas	Resultados
RMSE	1,45
MAE	1,161
MAPE	26,58%

Tabela 5.5: Métricas de comparação

A figura 5.14 abaixo apresenta os três resultados de maneira conjunta.

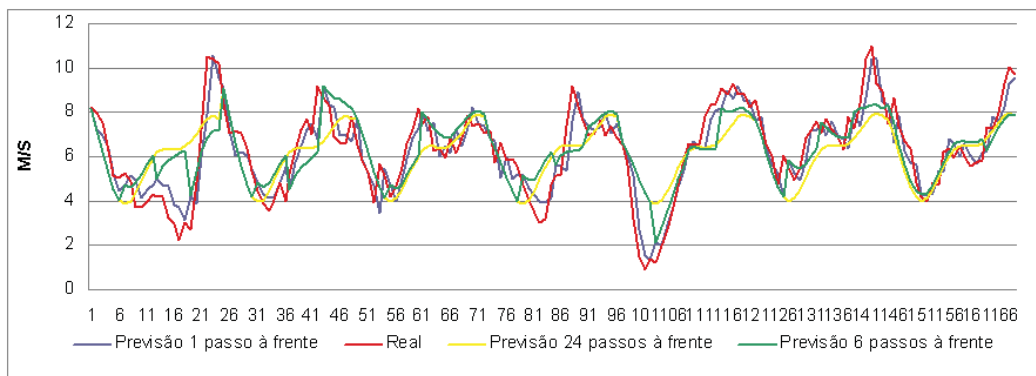


Figura 5.14: Previsões 1, 6 e 24 passo à frente e valores observados

## 5.2 Comparação dos resultados

O modelo proposto apresentou bons resultado quando comparado com outros tipos de modelos dentro da semana de teste. As previsões apresentadas foram comparadas com as geradas por modelos Persistence (ou ingênuo), ARFIMA, Harmônico e Híbrido Neuro-Fuzzy<sup>1</sup>.

Os resultados mostram que o modelo proposto obteve melhor desempenho na semana em estudo. A figura 5.15, 5.16 e 5.17 abaixo apresenta os resultados para os modelos Persistence, ARFIMA e Harmônico para previsões um passo à frente.

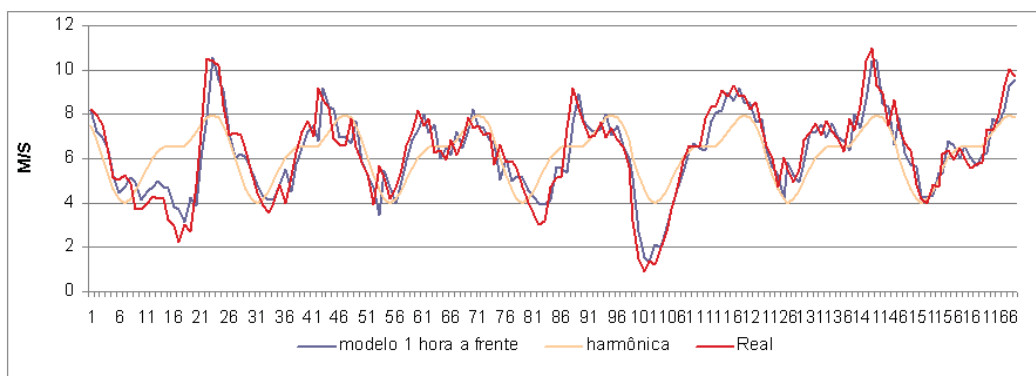


Figura 5.15: Comparação: Modelo Proposto e Modelo harmônico

A tabela 5.6 abaixo apresenta o resultado das métricas de comparação MAPE e RMSE para o modelo proposto e para os modelos ARFIMA, Persistence e Neuro-Fuzzy, considerando um passo à frente (uma hora).

<sup>1</sup>O modelo Neuro-Fuzzy não foi estimado nessa dissertação. Portanto a comparação veio através dos resultados (métricas de comparação) das previsões um passo à frente expostos para mesma semana de teste no artigo Previsão de Velocidade de Vento com Sistema de Inferência Neuro-Fuzzy Adaptativo. Pessanha, Silva e Souza - SPOLM 2010

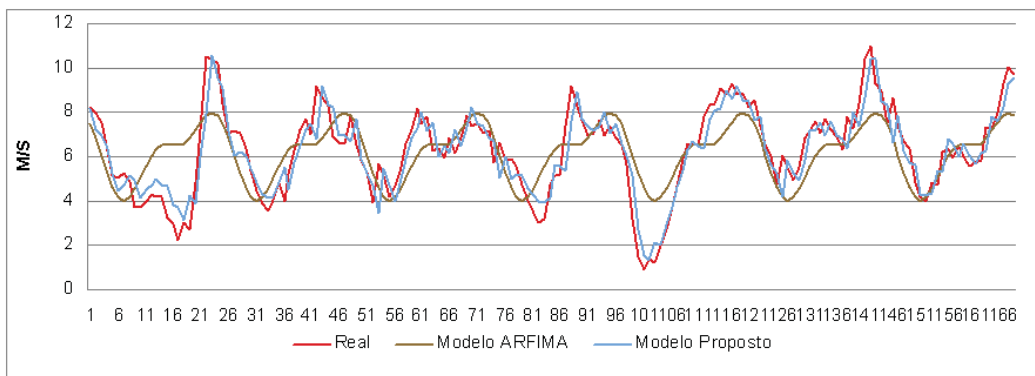


Figura 5.16: Comparação: Modelo Proposto e Modelo ARFIMA

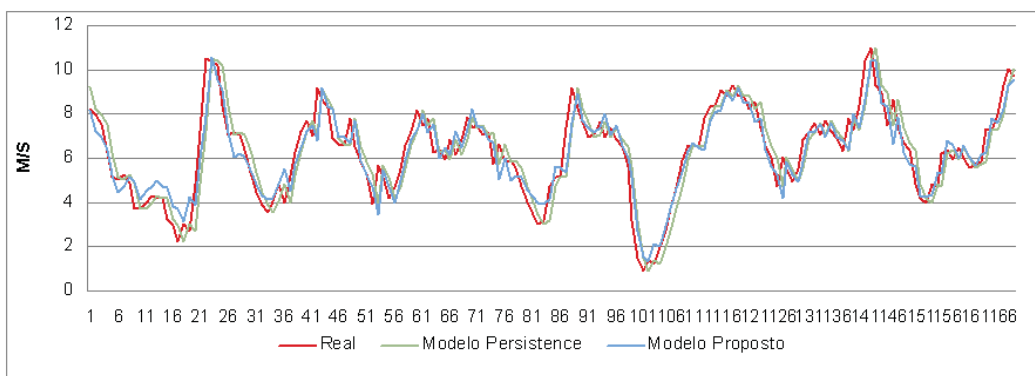


Figura 5.17: Comparação: Modelo Proposto e Modelo Persistence

Modelos	MAPE	Rmse
Persistence	13,74%	0,9333 <i>m/s</i>
Modelo proposto	12,05%	0,8066 <i>m/s</i>
Modelo Neuro-Fuzzy	16,63%	1,1449 <i>m/s</i>
Modelo Arfima	18,04%	1,2079 <i>m/s</i>

Tabela 5.6: Métricas de comparação entre os modelos

O coeficiente U de Theil apresentou valor menor que 1 (0,853), também mostrando que o modelo proposto é melhor que o modelo ingênuo (persistence). Em geral o modelo ingênuo apresenta maior competitividade um passo à frente.

No contexto do problema estudado, previsão de velocidade de vento, o modelo apresentou ganhos na qualidade da previsão. Fica aqui evidenciado que previsões de períodos mais longos, por exemplo 24 horas, utilizando esse tipo de modelagem, sofrem da piora significativa da habilidade de previsão.

Devemos aqui ressaltar que a definição de utilizar modelos ARFIMA

para os termos de erro teve desempenho fundamental para que se obtivesse melhor desempenho nas previsões do modelo proposto. Fica evidente, quando são analisadas as previsões do modelo harmônico e do modelo proposto, que o componente ARFIMA capta as variações de curto prazo e o comportamento de memória longa, enquanto o componente Harmônico capta os movimentos periódicos da série (vinte quatro horas e doze horas).