

5

Resultados

5.1

Descrição dos Dados

A partir do modelo proposto pretende-se desenvolver uma metodologia de estimação conjunta de vento e vazão, que auxilie o setor elétrico a considerar o despacho das usinas eólicas no planejamento. Porém, os avanços na produção de energia eólica se encontram em estágio inicial, o que dificulta muito a obtenção de dados de vento para pesquisa. Felizmente, tivemos acesso à série mensal de fator de capacidade eólico de uma usina não divulgada no Nordeste. O fator de capacidade eólico significa quanto da capacidade nominal do aerogerador está sendo aproveitada, que representa uma boa aproximação do potencial eólico da região. Além disso, seu cálculo (Seção 2.5.3) envolve direta ou indiretamente as características do aerogerador e as características físicas locais que interferem na produção, que se traduz em informações úteis ao processo de modelagem. A título de curiosidade, atualmente a potência nominal média dos aerogeradores no Brasil varia de 1,8 MW a 2,4 MW, [29].

Os dados referentes à vazão são representados pela série de vazão afluente média mensal (m^3/s) de um posto de avaliação localizado em Paraibuna no Sudeste. Os dados coletados têm a finalidade de avaliar a instalação de uma PCH (Pequena Central Hidrelétrica) neste ponto do rio, e como uma PCH não possui reservatório, toda a vazão afluente será turbinada. Sendo assim, a energia produzida pode ser obtida pela multiplicação entre a vazão afluente e um coeficiente de produtividade da PCH.

Os modelos desenvolvidos incluem variáveis exógenas: as séries de energia natural afluente (ENA) calculadas no *módulo1* do Newave (Seção 3.3.1.1) por subsistema. Os dados representam as médias mensais em $MW/m^3/s$. Essas variáveis funcionam como variáveis de controle que permitem aperfeiçoar as

especificações do modelo, capturando as variações na produção mensal de energia acumulada de cada subsistema. A inclusão dessas variáveis nos modelos resultou em resíduos com propriedades mais adequadas, e reduziu a variância dos parâmetros estimados.

Todas as séries possuem discretização mensal com data inicial em Janeiro de 1976 e data final em Julho de 2009, totalizando 403 registros.

5.1.1

Fator de Capacidade Eólico e Vazão

Para a série fator de capacidade eólico utilizada neste estudo, encontrou-se um fator de capacidade médio anual de 40,11%, e conforme já comentado em seções anteriores, este valor está muito acima da média mundial que varia em torno de 27%. Por se tratar de uma usina localizada no Nordeste, este resultado corrobora com a afirmação que esta é uma das regiões mais promissoras para produção de energia eólica no Brasil.

Para facilitar a nomenclatura, de agora em diante, a variável fator de capacidade eólico será denominada FC e a variável vazão, Vz.

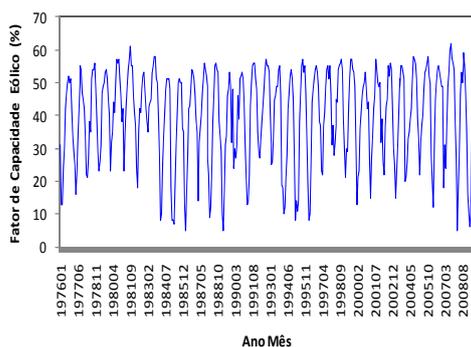


Figura 5-1 – Série histórica FC.

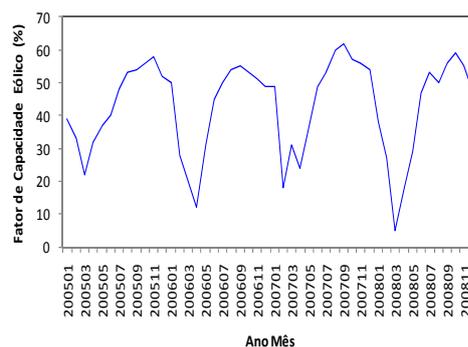


Figura 5-2 – Série histórica FC (2005-2008).

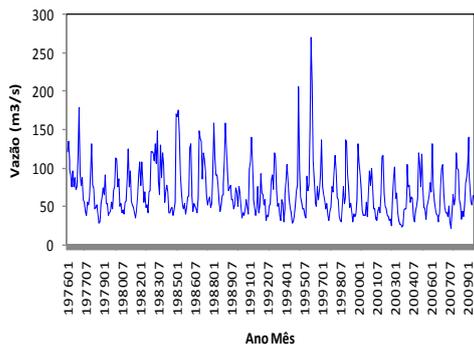


Figura 5-3 – Série histórica Vz.

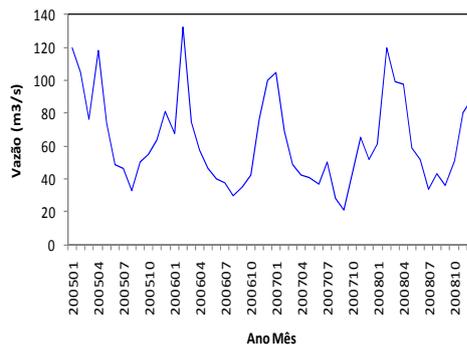


Figura 5-4 – Série histórica Vz (2005-2008).

A análise gráfica de uma série temporal é o ponto de partida para a aplicação de testes mais formais. A evolução das séries no tempo indica comportamento sazonal, já esperado, sem apresentar deslocamento de nível nem tendência, o que é um bom indicativo de estacionariedade. Na série Vz (Figura 5-3), os anos 1995 e 1996, se destacam dos demais por seus valores bastante elevados.

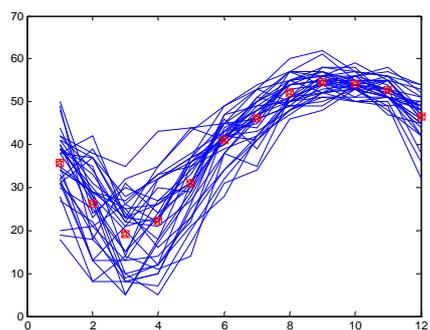


Figura 5-5 – Comparativo anual FC.

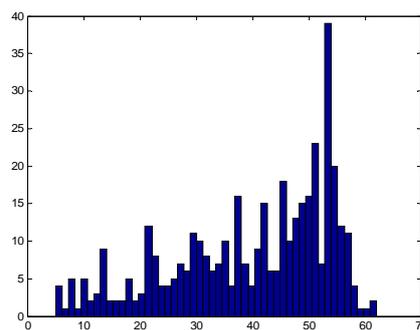


Figura 5-6 – Histograma FC.

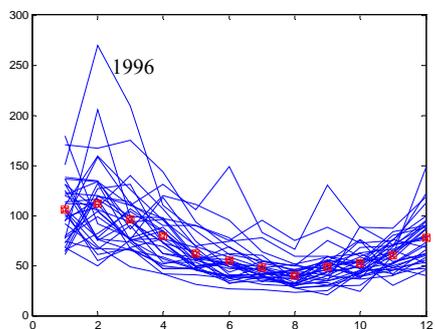


Figura 5-7 – Comparativo anual Vz.

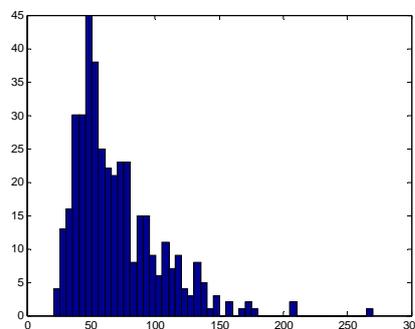


Figura 5-8 – Histograma Vz.

Acima, as Figura 5-5 e Figura 5-7 ilustram o comportamento ano a ano, como uma forma de avaliar a volatilidade contida nos meses. Uma análise estatística dos resultados indica que os meses de Janeiro a Abril possuem uma variância maior comparada aos demais meses. Isso foi observado tanto na série FC como na série Vz. Este resultado é facilmente observado através da dispersão dos dados em torno do ponto em destaque, que representa a média do mês. Nos histogramas das séries (Figura 5-6 e Figura 5-8) é interessante observar a assimetria inversa das séries FC e Vz.

Tabela 5-1– Estatísticas descritivas FC e Vz.

Estatísticas Descritivas	FC	Vz
Média	39,79	69,95
Mediana	43	61
Mínimo	5	21
Máximo	62	270
Desvio padrão	14,05	33,78
C.V.	0,35	0,48
Assimetria	-0,67	1,56
Curtose	-0,58	3,90

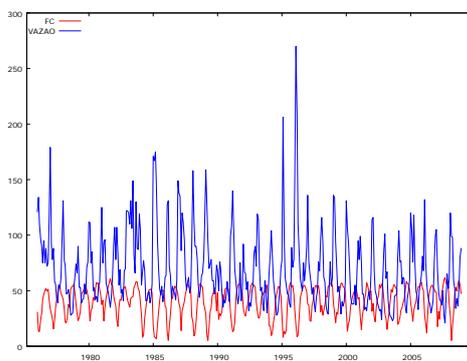


Figura 5-9 – Séries históricas Vz e FC.

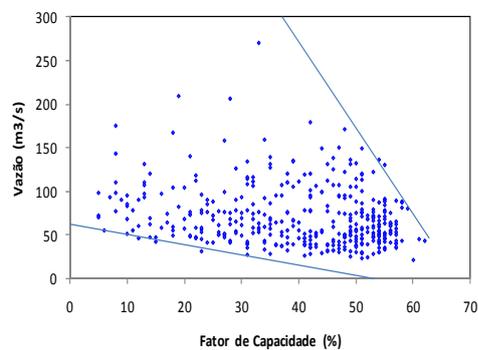


Figura 5-10 – Gráfico de dispersão (FC X Vz).

O comportamento complementar entre as séries chama atenção: os pontos de depressão de uma curva coincidem com os picos da outra curva (Figura 5-9). A Figura 5-10 mostra a relação entre as séries estudadas e como os maiores valores de FC estão concentrados no intervalo de baixa vazão.

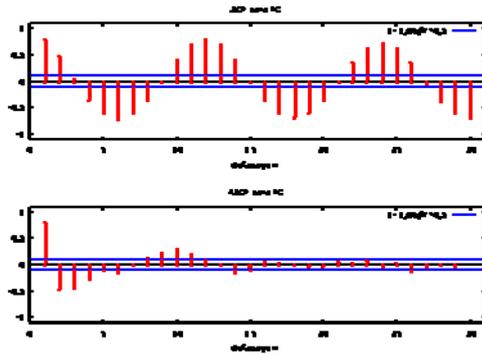


Figura 5-11 – Correlograma da série FC.

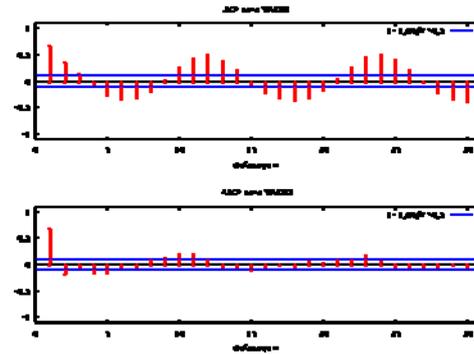


Figura 5-12 – Correlograma da série Vz.

Nas Figura 5-11 e Figura 5-12, o movimento infinito senoidal na função de autocorrelação (FAC) em conjunto com o movimento finito senoidal na função de autocorrelação parcial (FACP) sugerem modelos autoregressivos periódicos, tanto para a série FC quanto para a série Vz.

A aplicação do teste ADF (Seção 4.2.1.1) na série FC, resultou no *p-valor* $< 0,00001$. O teste aplicado à série Vz, também resultou em *p-valor* $< 0,00001$. Sendo assim, rejeita-se a hipótese nula em ambos os testes, indicando evidência estatística que as séries FC e Vz são estacionárias.

5.1.2

Variáveis Exógenas

As variáveis ENA's serão incluídas no modelo como variáveis exógenas, ou variáveis de controle da heterogeneidade estrutural de cada subsistema brasileiro.

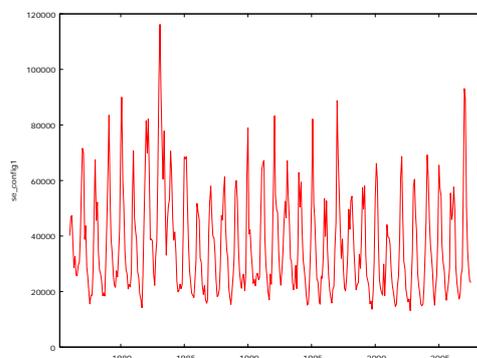


Figura 5-13 – Série histórica ENA-SE.

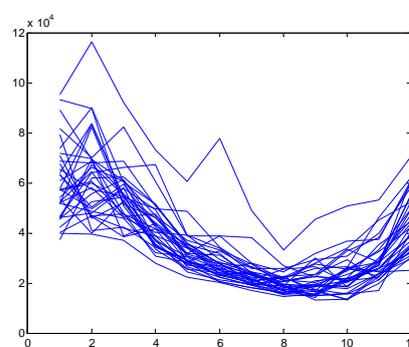


Figura 5-14 – Comparativo anual ENA-SE.

O Sudeste é o subsistema com a maior produção de energia acumulada no país. A evolução da série ENA-SE no tempo (Figura 5-13), apresenta bons indícios de estacionariedade. Como já era esperado, o comportamento da série deste subsistema é o que mais se aproxima do comportamento da série Vz também no Sudeste (Paraibuna). A Figura 5-14 reforça a idéia que os primeiros meses do ano apresentam maior volatilidade. A linha acima das demais e, portanto descolada da faixa de observações, representa o ano 1983.

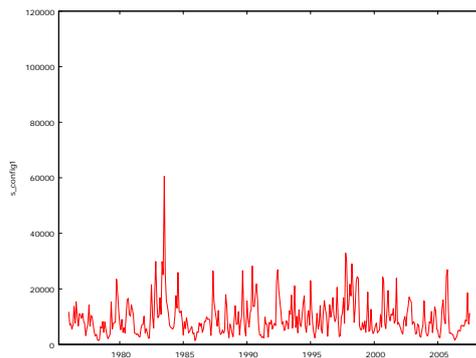


Figura 5-15 – Série histórica ENA-S.

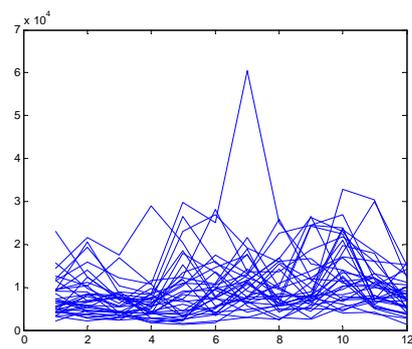


Figura 5-16 – Comparativo anual ENA-S.

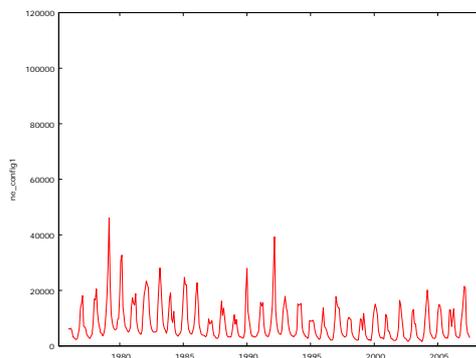


Figura 5-17 – Série histórica ENA-NE.

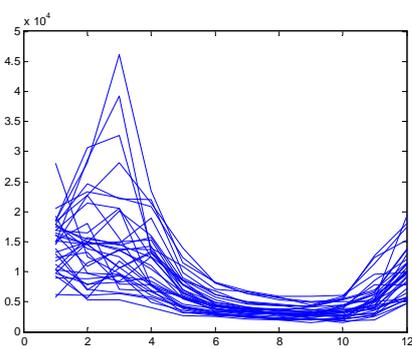


Figura 5-18 – Comparativo anual ENA-NE.

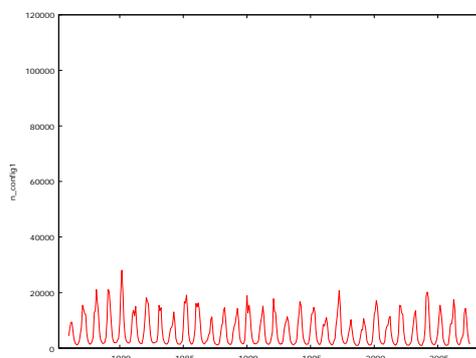


Figura 5-19 – Série histórica ENA-N.

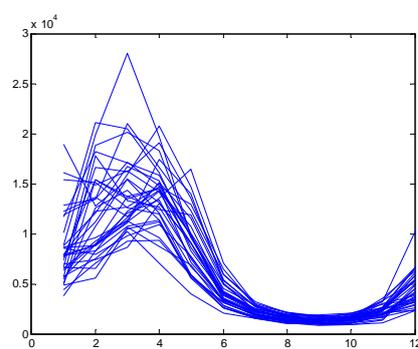


Figura 5-20 – Comparativo anual ENA-N.

Todas as séries ENA's apresentam bons indícios de estacionariedade, que serão verificados a seguir pelo teste ADF. A série ENA-S (Figura 5-13), apresenta um comportamento mais “nervoso”, com destaque para o pico da curva no ano 1983.

A série ENA-NE e ENA-N, nas Figura 5-17 e Figura 5-19, apresentam comportamento bastante regular com ciclos bem definidos. Além disso, ambas as séries apresentam variância muito pequena no período que vai de Junho a Outubro, o mesmo não ocorre nos primeiros meses dos anos (Figura 5-18 e Figura 5-20). Assim como nas séries FC e Vz, aplicou-se o teste ADF em cada série ENA, e todas apresentaram $p\text{-valor} < 0,00001$. Logo, rejeita-se a hipótese nula em todos os casos, indicando que as séries ENA's também são estacionárias.

5.2

Modelo Proposto

O desafio na escolha do modelo estatístico para as séries FC e Vz, observadas no tempo t , consiste em estabelecer uma proposta capaz de considerar a análise de cada série individual, onde a autocorrelação contida em cada série é fundamental, a análise das relações dinâmicas entre as séries de interesse e a análise do impacto exercido pelas séries ENA's. Com este objetivo em mente, a elaboração do modelo baseou-se em técnicas multivariadas autoregressivas de ordem (p), através do modelo VARX(p) (*Vector Autoregressive with eXternal Variables*).

O modelo VAR (*Vector Autoregressive*) é uma generalização do modelo univariado autoregressivo para séries temporais múltiplas e dinâmicas, onde as regressões obedecem à mesma estrutura para todas as equações individuais. Esta classe de modelos lineares também é considerada uma das mais bem sucedidas e flexíveis para a análise de séries temporais multivariada, provando ser especialmente útil para previsão e descrição do comportamento dinâmico das séries, [25], [38].

Este modelo é um sistema multiequacional onde cada variável endógena é explicada por uma constante, por seus valores defasados e pelos valores defasados

das demais variáveis endógenas, permitindo a inclusão de variáveis exógenas¹⁴, com ou sem defasagem.

O modelo VARX desenvolvido para as séries FC_t e Vz_t , representadas respectivamente por y_{1t} e y_{2t} , considera como variáveis exógenas as séries x_{1t} , x_{2t} , x_{3t} e x_{4t} , e têm as seguintes equações:

$$y_{1t} = c_1 + \phi_{111}y_{1t-1} + \phi_{121}y_{2t-1} + \dots + \phi_{11p}y_{1t-p} + \phi_{12p}y_{2t-p} + \theta_{111}x_{1t} + \theta_{121}x_{2t} + \theta_{131}x_{3t} + \theta_{141}x_{4t} + \dots + \theta_{11q}x_{1t-q+1} + \theta_{12q}x_{2t-q+1} + \theta_{13q}x_{3t-q+1} + \theta_{14q}x_{4t-q+1} + \varepsilon_{1t} \quad (5.1)$$

$$y_{2t} = c_2 + \phi_{211}y_{1t-1} + \phi_{221}y_{2t-1} + \dots + \phi_{21p}y_{1t-p} + \phi_{22p}y_{2t-p} + \theta_{211}x_{1t} + \theta_{221}x_{2t} + \theta_{231}x_{3t} + \theta_{241}x_{4t} + \dots + \theta_{21q}x_{1t-q+1} + \theta_{22q}x_{2t-q+1} + \theta_{23q}x_{3t-q+1} + \theta_{24q}x_{4t-q+1} + \varepsilon_{2t} \quad (5.2)$$

Cada equação do modelo possui um termo correspondente ao choque aleatório da i -ésima variável e é representado por ε_{it} . O coeficiente ϕ_{ijk} refere-se às variáveis endógenas, e traduz a sensibilidade da i -ésima variável do modelo com relação à k -ésima defasagem da j -ésima variável. Seguindo o mesmo raciocínio, o coeficiente θ_{ijl} refere-se às variáveis exógenas do sistema.

A seguir, as equações Eq.(5.1) e Eq.(5.2) estão escritas sob a forma matricial:

$$\begin{aligned} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} &= \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{111} & \phi_{121} \\ \phi_{211} & \phi_{221} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \dots + \begin{bmatrix} \phi_{11p} & \phi_{12p} \\ \phi_{21p} & \phi_{22p} \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \end{bmatrix} + \\ &+ \begin{bmatrix} \theta_{111} & \theta_{121} & \theta_{131} & \theta_{141} \\ \theta_{211} & \theta_{221} & \theta_{231} & \theta_{241} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \end{bmatrix} + \dots + \\ &+ \begin{bmatrix} \theta_{11q} & \theta_{12q} & \theta_{13q} & \theta_{14q} \\ \theta_{21q} & \theta_{22q} & \theta_{23q} & \theta_{24q} \end{bmatrix} \begin{bmatrix} x_{1t-q+1} \\ x_{2t-q+1} \\ x_{3t-q+1} \\ x_{4t-q+1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \end{aligned} \quad (5.3)$$

$$\Downarrow \quad (5.4)$$

$$Y_t = C + \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + \Theta_1 X_t + \dots + \Theta_q X_{t-q+1} + \varepsilon_t$$

¹⁴Ao acrescentar um vetor de variáveis exógenas ao um modelo VAR puro, este passará a ser denominado como modelo VARX.

$$Y_t = C + \sum_{k=1}^p \Phi_k Y_{t-k} + \sum_{l=1}^q \theta_l X_{t-l+1} + \epsilon_t \quad (5.5)$$

Onde:

Y_t = vetor 2×1 contendo o valor que as variáveis endógenas $[FC_t \ VZ_t]^T$ assumem no tempo t ;

C = vetor de constantes 2×1 ;

Φ_k = matriz 2×2 de coeficientes autoregressivos para as variáveis endógenas, e $k = 1, 2, \dots, p$.

X_t = vetor 4×1 contendo o valor que as variáveis exógenas $[SE_t \ S_t \ NE_t \ N_t]^T$ assumem no tempo t ;

θ_l = matriz 2×4 de coeficientes autoregressivos para as variáveis exógenas, e $l = 1, 2, \dots, q$.

ϵ_t = vetor 2×1 de ruídos brancos, com:

$$E(\epsilon_t) = 0$$

$$E(\epsilon_t \epsilon_\tau) = \begin{cases} \Omega, & t = \tau \\ 0, & t \neq \tau \end{cases}$$

Sendo Ω uma matriz simétrica 2×2 definida positiva.

A formulação precisou ser definida de forma ampla, incluindo todo o movimento sazonal das variáveis representado pela ordem de defasagem $p = 12$ das variáveis endógenas. A necessidade de manter todos os coeficientes, mesmo quando a significância de alguns não seja comprovada, surge da inserção deste estudo no projeto real que ocorre simultaneamente. O projeto, detalhado em (Seção 1.1), envolverá em etapas posteriores, a inclusão de outras séries no modelo. E como a significância dos coeficientes varia de série para série, adotou-se a solução mais geral.

A inclusão das quatro variáveis exógenas, que representam as séries ENA's, no modelo revelou aumento no poder preditivo e maior coerência das soluções encontradas. Para estas variáveis o critério escolhido de defasagem é representado por $q = 2$.

O modelo VARX pode atuar como um poderoso instrumento de planejamento, pois apresenta resultados que fornecem uma descrição estatística das relações dinâmicas entre as variáveis que compõem o sistema.

O modelo definido na Eq.(5.5) tem características autoregressivas com ênfase na heterocedasticidade sazonal e utiliza o método de mínimos quadrados ponderados (MQP) para estimação dos parâmetros. A escolha deste método de estimação começa com a verificação de indícios de heterocedasticidade apontados, principalmente, pelos comparativos anuais que mostraram variância não constante entre os meses. Porém, esses indícios devem ser testados nos resíduos do modelo. Para isso, os parâmetros foram estimados por MQO, supondo homocedasticidade. O exame do quadrado dos resíduos revela padrões sistemáticos com relação ao \hat{Y}_t estimado, conforme as figuras abaixo.

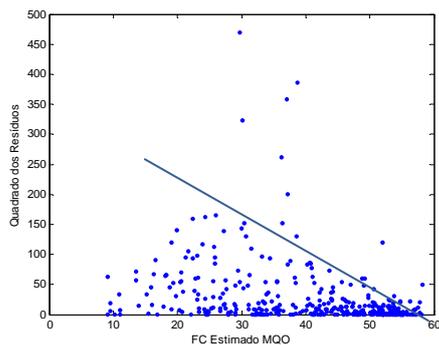


Figura 5-21 – FC estimado por MQO $\times \hat{\varepsilon}_{FC}^2$.

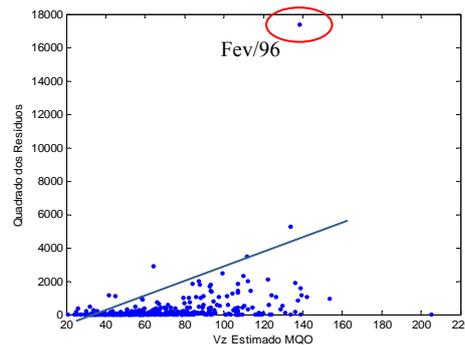


Figura 5-22 – Vz estimado por MQO $\times \hat{\varepsilon}_{Vz}^2$.

O teste formal de heterocedasticidade de Koenker-Basset (Seção 4.2.4.2) confirma a suspeita, pois avalia a significância do coeficiente de \hat{Y}_t em um modelo de regressão para $\hat{\varepsilon}^2$. Em ambos os casos, FC e Vz, rejeita-se a hipótese que este coeficiente seja igual a zero com $p\text{-valor} < 0,00001$. Logo, há evidência que a heterocedasticidade está presente nas séries estudadas. Contudo, não é a primeira vez que se observa o descolamento do ano 1996 do restante dos dados de vazão (Figura 5-22). O resíduo gerado pela observação de Fev/1996 também aparece como um ponto discrepante. Nota-se que os gráficos estão em escalas diferentes, para melhor visualização da dispersão dos pontos, por isso a maior discrepância é observada na série Vz, indicando presença de *outliers*.

A análise das séries sugere que a heterocedasticidade, já constatada, tem origem na sazonalidade natural de dados climáticos, portanto a estrutura de variâncias desiguais será dada por σ_m , onde m é o número de meses. Na prática, esta estrutura foi encontrada nos resíduos gerados por MQO.

A aplicação do método MQP implica na estimação da matriz W de pesos, que será realizada pelo algoritmo recursivo. A princípio, admitiu-se homocedasticidade e os resíduos foram analisados revelando uma estrutura mensal. Os pesos, então, assumem o inverso da variância mensal observada nos resíduos $\frac{1}{\sigma_m^2}$, os parâmetros são novamente estimados, e novos resíduos são gerados. As variâncias residuais encontradas são introduzidas na fórmula produzindo pesos diferenciados e os parâmetros são re-estimados. O processo é repetido até que não exista diferença significativa entre as estimativas.

5.2.1

Avaliação

O modelo alcançou a estabilidade das estimativas após 10 iterações. As estatísticas do modelo apresentaram valores satisfatórios. Os principais resultados encontram-se na tabela abaixo:

Tabela 5-2– Principais resultados modelos FC e Vz.

Principais Resultados		FC	Vz
R-quadrado		0.8404	0.7031
R-quadrado ajustado		0.8251	0.6747
Erro padrão dos resíduos		1.0004	1.0009
Resíduos	Teste Jarque- Bera (<i>p-valor</i>)	0.3833	<0.0001
	Teste Kolmogorov-Smirnov (<i>p-valor</i>)	0.7087	0.5030
	Teste Box Pierce Ljung (<i>p-valor</i>)	0.9659	0.8199
	Teste LM para autocorrelação (<i>p-valor</i>):	0.8538	0.7838

O modelo FC consegue explicar 84% da variação da série FC_t , e esse valor penalizado pelo número de variáveis explicativas ainda é superior a 80%, $R_{aj}^2 = 0,8251$. O modelo Vz consegue explicar 70% da variação da série Vz_t , e esse valor penalizado pelo número de variáveis explicativas cai para 67%.

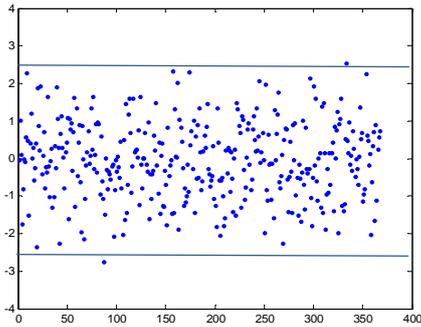


Figura 5-23 – Dispersão dos resíduos FC.

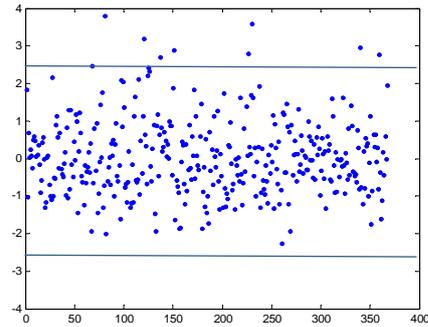


Figura 5-24 – Dispersão dos resíduos Vz.

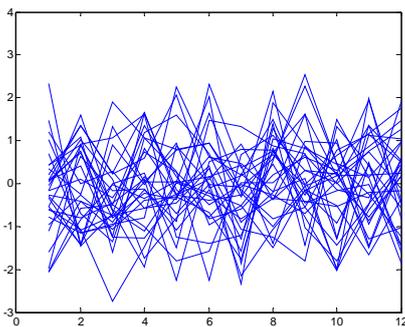


Figura 5-25 – Comparativo anual resíduos FC.

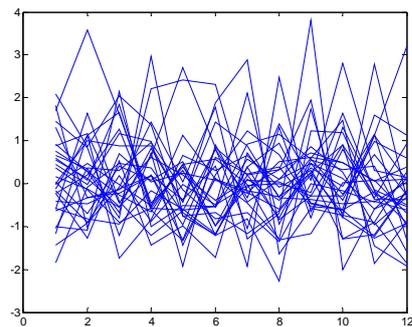


Figura 5-26 – Comparativo anual resíduos Vz.

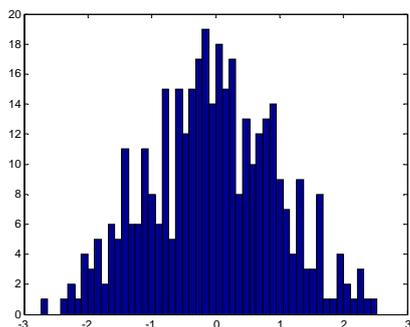


Figura 5-27 – Histograma resíduos FC.

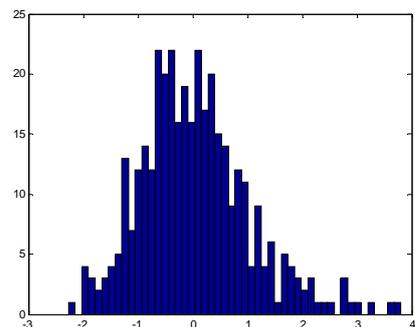


Figura 5-28 – Histograma resíduos Vz.

O diagnóstico do modelo é obtido, principalmente, através da análise dos resíduos gerados. Os gráficos de dispersão dos resíduos (Figura 5-23 e Figura 5-24) não apresentam um comportamento sistemático, embora existam alguns pontos no conjunto dos resíduos Vz com valores muito elevados. Os comparativos anuais dos resíduos (Figura 5-25 e Figura 5-26) mostram que as variâncias desiguais foram compensadas com a ponderação utilizada.

Os histogramas dos resíduos (Figura 5-27 e Figura 5-28), juntamente com os testes de normalidade informam o quanto os choques aleatórios se aproximam da distribuição Normal. A distribuição dos resíduos gerados pelo modelo FC, pode

ser aceita como Normal, pois os testes de Jarque-Bera e de Kolmogorov-Smirnov não mostram evidências para rejeitar a hipótese nula de normalidade (Tabela 5-2). Por outro lado, o teste de Jarque-Bera aplicado aos resíduos do modelo Vz, rejeita a hipótese de normalidade. O histograma dos resíduos Vz, apesar da sua forma regular, apresenta a cauda direita bastante acentuada. O teste de Kolmogorov-Smirnov não rejeita a hipótese de normalidade para os resíduos Vz, porém apresenta baixo poder de teste, como citado na (Seção 4.2.2.2).

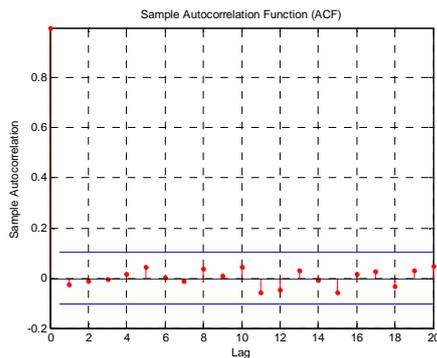


Figura 5-29 – Autocorrelação resíduos FC.

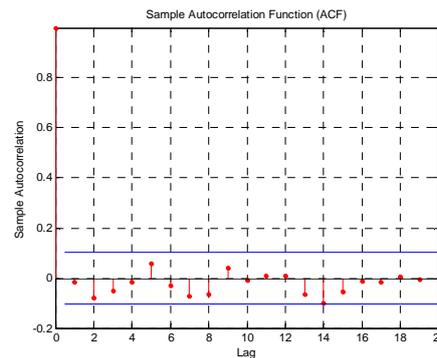


Figura 5-30 – Autocorrelação quadrado dos resíduos FC.

Os resíduos FC não possuem autocorrelações significativas nas Figura 5-29 e Figura 5-30. Este resultado é confirmado pelos testes de autocorrelação Box-Pierce-Ljung e LM (Tabela 5-2), que não rejeitam a hipótese nula (H_0 : não existe autocorrelação residual).

A função de autocorrelação (FAC) dos resíduos ao quadrado é um bom indicativo de heterocedasticidade condicional autoregressiva, ou efeito ARCH. Este efeito é verificado quando a variância do erro no período t se relaciona com o quadrado dos termos de erro do período anterior.

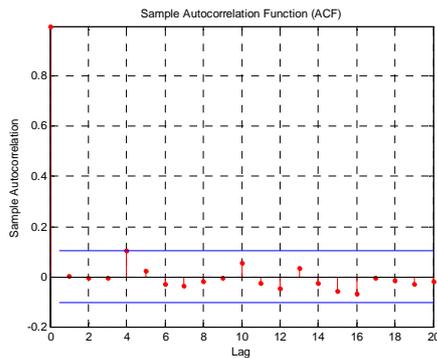


Figura 5-31 – Autocorrelação resíduos Vz.

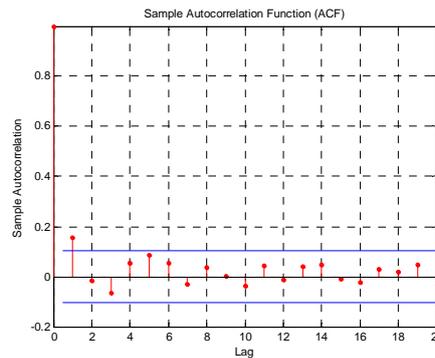


Figura 5-32 – Autocorrelação quadrado dos resíduos Vz.

Os testes Box-Pierce-Ljung e LM não indicam autocorrelação nos resíduos Vz (Tabela 5-2). Por outro lado, a autocorrelação na primeira defasagem do quadrado dos resíduos (Figura 5-32) é estatisticamente não-nula, sugerindo que os resíduos apresentam heterocedasticidade condicional. Essa característica nos levou a modelar a série com outras metodologias, voltadas para a modelagem da volatilidade, como o modelo ARCH¹⁵. Porém, a heterocedasticidade condicional persistiu após várias tentativas e os resultados foram inferiores ao resultado da modelagem atual. A adversidade percebida pode ser, portanto, um indício da presença de *outliers* na série, que podem estar conduzindo os resíduos a efeitos espúrios.

5.2.2

Identificação de Outliers

Os resíduos discrepantes concentrados nos primeiros meses do ano e o efeito na FAC do quadrado dos resíduos podem ser consequências da presença de *outliers* na série Vz. O gráfico *boxplot* permite avaliar a simetria dos dados e a dispersão, e é um bom indicativo de presença de *outliers*, além de ser uma ferramenta muito útil na comparação de dois ou mais conjuntos de dados.

¹⁵Para mais detalhes sobre modelos de volatilidade ver [38][38].

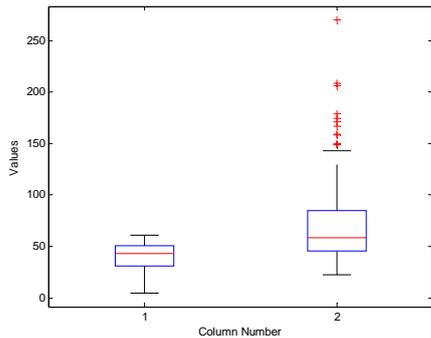


Figura 5-33 – Boxplot FC e Vz.

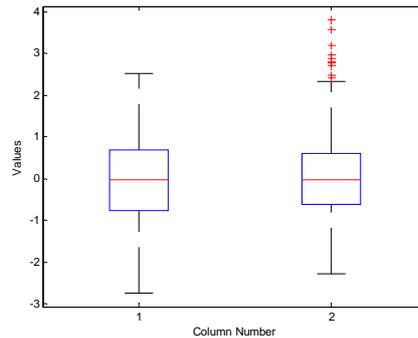


Figura 5-34 – Boxplot dos resíduos FC e Vz.

Na Figura 5-33, o *boxplot* da série Vz ilustra alguns pontos suspeitos fora do intervalo demarcado¹⁶. A maior parte desses pontos foi observada nos primeiros meses do ano, responsáveis por gerar resíduos discrepantes (Figura 5-34). Em séries temporais, os *outliers* têm efeito prolongado no tempo, seja na própria série de observações ou na série residual, portanto não basta detectá-los simplesmente através dos valores mais elevados. A verificação formal será realizada através da aplicação do teste de detecção de *outliers* em séries temporais, baseado na razão de verossimilhança (Seção 4.2.5.2). O teste será aplicado nas séries FC e Vz.

Para o desenvolvimento do teste, cada série foi modelada seguindo processo univariado autoregressivo, e os modelos com melhor desempenho foram:

$$FC_t = \phi_1 FC_{t-1} + \phi_4 FC_{t-4} + \phi_{10} FC_{t-10} + \phi_{11} FC_{t-11} + \phi_{12} FC_{t-12} + a_t \quad (5.6)$$

$$Vz_t = \phi_1 Vz_{t-1} + \phi_{10} Vz_{t-10} + \phi_{11} Vz_{t-11} + \phi_{12} Vz_{t-12} + e_t \quad (5.7)$$

Em seguida, os resíduos e os parâmetros estimados foram inseridos no procedimento iterativo que detecta os *outliers* um a um, por ordem de magnitude estimada. Sempre que um *outlier* é detectado, os resíduos impactados são ajustados, e a nova série residual é usada na etapa seguinte para detectar outro *outlier*. Esse método revela a quantidade, o tipo e a posição no tempo dos *outliers* da série estudada. O teste de detecção foi realizado com grau médio de

¹⁶O *boxplot* dispõe graficamente de um critério objetivo para identificação de *outliers*, que são identificados pelos pontos fora do intervalo demarcado $[(Q_1 - 1,5 \times (Q_3 - Q_1)); (Q_3 + 1,5 \times Q_3 - Q_1)]$. Seja Q_1 o 1º quartil e Q_3 o 3º quartil.

sensibilidade. Os resultados confirmam as suspeitas iniciais, e estão detalhados nas tabelas abaixo:

Tabela 5-3– Identificação de outliers na série FC.

Tipo	Estatística de Teste	Observação	Tempo
IO	4.0479	99	Mar/1984

Tabela 5-4– Identificação de outliers na série Vz.

Tipo	Estatística de Teste	Observação	Tempo
IO	5.584	242	Fev/1996
AO	5.9148	230	Fev/1995
IO	5.6153	109	Jan/1985
AO	4.326	13	Jan/1977
AO	4.0093	89	Mai/1983
AO	4.0048	362	Fev/2006
AO	3.7982	134	Fev/1987
AO	3.5381	387	Mar/2008
AO	3.2801	388	Abr/2008

A série FC revelou um único *outlier* de inovação, enquanto a série Vz sofre a influência de um conjunto de observações discrepantes, cujas principais estão descritas na

Tabela 5-4. A observação na data Fev/1996, não deixa mais dúvidas quanto à sua particularidade em relação ao restante dos dados. A maioria dos *outliers* é do tipo aditivo, mas existem dois de inovação, que exigem medidas corretivas diferentes. Contudo, vale lembrar que a ocorrência de *outliers* aditivos, na maioria das vezes, indica um erro de medição.

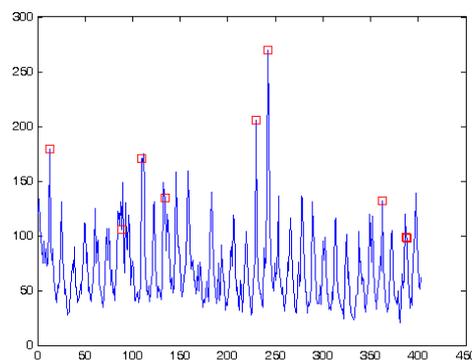


Figura 5-35 – Outliers série Vz.

Na Figura 5-35, fica claro que os pontos *outliers* não precisam ser os valores mais elevados da série, quando na verdade, precisam ser classificados como anomalias comparadas aos pontos vizinhos e ao comportamento global da série.

5.3

Modelo com Correção de Outliers

Uma vez identificados *outliers* na série, a próxima etapa consiste em empregar o algoritmo de correção que realiza uma estimação robusta da série (Seção 4.3), eliminando a interferência que essas observações podem causar no processo de modelagem. A correção foi aplicada apenas na série Vz.

Tabela 5-5– Estatísticas descritivas Vz corrigida.

Estatísticas Descritivas	Vz Original	Vz Corrigida
Média	69.95	69.17
Mediana	61	62.56
Mínimo	21	19
Máximo	270	223
Desvio padrão	33.78	29.52
C.V.	0.48	0.43
Assimetria	1.56	1.02
Curtose	3.90	4.71

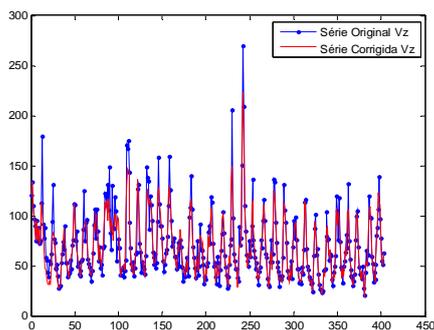


Figura 5-36 – Vz original e Vz corrigida.

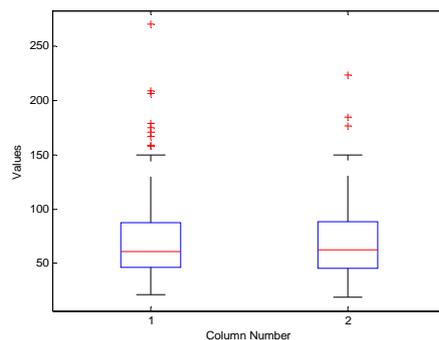


Figura 5-37 – Boxplot Vz original e Vz corrigida.

Os movimentos e as características originais da série Vz foram mantidos, e as principais alterações foram nos quantis (Tabela 5-5 e Figura 5-36), e na suavização dos picos discrepantes (Figura 5-37). A aplicação do teste ADF na série corrigida Vz, resultou no $p\text{-valor} < 0,00001$, indicando que após a correção a série permaneceu estacionária.

5.3.1

Avaliação

O modelo alcançou a estabilidade das estimativas após 17 iterações. As estatísticas do modelo apresentaram valores satisfatórios. Os principais resultados encontram-se na tabela abaixo:

Tabela 5-6– Principais resultados modelos FC e Vz.

Principais Resultados		FC	Vz
R-quadrado		0.8458	0.8310
R-quadrado ajustado		0.8206	0.8035
Erro padrão dos resíduos		1.0004	0.9997
Resíduos	Teste Jarque- Bera (<i>p-valor</i>)	0.4027	0.0031
	Teste Kolmogorov-Smirnov (<i>p-valor</i>)	0.4574	0.7617
	Teste Box Pierce Ljung (<i>p-valor</i>)	0.9825	0.1836
	Teste LM para autocorrelação (<i>p-valor</i>):	0.8259	0.2226

Os resultados do modelo FC tiveram alterações mínimas, permanecendo com R^2 e R_{aj}^2 próximos a 83%. Enquanto, o modelo Vz apresentou resultados superiores, e agora consegue explicar 83% da variação da série Vz_t . Esse valor penalizado pelo número de variáveis explicativas ainda fica em torno de 80%.

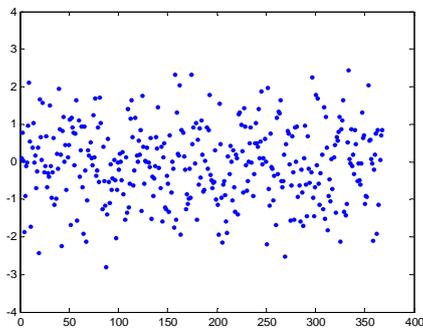


Figura 5-38 – Dispersão dos resíduos FC.

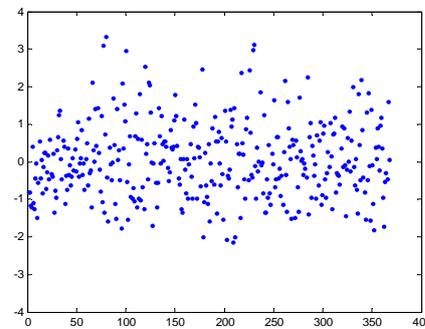


Figura 5-39 – Dispersão dos resíduos Vz.

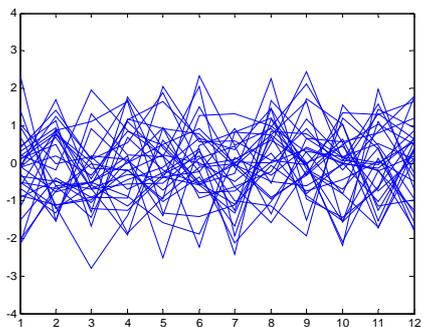


Figura 5-40 – Comparativo anual resíduos FC.

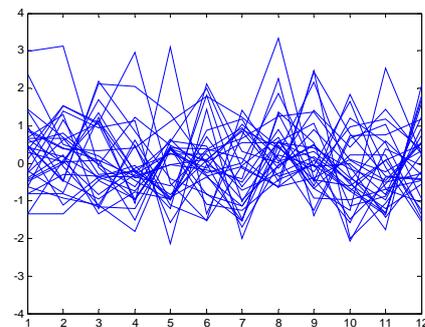


Figura 5-41 – Comparativo anual resíduos Vz.

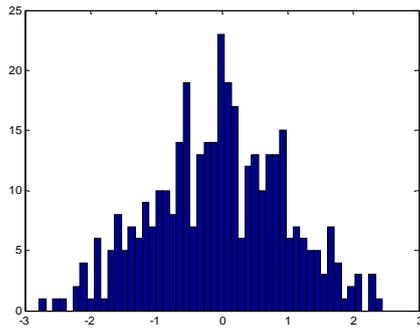


Figura 5-42 – Histograma resíduos FC.

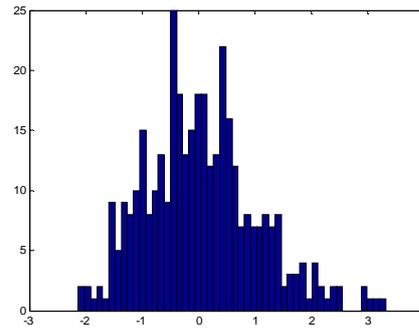


Figura 5-43 – Histograma resíduos Vz.

Os gráficos de dispersão dos resíduos (Figura 5-38 e Figura 5-39) e os comparativos anuais (Figura 5-40 e Figura 5-41) apresentam comportamento aleatório em torno do eixo, bastante satisfatório. O histograma dos resíduos FC (Figura 5-42), juntamente com os testes de normalidade (Tabela 5-6), admitem distribuição Normal para os resíduos gerados pelo modelo FC. Enquanto o teste de Jarque-Bera continua rejeitando a hipótese nula de normalidade para os resíduos Vz. Note que no caso de rejeição da hipótese de normalidade¹⁷, os testes de hipótese e os valores da estatística t do modelo podem ser afetados, e por isso, passam a ter justificativa apenas assintótica para grandes amostras.

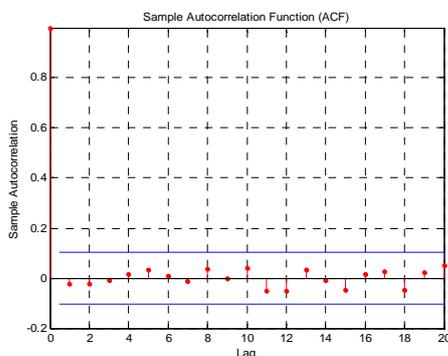


Figura 5-44 – Autocorrelação resíduos FC.

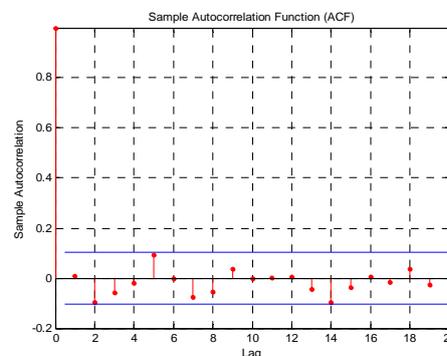


Figura 5-45 – Autocorrelação quadrado dos resíduos FC.

Os resíduos FC não possuem autocorrelações significativas nas Figura 5-44 e Figura 5-45. Este resultado é confirmado pelos testes de autocorrelação Box-Pierce-Ljung e LM (Tabela 5-6).

¹⁷Na prática, dificilmente todas as pressuposições serão satisfeitas, e pequenas transgressões à pressuposição de que os erros têm distribuição Normal são usuais. Em situações com tamanho de amostra grande, as inferências sobre médias não são substancialmente afetadas.

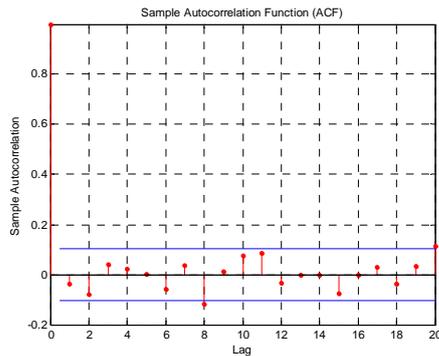


Figura 5-46 – Autocorrelação resíduos Vz.

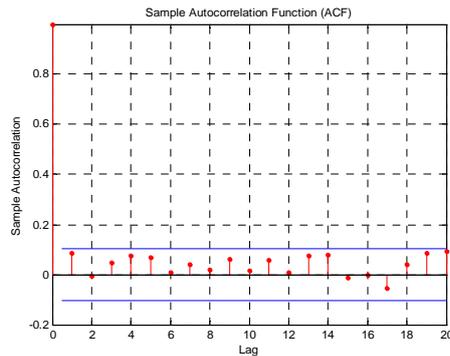


Figura 5-47 – Autocorrelação quadrado dos resíduos Vz.

A FAC do quadrado dos resíduos Vz indicava autocorrelação significativa na primeira defasagem, e conseqüentemente, efeito de heterocedasticidade condicional nos resíduos. Essa hipótese, finalmente, foi descartada pela série corrigida (Figura 5-32 e Figura 5-47), confirmando que o efeito foi provocado pela presença de *outliers* na série.

Como conclusão, a correção da série Vz foi importante na eliminação de efeitos espúrios causados por *outliers*. Um grau de sensibilidade maior no teste de detecção de *outliers* selecionaria uma quantidade maior de pontos e possibilitaria o algoritmo de correção a executar maiores alterações na série. Acredita-se que, desta forma, a distribuição dos resíduos perderia a cauda acentuada à direita. No entanto, preferiu-se realizar as alterações mínimas necessárias na série Vz.

5.3.2

Validação

Todo o processo de modelagem e ajuste dos modelos desenvolvido até aqui, utilizou o período entre Jan/1976 a Jul/2007. O período entre Ago/2007 e Jul/2009 foi reservado para testar a capacidade preditiva do modelo. Este também é denominado período *out-of-sample* ou período de validação.

Nesta etapa, as medidas de aderência (Seção 4.2.7) auxiliam a quantificar a distância entre os valores estimados e o valor real observado, e avaliar o desempenho de um modelo contra uma variedade de critérios, em particular o cálculo do RMSE das previsões será enfatizado. Essa medida está estabelecida na

literatura de modelos de previsão de energia eólica entre 10-15% da capacidade instalada para um horizonte de 36 horas. A comparação entre modelos não é trivial e nem sempre é viável. Por isso este valor deve servir apenas como uma medida base de avaliação, devido à escassez de informações acerca de modelos preditivos nesta área. No caso particular deste estudo, a informação da capacidade instalada da usina não está disponível, mas fica a observação para trabalhos futuros. Abaixo, as tabelas exibem os valores previstos e seus intervalos de confiança ao nível de 95%.

Tabela 5-7– Validação do modelo FC.

Data	Limite Inferior IC 95%	Previsão FC	Limite Superior IC 95%	Valor Real FC
2007/08	54.87	57.78	60.69	60.00
2007/09	53.63	56.49	59.35	62.00
2007/10	56.50	59.42	62.33	57.00
2007/11	50.82	53.99	57.16	56.00
2007/12	48.09	51.61	55.13	54.00
2008/01	41.11	44.89	48.66	38.00
2008/02	30.04	33.73	37.41	27.00
2008/03	22.20	25.66	29.11	5.00
2008/04	9.98	13.69	17.39	18.00
2008/05	24.77	28.07	31.38	29.00
2008/06	38.81	41.86	44.91	47.00
2008/07	49.93	52.92	55.91	53.00
2008/08	53.09	55.96	58.83	50.00
2008/09	54.03	56.88	59.73	56.00
2008/10	52.69	55.35	58.00	59.00
2008/11	53.93	56.65	59.38	55.00
2008/12	46.32	49.19	52.07	47.00
2009/01	36.81	39.95	43.10	35.00
2009/02	23.15	26.70	30.25	16.00
2009/03	14.30	18.04	21.78	11.00
2009/04	10.63	13.95	17.26	6.00
2009/05	12.63	16.50	20.37	10.00
2009/06	28.87	32.32	35.77	33.00
2009/07	37.96	41.03	44.10	42.00

Tabela 5-8– Validação do modelo Vz.

Data	Limite Inferior IC 95%	Previsão Vazao	Limite Superior IC 95%	Valor Real Vazao
2007/08	25.78	31.38	36.99	28.00
2007/09	21.36	27.76	34.16	21.00
2007/10	24.85	32.00	39.16	43.00
2007/11	63.36	71.49	79.61	65.00
2007/12	60.99	69.93	78.87	52.00
2008/01	49.70	58.62	67.55	61.00
2008/02	79.37	87.02	94.67	120.00
2008/03	78.84	85.20	91.56	99.00
2008/04	75.62	82.29	88.97	98.00
2008/05	58.17	64.24	70.32	59.00
2008/06	45.59	50.65	55.70	52.00
2008/07	24.94	29.88	34.82	34.00
2008/08	32.59	37.73	42.87	43.00
2008/09	32.52	37.89	43.25	36.00
2008/10	35.60	41.19	46.78	51.00
2008/11	61.82	67.91	74.00	80.00
2008/12	77.26	83.96	90.66	88.00
2009/01	92.33	99.43	106.53	107.00
2009/02	115.27	122.30	129.34	139.00
2009/03	86.45	92.94	99.42	97.00
2009/04	84.13	89.98	95.83	78.00
2009/05	42.05	48.95	55.85	55.00
2009/06	37.71	43.92	50.13	51.00
2009/07	37.76	42.69	47.63	63.00

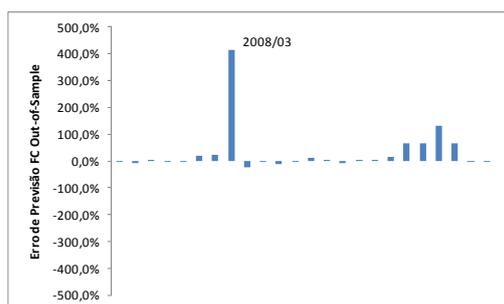


Figura 5-48 – Erros de previsão (%) FC.

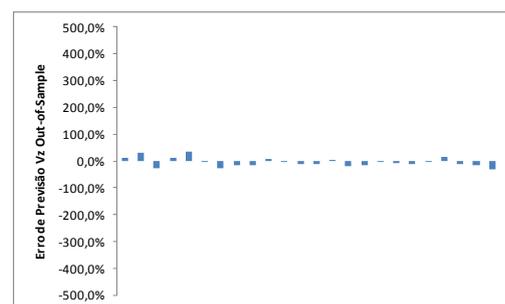


Figura 5-49 – Erros de previsão (%) Vz.

Na Tabela 5-7 e na Figura 5-48, o erro de previsão na observação de Mar/2008 merece atenção, pois alcança quase 400% em relação ao valor real. Nesta data, o fator de capacidade observado atingiu um valor atípico muito baixo, o que justifica a inabilidade do modelo em prevê-lo. Outros valores baixos observados

também são responsáveis por erros de previsão mais elevados, sugerindo que o modelo de previsão para FC apresenta alguma dificuldade em capturar os valores muito baixos. A título de comparação, o modelo FC, antes da correção de *outliers* aplicada à série Vz, apresentava o mesmo problema em proporções maiores. Em geral, as previsões, tanto para FC quanto para Vz, apresentaram melhoras no período de validação após a correção de *outliers*.

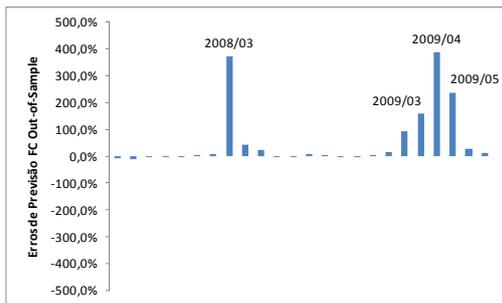


Figura 5-50 – Erros de previsão (%) FC- antes da correção.

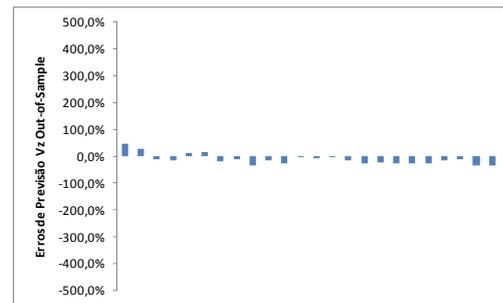


Figura 5-51 – Erros de previsão (%) Vz- antes da correção.

O fator que quantifica a média percentual dos erros de previsão (MPE) é alto para FC, aproximadamente 31,4%. Esse valor é extremamente influenciado pela observação irregular de Mar/2008. Ao retirar essa observação do cálculo, o MPE cai para 14,8%. Para a série Vz, o MPE é -4,8%. Outras medidas estão descritas na tabela abaixo:

Tabela 5-9– Medidas de aderência FC e Vz.

	FC	Vz
RMSE (Root Mean Square Error)	6.3532	11.9087
MAE (Mean Absolute Error)	4.6834	9.4996
MAPE (Mean Absolute Percentage Error)	37.23%	14.65%
MPE (Mean Percentage Error)	31.44%	-4.79%
SDPE (Standard Deviation Percentage Error)	0.8649	0.1662

O coeficiente de determinação R^2 , definido como o grau de ajuste do modelo ao conjunto de dados, pode ser calculado para o período de validação e é uma medida muito útil para quantificar a proporção da variação da série real representada nas previsões do modelo. O R^2 no período de validação de FC é 69%, e para Vz essa medida sobe para 77%.

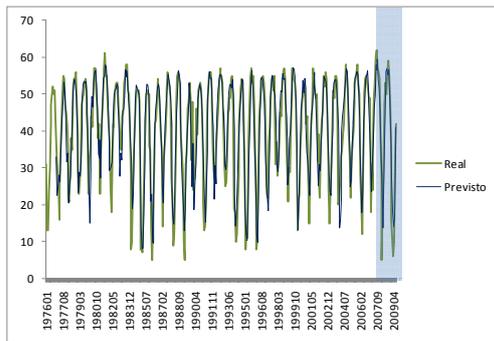


Figura 5-52 – Desempenho do modelo FC - total.

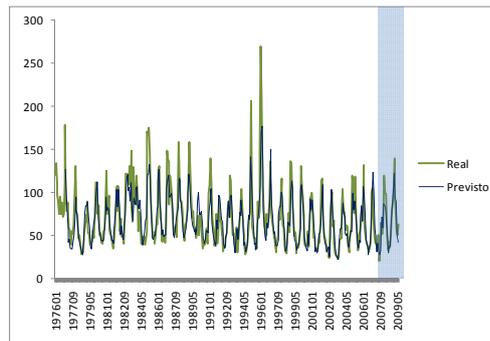


Figura 5-53 – Desempenho do modelo Vz - total.

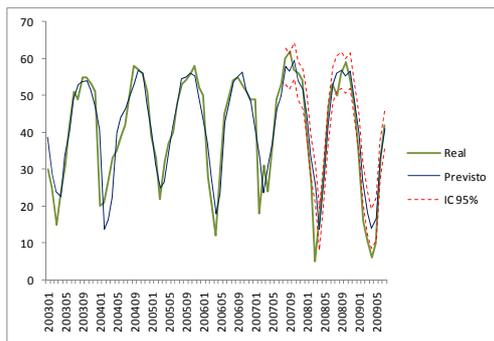


Figura 5-54 – Desempenho do modelo FC (IC 95%).

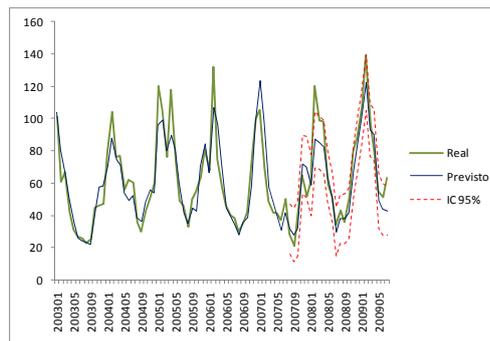


Figura 5-55 – Desempenho do modelo Vz (IC 95%).

Um modelo construído para fins de simulação deve ter sua capacidade preditiva avaliada antes de ser utilizado na geração de séries sintéticas. Em linhas gerais, o modelo proposto VARX, estimado por MQP e com correção de *outliers*, apresentou resultados satisfatórios e um bom desempenho para o horizonte de previsão de dois anos, correspondente ao período de validação adotado. O modelo concentra as principais características das variáveis de interesse e de suas interações, controlado pelo impacto da energia acumulada em cada subsistema do Brasil. Neste contexto, o modelo tem o objetivo de captar os efeitos de mudanças nos valores das variáveis e poderá ser utilizado na geração de cenários conjuntos, que será detalhado no capítulo seguinte.