

P&D UTE Norte Fluminense/PUC-Rio

Relatório 1: Modelo Estatístico de Vento e Vazão e Especificação de Contratos de Energia

Prof. Alexandre Street, *Coordenador*, Prof. Delberis A. Lima, *Gerente*, Prof. Álvaro Veiga, Bianca Amaral, Bruno Fânzeres, Lucas Freire.

Resumo — Os objetivos deste primeiro relatório são (i) descrever o escopo geral do projeto, (ii) apresentar o desenvolvimento e os resultados do modelo estatístico de vento e vazão, assim como a simulação de cenários para a evolução futura dessas grandezas, (iii) especificar contratos de energia a serem utilizados no modelo de otimização e (iv) apresentar e discutir brevemente o planejamento das atividades dos próximos quadrimestres.

Palavras-chave — contratos de energia, energia renovável, modelagem estatística.

I. INTRODUÇÃO

A principal motivação deste projeto decorre do estímulo que o aquecimento global tem provocado no uso de fontes renováveis de energia elétrica para a expansão da oferta de geração de energia. As três principais fontes renováveis de energia no Brasil – pequenas centrais hidrelétricas (PCHs), eólica e cogeração a biomassa (a partir do bagaço da cana-de-açúcar) – têm dificuldades em estabelecer contratos com consumidores no ambiente de comercialização livre (ACL) em função da sazonalidade e incerteza presentes na disponibilidade de seus recursos/insumos (vento e vazão para o caso da PCH e eólica). Essa sazonalidade pode gerar cenários de indisponibilidade de recursos, fazendo com que o gerador contratado tenha que comprar o déficit de produção no mercado de curto prazo, se expondo a alta volatilidade de seus preços. Contudo, essas fontes apresentam uma forte complementaridade sazonal de seus recursos, podendo ser "combinadas" de maneira a mitigar esse risco e viabilizar a venda de contratos no ACL. O efeito prático e de médio e longo prazo dessa estratégia é o aumento da competitividade dessas fontes e, por consequência, uma maior penetração na matriz energética brasileira.

O objetivo deste projeto é desenvolver uma metodologia de comercialização ótima de energia proveniente das três fontes renováveis supracitadas de maneira integrada. Essa metodologia tem como foco a comercialização de energia no ACL e será implementada de modo a gerar uma ferramenta (software) de apoio à decisão. Essa ferramenta levará em conta as principais características dessas fontes e indicará, como principal resultado (saída do modelo), um percentual de contratação das oportunidades (contratos) cadastradas pelo usuário.

O primeiro dos dois principais desafios presentes no escopo deste projeto é (i) a elaboração de uma metodologia de especificação de um modelo estatístico bivariado que relate vento e vazão incluindo como variável explicativa a Energia Natural Afluente (*ENA*) do sistema, o que representa uma forma de obtermos uma relação entre o modelo e o Newave. A partir desse modelo, realizaremos a geração de cenários coerentes de produção, isto é, cenários de disponibilidade dos recursos naturais (vento e vazão) das usinas que sejam correlacionados com os cenários de preço de liquidação de diferenças (*PLD*) ou preço de curto prazo advindos do Newave. O segundo desafio passa pela (ii) especificação de expressões matemáticas que modelem a renda de quatro modalidades de contratos bilaterais de compra e venda de energia: contratos de quantidade, contratos de quantidade com cláusulas *Take or Pay* (*ToP*), contratos de quantidade baseados no *PLD* com piso e teto e contratos de disponibilidade.

Os cenários gerados, junto com o preço *spot*, alimentarão um modelo de programação matemática que determinará a melhor estratégia de comercialização dos contratos especificados (cadastrados pelo usuário), maximizando uma função de valor da renda líquida do comercializador. O perfil de risco será incorporado de diferentes maneiras: através de restrições de risco máximo aceitável e/ou através de uma penalidade na função objetivo.

Além disso, a partir da simulação de vento e vazão coerentes com o preço de curto prazo, outras decisões que não estejam contempladas no escopo do modelo poderão ser implementadas e analisadas pelo decisor com facilidade. Por fim, outras aplicações podem ser pensadas com o objetivo de gerar impactos muito além do resultado associado à venda da energia. Podemos citar o

caso de um futuro amadurecimento do mercado de créditos de carbono em que estes representariam uma renda extra para a empresa. Nesse caso, o modelo estatístico se torna peça-chave dessa avaliação, pois é capaz de gerar cenários de redução de emissões casados (estatisticamente correlacionados) com os cenários de despacho das usinas do sistema.

O restante deste documento está organizado da seguinte forma: na seção II, apresentamos o fluxograma do funcionamento do modelo proposto; na seção III, apresentamos o modelo estatístico para vento e vazão e mostramos os resultados parciais de aderência e simulação de cenários para um conjunto de dados específicos; na seção IV, falamos sobre os quatro tipos de contrato que serão considerados pelo modelo de otimização, suas respectivas expressões de renda líquida e a expressão final da renda da comercializadora; por fim, na seção V, apresentamos o cronograma de tarefas, falamos sobre os problemas encontrados nesta primeira etapa e discutimos os próximos passos do projeto.

II. FLUXOGRAMA DA PLATAFORMA

O produto decorrente da metodologia desenvolvida neste projeto é a plataforma de apoio à decisão que visa auxiliar na contratação ótima de portfólios que contenham contratos lastreados em fontes de energia renovável. Essa plataforma é constituída por duas ferramentas: um simulador de cenários de disponibilidade dos recursos naturais de fontes renováveis (vento e vazão) e um modelo de otimização de contratos com aversão a risco. A plataforma opera como um programa de computador, sendo capaz de receber dados externos, inseridos pelo usuário; gravar estudos de caso; e ler estudos antigos para serem editados e reestudados.

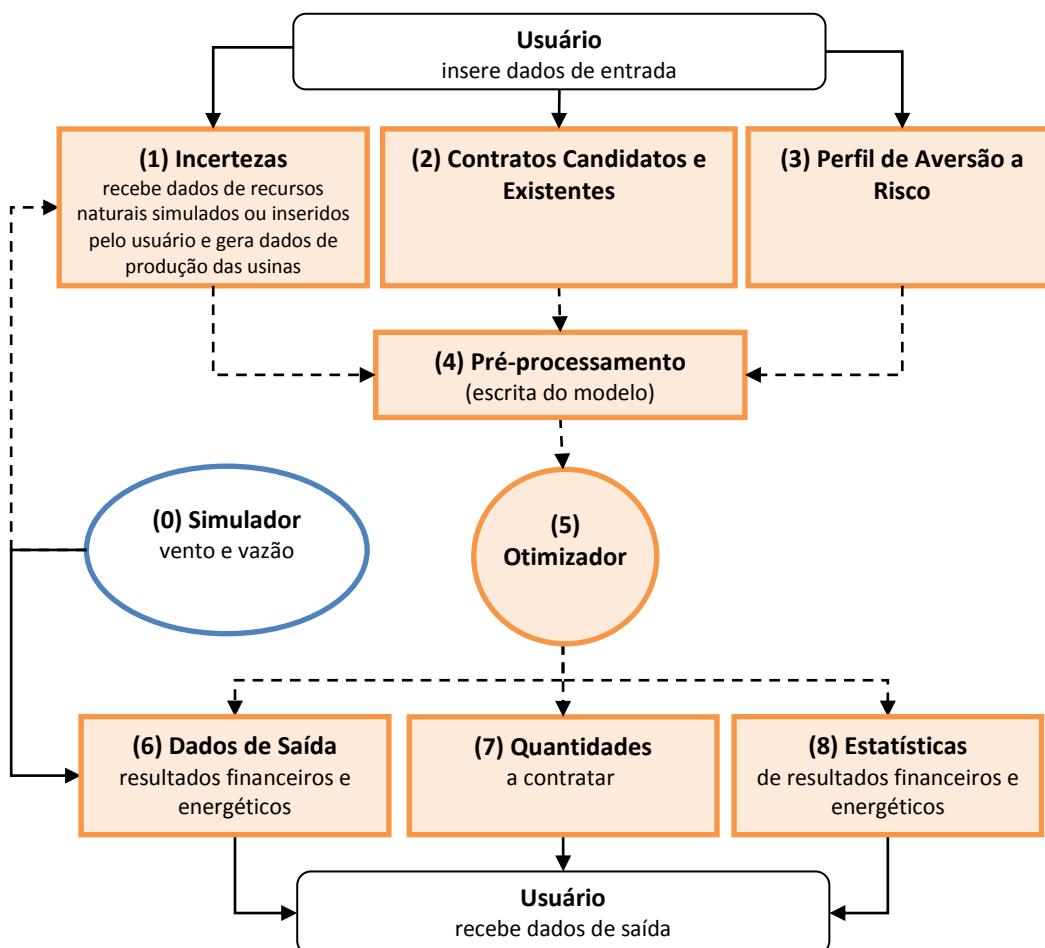


Figura II-1: Fluxograma da Plataforma de Apoio à Decisão.

A Figura II-1 detalha a plataforma de apoio à decisão que será desenvolvida como resultado final do projeto. Os blocos (1)-(3) representam a interação entre a plataforma e o usuário. Nele, o usuário insere os dados relativos às oportunidades de contrato que deseja analisar se inclui ou não em seu portfolio e os dados dos contratos já firmados, bloco (2), o perfil de aversão a risco que deseja adotar, bloco (3), e os cenários que caracterizam as incertezas que afetam os contratos considerados, bloco (1). Os dados de entrada podem ser gravados e posteriormente editados para um próximo estudo.

O bloco (0) representa o modelo de geração de cenários de vento e vazão que pode ou não ser utilizado para alimentar o bloco (1). O bloco (0) constitui um módulo separado do modelo de comercialização ótima, mas integrado à plataforma de apoio à decisão. Nele, o usuário pode cadastrar diversas séries de dados referentes a diferentes usinas e gerar cenários futuros (simulações) da disponibilidade dos seus respectivos recursos (vento e vazão). Posteriormente, no módulo de comercialização, dentro do escopo do bloco (1), o usuário pode selecionar um conjunto de dados simulados e armazenados pelo bloco (0).

As setas pontilhadas representam o fluxo de informação que ocorre automaticamente dentro da plataforma e não se destinam ou originam do usuário. As demais setas representam saídas ou entradas de dados cuja interferência do usuário é necessária. No contexto do fluxo de informações internas ao modelo, o bloco (4) desempenha um papel preponderante. Ele recebe todas as informações do estudo e escreve, de maneira automática, as informações que servirão de entrada para o modelo de otimização. Nessa etapa, todos os dados de entrada dos contratos informados pelo usuário serão convertidos em cenários de futuras realizações de fluxo de caixa de acordo com as respectivas modalidades de cada contrato. Esses dados serão entregues, através de arquivos, para o módulo de otimização, bloco (5), que por sua vez retornará os resultados de saída ao usuário, blocos (6)-(8).

O bloco (6) é constituído pelos arquivos de saída: arquivos com cenários de futuras realizações de fluxo de caixa do portfolio ótimo e as respectivas quantidades de energia liquidadas no curto prazo. O bloco (7) apresenta os resultados de maneira visual (gráficos e tabelas) para o usuário e, essencialmente, mostra a decisão ótima de compra ou venda (quantidade) de cada contrato candidato. Por fim, o bloco (8) apresenta algumas estatísticas relativas aos resultados financeiros e energéticos do portfolio, também de maneira visual, através de gráficos e tabelas.

Na próxima seção, o modelo estatístico relativo ao bloco (0) será apresentado e testado para um conjunto de dados realistas do setor elétrico.

III. MODELO ESTATÍSTICO DE VENTO E VAZÃO

Esta seção apresenta o desenvolvimento de um modelo de séries temporais bivariado para a vazão afluente (V_z) a uma PCH e o fator de capacidade de uma usina eólica (FC), sendo que o FC é proporcional à energia produzida pela força do vento incidente na turbina eólica. Esse modelo será utilizado para gerar cenários de evolução futura dessas grandezas. Esses cenários, juntamente com os cenários correspondentes do preço da energia no mercado aberto – o chamado preço spot – vão alimentar o módulo de otimização estocástica, que determina a melhor forma de contratação dessa energia.

Os cenários de preço spot podem ser obtidos através do programa Newave, utilizado pelo ONS. Esse programa se destina a determinar a geração ótima de cada classe de gerador em cada um dos quatro submercados do sistema brasileiro. Para isso, o Newave gera cenários de trajetórias futuras para a Energia Natural Afluente (ENA), otimiza a operação do sistema para cada trajetória e calcula o preço spot semanal, dado pelo custo de produção de 1MW adicional.

O preço tenderá a ser menor para as trajetórias com ENAs maiores, já que a energia proveniente de centrais hidroelétricas, que são despachadas primeiro por ordem de mérito, tem um preço menor. Ao mesmo tempo, como a ENA representa a energia que pode ser obtida pelo turbinamento de toda a vazão do período pelas usinas hidrelétricas existentes no sistema, espera-se poder observar uma correlação entre as ENAs, a vazão de PCHs e o vento, ou o FC , em pontos específicos. Sendo assim, para cada trajetória das ENAs teremos uma trajetória para o preço spot, ambas produzidas pelo Newave, e uma ou mais trajetórias para o par V_z-FC , produzidas pelo modelo desenvolvido nesta seção.

Na subseção A, descrevemos as variáveis dependentes, enquanto, na subseção B, descrevemos as variáveis exógenas. Na subseção C, as componentes do modelo serão analisadas individualmente. Na subseção D, o modelo VARX será abordado como a primeira tentativa de modelagem, seguido por

seus resultados e testes de aderência (na *subseção E*). Na *subseção F*, desenvolvemos uma modelagem alternativa que aplica a lei de variância na estimativa dos parâmetros em resposta aos problemas encontrados. Na *subseção G*, obtém-se seus resultados e seus respectivos testes de aderência. Por fim, na *subseção H*, os cenários simulados são ilustrados.

A. Descrição das variáveis dependentes: Vz e FC

O modelo desenvolvido nesta subseção tem duas variáveis dependentes: a vazão afluente a uma PCH, Vz , e o fator de capacidade eólica, FC , referente a alguma localidade onde há ou poderá ser instalada uma usina eólica (UE).

- Vazão – Vz

Vz representa a vazão média mensal medida em m^3/s em um ponto específico de um rio onde existe ou pode vir a existir uma PCH. Como uma PCH não tem reservatório, toda a vazão afluente será turbinada. Sendo assim, a energia produzida será dada pela multiplicação da vazão por um fator de produtividade (FP), em MWs/m^3 , da PCH.

- Fator de Capacidade Eólica – FC

A relação entre a força do vento e a energia produzida por uma UE é complexa. Por isso, não iremos modelar diretamente a força do vento, usaremos o Fator de Capacidade Eólica, FC , que é a razão entre a energia total produzida (E_a) num período T (730 horas em um mês, por exemplo) de uma UE e a energia que seria produzida caso a turbina eólica trabalhasse 100% do tempo em sua potência nominal (P_n) ao longo desse mesmo período T , isto é, o produto $T \cdot P_n$. O FC é, portanto, uma variável adimensional.

O Fator de Capacidade Eólica é definido como (Martinez, 2003):

$$FC = \frac{E_a}{T \cdot P_n}. \quad (1)$$

Portanto, para obter a energia produzida por uma turbina ou um conjunto de turbinas específicas ao longo de um período T , basta multiplicar o FC pela potência nominal e pelo período correspondente.

B. Variáveis exógenas: ENA

No modelo proposto, as variáveis dependentes Vz e FC são expressas como uma função das ENAs observadas no mesmo instante nos quatro submercados do sistema brasileiro. A ENA de um submercado referente a um dado período, no nosso caso uma semana, é calculada como a energia que se teria obtido caso toda a vazão do período fosse turbinada pelas hidrelétricas existentes naquele submercado.

É importante notar que a ENA depende da configuração das hidrelétricas consideradas. Sendo assim, é fundamental que, na etapa de simulação das ENAs, o Newave gere cenários sob a mesma configuração usada para calcular as ENAs históricas utilizadas na etapa de estimativa dos parâmetros do modelo vento-vazão. Ao mesmo tempo, para gerar cenários de preços *spots* futuros, o Newave deverá considerar a entrada de novas usinas, pois o preço, naturalmente, depende da capacidade instalada das hidrelétricas.

Em resumo, o Newave será utilizado duas vezes. Primeiramente, considerando uma configuração fixa e igual àquela utilizada para transformar as vazões históricas em ENAs. Os cenários de ENAs serão então utilizados para gerar os cenários condicionais de Vz e FC . Numa segunda etapa, o Newave considera a entrada de novas usinas e gera cenários de ENAs, que serão ignorados, e de preços, que serão utilizados no modelo de otimização. Todo cuidado deve ser tomado para que os cenários de ENAs oriundos da configuração fixa sejam associados de forma correta com os preços gerados a partir de uma configuração dinâmica, de maneira a manter a correspondência entre os cenários de ENAs e de preços.

C. Dados utilizados para ajustar o modelo

Para o desenvolvimento do modelo, utilizamos as seguintes séries: *FC* de uma usina eólica em local não-determinado no Nordeste e a vazão da região de Paraibuna. Os dados têm frequência mensal com data inicial em janeiro de 1976 e data final em julho de 2009, totalizando 403 registros. As variáveis exógenas empregadas correspondem a quatro séries denominadas *ENAs*, que representam os valores mensais acumulados de energia produzida nos quatro subsistemas (Sudeste, Sul, Nordeste e Norte) no mesmo período.

Para o ajuste dos parâmetros dos modelos que serão vistos a seguir, utilizamos o período de janeiro de 1976 a julho de 2007, que chamamos de Dentro da Amostra. Os dados correspondentes ao período restante – agosto de 2007 a julho de 2009 – foram reservados para testar a capacidade preditiva do modelo, logo, não foram utilizados no ajuste dos modelos. Nós nos referimos a esse período como Fora da Amostra.

a) Fator de Capacidade Eólica e Vazão

O primeiro passo na análise de uma série temporal é, em geral, a sua representação gráfica. O gráfico dá uma ideia inicial da provável natureza das séries temporais. Essa percepção intuitiva é o ponto de partida para a aplicação de testes mais formais.

Na Figura III-1 é possível notar que as séries não apresentam tendência de crescimento ou queda, o que é um bom indicativo de estacionariedade. Além disso, uma rápida análise indica a repetição de um padrão ao longo de um ano. Esse comportamento sazonal já era esperado, pois estamos tratando séries climáticas. Contudo, o que nos chama a atenção é o comportamento complementar entre as séries: a queda de produção de uma coincide com o pico da outra. A Figura III-2 mostra a relação entre as séries estudadas e como os maiores valores de *FC* estão concentrados no intervalo de baixa vazão.

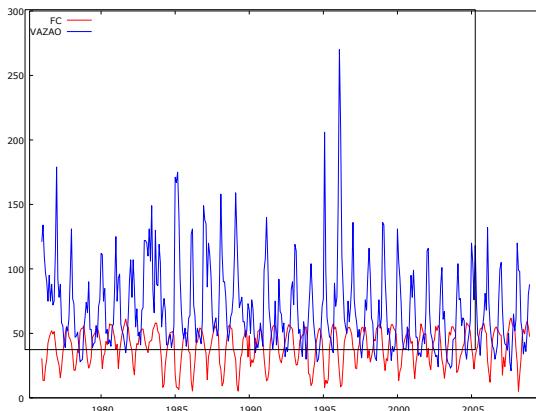


Figura III-1: *Vz* e *FC*, Paraibuna, 1976-2009 (mensais).

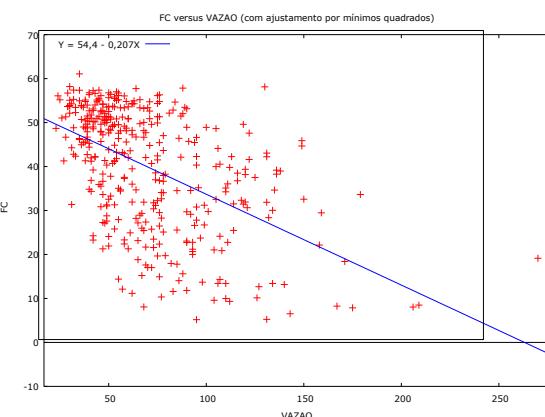


Figura III-2: *FC* x *Vz*, com ajuste por mínimos quadrados.

As figuras abaixo (Figuras III-3 e III-4) representam os histogramas das séries. Só de observar o formato desses histogramas, podemos intuir que o processo gerador das séries não segue uma distribuição Normal. Formalmente, utilizamos o teste de normalidade Jarque-Bera (Hamilton, 1994) para verificar essa intuição. No presente caso, o teste foi aplicado e a Hipótese Nula, de normalidade, foi rejeitada em ambos os casos (*p*-valor insignificante, menor que 0,05).

Nas figuras seguintes (Figuras III-5 e III-6), em que cada linha representa a trajetória da grandeza ao longo dos 12 meses, mostramos o comportamento anual das séries. Podemos observar uma variância mais alta nos meses iniciais do ano, tanto para a *Vz* quanto para o *FC*, além da já mencionada complementaridade entre as fontes.

Mais adiante (Figuras III-7 e III-8), apresentamos a função de autocorrelação (fac) e autocorrelação parcial (facp) para as séries de *FC* e *Vz*. Essa função é de extrema importância na especificação do modelo a ser adotado e para a análise do comportamento temporal da série. Dessa forma, o movimento infinito senoidal da fac em conjunto com o movimento finito senoidal da facp sugere modelos autorregressivos, tanto para o *FC* quanto para a *Vz*.

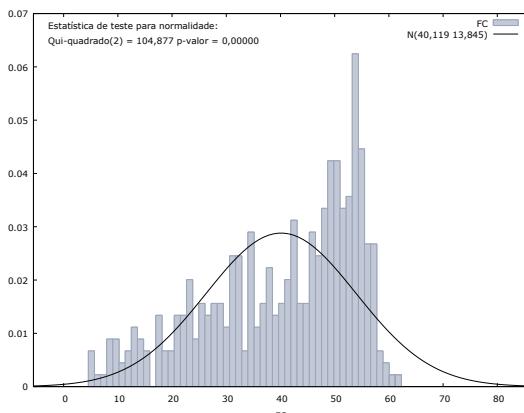


Figura III-3: Histograma e Teste de Normalidade para FC .

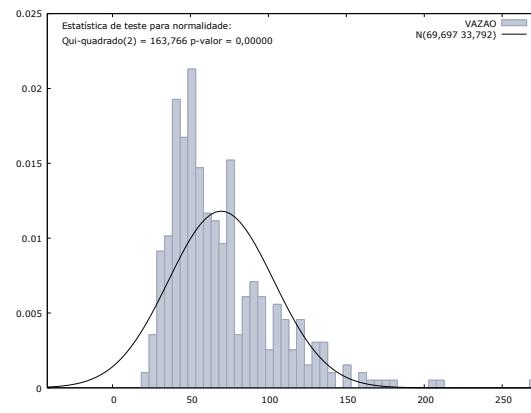


Figura III-4: Histograma e Teste de Normalidade para Vz .

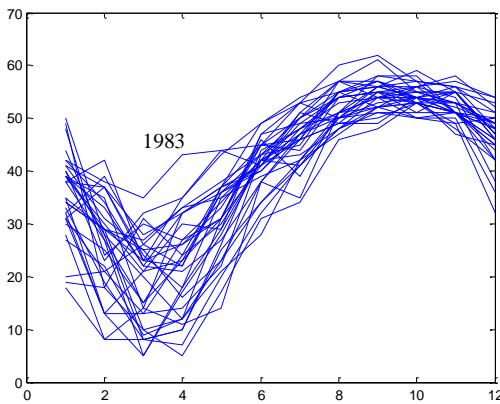


Figura III-5: Comparativo Anual FC .

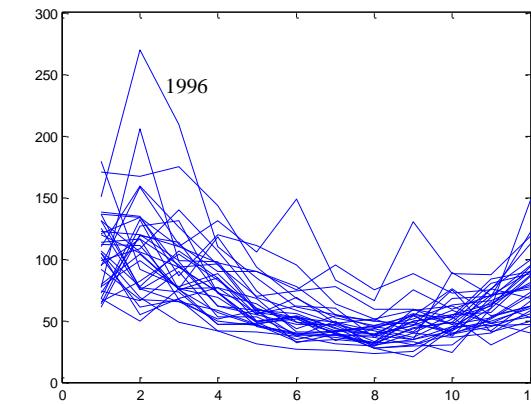


Figura III-6: Comparativo Anual Vz .

Estatísticas Descritivas	FC	Vz
Média	40,119	69,697
Mediana	43,421	60
Mínimo	51,223	21
Máximo	61,738	270
Desvio padrão	13,845	33,792
C.V.	0,34509	0,48484
Assimetria	-0,68291	15817
Curtose	-0,55989	40115

Quadro III-1: Estatísticas Descritivas

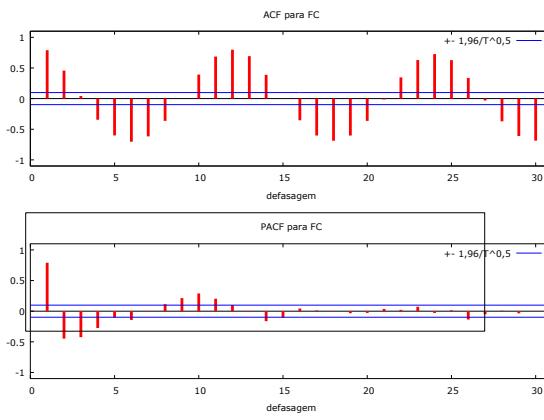


Figura III-7: Correlograma para série FC .

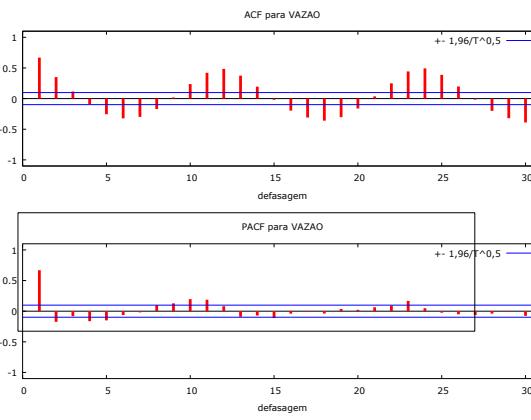


Figura III-8: Correlograma para série Vz .

b) Variáveis Exógenas – ENAs

As ENAs representam a produtibilidade acumulada de energia afluente em cada região do Brasil. Mais uma vez, faremos uma rápida análise do comportamento das variáveis envolvidas, dessa vez das variáveis exógenas analisadas na configuração fixa.

i) Sudeste – SE

A Figura III-9 representa o histórico da ENA-Sudeste e apresenta bons indícios de estacionariedade. Como já eram esperados, os dados desse subsistema são os que mais se aproximam do comportamento da série V_z , uma vez que esses dados foram coletados da região de Paraibuna, que fica no Sudeste. A Figura III-10 reforça a ideia que apresentamos na subseção anterior: os primeiros meses do ano possuem maior variabilidade.

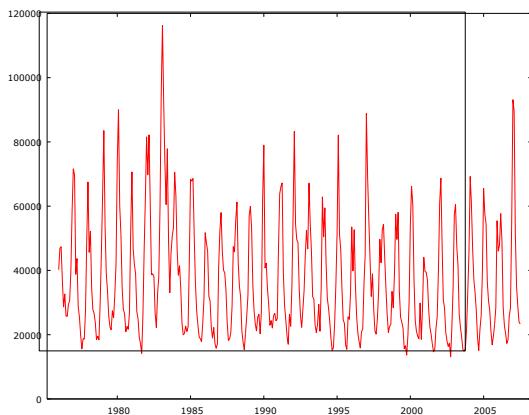


Figura III-9: Histórico ENA-SE.

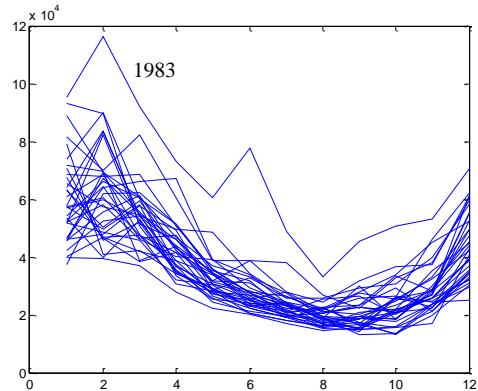


Figura III-10: Comparativo Anual ENA-SE.

ii) Sul – S

A Figura III-11 representa o histórico da ENA-Sul, também com indícios de estacionariedade. A Figura III-12 mostra um comportamento mais regular ao longo dos anos, com exceção do pico em 1983.

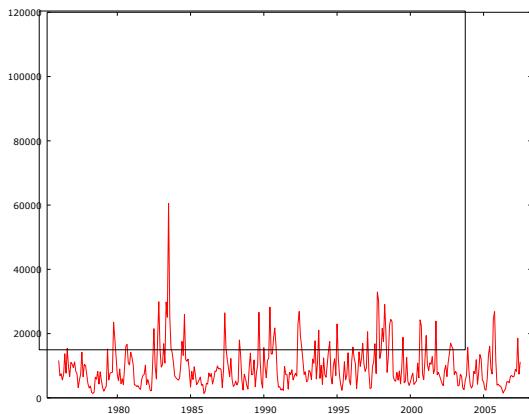


Figura III-11: Histórico ENA-S.

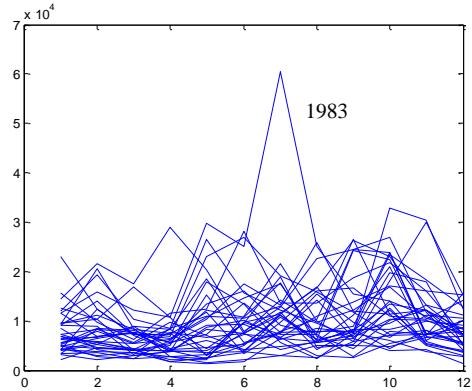


Figura III-12: Comparativo Anual ENA-S.

iii) Nordeste – NE

A Figura III-13 representa o histórico da ENA-Nordeste, também com indícios de estacionariedade. A Figura III-14 mostra uma variância muito pequena no período que vai de junho a outubro e muita alta nos meses iniciais do ano.

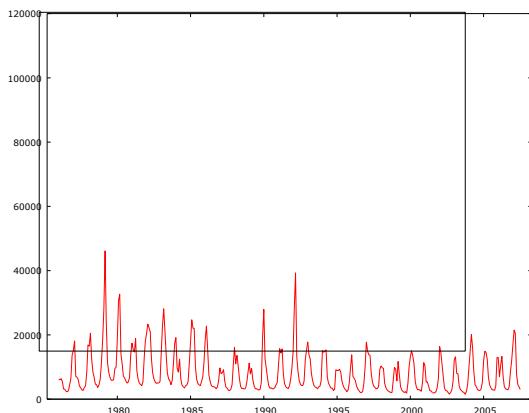


Figura III-13: Histórico ENA-NE.

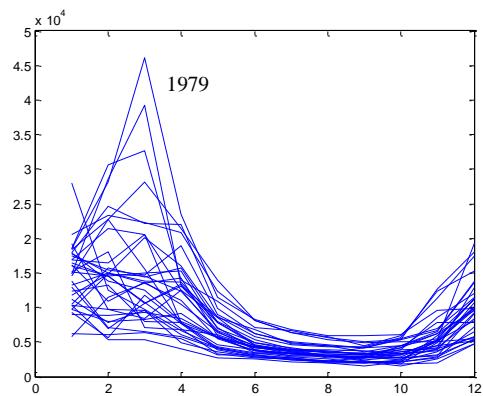


Figura III-14: Comparativo Anual ENA-NE.

iv) Norte - N

A Figura III-15 representa o histórico da ENA-Norte, também com indícios de estacionariedade. A Figura III-16 também chama atenção para o período que vai de junho a outubro com uma variância muito pequena, o mesmo não ocorre nos primeiros meses dos anos.

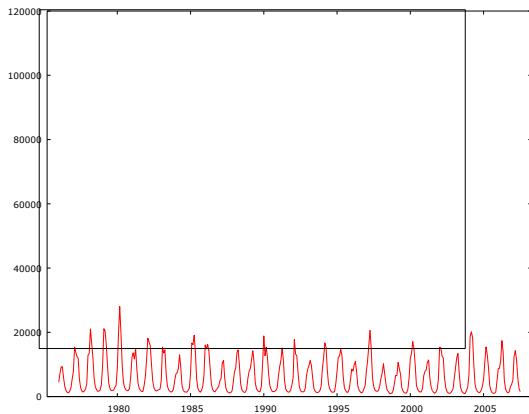


Figura III-15: Histórico ENA-N.

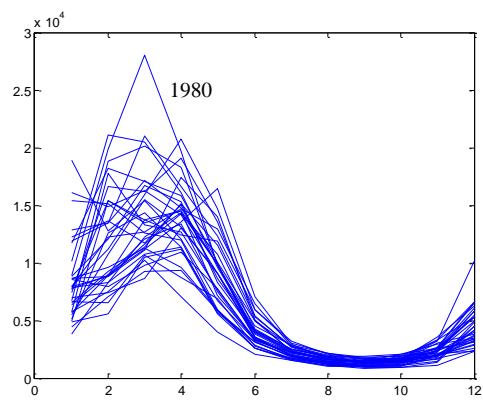


Figura III-16: Comparativo Anual ENA-N.

c) Teste Dickey-Fuller Aumentado (ADF)

Uma série temporal pode ser encarada como uma realização de um processo estocástico. Portanto, inferir sobre todos os parâmetros envolvidos no modelo com base em apenas uma realização desse processo é uma tarefa arriscada, sendo necessárias algumas simplificações. A principal hipótese simplificadora corresponde à estacionariedade, que estabelece que as propriedades estocásticas não se alteram ao longo do tempo.

Como a maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias, um procedimento comum para transformar uma série não estacionária em estacionária é a sua diferenciação sucessiva. A primeira diferença de uma série y_t é definida por:

$$\Delta y_t = y_t - y_{t-1}. \quad (2)$$

A segunda diferença é:

$$\begin{aligned} \Delta^2 y_t &= \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}); \\ \Delta^2 y_t &= \Delta y_t - \Delta y_{t-1}; \\ \Delta^2 y_t &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned} \quad (3)$$

De modo geral, a n -ésima diferença de y_t é $\Delta^n y_t = \Delta(\Delta^{n-1} y_t)$. Normalmente, uma diferenciação apenas basta para se obter uma série estacionária.

Se uma série deve ser diferenciada d vezes antes de tornar-se estacionária, então ela contém, originalmente, d raízes unitárias. O teste mais comum na literatura para identificar a presença de raízes unitárias em uma série temporal é o *Teste Dickey-Fuller Aumentado (ADF)* (Hamilton, 1994). Com ele, testamos a presença de raiz unitária na série (H_0) contra a ausência (H_a) indicando estacionariedade.

O teste aplicado às séries de Vz, FC e ENAs resultou em um *p-valor* inferior a 0,00001. Consequentemente, rejeitamos fortemente a Hipótese Nula de não estacionariedade das séries.

D. Modelagem estatística - VARX

O primeiro teste de modelagem recorre à técnica estatística que incorpora o uso de modelos multivariados. Essa técnica vem da classe de modelos autorregressivos vetoriais de ordem p com variáveis exógenas, denotado pelo acrônimo *VARX(p)* (*vector autoregression with external variables*).

O modelo *VARX* é uma generalização para séries temporais múltiplas do modelo autorregressivo univariado, em que as regressões obedecem à mesma estrutura para todas as equações individuais. Nesse modelo, cada variável endógena, ou dependente, é explicada por seus valores defasados e pelos valores defasados das demais variáveis, endógenas e exógenas, presentes.

Analiticamente é possível representar o modelo *VARX(p)*, através de notação matricial, da seguinte forma:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \delta_0 w_t + \cdots + \delta_q w_{t-q} + \varepsilon_t, \quad (4)$$

onde:

$y_t = (y_{1t}, y_{2t}, y_{3t}, \dots, y_{nt})'$ é um vetor $(n \times 1)$ contendo o valor das séries no instante t ;

c é o vetor de constantes $(n \times 1)$;

ϕ_i $i = 1, 2, \dots, p$ são matrizes $(n \times n)$ de coeficientes autorregressivos;

$w_t = (w_{1t}, w_{2t}, w_{3t}, \dots, w_{mt})'$ representa um conjunto de m variáveis exógenas dispostas em um vetor $(m \times 1)$ contendo o valor dessas variáveis no tempo t ;

δ_j é uma matriz $(n \times m)$ de coeficientes para $j = 1, 2, \dots, p$;

ε_t forma uma sequência de vetores iid $(n \times 1)$ com:

$$E(\varepsilon_t) = 0;$$

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

Sendo Ω uma matriz simétrica $(n \times n)$ definida positiva.

Uma condição essencial para a convergência dos estimadores dos parâmetros do modelo é a estacionariedade das séries (Hamilton, 1994).

E. Metodologia e testes de aderência - VARX

A construção do modelo seguiu o ciclo de identificação, estimação e diagnóstico. O modelo identificado com base nos dados apresentados pode ser escrito da seguinte forma:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_{12} y_{t-12} + \delta_0 w_t + \delta_1 w_{t-1} + \varepsilon_t, \quad (5)$$

onde:

$y_t = (y_{1t}, y_{2t})'$ representa o conjunto de séries (FC e Vz) dispostas em um vetor (2×1) ;

c é o vetor de constantes (2×1) ;

ϕ_i é uma matriz (2×2) de coeficientes autorregressivos para $i = 1, 2, \dots, 12$;

$w_t = (w_{1t}, w_{2t}, w_{3t}, w_{4t})'$ representa um conjunto de variáveis exógenas (ENA-SE, ENA-S, ENA-NE e ENA-N) dispostas em um vetor (4×1) ;

δ_j é uma matriz (2×4) de coeficientes autorregressivos para $j = 0, 1$;

ε_t é um ruído branco vetorial de dimensão 2 com média nula e matriz de variâncias-covariâncias Ω .

Ao estimarmos o modelo proposto, obtemos as seguintes estatísticas apresentadas no Quadro III-2. As estatísticas de ajuste do modelo apresentam valores satisfatórios. Note que o modelo consegue explicar 83% da variação dos dados para *FC* e 69% para *Vz*.

	FC	Vazao
Soma dos resíduos quadrados =	10320	117841
Erro padrão dos resíduos =	5,55863	18,7834
R-quadrado não-ajustado=	0,849373	0,722555
R-quadrado ajustado=	0,834941	0,695974
Estatística F(32,334)=	58,8561*	27,18268*
Estatística de Durbin-Watson =	2,02757	1,94362
Coeficiente de autocorrelação de primeira-ordem =	-0,014301	0,0202622
Critério de informação de Akaike (AIC) =	2331,99	3225,73
Critério Bayesiano de Schwarz (BIC) =	2460,87	3354,6
Critério de Hannan-Quinn (HQC) =	2383,2	3276,93

* p-valor<0,00001

Quadro III-2: Resultados do Modelo Final.

	Resíduos FC	Resíduos Vz
Resíduos FC	$\sigma_{1,1}^2 = 30,89833$	$\sigma_{1,2}^2 = -24,56453$
Resíduos Vz	$\sigma_{2,1}^2 = -24,56453$	$\sigma_{2,2}^2 = 352,8180$

Quadro III-3: Matriz de Covariância Residual Ω .

	Resíduos FC	Resíduos Vz
Resíduos FC	$\sigma_{1,1}^2 = 1,000000$	$\sigma_{1,2}^2 = -0,235269$
Resíduos Vz	$\sigma_{2,1}^2 = -0,235269$	$\sigma_{2,2}^2 = 1,000000$

Quadro III-4: Matriz de Correlação Residual.

Para o período Fora da Amostra de avaliação (agosto de 2007 a julho de 2009) foram calculados os intervalos de confiança (ao nível de 95%) para as estimativas de previsão. Como podemos observar nas Figuras III-17 e III-18, as previsões para *Vz* mostraram-se mais precisas com relação à proximidade dos valores reais e preditos, e ainda em relação aos intervalos de confiança. A Figura III-17 sugere que o modelo de previsão para *FC* apresenta dificuldades para capturar os valores muito baixos.

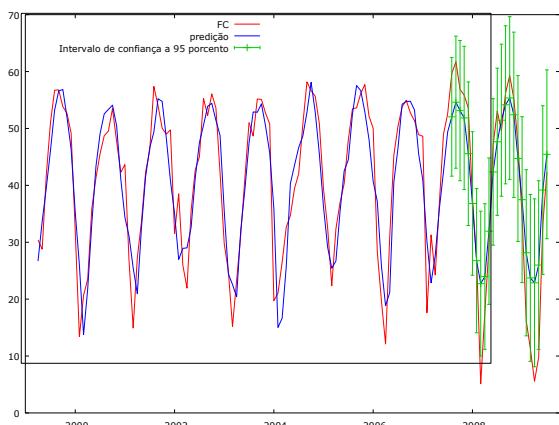


Figura III-17: Desempenho do Modelo para *FC*.

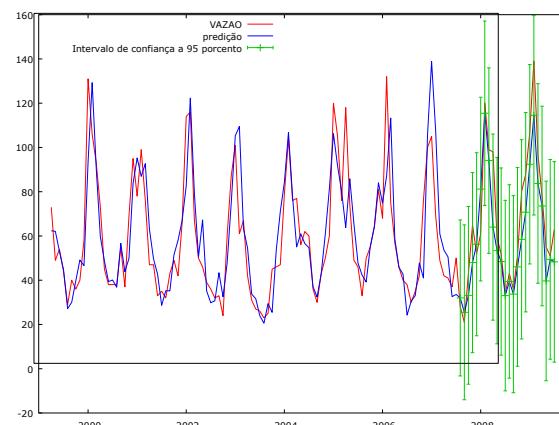


Figura III-18: Desempenho do Modelo para *Vz*.

Para quantificar a precisão do modelo na previsão das variáveis de interesse (*FC* e *Vz*), usaremos as seguintes medidas de aderência: a média dos módulos dos erros percentuais (*MAPE*), o erro absoluto médio (*MAE* – *Mean Absolute Error*), a raiz do erro quadrático médio (*RMS* – *Root Mean Square*), o erro percentual médio (*MPE* – *Mean Absolute Percent Error*) e o desvio-padrão da média

porcentual dos erros (*SDPE – Standard Deviation Percentage Error*). Essas medidas são definidas como:

$$MAE = \frac{1}{T} \sum_{i=1}^T |y_t^{prev} - y_t^{real}|; \quad (6)$$

$$MAPE = \left(\frac{1}{T} \sum_{i=1}^T \left| \frac{y_t^{prev} - y_t^{real}}{y_t^{real}} \right| \right) \times 100; \quad (7)$$

$$RMS = \sqrt{\frac{1}{T} \sum_{i=1}^T (y_t^{prev} - y_t^{real})^2}; \quad (8)$$

$$MPE = \frac{1}{T} \sum_{i=1}^T \left(\frac{y_t^{prev} - y_t^{real}}{y_t^{real}} \right); \quad (9)$$

$$SDPE = \sqrt{\frac{1}{T} \sum_{i=1}^T \left(\frac{y_t^{prev} - y_t^{real}}{y_t^{real}} - MPE \right)^2}. \quad (10)$$

Além delas, Pindyck e Rubinfeld (1991) sugerem o uso do coeficiente de desigualdade (ou índice) de Theil (1961, 1966), que, em linhas gerais, é uma estatística relacionada ao erro *RMS* e aplicada em avaliação de simulações históricas ou previsões Fora da Amostra. Esse coeficiente é definido como:

$$U = \frac{\sqrt{\frac{1}{T} \sum_{i=1}^T (y_t^{prev} - y_t^{real})^2}}{\sqrt{\frac{1}{T} \sum_{i=1}^T (y_t^{prev})^2} + \sqrt{\frac{1}{T} \sum_{i=1}^T (y_t^{real})^2}}. \quad (11)$$

O numerador é o erro *RMS*. O denominador foi definido de modo a fazer com que a medida *U* fique sempre entre o valor zero e o valor um. Caso a aderência entre os dados seja perfeita ($y_t^{prev} = y_t^{real}$), o valor de *U* será zero. Por outro lado, se *U* for igual a 1, a performance do modelo será a pior possível.

A medida encontrada para o *MAPE* é bastante alta para *FC*, 49,4%, indicando uma fraca precisão do modelo na previsão dessa variável. Contudo, esse valor é extremamente influenciado pelas observações 2008/03 e 2009/04 que apresentam valores muito abaixo do esperado. Retirando essas duas observações, o *MAPE* cai para 24,1%. Para a série *Vz*, o *MAPE* é 15,2%.

Para o nosso conjunto de previsões Fora da Amostra, os valores calculados para as outras medidas de aderência foram:

	<i>FC</i>	<i>Vz</i>
MAE	6,330417	10,47417
RMS	8,07778151	13,4793997
MPE	0,412201304	-0,087737217
SDPE	9,529211123	0,157239091
U	0,095157	0,09721

Quadro III-5: Medidas de Aderência – Fora da Amostra

Uma medida clássica de aderência é o coeficiente de determinação (R^2), que pode ser definido como o grau de ajuste do modelo ao conjunto de dados e indica a proporção da variação total da variável dependente, que é explicada pelo modelo ajustado. O R^2 no conjunto de validação (Fora da Amostra) para *FC* é 40%, ao passo que para *Vz* essa medida sobe para 79%.

A análise dos resíduos é outra etapa importante para detectar problemas com o ajuste do modelo. Os resíduos são definidos como a diferença entre o valor real observado e o valor previsto

pelo modelo, isto é, $r_t = y_t^{real} - y_t^{prev}$. Eles podem ser considerados erros observados ou erro de previsão.

As Figuras III-19 e III-20 mostram os histogramas dos resíduos das equações para FC e Vz respectivamente. Utilizando o teste Jarque-Bera, rejeitamos a hipótese de normalidade em ambos os casos.

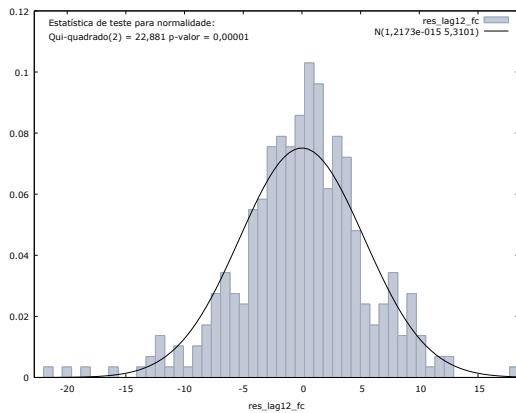


Figura III-19: Histograma e Teste de Normalidade para os Resíduos – FC .

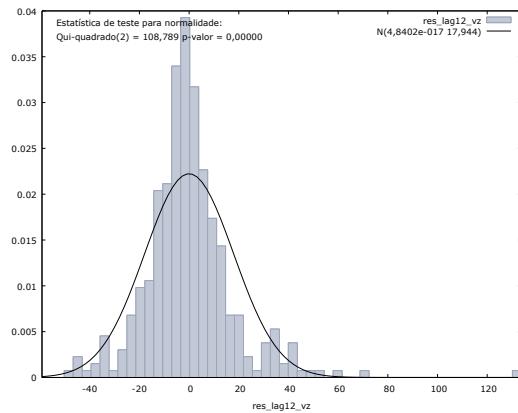


Figura III-20: Histograma e Teste de Normalidade para os Resíduos – Vz .

As figuras III-21 e III-22 mostram os comparativos anuais dos resíduos. Na equação do modelo para FC , os maiores resíduos foram gerados nos primeiros meses dos anos, aproximadamente até abril. O mesmo acontece com os resíduos da equação do modelo para Vz . Esse resultado já era esperado uma vez que fora identificada uma variância mais alta nos dados neste período.

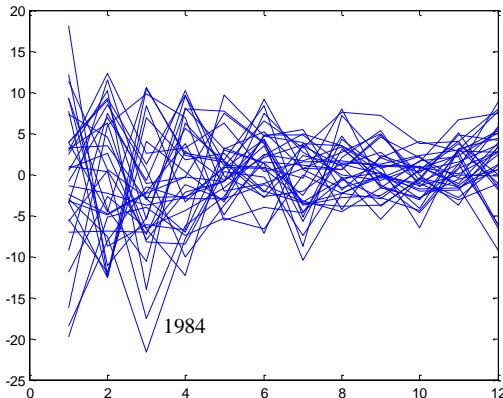


Figura III-21: Comparativo Anual Resíduos – FC .

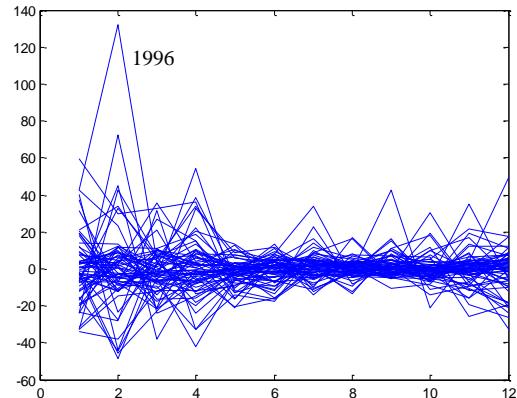


Figura III-22: Comparativo Anual Resíduos – Vz .

A Figura III-23 apresenta a fac e a facp dos resíduos do modelo para FC , com as estatísticas de Box-Pierce, com p-valores maiores do que 0,750, indicando que o modelo é adequado para modelar a dependência linear entre os sucessivos fatores de capacidade (FC).

A Figura III-24 apresenta a fac e a facp dos resíduos do modelo para Vz , com as estatísticas de Box-Pierce, com p-valores maiores do que 0,697, indicando mais uma vez que o modelo é adequado para modelar a dependência linear entre as sucessivas vazões.

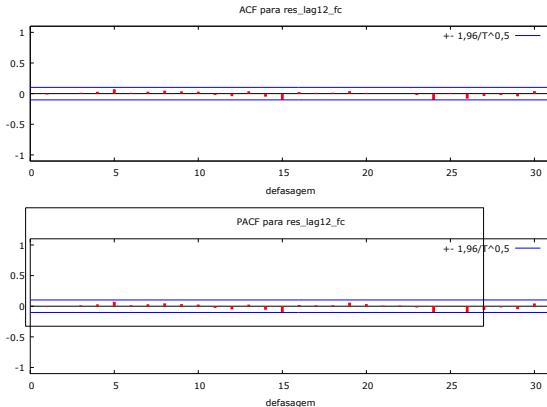


Figura III-23: Correlograma para os Resíduos – FC.

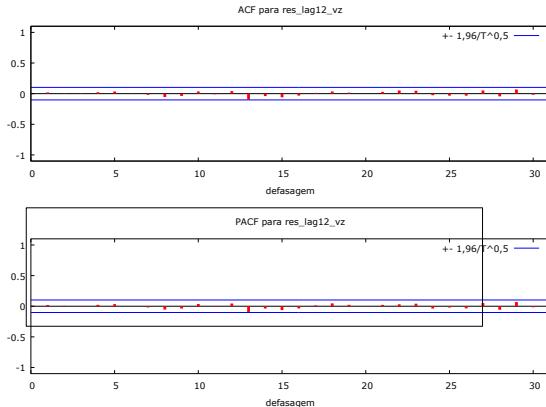


Figura III-24: Correlograma para os Resíduos – Vz.

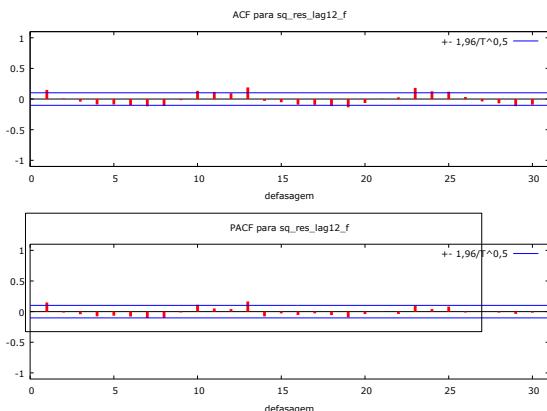


Figura III-25: Correlograma para os Quadrados dos Resíduos – FC.

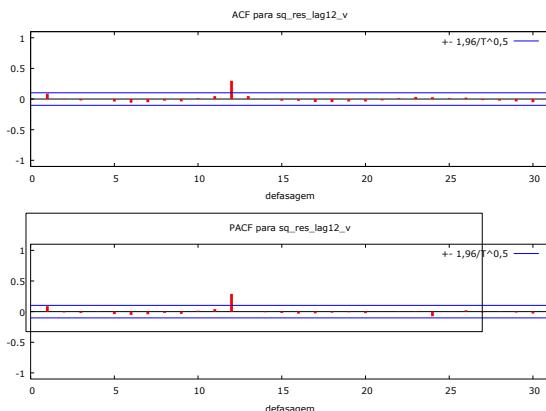


Figura III-26: Correlograma para os Quadrados dos Resíduos – Vz.

Verificamos que a série apresenta heterocedasticidade condicional através da análise da fac e da facp dos resíduos ao quadrado. A Figura III-25, referente aos resíduos do FC, apresenta heterocedasticidade condicional na defasagem 13. Adicionalmente, a Figura III-26 (referente aos resíduos de Vz) apresenta uma autocorrelação parcial não-nula na defasagem 12. Esses dois efeitos sugerem a presença de um efeito ARCH. Esse efeito pode ser confirmado através da aplicação do teste ARCH (Hamilton, 1994) (Quadro III-8), que caracterizou ambas as séries como condicionalmente heterocédasticas. Visando corrigir esse efeito, mais adiante, será testada outra metodologia que pondera os resíduos através da aplicação de uma lei de variância.

Equação FC					Equação Vazao					
Teste ARCH de ordem 12					Teste ARCH de ordem 12					
	coeficiente	erro padrão	razão-t	p-valor		coeficiente	erro padrão	razão-t	p-valor	
alpha(0)	31,16160	6,30944	4,93900	0,00000	***	alpha(0)	242,75800	79,20360	3,06500	0,00230
alpha(1)	0,11563	0,05402	2,14100	0,03300		alpha(1)	0,06583	0,05170	1,27300	0,20380
alpha(2)	-0,02869	0,05434	-0,52810	0,59780		alpha(2)	-0,01527	0,05181	-0,29480	0,76830
alpha(3)	-0,02652	0,05403	-0,49090	0,62380		alpha(3)	-0,01257	0,05180	-0,24260	0,80850
alpha(4)	-0,06306	0,05405	-1,16700	0,24420		alpha(4)	0,01321	0,05178	0,25510	0,79880
alpha(5)	-0,04918	0,05393	-0,91200	0,36240		alpha(5)	-0,02113	0,05178	-0,40810	0,68340
alpha(6)	-0,05368	0,05418	-0,99080	0,32250		alpha(6)	-0,03497	0,05184	-0,67460	0,50040
alpha(7)	-0,08402	0,05417	-1,55100	0,12180		alpha(7)	-0,03183	0,05187	-0,61370	0,53980
alpha(8)	-0,09135	0,05430	-1,68200	0,09340	*	alpha(8)	-0,01893	0,05190	-0,36480	0,71540
alpha(9)	-0,02432	0,05427	-0,44820	0,65430		alpha(9)	-0,02903	0,05196	-0,55860	0,57680
alpha(10)	0,11075	0,05419	2,04400	0,04180	**	alpha(10)	0,01835	0,05197	0,35300	0,72430
alpha(11)	0,04885	0,05446	0,89700	0,37030		alpha(11)	0,01812	0,05184	0,34960	0,72690
alpha(12)	0,04367	0,05421	0,80560	0,42110		alpha(12)	0,29216	0,05162	5,65900	0,00000

Hipótese nula: efeito ARCH não está presente
Estatística de teste: LM = 28,286
p-valor = 0,00502304

Hipótese nula: efeito ARCH não está presente
Estatística de teste: LM = 36,056
p-valor = 0,000317363

Quadro III- 6: Teste para efeito ARCH

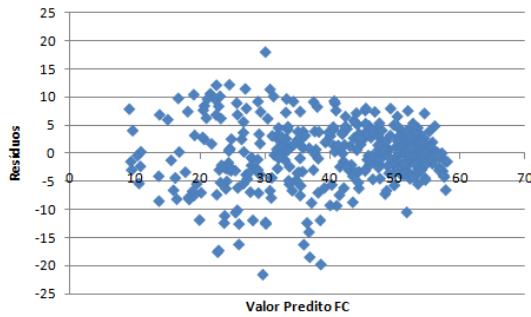


Figura III-27: Resíduos x Valor Preditivo – FC .

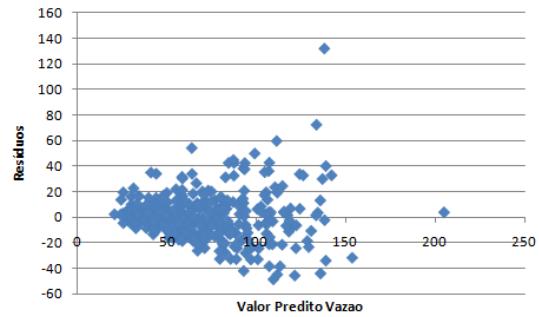


Figura III-28: Resíduos x Valor Preditivo – Vz .

Nota-se, através da Figura III-27, que a equação do modelo para FC concentra os maiores valores residuais em intervalos de baixa previsão. Já na equação do modelo para Vz , onde os maiores resíduos foram gerados nas previsões de vazão alta, ocorre o inverso, conforme demonstra a Figura III-28.

F. Modelagem Estatística - Lei de Variância

Em resposta aos resultados apresentados, principalmente no correlograma dos quadrados dos resíduos FC e Vz (Figuras III-25 e III-26) e no teste ARCH, desenvolveremos um modelo que aplicará uma lei de variância para modelar o fato de que a variância da componente irregular se altera de acordo com o mês observado. Com isso, teremos um modelo para cada uma das séries, mas que ainda considera a energia afluente acumulada em cada região do Brasil ($ENAs$) e o efeito que uma variável exerce sobre a outra. Por exemplo, o modelo desenvolvido para FC (variável dependente) incluirá defasagens da própria variável dependente, das $ENAs$ e das defasagens da variável Vz . Para modelo de Vz , o inverso também é válido.

O modelo desenvolvido tem características autorregressivas com ênfase na heterocedasticidade sazonal identificada e utiliza o método de mínimos quadrados ponderados (MQP) para estimação dos parâmetros. Quando os choques aleatórios são gaussianos, o MQP equivale ao método da máxima verossimilhança. Seus resíduos são ponderados com base na estimativa do desvio-padrão mensal da série. A ideia básica é que os dados de FC ou Vz são não-correlacionados serialmente, mas a volatilidade depende da variação dos dados em cada mês. O modelo para a série y_t tem a seguinte estrutura:

$$y_t = c + \delta_{1,1}SE_t + \delta_{1,2}S_t + \delta_{1,3}NE_t + \delta_{1,4}N_t + \delta_{2,1}SE_{t-1} + \delta_{2,2}S_{t-1} + \delta_{2,3}NE_{t-1} \\ + \delta_{2,4}N_{t-1} + \dots + \varphi_1FC_{t-1} + \dots + \varphi_{12}FC_{t-12} + \theta_1Vz_{t-1} + \dots \\ + \theta_{12}Vz_{t-12} + a_t; \quad (12)$$

$$a_t = \sqrt{h_t}\varepsilon_t; \quad (13)$$

$$h_t = \sigma_{jan}^2 d_{jan} + \sigma_{fev}^2 d_{fev} + \sigma_{mar}^2 d_{mar} + \dots + \sigma_{nov}^2 d_{nov} + \sigma_{dez}^2 d_{dez}; \quad (14)$$

onde:

y_t representa ora a variável FC_t , ora a variável Vz_t no tempo t ;

c representa a constante;

SE_t, S_t, NE_t, N_t representam os 4 subsistemas $ENAs$ – variáveis exógenas – no tempo t ;

$\delta_{1,i}$ são os respectivos coeficientes das $ENAs$ para $i=1,2,3,4$ no tempo t ;

$SE_{t-1}, S_{t-1}, NE_{t-1}, N_{t-1}$ representam os 4 subsistemas $ENAs$ defasados no tempo $t-1$;

$\delta_{2,i}$ são os respectivos coeficientes das $ENAs$ defasadas no tempo $t-1$;

$\varphi_m FC_{t-m}$ e $\theta_m Vz_{t-m}$ representam os coeficientes com suas respectivas defasagens da variável dependente e da variável cuja relação com a variável dependente será estudada. De acordo com a equação analisada, essas definições se invertem. Em ambas as equações essas variáveis são inseridas como variáveis independentes ou exógenas para $m=1, \dots, 12$;

σ^2_i é a variância observada nos resíduos no mês i e d_i são variáveis indicadoras que identificam o mês i , para todo $i = \text{jan}, \dots, \text{dez}$.

Formulamos um modelo geral que inclui as 12 defasagens das variáveis, como descrito acima, e que permitirá uma comparação com o modelo anterior. As estimativas são calculadas em várias etapas, sendo que a etapa inicial consiste em supor variâncias iguais para todos os meses e estimar os coeficientes. Em seguida, os resíduos são analisados e as variâncias encontradas são introduzidas na fórmula produzindo pesos diferenciados e novamente os parâmetros são re-estimados. Esse processo de iteração só é interrompido quando a diferença entre os coeficientes estimados é muito pequena ($<10^{-7}$), ou seja, as estimativas convergirem.

a) Especificação - Fator de Capacidade Eólica (FC)

FC	Coeficiente	Erro Padrão	IC - 95%		t-Statistic	Prob.
			Lim. Inferior	Lim. Superior		
C	15,1518	10,3376	8,8271	21,4764	4,7125	<0,0001 **
SE	-0,0001	1,1706e-009	-0,0002	-0,0000	-2,6117	0,0094 **
SE(-1)	0,0001	1,8389e-009	-0,0000	0,0001	1,3565	0,1758
S	-0,0000	8,6534e-010	-0,0001	0,0000	-1,2790	0,2018
S(-1)	0,0001	8,8590e-010	0,0000	0,0001	2,6374	0,0087 **
NE	0,0002	1,6452e-008	-0,0000	0,0005	1,8554	0,0644 *
NE(-1)	-0,0000	1,3003e-008	-0,0002	0,0002	-0,1981	0,8431
N	-0,0012	3,6071e-008	-0,0015	-0,0008	-6,1434	<0,0001 **
N(-1)	0,0005	2,5991e-008	0,0001	0,0008	2,8644	0,0044 **
FC(-1)	0,4580	0,0025	0,3598	0,5562	9,1752	<0,0001 **
FC(-2)	0,0963	0,0023	0,0020	0,1906	2,0096	0,0453 *
FC(-3)	0,0162	0,0016	-0,0623	0,0947	0,4061	0,6849 *
FC(-4)	-0,0554	0,0013	-0,1275	0,0167	-1,5120	0,1315
FC(-5)	0,0648	0,0012	-0,0025	0,1322	1,8929	0,0592 *
FC(-6)	-0,0617	0,0010	-0,1243	0,0010	-1,9372	0,0536 *
FC(-7)	0,0311	0,0009	-0,0276	0,0898	1,0413	0,2985
FC(-8)	0,0019	0,0008	-0,0545	0,0583	0,0674	0,9463
FC(-9)	0,0048	0,0009	-0,0540	0,0636	0,1600	0,8730
FC(-10)	-0,0094	0,0013	-0,0796	0,0608	-0,2636	0,7922
FC(-11)	0,1706	0,0021	0,0808	0,2603	3,7398	0,0002 **
FC(-12)	0,0520	0,0022	-0,0410	0,1450	1,1004	0,2720
Vz(-1)	-0,0044	0,0002	-0,0334	0,0245	-0,2994	0,7648
Vz(-2)	-0,0312	0,0002	-0,0578	-0,0045	-2,2993	0,0221 *
Vz(-3)	0,0147	0,0001	-0,0081	0,0374	1,2682	0,2056
Vz(-4)	0,0194	0,0001	-0,0017	0,0405	1,8089	0,0714 *
Vz(-5)	-0,0114	0,0001	-0,0308	0,0081	-1,1492	0,2513
Vz(-6)	0,0048	0,0001	-0,0124	0,0221	0,5516	0,5816
Vz(-7)	0,0026	0,0001	-0,0119	0,0170	0,3517	0,7253
Vz(-8)	0,0033	4,2656e-005	-0,0095	0,0161	0,5054	0,6136
Vz(-9)	-0,0050	0,0001	-0,0190	0,0089	-0,7094	0,4786
Vz(-10)	-0,0067	0,0001	-0,0242	0,0107	-0,7621	0,4465
Vz(-11)	0,0117	0,0001	-0,0106	0,0340	1,0308	0,3034
Vz(-12)	-0,0245	0,0001	-0,0465	-0,0026	-2,1970	0,0287 *

Quadro III-7: Ajuste Final Modelo FC

Após 10 iterações, o modelo alcançou a estabilidade dos coeficientes, definindo a matriz de pesos formada pelos termos h_t . As estatísticas do modelo apresentaram valores satisfatórios, uma vez que o modelo consegue explicar 84% ($R^2 = 0,8404$) da variação dos dados FC.

A inclusão de inúmeras variáveis, mesmo que tenham muito pouco poder explicativo sobre a variável resposta, aumentarão o valor de R^2 , o que é ruim, pois incentiva a inclusão indiscriminada de variáveis, indo contra o princípio da parcimônia, que sugere a adequação do modelo com o mínimo de variáveis explicativas possível, descartando variáveis que incrementem de forma não significativa as estatísticas de aderência. Para resolver esse problema, podemos usar uma medida ajustada do coeficiente de determinação que é penalizada quando incluímos variáveis pouco explicativas. Esse é o R^2 ajustado: $R_{aj}^2 = 0,8251$.

A especificação do modelo inclui, entre outros testes, a avaliação da significância das variáveis e a análise dos resíduos. Optou-se por não interferir nas variáveis, apesar de algumas serem não-significantes. A permanência das ENAs servirá como uma ligação entre o modelo estatístico especificado e os dados do Newave na geração de cenários. Já as demais variáveis permaneceram

devido à necessidade de retratar um modelo com uma fórmula mais geral onde, em etapas posteriores, outros conjuntos de dados (séries) serão inseridos.

A análise dos resíduos consiste, principalmente, em três passos: (i) no resultado do teste de Normalidade de Jarque-Bera realizado sobre os resíduos padronizados ε_t , que teve p-valor 0,3833, o que indica que eles podem ser provenientes de uma distribuição Normal (ver Figura III-29); (ii) no comparativo anual das sequências de resíduos, mostrando que as diferenças entre as variâncias foram compensadas pelo ajuste do modelo (Figura III-30); e, por fim, (iii) nos correlogramas dos resíduos (Figura III-33) e dos resíduos ao quadrado (Figura III-34), mostrando que os resíduos gerados não possuem autocorrelações significativas e que a série não mais apresenta heterocedasticidade condicional.

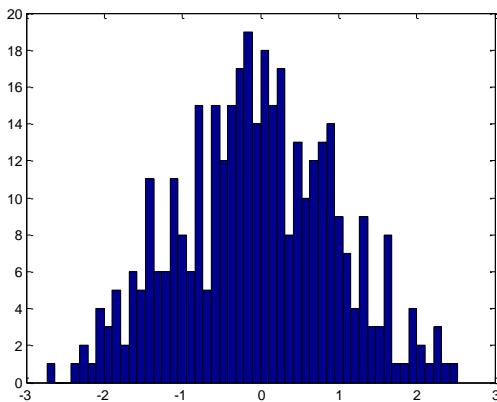


Figura III-29: Histograma dos Resíduos Padronizados – FC.

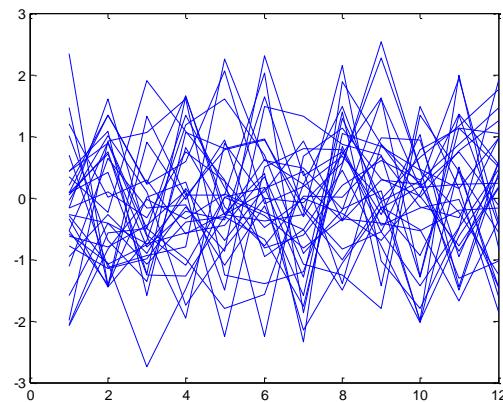


Figura III-30: Comparativo Anual dos Resíduos Padronizados – FC.

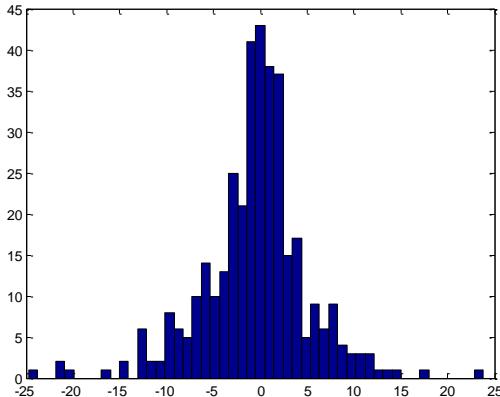


Figura III-31: Histograma dos Resíduos – FC.

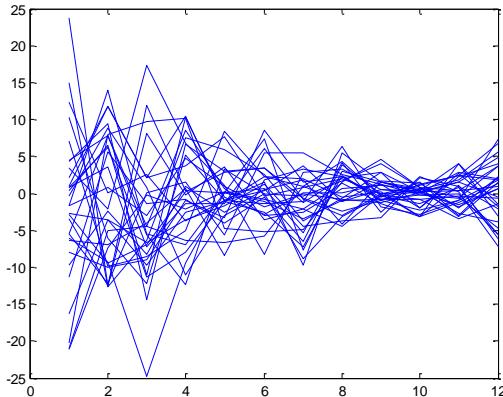


Figura III-32: Comparativo Anual dos Resíduos – FC.

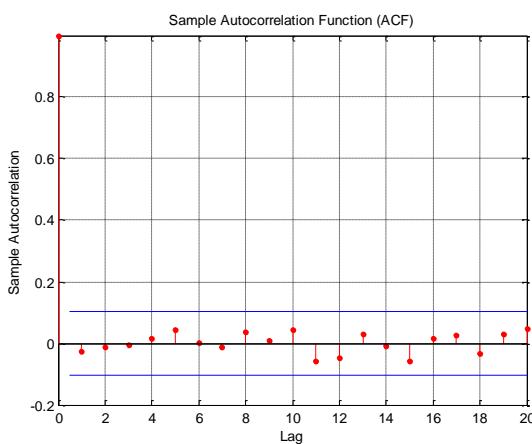


Figura III-33: Correlograma para os Resíduos Padronizados – FC.

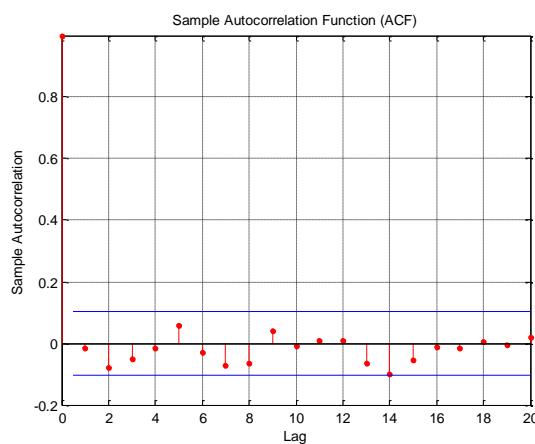


Figura III-34: Correlograma para os Quadrados dos Resíduos Padronizados – FC.

b) Especificação – Vazão (V_z)

V_z			IC – 95%			
	Coeficiente	Erro Padrão	Lim. Inferior	Lim. Superior	t-Statistic	Prob.
c	8,9726	73,3759	-7,8776	25,8229	1,0475	0,2956
SE	0,0015	1,1712e-008	0,0013	0,0018	14,2676	<0,0001**
SE(-1)	-0,0009	1,0815e-008	-0,0011	-0,0007	-8,5084	<0,0001**
S	0,0001	8,1459e-009	-0,0001	0,0003	1,0807	0,2806
S(-1)	0,0001	6,8839e-009	-0,0001	0,0002	0,6985	0,4854
NE	-0,0003	1,2960e-007	-0,0010	0,0004	-0,7531	0,4519
NE(-1)	0,0005	7,5470e-008	-0,0000	0,0010	1,8445	0,0660*
N	0,0001	2,1237e-007	-0,0008	0,0010	0,1810	0,8565
N(-1)	-0,0002	1,2441e-007	-0,0009	0,0005	-0,5212	0,6025
FC(-1)	-0,0168	0,0118	-0,2305	0,1969	-0,1547	0,8772
FC(-2)	-0,1515	0,0123	-0,3701	0,0671	-1,3637	0,1736
FC(-3)	0,1824	0,0079	0,0079	0,3569	2,0566	0,0405*
FC(-4)	-0,0108	0,0063	-0,1668	0,1452	-0,1359	0,8920
FC(-5)	0,0163	0,0067	-0,1450	0,1777	0,1991	0,8423
FC(-6)	-0,1714	0,0074	-0,3410	-0,0018	-1,9885	0,0476*
FC(-7)	0,1243	0,0076	-0,0471	0,2957	1,4263	0,1547
FC(-8)	-0,0985	0,0117	-0,3114	0,1143	-0,107	0,3631
FC(-9)	-0,1276	0,0167	-0,3822	0,1269	-0,9864	0,3246
FC(-10)	-0,1309	0,0196	-0,4062	0,1444	-0,9356	0,3502
FC(-11)	0,0057	0,0223	-0,2882	0,2995	0,0381	0,9696
FC(-12)	0,1003	0,0142	-0,1342	0,3347	0,8414	0,4007*
$V_z(-1)$	0,3917	0,0013	0,3215	0,4619	10,9723	<0,0001**
$V_z(-2)$	0,0468	0,0012	-0,0221	0,1156	1,3360	0,1825
$V_z(-3)$	0,0870	0,0006	0,0386	0,1354	3,5348	0,0005**
$V_z(-4)$	-0,0021	0,0005	-0,0462	0,0420	-0,0924	0,9264
$V_z(-5)$	0,0461	0,0005	0,0026	0,0896	2,0857	0,0378*
$V_z(-6)$	-0,0059	0,0004	-0,0443	0,0324	-0,3040	0,7613
$V_z(-7)$	0,0051	0,0005	-0,0374	0,0476	0,2376	0,8124
$V_z(-8)$	0,0405	0,0007	-0,0115	0,0926	1,5327	0,1263
$V_z(-9)$	-0,0176	0,0010	-0,0783	0,0431	-0,5693	0,5695
$V_z(-10)$	0,0013	0,0012	-0,0657	0,0684	0,0396	0,9684
$V_z(-11)$	-0,0403	0,0013	-0,1109	0,0302	-1,1240	0,2618
$V_z(-12)$	0,0669	0,0014	-0,0058	0,1397	1,8101	0,0712*

Quadro III-8: Ajuste Final Modelo V_z

Após 9 iterações, a matriz de pesos formada pelos termos h_t foi definida, e o modelo final consegue explicar 70% ($R^2 = 0,7031$) da variação dos dados V_z e o R^2 ajustado é 0,6747. Mais uma vez, não interferimos nas variáveis ENAs e no uso da forma geral do modelo pelas razões já comentadas.

O teste de Jarque-Bera realizado sobre os resíduos padronizados ε_t tem p-valor menor que 0,001 e, portanto, rejeitamos a hipótese de normalidade. A partir da Figura III-35, nota-se que possivelmente a cauda mais alongada para a direita não permite que os dados se aproximem de uma distribuição normal.

O comparativo anual das sequências de resíduos mostra que as diferenças entre as variâncias foram eliminadas com a ponderação utilizada (Figura III-36).

O correlograma dos resíduos (Figura III-39) mostra que os resíduos gerados não possuem autocorrelações significativas, enquanto o correlograma dos resíduos ao quadrado (Figura III-40) mostra que o efeito de heterocedasticidade condicional persiste, agora na correlação de ordem 1.

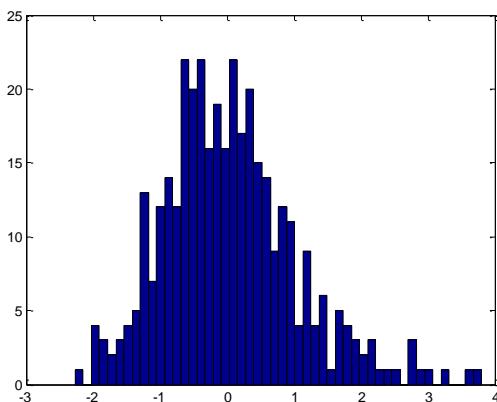


Figura III-35: Histograma dos Resíduos Padronizados – Vz.

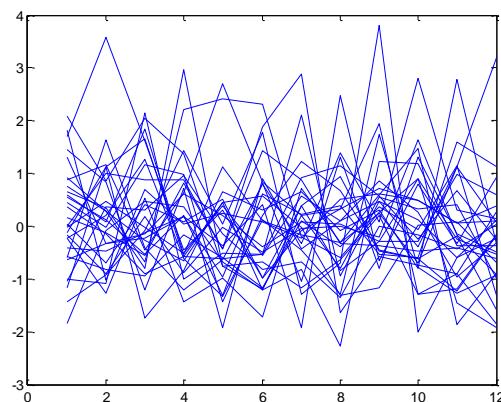


Figura III-36: Comparativo Anual dos Resíduos Padronizados – Vz.

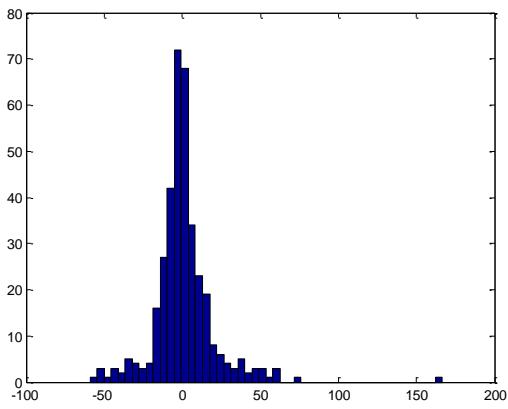


Figura III-37: Histograma dos Resíduos – Vz.

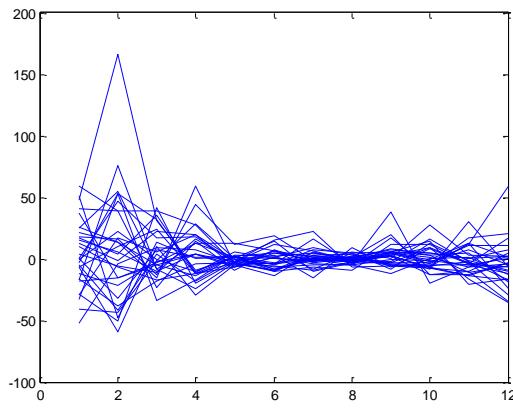


Figura III-38: Comparativo Anual dos Resíduos – Vz.

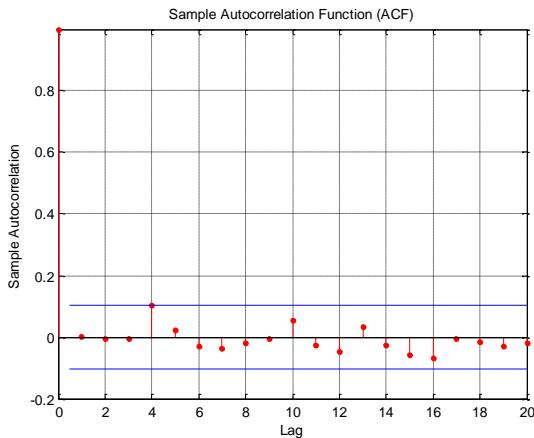


Figura III-39: Correlograma para os Resíduos Padronizados – Vz.

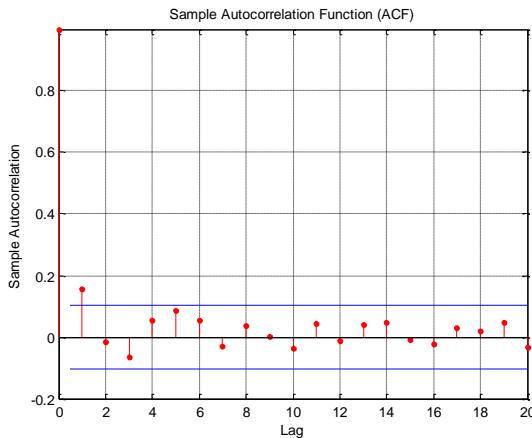


Figura III-40: Correlograma para os Quadrados dos Resíduos Padronizados – Vz.

A persistência do efeito ARCH, os altos valores residuais nos primeiros meses do ano e o formato do histograma dos resíduos podem ser indicativos da presença de observações aberrantes na série. Isso pode ser constatado no gráfico *boxplot* abaixo:

Na Figura III-41 é possível detectar 12 pontos fora do intervalo demarcado entre o maior e o menor valor não outlier [$Q_3 + 1,5 \cdot (Q_3 - Q_1)$, $Q_1 - 1,5 \cdot (Q_3 - Q_1)$], onde Q_1 corresponde ao 1º quartil e Q_3 , ao 3º quartil. Dez desses pontos foram observados nos primeiros meses do ano: janeiro (3 pontos), fevereiro (5 pontos) e março (2 pontos).

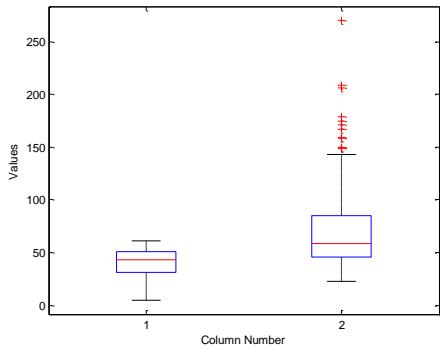


Figura III-41: Boxplot – FC e Vz.

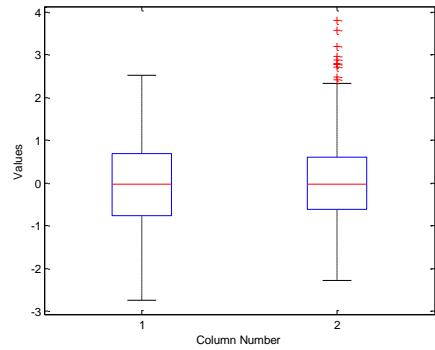


Figura III-42: Boxplot dos Resíduos Padronizados – FC e Vz.

Na Figura III-42 também é possível detectar 10 pontos aberrantes nos resíduos. Ao analisarmos esses resíduos discrepantes, vimos que foram gerados a partir de observações ao longo do ano com exceção dos primeiros meses (janeiro, fevereiro e março). Isso acaba interferindo nos resultados e gerando um efeito ARCH espúrio.

Para melhor retratar esse fato, substituímos os 10 resíduos *outliers* pelo valor da média da sequência de resíduos a_t . Em seguida, recalculamos os resíduos padronizados ε_t . A análise dos resíduos transformados mostra que a substituição de apenas 10 observações significou uma grande mudança nos resultados. O teste de Jarque-Bera teve p-valor 0,1761, não rejeitando a hipótese de normalidade dos resíduos (ver Figura III-44). Os correlogramas dos resíduos (Figura III-45) e dos resíduos ao quadrado (Figura III-46) mostram que os resíduos gerados não possuem autocorrelações significativas e que a série não mais apresenta heterocedasticidade condicional.

Daremos procedimento com o uso desse modelo, e futuramente esse assunto será abordado de forma mais detalhada.

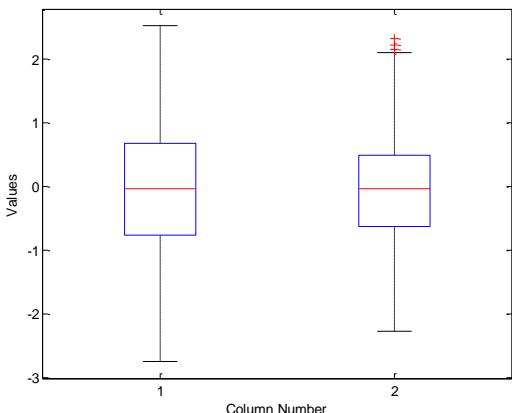


Figura III-43: Boxplot dos Resíduos Padronizados FC e Resíduos Pad. Transformados Vz.

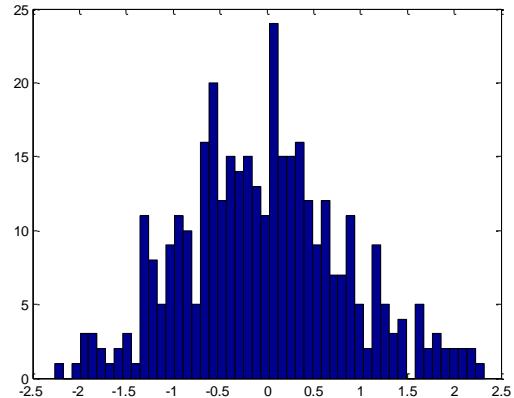


Figura III-44: Histograma dos Resíduos Pad. Transformados – Vz.

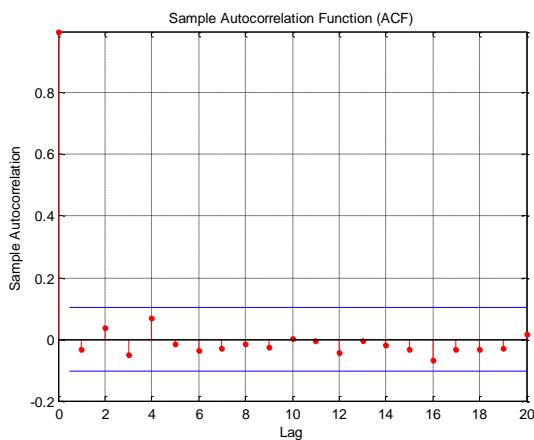


Figura III-45: Correlograma para os Resíduos Padronizados Transformados – Vz.

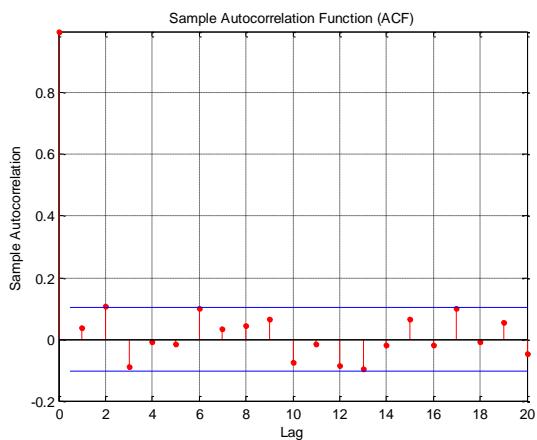


Figura III-46: Correlograma para os Quadrados dos Resíduos

G. Metodologia e testes de aderência - Lei de Variância

As séries foram modeladas separadamente, contudo ainda considerando a relação entre elas através da inclusão da outra série como variável explicativa. É necessário avaliar também a relação entre os resíduos (Quadro III-9 e III-10), pois esse resultado será utilizado mais adiante na geração de cenários conjuntos de FC e Vz .

	Resíduos FC	Resíduos Vz
Resíduos FC	$\sigma_{1,1}^2 = 1,0004$	$\sigma_{1,2}^2 = -0,1240$
Resíduos Vz	$\sigma_{2,1}^2 = -0,1240$	$\sigma_{2,2}^2 = 1,0009$

Quadro III-9: Matriz de Covariância Residual ε_t .

	Resíduos FC	Resíduos Vz
Resíduos FC	$\sigma_{1,1}^2 = 1,0000$	$\sigma_{1,2}^2 = -0,1239$
Resíduos Vz	$\sigma_{2,1}^2 = -0,1239$	$\sigma_{2,2}^2 = 1,0000$

Quadro III-10: Matriz de Correlação Residual ε_t .

Novamente, os últimos 24 dados da série, correspondentes ao período de agosto de 2007 a julho de 2009, foram reservados para validação. Nas figuras abaixo, é possível notar que as previsões Fora da Amostra para FC não acompanham totalmente as quedas da série real e estão próximas ao limite inferior do intervalo de confiança, ultrapassando algumas vezes esse limite.

As previsões para Vz (Figura III-48) apresentam maiores erros (resíduos) em picos periódicos que ocorrem de 12 em 12 meses aproximadamente.

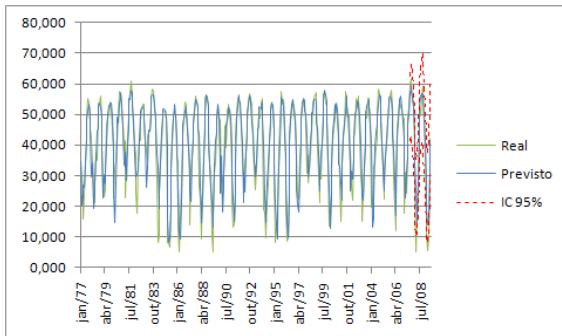


Figura III-47: Desempenho do Modelo para FC – Total.

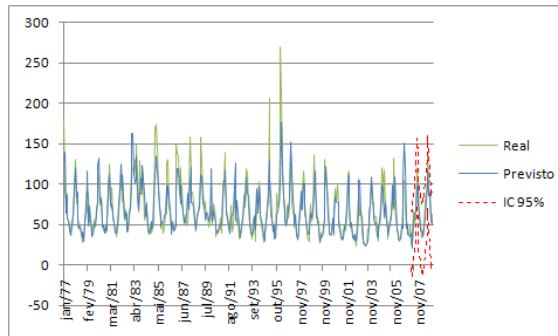


Figura III-48: Desempenho do Modelo para Vz – Total.

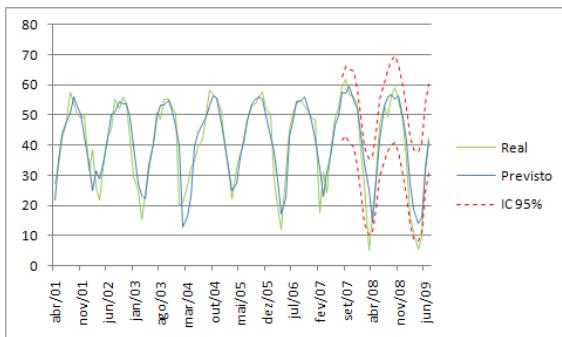


Figura III-49: Desempenho do Modelo para FC (com 100 dados).

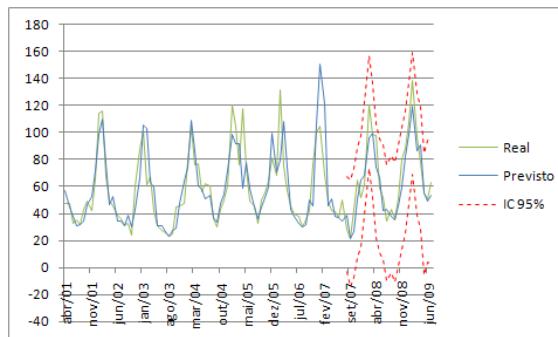


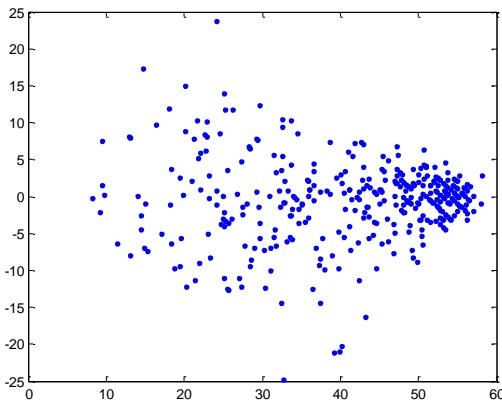
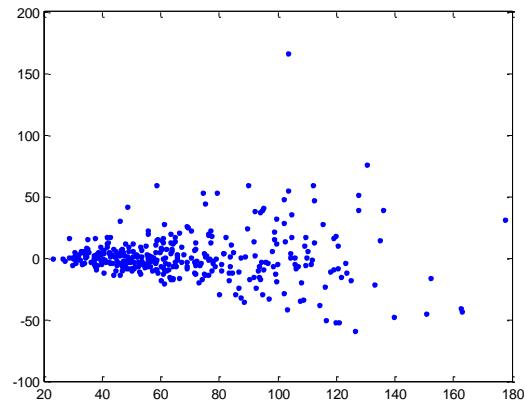
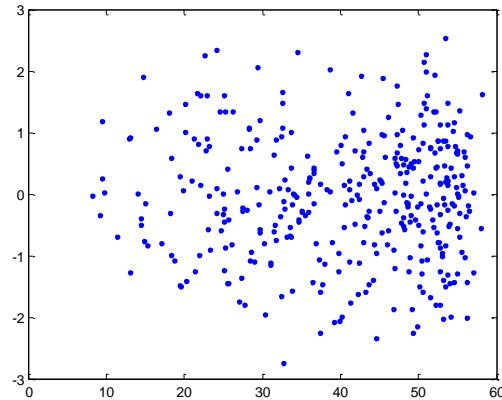
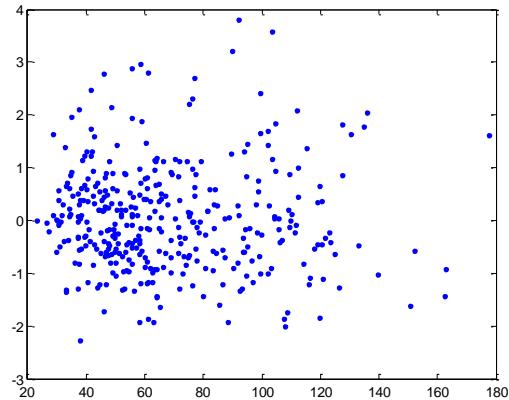
Figura III-50: Desempenho do Modelo para Vz (com 100 dados).

O MAPE desse modelo é menor que o MAPE encontrado no modelo anterior, porém ainda é alto para FC , 36,8%. Esse valor continua sendo influenciado pelas observações 2008/03 e 2009/04 que apresentam valores muito abaixo do esperado. Retirando apenas a observação 2008/03, o MAPE cai para 20,6%. Para a série Vz , o MAPE, 15,1%, não variou em relação ao modelo anterior.

O R^2 no conjunto de validação (Fora da Amostra) sofreu aumento significativo em relação ao modelo anterior, principalmente para FC , cuja medida sobe para 69%, ao passo que para Vz a medida é de 75%. As outras medidas de aderência encontram-se no Quadro III-15, abaixo.

	<i>FC</i>	<i>Vz</i>
MAE	4,573457	9,876546
RMS	6,31500946	13,4793997
MPE	0,319504278	-0,042010898
SDPE	7,396585192	0,179921389
U	0,072896	0,085464

Quadro III-11: Medidas de Aderência – Fora da Amostra.

Figura III-51: Resíduos X Valor Preditivo para FC .Figura III-52: Resíduos X Valor Preditivo para Vz .Figura III-53: Resíduos Padronizados X Valor Preditivo para FC .Figura III-54: Resíduos Padronizados X Valor Preditivo para Vz .

Das figuras acima, nota-se principalmente que os resíduos padronizados apresentam comportamento aleatório em torno do zero, o que indica a eliminação da dependência entre os resíduos e as variáveis explicativas do modelo. Nota-se também que o modelo para Vz concentra os maiores valores do resíduo não padronizado em intervalos de previsão alta, destacando-se o resíduo gerado em fevereiro de 96.

H. Simulações e análises dos resultados

O modelo Newave visa o planejamento da operação a médio prazo de subsistemas hidrotérmicos interligados e um dos seus resultados é a simulação das ENAs e dos preços de curto prazo, para diversos cenários de afluências, sintéticos ou históricos, ao longo do período de planejamento. A partir dessa simulação, podemos compor uma amostra do tamanho necessário para estimar índices de risco com incertezas aceitáveis.

Foram simuladas 2.000 séries, 5 anos a frente (2011 a 2015), para cada subsistema. As figuras abaixo apresentam o comparativo anual (energia afluente real) seguidas das simulações realizadas

no Newave dentro do período de planejamento. Podemos notar que os formatos das séries reais são mantidos nas simulações, observando-se algum aumento na variância.

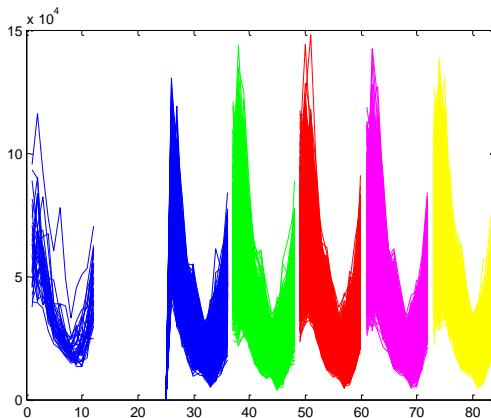


Figura III-55: Comparativo Anual + Simulações (2011 a 2015) – ENA-SE.

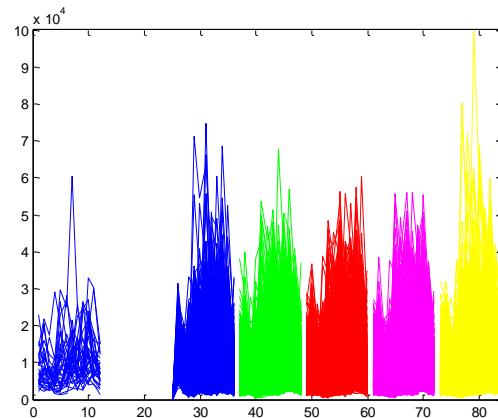


Figura III-56: Comparativo Anual + Simulações (2011 a 2015) – ENA-S.

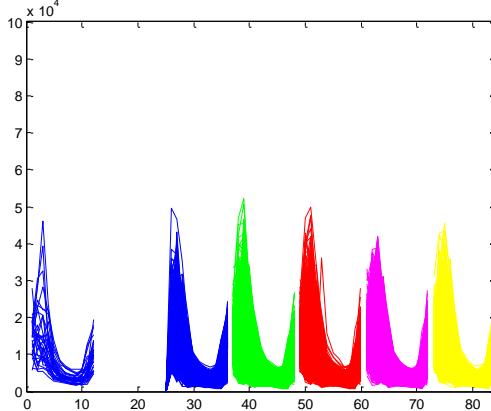


Figura III-57: Comparativo Anual + Simulações (2011 a 2015) – ENA-NE.

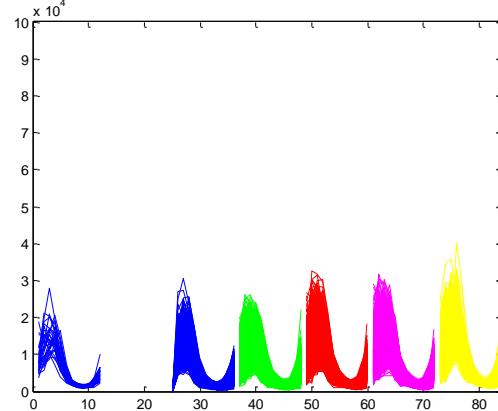


Figura III-58: Comparativo Anual + Simulações (2011 a 2015) – ENA-N.

Utilizando as simulações geradas pelo Newave, construímos um mecanismo de simulação para o *FC* e a *Vz*, respeitando a correlação residual observada no processo de modelagem.

O objetivo da simulação é descrever a distribuição e as características dos possíveis valores das variáveis dependentes (*FC* e *Vz*) depois de determinar os possíveis valores e comportamentos das variáveis independentes (*ENAs*).

Após a geração de números aleatórios atribuídos aos resíduos $\varepsilon_t \sim N(0, \Omega)$, inserimos as *ENAs* simuladas na fórmula de geração, efetuamos a leitura do período de agosto de 2005 a julho de 2006 para cobrir as defasagens das variáveis necessárias ao modelo e simulamos as séries *FC* e *Vz* para o período de um ano: agosto de 2006 a julho de 2007 – Dentro da Amostra. Os gráficos abaixo assumem valor zero quando a simulação de *FC* ou *Vz* gerou algum número negativo.

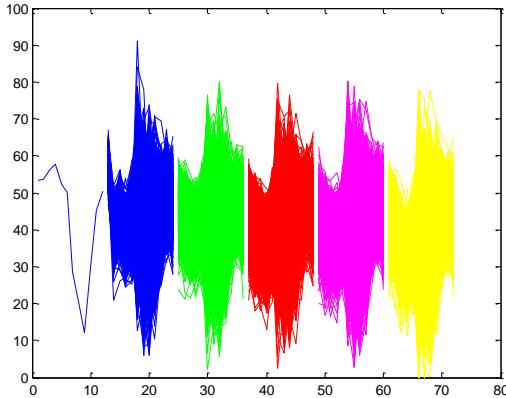


Figura III-59: 2000 Simulações - FC.

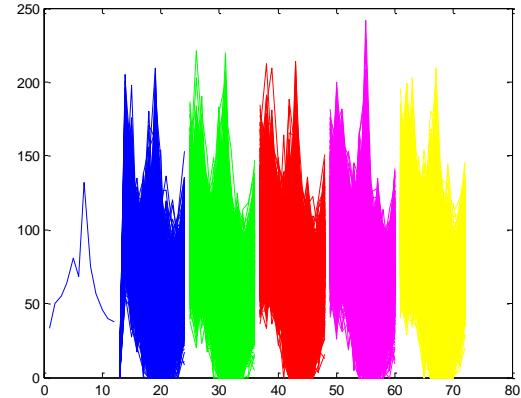


Figura III-60: 2000 Simulações - Vz.

Abaixo, as duas séries foram plotadas na mesma escala, evidenciando a diferença na ordem de grandeza de seus valores.

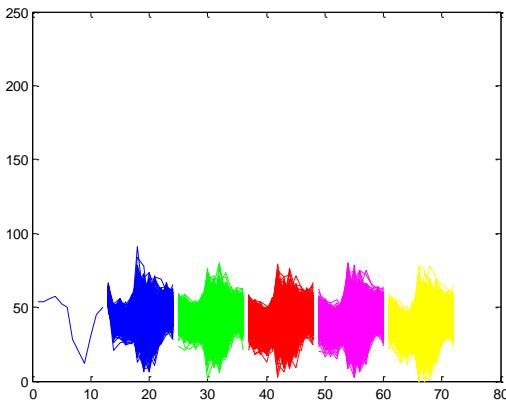


Figura III-61: 2000 Simulações - FC.

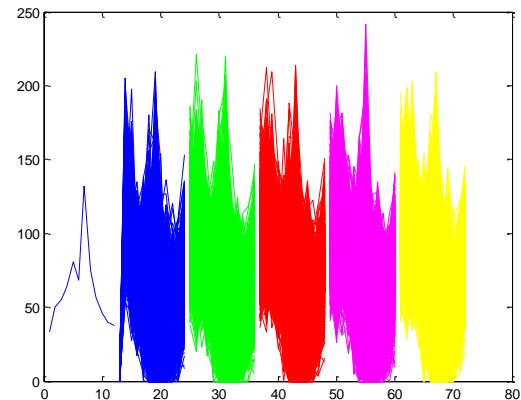


Figura III-62: 2000 Simulações - Vz.

As simulações foram novamente geradas, contudo efetuamos a leitura do período de agosto de 2006 a julho de 2007 para cobrir as defasagens das variáveis, e as novas séries FC e Vz abrangem o período de um ano: agosto de 2007 a julho de - Fora da Amostra.

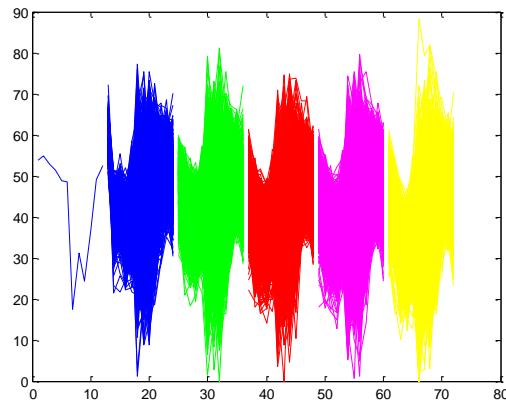


Figura III-63: 2000 Simulações Fora da Amostra - FC.

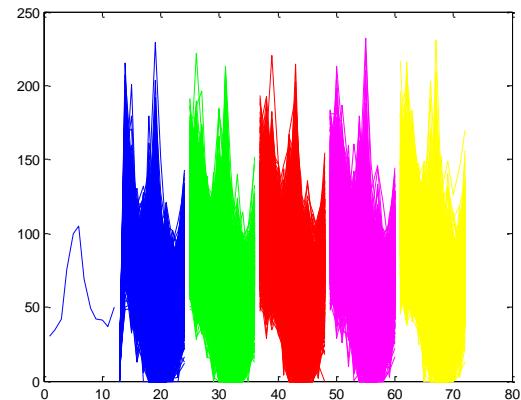


Figura III-64: 2000 Simulações Fora da Amostra - Vz.

Com base nos resultados obtidos, os próximos passos incluem uma análise mais detalhada a respeito da autenticidade do efeito ARCH encontrado no modelo de Vazão (Vz) que implicará manter o modelo desenvolvido (Lei de Variância) ou testar nova metodologia.

IV. CONTRATOS DE ENERGIA

Nesta seção, são apresentadas as expressões de renda de quatro modalidades de contratos bilaterais de compra e venda de energia elétrica. São eles: contratos de quantidade, contratos de quantidade com cláusulas *Take or Pay (ToP)*, contratos de quantidade baseados no *PLD* com piso e teto e contratos de disponibilidade.

O objetivo desta seção é obter uma expressão final de renda que reproduza os ganhos de uma comercializadora de energia, levando em conta os diferentes cenários de geração e de preço *spot*. Essas expressões serão utilizadas no modelo de otimização, que visa definir a quantidade ótima de energia a ser contratada em cada oportunidade cadastrada como contrato candidato.

Organizamos a sessão da seguinte forma: nas subseções *A* e *B* são apresentadas as expressões de renda para cada modalidade de contrato associadas a um período *t* e a um cenário *s*; na subseção *C* é apresentada a expressão do saldo da contabilização dos contratos de compra e venda de energia na Câmara de Comercialização de Energia Elétrica (CCEE); na subseção *D* é apresentada a expressão geral de renda da comercializadora, também para um dado período *t* e cenário *s*. Além disso, nessa mesma subseção *D*, é apresentado o balanço de compra e venda de contratos de energia que a comercializadora deve respeitar. Esse balanço garante que a comercializadora tenha o mínimo de contratos de compra necessários para lastrear os seus contratos de venda e, portanto, não sofrer penalidade. Por fim, na subseção *E*, é apresentada de forma resumida a interface de toda a seção de contratos de energia com as demais seções do trabalho, introduzindo os próximos passos a serem seguidos, em especial o otimizador, e mostrando o modelo geral do trabalho.

A. Contratos de Quantidade

Em contratos de quantidade, o vendedor compromete-se a entregar ao comprador, para cada período de vigência do contrato, uma determinada quantidade de energia Q_t (MWh). A expressão da receita devida à venda de um contrato *i*, de quantidade, no período *t* é definida como:

$$R_{i,t,s}^Q = P_{i,t,s} \cdot Q_{i,t,s}. \quad (1)$$

Sendo:

$Q_{i,t,s}$ – Quantidade de energia do contrato *i* sazonizada em *t* no cenário *s* (em MWh);

$P_{i,t,s}$ – Preço da energia contratada através do contrato *i* no período *t* no cenário *s* (em R\$/MWh).

A expressão de custo devido à compra de um contrato *i*, de quantidade, no período *t* é definida como:

$$C_{i,t,s}^Q = P_{i,t,s} \cdot Q_{i,t,s}. \quad (2)$$

a) Contratos de Quantidade com cláusulas de Take or Pay (ToP)

Contratos de quantidade com cláusulas de *Take or Pay (ToP)* permitem flexibilizar a quantidade de energia contratada para uma quantidade mínima ($Q_{i,t}^{min}$) ou uma quantidade máxima ($Q_{i,t}^{max}$), definidas em contrato em função dos interesses do comprador.

Assim, a expressão de receita de venda de quantidade *ToP* é a mesma que em (1), contudo a quantidade de energia contratada passa a ser função do preço da energia do contrato *i* com relação ao preço *spot* ($\pi_{t,s}^{sub,i}$) no submercado do contrato *i*, no período *t* e no cenário *s*. Dessa forma:

$$Q_{i,t}(\pi_{t,s}^{sub,i}) = \begin{cases} Q_{i,t}^{max}, & \pi_{t,s}^{sub,i} \geq P_{i,t}^{ToP}; \\ Q_{i,t}^{min}, & \pi_{t,s}^{sub,i} < P_{i,t}^{ToP}. \end{cases} \quad (3)$$

Sendo:

$\pi_{t,s}^{sub,i}$ – Preço *spot* (em R\$/MWh) do submercado onde está localizado o contrato *i* no período *t*;

$Q_{i,t}^{max}$ – Quantidade de energia máxima contratada (MWh) em contratos de venda de quantidade *ToP* do contrato *i* no período *t*;

$Q_{i,t}^{min}$ – Quantidade de energia mínima contratada (MWh) em contratos de venda de quantidade *ToP*

do contrato i no período t ;

$P_{i,t}^{ToP}$ – Preço da energia contratada (R\$/MWh) por meio do contrato i , no contrato de venda de quantidade *ToP*, no período t ;

Nos contratos de compra de quantidade *ToP* pode-se usar a expressão (2) com as condições descritas em (3).

b) Contratos de Quantidade baseados no PLD com piso e teto

Nos contratos bilaterais de quantidade de energia baseados no PLD com piso e teto, o preço da energia contratada é igual ao preço *spot* (R\$/MWh) do submercado do agente vendedor do contrato i , acrescentado de um ágio a_i com uma limitação de piso ($\pi_{i,t}^{min}$) e teto ($\pi_{i,t}^{max}$) no preço dessa energia.

Como nos contratos de quantidade *ToP*, a expressão de receita devida à venda de contratos de quantidade baseados no PLD com piso e teto também é a mesma que em (1). Neste caso, porém, o preço da energia contratada é função de $\pi_{t,s}^{sub_i}$. Assim:

$$P_{i,t}(\pi_{t,s}^{sub_i}) = \begin{cases} \pi_{i,t}^{max}, & \pi_{t,s}^{sub_i}(1 + a_i) \geq \pi_{i,t}^{max}; \\ \pi_{i,t}^{min}, & \pi_{t,s}^{sub_i}(1 + a_i) \leq \pi_{i,t}^{min}; \\ \pi_{t,s}^{sub_i}(1 + a_i), & \pi_{i,t}^{min} \leq \pi_{t,s}^{sub_i}(1 + a_i) \leq \pi_{i,t}^{max}. \end{cases} \quad (4)$$

Sendo:

a_i – Valor do ágio aplicado sobre o preço *spot* (em %) do submercado onde está localizado o contrato i .

$\pi_{i,t}^{max}$ – Preço *spot* teto (R\$/MWh) do contrato i no período t ;

$\pi_{i,t}^{min}$ – Preço *spot* piso (R\$/MWh) do contrato i no período t .

Novamente, nos contratos de compra de quantidade *baseados no PLD com piso e teto* pode-se usar a expressão (2) com as condições descritas em (4).

c) Exemplo

Neste exemplo, analisaremos o fluxo de cada modalidade de contrato apresentado até agora, mostrando como se comporta cada um desses contratos frente a diferentes cenários de preço *spot*. Para fins didáticos, vamos considerar dois cenários extremos de PLD, 10 R\$/MWh (1º Período) e 500 R\$/MWh (2º Período). Dessa forma, consideraremos uma comercializadora sem geração própria, apenas vendendo contratos a consumidores.

i) Quantidade

Os parâmetros do primeiro tipo de contrato apresentado são:

P (R\$/MWh)	Q (MWh)
100	1

O fluxo dessa modalidade de contrato, para os dois cenários de preço *spot*, é:

Período	PLD (R\$/MWh)	Custo Spot	Renda Contrato	Total (Contrato - Spot)
1	10	$Q \cdot \pi = R\$ 10$	$Q \cdot P = R\$ 100$	$100 - 10 = R\$ 90$
2	500	$Q \cdot \pi = R\$ 500$	$Q \cdot P = R\$ 100$	$100 - 500 = - R\$ 400$
Período 1 + 2	R\$ 510		R\$ 200	- R\$ 310

ii) Quantidade com cláusulas de Take or Pay (ToP)

Nos contratos de quantidade com cláusula *Take or Pay*, a quantidade de energia a ser contratada é função da relação entre o preço de referência do contrato e o preço *spot*, segundo (3). Vemos, porém, que há um favorecimento ao comprador (*game do comprador*) uma vez que, para um PLD acima do preço do contrato, a quantidade é máxima e para um PLD abaixo do preço do contrato, a quantidade é mínima. Isso maximiza a renda do comprador uma vez que ele adquire uma quantidade máxima, podendo liquidar mais energia em um preço *spot* alto. Já para um *spot* baixo, a quantidade é mínima, valendo o mesmo raciocínio.

Como estamos analisando um contrato de venda, ou seja, estamos assumindo a contraparte vendedora do contrato, esperamos que esta modalidade de contrato tenha o pior resultado dentre as analisadas até agora. Vamos supor um contrato *ToP* com os seguintes parâmetros:

P (R\$/MWh)	Q^{max} (MWh)	Q^{min} (MWh)
100	1,5	0,5

Temos de (3):

- Para o primeiro período, $\pi = 10 \text{ R\$/MWh} < P = 100 \text{ R\$/MWh}$. Assim, a quantidade de energia contratada passa a ser $Q^{min} = 0,5 \text{ MWh}$.
- Já para o segundo período, $\pi = 500 \text{ R\$/MWh} > P = 100 \text{ R\$/MWh}$. Dessa forma, a quantidade de energia contratada passa a ser $Q^{max} = 1,5 \text{ MWh}$.

Assim, o fluxo dessa modalidade de contrato, para os dois cenários de preço *spot*, é:

Período	PLD (R\$/MWh)	Custo Spot	Renda Contrato	Total (Contrato - Spot)
1	10	$Q^{min} \cdot \pi = R\$ 5$	$Q^{min} \cdot P = R\$ 50$	$50 - 5 = R\$ 45$
2	500	$Q^{max} \cdot \pi = R\$ 750$	$Q^{max} \cdot P = R\$ 150$	$150 - 750 = - R\$ 600$
Período 1 + 2		R\$ 755	R\$ 200	- R\$ 555

iii) Quantidade baseados no PLD com piso e teto

Nos contratos de quantidade baseados no PLD com piso e teto, o preço de referência dos contratos pode assumir diferentes valores em função do preço *spot*, como visto em (4). Assim, vamos supor um contrato PLD com os seguintes parâmetros:

Q (MWh)	π^{max} (R\$/MWh)	π^{min} (R\$/MWh)	Ágio
100	150	50	0

De acordo com (4), temos que:

- Para o primeiro período, $\pi = 10 \text{ R\$/MWh} < \pi^{min} = 50 \text{ R\$/MWh}$. Assim, o preço de referência do contrato passa a ser $P = \pi^{min} = 50 \text{ R\$/MWh}$.
- Já para o segundo período, $\pi = 500 \text{ R\$/MWh} > \pi^{max} = 100 \text{ R\$/MWh}$. Dessa forma, o preço de referência do contrato passa a ser $P = \pi^{max} = 150 \text{ R\$/MWh}$.

O fluxo dessa modalidade de contrato, para os dois cenários de preço *spot*, é:

Período	PLD (R\$/MWh)	Custo Spot	Renda Contrato	Total (Contrato - Spot)
1	10	$Q \cdot \pi = R\$ 10$	$Q \cdot \pi^{min} = R\$ 50$	$50 - 10 = R\$ 40$
2	500	$Q \cdot \pi = R\$ 500$	$Q \cdot \pi^{max} = R\$ 150$	$150 - 500 = - R\$ 350$
Período 1 + 2		R\$ 510	R\$ 200	- R\$ 310

iv) Comparação entre os fluxos dos contratos

Na tabela abaixo, podemos comparar os fluxos de cada modalidade de contrato para o mesmo cenário de preço *spot*.

	Período (R\$/MWh)	Custo Spot	Renda Contrato	Total (Contrato - Spot)
<i>Qt.</i>	1 10	$Q \cdot \pi = R\$ 10$	$Q \cdot P = R\$ 100$	$100 - 10 = R\$ 90$
	2 500	$Q \cdot \pi = R\$ 500$	$Q \cdot P = R\$ 100$	$100 - 500 = - R\$ 400$
	Período 1 + 2	R\$ 510	R\$ 200	- R\$ 310
<i>ToP</i>	1 10	$Q^{min} \cdot \pi = R\$ 5$	$Q^{min} \cdot P = R\$ 50$	$50 - 5 = R\$ 45$
	2 500	$Q^{max} \cdot \pi = R\$ 750$	$Q^{max} \cdot P = R\$ 150$	$150 - 750 = - R\$ 600$
	Período 1 + 2	R\$ 755	R\$ 200	- R\$ 555
<i>PLD</i>	1 10	$Q \cdot \pi = R\$ 10$	$Q \cdot \pi^{min} = R\$ 50$	$50 - 10 = R\$ 40$
	2 500	$Q \cdot \pi = R\$ 500$	$Q \cdot \pi^{max} = R\$ 150$	$150 - 500 = - R\$ 350$
	Período 1 + 2	R\$ 510	R\$ 200	- R\$ 310

Primeiramente, comprova-se a expectativa de que o ToP seria a modalidade que traria o pior rendimento à comercializadora, por conta de estarmos analisando um contrato de venda.

Já a modalidade de PLD com piso e teto, apesar da coincidência de obter um resultado final igual ao de quantidade padrão, possui uma menor volatilidade devido ao preço se ajustar dentro de um intervalo pré-definido em contrato (π^{max}, π^{min}).

No fim, a ideia central do projeto é desenvolver uma plataforma de apoio à decisão que tenha como entrada os parâmetros das oportunidades disponíveis (contratos) cadastradas pelo usuário e, como saída, o percentual da quantidade a ser contratada de cada uma dessas oportunidades, maximizando uma função de valor da renda líquida do comercializador.

B. Contratos de Disponibilidade

Contratos de disponibilidade funcionam como contratos de aluguel em que o comprador paga um valor mensal a uma dada usina geradora pelo direito de comercializar todo ou parte do seu lastro. Nas ocasiões em que o preço *spot* supera o valor do custo variável unitário (CVU) estabelecido no contrato, o comprador deve reembolsar o gerador pelo custo variável incorrido com a geração acima de G_i^{min} . O custo da compra de contratos nessa modalidade é, portanto, função do preço e da quantidade de energia contratada, e também dos níveis mínimo e máximo de geração da usina e do CVU. Portanto, a expressão de custo devido à compra de um contrato *i* de disponibilidade no período *t* e no cenário *s* é definida como:

$$C_{i,t,s}^D = Q_{i,t}^D \cdot P_{i,t}^D + G_{i,t,s}^{D,F}(CVU_i, \pi_{t,s}^{sub_i}) \cdot CVU_i. \quad (5)$$

Em que:

$$G_{i,t,s}^{D,F}(CVU_i, \pi_{t,s}^{sub_i}) = \begin{cases} G_i^{max} - G_i^{min}, & \pi_{t,s}^{sub_i} \geq CVU_i; \\ 0, & cc. \end{cases} \quad (6)$$

Sendo:

$Q_{i,t}^D$ – Quantidade de energia do contrato *i*, no período *t*, na modalidade disponibilidade (em MWh);

$P_{i,t}^D$ – Preço do contrato *i* no período *t* na modalidade disponibilidade (em R\$/MWh);

$G_{i,t,s}^{D,F}(CVU_i, \pi_{t,s}^{sub_i})$ – Geração flexível, acima do G_i^{min} , (em MWh) associada ao contrato *i* de disponibilidade no instante *t* e cenário *s*;

CVU_i – Custo variável unitário de geração associado ao contrato *i* (em R\$/MWh);

C. Expressão de Renda Líquida na CCEE – Câmara de Comercialização de Energia Elétrica

A nomenclatura usada para definir a expressão de renda da comercializadora é a seguinte:

\mathbf{x} – Vetor de variáveis de decisão. O elemento i do vetor \mathbf{x} representa o percentual da quantidade $Q_{i,t}$ a ser contratada, sendo $x_i \in [0,1]$;

CC_j : Conjunto de contratos candidatos de compra pertencentes à modalidade j ;

CE_j : Conjunto de contratos existentes de compra pertencentes à modalidade j ;

VC_j : Conjunto de contratos candidatos de venda pertencentes à modalidade j ;

VE_j : Conjunto de contratos existentes de venda pertencentes à modalidade j ;

DC : Conjunto de contratos candidatos pertencentes à modalidade *Disponibilidade*;

DE : Conjunto de contratos existentes pertencentes à modalidade *Disponibilidade*.

Sendo a modalidade j tal que $j \in \{Q, ToP, PLD\}$

Q – contratos de *Quantidade*;

ToP – contratos de quantidade com cláusulas de *Take or Pay*;

PLD – contratos de quantidade com preços *baseados no PLD com piso e teto*.

De acordo com o modelo vigente do setor elétrico, todas as compras e vendas no ACL devem ser registradas na *Câmara de Comercialização de Energia Elétrica* (CCEE). Na prática, a energia comprada e a energia vendida por uma comercializadora são contabilizadas na CCEE ao preço *spot*.

Os direitos e deveres da comercializadora contabilizados na CCEE possuem as seguintes expressões de receita e despesa, respectivamente:

$$R_{t,s}^{CCEE} = \sum_{\substack{i \in CC_Q \\ \cup CC_{PLD}}} x_i \cdot Q_{i,t} \cdot \pi_{t,s}^{sub_i} + \sum_{i \in DC} x_i \cdot (G_{i,t,s}^{D,F}(CVU_i, \pi_{t,s}^{sub_i}) + G_i^{min}) \cdot \pi_{t,s}^{sub_i} \\ + \sum_{i \in CC_{ToP}} x_i \cdot Q_{i,t}(\pi_{t,s}^{sub_i}) \cdot \pi_{t,s}^{sub_i} + \sum_{\substack{i \in CE_Q \\ \cup CE_{PLD}}} Q_{i,t} \cdot \pi_{t,s}^{sub_i} \\ + \sum_{i \in DE} (G_{i,t,s}^{D,F}(CVU_i, \pi_{t,s}^{sub_i}) + G_i^{min}) \cdot \pi_{t,s}^{sub_i} + \sum_{i \in CE_{ToP}} Q_{i,t}(\pi_{t,s}^{sub_i}) \cdot \pi_{t,s}^{sub_i}; \quad (7)$$

$$C_{t,s}^{CCEE} = \sum_{\substack{i \in VC_Q \\ \cup VC_{PLD}}} x_i \cdot Q_{i,t} \cdot \pi_{t,s}^{sub_i} + \sum_{\substack{i \in VE_Q \\ \cup VE_{PLD}}} Q_{i,t} \cdot \pi_{t,s}^{sub_i} \\ + \sum_{VC_{ToP}} x_i \cdot Q_{i,t}(\pi_{t,s}^{sub_i}) \cdot \pi_{t,s}^{sub_i} + \sum_{VE_{ToP}} Q_{i,t}(\pi_{t,s}^{sub_i}) \cdot \pi_{t,s}^{sub_i}. \quad (8)$$

Sendo:

$R_{t,s}^{CCEE}$ – Receita proveniente da contabilização dos direitos da comercializadora junto à CCEE (R\$);

$C_{t,s}^{CCEE}$ – Custo proveniente da contabilização dos deveres da comercializadora junto à CCEE (R\$).

Importante ressaltar que toda quantidade de contrato comprada passa a ser um direito da comercializadora, assim como toda quantidade vendida passa a ser uma obrigação. A CCEE é o ambiente em que todos estes direitos e obrigações são liquidados ao preço *spot*. Desta forma, os contratos de compra geram uma liquidação positiva e os de venda uma liquidação negativa, ambos na proporção da quantidade contratada. A exceção são os contratos de disponibilidade. Nestes, o direito é sobre o despacho de uma usina.

Portanto, o saldo final da contabilização dos direitos e deveres da comercializadora na CCEE possui a seguinte expressão:

$$S_{t,s}^{CCEE} = R_{t,s}^{CCEE} - C_{t,s}^{CCEE}. \quad (9)$$

D. Expressão Final de Renda da Comercializadora

Por fim, a expressão de renda final da comercializadora para um período t e um cenário s , considerando os contratos existentes e candidatos, pode ser obtida por:

$$\begin{aligned}
 R_{t,s} = & S_{t,s}^{CCEE} - \sum_{i \in DC} x_i \cdot C_{i,t}^D - \sum_{i \in DE} C_{i,t}^D + \sum_{\substack{i \in VC_Q \\ \cup VC_{ToP} \\ \cup VC_{PLD}}} x_i \cdot R_{i,t,s}^Q + \sum_{\substack{i \in VE_Q \\ \cup VE_{ToP} \\ \cup VE_{PLD}}} R_{i,t,s}^Q \\
 & - \sum_{\substack{i \in CC_Q \\ \cup CC_{ToP} \\ \cup CC_{PLD}}} x_i \cdot C_{i,t,s}^Q - \sum_{\substack{i \in CE_Q \\ \cup CE_{ToP} \\ \cup CE_{PLD}}} C_{i,t,s}^Q
 \end{aligned} \tag{10}$$

A relação entre contratos de compra e venda no período t deve satisfazer a restrição de balanço de energia que garante que a comercializadora não sofrerá penalidade¹. Assim:

$$\begin{aligned}
 & \sum_{\substack{i \in VC_Q \\ \cup VC_{PLD}}} x_i \cdot Q_{i,t} + \sum_{\substack{i \in VE_Q \\ \cup VE_{PLD}}} Q_{i,t} + \sum_{VC_{ToP}} x_i \cdot Q_{i,t}(\pi_{t,s}^{sub_i}) + \sum_{VE_{ToP}} Q_{i,t}(\pi_{t,s}^{sub_i}) \\
 & \leq \sum_{\substack{i \in CC_Q \\ \cup CC_{PLD}}} x_i \cdot Q_{i,t} + \sum_{CC_{ToP}} x_i \cdot Q_{i,t}(\pi_{t,s}^{sub_i}) + \sum_{\substack{i \in CE_Q \\ \cup CE_{PLD}}} Q_{i,t} \\
 & \quad + \sum_{CE_{ToP}} Q_{i,t}(\pi_{t,s}^{sub_i}) + \sum_{i \in DC} x_i \cdot Q_{i,t}^C + \sum_{i \in DE} Q_{i,t}^C.
 \end{aligned} \tag{11}$$

E. Modelo de Otimização

O próximo passo do trabalho consiste em modelar o problema apresentado pelas expressões acima como um problema de otimização linear sob incerteza. A expressão de renda apresentada na equação (10) fará parte da função objetivo a ser maximizada no modelo de otimização de contratação de energia da comercializadora. Para ilustrar o fluxo de dados do modelo, considere novamente o Fluxograma da Plataforma de Apoio à Decisão (Figura IV-1).

A função a ser otimizada será processada no otimizador (bloco 5). Esse bloco será alimentado por um bloco de pré-processamento (bloco 4), que organizará os dados de entrada do otimizador. Esses dados de entrada podem ser divididos nas seguintes categorias:

- i. Incertezas (bloco 1): São cenários de geração de energia elétrica das fontes renováveis e cenários de preço spot, ambos obtidos a partir do simulador de vento e vazão (bloco 5) ou de dados externos ao modelo;
- ii. Contratos candidatos e existentes (bloco 2): Formam o conjunto de contratos existentes da comercializadora e os ainda passíveis de serem contratados pela comercializadora;
- iii. Dados de perfil de aversão a risco do usuário (bloco 3): Define a forma com que os diferentes cenários de renda provenientes de (10) serão considerados na otimização.

Com as possibilidades de contratação de energia e levando em consideração o portfólio de contratos existentes (bloco 2), bem como as incertezas do modelo (bloco 1), o otimizador (bloco 5) definirá a quantidade ótima de energia a ser contratada (bloco 7) de cada um dos possíveis contratos candidatos. O vetor de decisão x , que estabelece o percentual comprado ou vendido de cada contrato candidato, será a variável de decisão do modelo.

Os dados de saída do modelo (bloco 6) apresentarão os resultados financeiros do processo de otimização e do simulador de vento e vazão. Como se trata de um modelo de otimização estocástica, faz-se necessária a utilização de uma medida de valor que traduza o perfil de risco do decisão. A medida a ser usada para representar o resultado financeiro será o CVaR (*Condition Value at Risk*), que leva em conta a média dos piores cenários de resultados obtidos pelo modelo de otimização. A quantidade de piores cenários de resultados será uma escolha dos usuários do modelo. Essa escolha deve ser feita considerando o nível de aversão a risco de cada usuário (bloco 3) [Street, 2009].

O resultado financeiro (bloco 6) permite que, a partir de uma única medida (o CVaR), seja possível avaliar a decisão de contratação do modelo de otimização. Entretanto, para que se tenha

¹ De fato, a restrição de lastro contratual é verificada através de uma média móvel. Entretanto, aqui adotamos uma restrição mais conservadora.

uma visão mais abrangente do resultado produzido, são apresentadas as estatísticas do resultado (bloco 8). Por exemplo, o usuário pode, com um histograma de frequência, determinar a probabilidade da comercializadora obter lucro ou prejuízo. Em um modelo de decisão sob incerteza, esses resultados são altamente desejáveis.

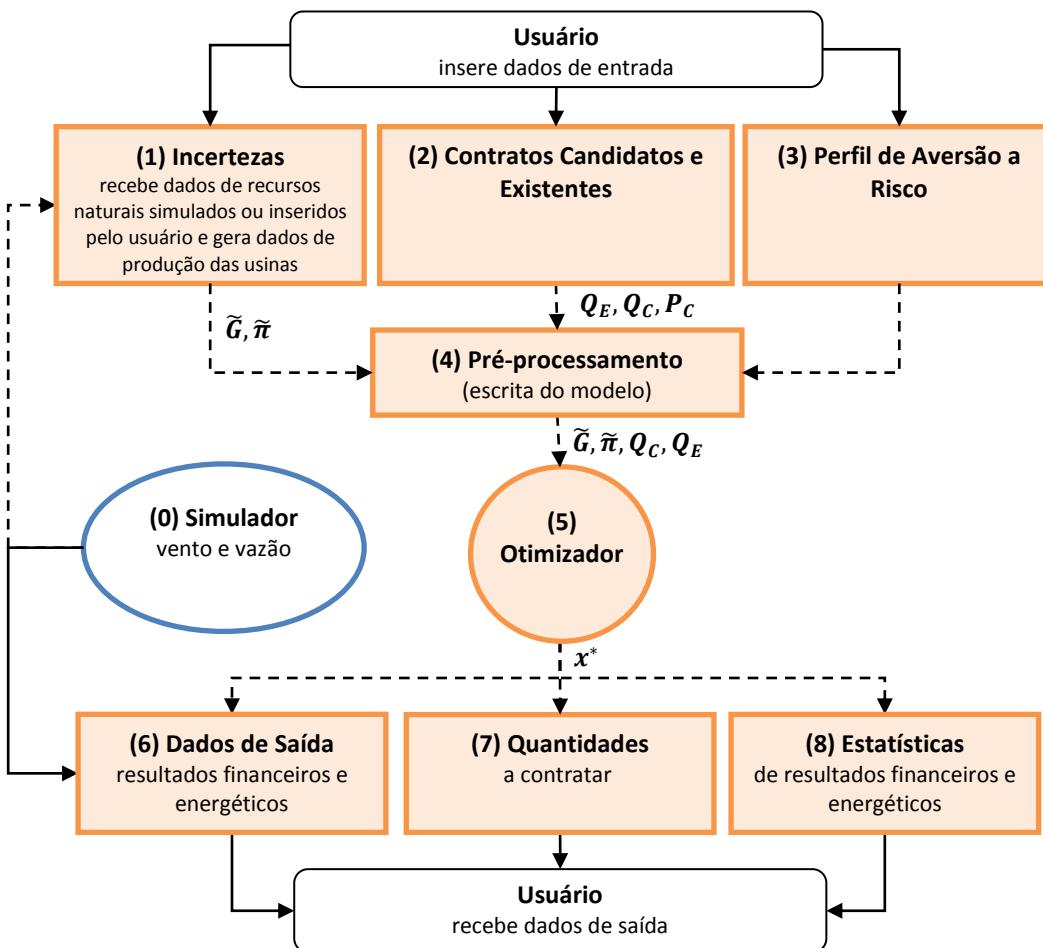


Figura IV-1: Fluxograma da Plataforma de Apoio à Decisão.

Sendo:

\tilde{G} – Matriz de cenários de geração de energia;

$\tilde{\pi}$ – Matriz de cenários de preço spot de energia;

Q_C – Vetor de contratos de energia candidatos;

Q_E – Vetor de contratos de energia existentes;

x^* – Vetor de quantidades ótimas da energia a ser contratada em cada contrato candidato.

V. CRONOGRAMA DE ATIVIDADES

O primeiro quadrimestre, que se encerra na data de entrega do presente relatório, foi realizado dentro do previsto no cronograma de atividades (ver Tabela V-1). Neste quadrimestre, um primeiro protótipo de modelo estatístico foi especificado e estimado para um conjunto de dados de testes realistas do setor elétrico brasileiro. Além disso, rotinas programadas em Matlab foram criadas para executar a simulação de séries sintéticas de vento e vazão. O intuito dessas rotinas é alimentar o bloco (1) do fluxograma apresentado na Figura II-1. Por fim, especificamos os quatro principais contratos que serão incorporados na plataforma de comercialização, bem como suas respectivas expressões de renda – definidas de maneira formal.

Conforme demonstra a Tabela V-1, o cronograma de atividades do projeto prevê, para o próximo quadrimestre (06/11 – 09/11), a apresentação da metodologia de especificação do modelo estatístico, além de diferentes testes de robustez dessa metodologia/modelo. É importante ressaltar que a especificação de um modelo estatístico multivariado com variáveis explicativas é um procedimento não trivial e que requer a análise de diversos indicadores de aderência. No capítulo III deste relatório, realizamos um conjunto de análises para uma série única de vazão. Diversas análises e testes foram realizados para este conjunto de dados de maneira a proporcionar uma especificação adequada dos parâmetros do modelo. Nesse sentido, os testes de robustez se fazem necessários na medida em que diferentes conjuntos de dados podem proporcionar diferentes qualidades nos indicadores de ajuste. Assim sendo, outros modelos estatísticos devem ser testados de maneira a proporcionar uma base mínima de comparação entre as diferentes abordagens.

Ainda no escopo das tarefas a serem realizadas no próximo semestre, está prevista a especificação do modelo de otimização de portfolio e o critério de risco a ser implementado na plataforma de apoio à decisão. De maneira a uniformizar o conhecimento dentro da equipe, um curso de comercialização de energia, que visa capacitar e/ou reciclar o conhecimento do grupo da UTE Norte Fluminense que terá contato com a plataforma, será formulado e ministrado em data futura a ser acordada entre as partes.

Para os dois quadrimestres subsequentes (10/11 – 01/12 e 02/12 – 05/12), serão definidas questões referentes à ferramenta de apoio à decisão (produto final da pesquisa) como a estrutura de dados a ser utilizada, interface computacional com o usuário, além de testes da estrutura do software com análise de casos.

Na reunião de exposição do presente relatório, algumas tecnologias (linguagens de programação), serão apresentadas para o grupo da UTE como propostas para a implementação da interface da plataforma de apoio à decisão, produto final deste projeto. Os pontos a favor e contra de cada tecnologia serão ressaltados e discutidos. Basicamente, os fatores que deverão ser ponderado são: (i) adequabilidade para o tipo de aplicação desenvolvida, (ii) flexibilidade de exportação e importação de dados externos, (iii) capacidade de integração com outras ferramentas de uso comum do grupo da UTE, (iv) razoabilidade de custos e, por fim, (v) flexibilidade para alterações no código sem a dependência da assistência de terceiros.

ATIVIDADES	Meses do Ano 2011												Meses do Ano 2012					
	1	2*	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	
1 Quadrimestre 1: Metodologia Estatística Vento e Vazão e Especificação de Contratos																		
1.1 Especificação manual de um modelo conjunto de vento e vazão																		
1.2 Inclusão de variáveis explicativas (ENAs) inseridas no modelo de despacho (Newave)																		
1.3 Testes de aderência e especificação dos modelos																		
1.4 Simulador e testes de compatibilidade com as principais características das séries																		
1.5 Especificação dos contratos candidatos: quantidade, disponibilidade, quantidade com ToF e baseados no PLD com piso, teto e ágio																		
1.6 Definição da expressão de renda por unidade de quantidade contratada																		
1.7 Elaboração do Relatório 1																		
ENTREGA DE PRODUTO: Relatório 1																		
2 Quadrimestre 2: Testes de Robustez do Modelo Estatístico e Especificação do Modelo de Otimização																		
2.1 Comparação entre o PAR-p e o modelo especificado em (1.2)-(1.3)																		
2.2 Metodologia de especificação dos modelos definidos em (2.1)																		
2.3 Testes de robustez da metodologia de especificação definida em (2.2)																		
2.4 Definição dos critérios de risco a serem implementados no modelo de otimização de portfolio																		
2.5 Especificação do modelo de otimização de portfolio geral																		
2.6 Especificação dos arquivos de entrada para o modelo de otimização																		
2.7 Testes do modelo de otimização para validação																		

(continua)

ATIVIDADES	Meses do Ano 2011												Meses do Ano 2012				
	1	2*	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5
2.8 Curso de comercialização de energia																	
2.9 Elaboração do Relatório 2																	
	ENTREGA DE PRODUTO: Relatório 2																
3 Quadrimestre 3: Estruturas de Dados e Pré-processamento																	
3.1 Estudo de mercado para a seleção de alternativas para a implementação da interface computacional dos modelos com o usuário																	
3.2 Definição dos arquivos de dados de contratos e perfis de geração associados aos contratos de disponibilidade																	
3.3 Especificação do fluxo de informação da interface de gestão de comercialização ótima de contratos																	
3.4 Especificação das rotinas de processamento de dados necessárias para promover o item (3.3) e produzir os arquivos especificados em (2.5)																	
3.5 Elaboração do Relatório 3																	
	ENTREGA DE PRODUTO: Relatório 3																
4 Quadrimestre 4: Implementação da interface e suas rotinas de gerenciamento de dados																	
4.1 Construção da interface computacional																	
4.2 Confecção de rotinas de leitura e escrita de arquivos de contratos e de simulação de vento e vazão.																	
4.3 Testes da estrutura da interface e da gestão de casos																	
4.4 Elaboração do Relatório 4																	
	ENTREGA DE PRODUTO: Relatório 4																

Tabela V-1: Cronograma de atividades realizadas e a realizar dos quatro primeiros quadrimestres.

REFERÊNCIAS

- HAMILTON, JAMES D. *Time Series Analysis*. New Jersey: Princeton University Press, 1994.
- MARTÍNEZ, C. "Principios de Convercion de la Energia Eólica". In: Amenedo, J. L. R.; Gómez, S. A.; e Díaz, J. C. B. *Sistemas Eólicos de Procucción de Energia Electrica*. Madrid: Rueda, 2003.
- PINDYCK, R.S. e RUBINFELD, D.L. *Econometric Models and Economic Forecasts*. Nova York: McGraw-Hill International Editors, 1991.
- STREET, A. "On the conditional value-at-risk probability-dependent utility function". *Theory and Decision*, vol. 68(1), pp. 49-68, fevereiro de 2010.
- THEIL, H. *Economic Forecasts and Policy*. Amsterdã: North-Kluwer Academic Publishing Company, 1961.
- THEIL, H. *Applied Economic Forecasting*. Amsterdã: North- Holland Publishing Company, 1966.