



**Schaiane Nogueira Ouverney Barroso**

**Modelo HPA para a modelagem de dados de alta  
frequência: aplicação à previsão horária de carga elétrica**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para  
obtenção do título de Mestre pelo Programa de Pós-  
Graduação em Engenharia Elétrica da PUC-Rio.

Orientador: Prof. Reinaldo Castro Souza

Rio de Janeiro, agosto de 2010.

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



**Schaiane Nogueira Ouverney Barroso**

**Modelo HPA para a Modelagem de Dados de  
Alta Frequência: Aplicação à Previsão Horária  
de Carga Elétrica.**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Reinaldo Castro Souza**

Orientador

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. José Francisco Moreira Pessanha**

UERJ

**Prof. Ricardo Tanscheit**

Departamento de Engenharia Elétrica – PUC-Rio

**Prof. José Eugenio Leal**

Coordenador Setorial do Centro

Técnico Científico – PUC-Rio

Rio de Janeiro, 26 de agosto de 2010.

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

### **Schiane Nogueira Ouverney Barroso**

Graduou-se em Estatística pela Escola Nacional de Ciências Estatísticas ENCE/IBGE em 2008 com a monografia intitulada "Uma comparação do Modelo Box & Jenkins sugerido pela ANEEL com a Modelagem Estrutural na previsão do Consumo Residencial de Energia Elétrica".

#### Ficha Catalográfica

Barroso, Schiane Nogueira Ouverney

Modelo HPA para a modelagem de dados de alta frequência: aplicação à previsão horária de carga elétrica / Schiane Nogueira Ouverney Barroso; orientador: Reinaldo Castro Souza – 2010.

89 f. : il. ; 30 cm

Dissertação (Mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2010.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. HPA. 3. ARIMA. 4. Dados de alta frequência. 5. Previsão de carga. 6. Séries temporais. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

*Para o meu avô,  
Laurindo Aldeno Nogueira Peixoto,  
por simplesmente acreditar.*

## **Agradecimentos**

Agradeço, em primeiro lugar, a Deus que me deu a vida e por ser meu alicerce sempre.

Obrigada, também, à minha família pelo amor incondicional e apoio em todos os momentos. Agradeço pelas boas energias, pelo incentivo e amor que me dedicaram.

Aos meus pais, Wanderley e Nilcéia, que foram compreensivos e confiaram em mim quando eu decidi correr atrás dos meus sonhos. Em especial à minha mãe, que sempre foi meu exemplo de mulher e de ser humano por seu caráter e força.

À minha irmã, Rayssa, que me fez ir mais longe para mostrar que nada é impossível e que sempre vale à pena tentar.

Ao meu marido, Giovane, que me apoiou sempre com carinho, amor e atenção, e que participou diretamente desse trabalho revisando, pacientemente, os textos e me orientando em toda a programação. Sem ele eu não teria conseguido. Formamos uma dupla e tanto, na vida e no trabalho.

Ao meu orientador, Professor Reinaldo, que me sugeriu o tema, me guiou e me apoiou em todo o processo de desenvolvimento do trabalho.

Ao professor Hasan Al-Madfai, desenvolvedor da metodologia abordada, pela solicitude em tirar dúvidas por e-mail e pelos papers e apresentações cedidas.

À CAPES e à PUC-Rio pela oportunidade e apoio financeiro imprescindível à obtenção deste título.

Aos professores e funcionários do Departamento de Engenharia Elétrica da PUC-Rio pela dedicação. Em especial, agradeço à Ana Paiva pelo carinho e atenção.

Aos funcionários das bibliotecas da PUC, em especial à Mônica de Oliveira da BS/CTC pela atenção e eficiência em me ajudar com inúmeras solicitações de livros e artigos.

Aos membros da banca por aceitar o convite e pelas contribuições valiosas.

Aos queridos amigos da PUC-Rio: Fernando Luiz Cyrino de Oliveira, Paulo Roberto Bastos Maia, Alexandre José dos Santos e Ilitch Vitali Gomes da Silva pelas dicas, idéias, conversas constantes e amizade.

A todos que direta ou indiretamente me ajudaram, torceram por mim e me apoiaram para vencer mais esta etapa.

## Resumo

Barroso, Schaiane Nogueira Ouverney; Souza, Reinaldo Castro. **Modelo HPA para a modelagem de dados de alta frequência: aplicação à previsão horária de carga elétrica.** Rio de Janeiro, 2010. 89p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A previsão de curto prazo, que envolve dados de alta frequência, é essencial para a confiabilidade e eficiência da operação do setor elétrico, fazendo com que a alocação da carga seja feita de forma eficiente, além de indicar possíveis distorções nos próximos períodos (dias, horas, ou frações de hora). A fim de garantir a operação energética, diversas abordagens têm sido empregadas com vistas à previsão de carga de energia a curto prazo. Dentre elas, pode-se citar os modelos híbridos de Séries Temporais, Lógica Fuzzy e Redes Neurais e o Método Holt-Winters com múltiplos ciclos que é a principal ferramenta utilizada atualmente. O HPA (*Hierarchical Profiling Approach*) é um modelo que decompõe a variabilidade dos dados de séries temporais em três componentes: determinística, estocástica e ruído. A metodologia é capaz de tratar observações únicas, periódicas e aperiódicas, e ao mesmo tempo, serve como uma técnica de pré-branqueamento. Este trabalho tem por objetivo implementar o HPA e aplicá-lo a dados de carga de energia elétrica de 15 em 15 minutos para um estado da região Sudeste do Brasil. Também serão analisadas as previsões de curto prazo geradas pelo modelo para a série considerada, visto que a habilidade preditiva do HPA ainda é desconhecida para séries brasileiras. As previsões forneceram Coeficiente U de Theil igual a 0,36 e um Erro Percentual Absoluto Médio (MAPE, *Mean Absolute Percentage Error*) de 5,46%, o qual é bem inferior ao valor fornecido pelo Modelo Ingênuo usado para comparação (15,08%).

## Palavras-chave

HPA, Hierarchical Profiling Approach; ARIMA; dados de alta frequência; previsão de carga; Séries Temporais.

## **Abstract**

Barroso, Schaiane Nogueira Ouverney; Souza, Reinaldo Castro (Advisor). **HPA model for modeling high frequency data: application to forecast hourly electric load.** Rio de Janeiro, 2010. 89p. MSc. Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Short-term forecast, which involves high frequency data, is essential for a reliable and efficient electricity sector operation, enabling an efficient power load allocation and indicating possible distortions in the coming periods (days, hours, or hour fractions). To ensure the operation efficiency, several approaches have been employed in order to forecast the short-term load. Among them, one can mention the hybrid models of Time Series, Fuzzy Logic and Neural Networks and Holt-Winters Method with multiple cycles, which is the main tool used today. The HPA (Hierarchical Profiling Approach) model decomposes the variability of time series data into three components: deterministic, stochastic and noise. The model is capable of modeling single, periodic and aperiodic observations, and at the same time function as a pre-whitening technique. This work aims to implement the HPA and to apply it in 15 in 15 minutes load data of a Brazil's southeastern state, since the predictive ability of the HPA is still not known for the Brazilian series. The short-term forecasts estimated for the series considered are analyzed and provided a Theil-U Coefficient equal to 0.36 and a Mean Absolute Percentage Error (MAPE) of 5.46%, which is smaller than the value given by the Naive Model (15.08%).

## **Keywords**

HPA; Hierarchical Profiling Approach; ARIMA; high frequency data; load forecasting; Time Series.

## Sumário

|  |    |
|--|----|
| 1 Introdução   | 14 |
| 2 Revisão Bibliográfica  | 17 |
| 2.1. Séries Temporais  | 17 |
| 2.2. Processo Estocástico  | 20 |
| 2.3. Previsão de Séries Temporais                                    | 21 |
| 2.4. Método Ingênuo (Naive)  | 23 |
| 2.5. Metodologia de Box & Jenkins                                    | 24 |
| 2.5.1. Processo de modelagem dos modelos ARIMA                       | 27 |
| 2.5.2. Testes de adequação do modelo                                 | 30 |
| 2.6. Regressão Harmônica   | 31 |
| 2.6.1. Regressão Harmônica com Tendência Polinomial                  | 36 |
| 2.6.2. Periodograma  | 37 |
| 2.6.3. Teste de Fisher   | 38 |
| 2.7. Ajuste polinomial de curvas                                     | 40 |
| 3 Modelo Proposto  | 42 |
| 4 Aplicação do Modelo  | 48 |
| 4.1. Perfil de Nível 1 – Padrões para os dias da semana              | 52 |
| 4.2. Perfil de Nível 2 – Tratamento de feriados                      | 59 |
| 4.3. Perfil de Nível 3 – Identificação e modelagem do padrão sazonal | 70 |
| 4.4 Criação da Função Determinística                                 | 74 |
| 4.5 Modelagem da Componente Estocástica                              | 76 |
| 5 Resultados   | 80 |
| 6 Conclusões   | 83 |
| 6.1. Sugestões   | 84 |
| 7 Referências Bibliográficas   | 86 |

## **Lista de tabelas**

|                                     |    |
|-------------------------------------|----|
| Tabela 4.1 Lista de feriados fixos  | 62 |
| Tabela 4.2 Lista de feriados móveis | 62 |

## Lista de figuras

|  |    |
|--|----|
| Figura 2.1 Mecanismo de geração de uma série temporal  | 25 |
| Figura 2.2 Análise de uma série temporal   | 25 |
| Figura 4.1 Série de demanda de energia em todo o histórico   | 49 |
| Figura 4.2 Demanda de energia a cada quinze minutos de um ano  | 49 |
| Figura 4.3 Demanda de energia a cada quinze minutos de um trimestre  | 50 |
| Figura 4.4 Demanda de energia a cada quinze minutos de um mês  | 50 |
| Figura 4.5 Demanda de energia a cada quinze minutos de uma semana  | 51 |
| Figura 4.6 Demanda de energia a cada quinze minutos de um dia  | 51 |
| Figura 4.7 Padrão médio observado e polinômio estimado para o perfil de domingo  | 54 |
| Figura 4.8 Padrão observado e polinômio estimado para o perfil de segunda-feira  | 54 |
| Figura 4.9 Padrão observado e polinômio estimado para o perfil médio de terça, quarta e quinta-feira                         | 55 |
| Figura 4.10 Padrão observado e polinômio estimado para o perfil de sexta-feira   | 56 |
| Figura 4.11 Padrão observado e polinômio estimado para o perfil de sábado  | 56 |
| Figura 4.12 Série obtida no nível 1  | 57 |
| Figura 4.13 Série residual proveniente do ajuste de nível 1  | 58 |
| Figura 4.14 Demanda de carga elétrica em uma semana que ocorreu um feriado   | 60 |
| Figura 4.15 Demanda em uma semana de feriado comum, uma semana de ocorrência de Natal e em uma semana onde não houve feriado | 61 |
| Figura 4.16 Padrão observado e polinômio estimado para o perfil de domingo de feriado  | 63 |
| Figura 4.17 Padrão observado e polinômio estimado para o perfil dos feriados às segundas-feiras                              | 64 |
| Figura 4.18 Padrão observado e polinômio estimado para o perfil dos feriados às terças-feiras                                | 65 |
| Figura 4.19 Padrão observado e polinômio estimado para o perfil dos feriados   |    |

|  |    |
|--|----|
| às quartas-feiras  | 65 |
| Figura 4.20 Padrão observado e polinômio estimado para o perfil dos feriados às quintas-feiras | 66 |
| Figura 4.21 Padrão observado e polinômio estimado para o perfil dos feriados às sextas-feiras  | 67 |
| Figura 4.22 Padrão observado e polinômio estimado para o perfil dos feriados aos sábados       | 68 |
| Figura 4.23 Série de polinômios para os feriados obtida no nível 2                             | 69 |
| Figura 4.24 Série de resíduos obtida após o ajuste dos níveis 1 e 2                            | 70 |
| Figura 4.25 Periodograma da série de demanda de carga de energia elétrica                      | 71 |
| Figura 4.26 Ajuste da série resultantes dos níveis 1 e 2                                       | 73 |
| Figura 4.27 Série de resíduos obtidos após o ajuste  | 74 |
| Figura 4.28 Histograma dos dados   | 77 |
| Figura 4.29 Normal Q-Q Plot dos dados de demanda de carga elétrica                             | 77 |
| Figura 4.30 FAC e FACP da série de demanda de carga elétrica                                   | 78 |
| Figura 5.31 Comparação entre as previsões geradas pelo HPA e pelo Método Ingênuo               | 81 |

*“Entre pedras  
cresceu minha poesia.  
Minha vida...  
Quebrando pedras  
e plantando flores.”  
(Cora Coralina)*

# 1 Introdução

A energia, em suas mais diversas formas, é indispensável à sobrevivência da espécie humana. E mais do que sobreviver, o homem procurou sempre evoluir, descobrindo fontes e maneiras alternativas de adaptação ao ambiente em que vive e de atendimento às suas necessidades.

Dessa forma, a exaustão, a escassez ou a inconveniência de um dado recurso tendem a ser compensadas pelo surgimento de outro(s). Em termos de suprimento energético, a eletricidade se tornou uma das formas mais versáteis e convenientes de energia, passando a ser um recurso indispensável e estratégico para o desenvolvimento socioeconômico de muitos países e regiões. No Brasil, as plantas hidrelétricas são responsáveis por quase 80% do total de energia produzida (ANEEL, 2007).

A exemplo do que ocorre em todo o mundo, o setor elétrico brasileiro tem passado por importantes alterações de cunho estrutural e institucional, migrando de uma configuração centrada no monopólio estatal como provedor dos serviços e único investidor, para um modelo de mercado com a participação de múltiplos agentes e investimentos partilhados com o capital privado. Esta nova configuração sustentou também a execução da privatização de ativos de serviços de energia elétrica sob controle estadual e federal, onde se inserem as empresas de distribuição de energia elétrica.

A transmissão de energia elétrica entre as regiões de geração e consumo é um processo que envolve diversos fatores importantes como qualidade e confiabilidade. Por isso, um planejamento adequado é parte essencial do processo, no qual devem ser levadas em consideração necessidades futuras de distribuição, que podem ser projetadas através da utilização de técnicas de previsão.

Diante desse novo contexto, a previsão de carga de curtíssimo prazo, que envolve dados de alta frequência, assume uma importância vital, servindo de base para o cálculo do preço futuro da energia elétrica, para a programação do despacho ótimo das usinas do sistema, fazendo com que a alocação da carga seja feita de forma eficiente, além de indicar possíveis distorções nos próximos períodos (dias, horas, ou frações de horas).

Assim, é de suma importância produzir uma boa previsão para a variável “carga elétrica”, a qual depende exclusivamente do usuário e de fatores que interferem no cotidiano do mesmo, como o clima, a economia e o nível social.

Com uma previsão de qualidade para valores futuros da carga elétrica, pode-se melhorar, em muito, procedimentos intrínsecos ao processo global de distribuição de energia, além, é claro, de corrigir possíveis problemas pré-existentes. Além disso, a previsão de carga de energia evita o desperdício, a escassez e a má alocação de recursos.

Com isso, a fim de garantir a operação energética no país, diversas abordagens têm sido empregadas com vistas à previsão de carga a curto prazo. Entre elas os modelos híbridos de Séries Temporais, Lógica Fuzzy e Redes Neurais e o Método Holt-Winters com múltiplos ciclos que é a principal ferramenta utilizada atualmente.

Em 2001, Hasan Al-Madfai desenvolveu o HPA (*Hierarchical Profiling Approach*), uma metodologia de modelagem que promete fornecer previsões superiores às abordagens tradicionais. O HPA decompõe a variabilidade dos dados de séries temporais em três componentes: determinística, estocástica e ruído.

A identificação dos perfis de carga é, então, modelada por meio de funções contínuas acumuladas aditivamente na componente determinística. Assim, a componente estocástica pode ser modelada utilizando uma abordagem adequada, geralmente um modelo ARIMA, e o ruído é assumido como aleatório.

O HPA é capaz de modelar observações únicas, periódicas e aperiódicas, enquanto, ao mesmo tempo, serve como uma técnica de pré-branqueamento. Segundo Al-Madfai *et al.* (2001, 2004 e 2007), a aplicação desta técnica melhora significativamente a confiabilidade das previsões em termos das métricas MAPE e RMSE.

Este trabalho tem por objetivo implementar a metodologia e testar as previsões de curto prazo geradas pelo modelo HPA, utilizando dados de carga de energia de alta frequência, além verificar a adequação da metodologia a dados de energia brasileiros, visto que sua habilidade preditiva ainda é desconhecida frente a tais séries de tempo.

A dissertação está organizada em seis capítulos. No segundo capítulo serão descritos alguns conceitos básicos, utilizados ao longo deste trabalho, além de métodos estatísticos que são empregados na construção do modelo HPA. No terceiro capítulo será apresentada e discutida a metodologia *Hierarchical Profiling Approach*. Em seguida, no capítulo quatro, são apresentados alguns resultados da análise estatística da série demanda de energia elétrica de 15 em 15 minutos para um estado da região Sudeste do Brasil e a aplicação da modelagem HPA a estes dados. O capítulo cinco apresenta os resultados obtidos e uma comparação via critério MAPE e Coeficiente U de Theil entre as previsões para uma semana à frente fornecidas pelo HPA e pelo Método Ingênuo de previsão. Por fim, no capítulo seis são apresentadas as principais conclusões e sugestões para futuros trabalhos e aprofundamentos na área.

## 2 Revisão Bibliográfica

Este capítulo apresenta os principais conceitos, abordagens e a formulação básica das metodologias que estão incluídas no modelo HPA. Conceitos matemáticos e estatísticos mais detalhados podem ser encontrados nas referências bibliográficas citadas.

### 2.1. Séries Temporais

Uma série temporal é uma coleção de observações de uma variável ordenadas no tempo (Souza e Camargo, 2004). Há, basicamente, dois enfoques usados na análise de séries temporais. Em ambos, o objetivo é construir modelos para as séries, com propósitos determinados. No primeiro enfoque, a análise é feita no domínio temporal e os modelos propostos são modelos paramétricos. No segundo, a análise é conduzida no domínio de frequências e os modelos propostos são modelos não-paramétricos.

Segundo Morettin (2006), a natureza de uma série temporal e a estrutura de seu mecanismo gerador estão relacionadas ao intervalo de ocorrência das observações no tempo. Caso o levantamento das observações da série possa ser feito a qualquer momento do tempo, a série temporal é dita contínua. Entretanto, na maioria das séries, as observações são tomadas em intervalos de tempo discretos e equidistantes.

Uma série temporal discreta pode ser representada por  $Z(t_1), \dots, Z(t_n)$ , sendo que cada observação está associada a um instante de tempo distinto, existindo uma relação de dependência serial entre essas observações.

O objetivo inicial da análise de séries temporais é a realização de inferências sobre as propriedades ou características básicas do mecanismo gerador das observações da série. Assim, através da abstração de regularidades contidas nos fenômenos observáveis de uma série temporal, pode-se construir um modelo matemático como uma representação simplificada da realidade.

Quando o enfoque é feito no domínio do tempo considera-se a evolução temporal do processo, ou seja, o interesse está na dimensão do evento em cada instante de tempo. Para este tipo de avaliação, são usadas as funções de autocovariância e autocorrelação.

A autocovariância de *lag*  $k$  é a covariância entre duas observações defasadas de  $k$  intervalos de tempo entre si e pode ser definida por:

$$\gamma_k = Cov [Z_t, Z_{t+k}] = E[(Z_t - \mu)(Z_{t+k} - \mu)] \quad (2.1)$$

onde  $\mu$  é a média do processo.

A autocovariância de uma série temporal pode ser estimada por:

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}) \quad (2.2)$$

sendo  $\bar{Z}$  a média de  $Z_t$ , descrita por:

$$\bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t \quad (2.3)$$

A autocorrelação é a autocovariância padronizada e tem a função de medir a extensão para o qual um valor tomado no instante  $t$  depende daquele tomado no instante  $t-k$  (Casella, 2001).

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{Cov}[Z_t, Z_{t+k}]}{\sqrt{\text{Var}(Z_t)\text{Var}(Z_{t+k})}} \quad (2.4)$$

sendo:  $\gamma_0 = \text{Var}(Z_t) = \text{Var}(Z_{t+k})$  a variância do processo,  $\rho_0 = 1$  e  $\rho_k = \rho_{-k}$ .

A autocorrelação de uma série pode ser estimada pela fórmula:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{N-k} (Z_t - \bar{Z})^2} \quad (2.5)$$

Outra função importante é a Autocorrelação Parcial, que mede a correlação entre duas observações seriais  $Z_t$  e  $Z_{t+k}$  eliminando a dependência dos termos intermediários  $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$ , e é representada por:

$$\text{Cor}(Z_t, Z_{t+k} / Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}) \quad (2.6)$$

Na análise no domínio da frequência, o interesse está em verificar a frequência com que os eventos ocorrem em determinado período de tempo. A ferramenta utilizada para isso é a Análise Espectral, onde são estabelecidas as características de um processo estocástico em termos de frequências podendo, no caso das séries temporais, determinar as periodicidades existentes nas mesmas (Morettin, 2006). Como o espectro de um processo não é conhecido, este precisa ser estimado, em geral, por meio do Periodograma de janelas espectrais, que possui boas propriedades estatísticas.

## 2.2. Processo Estocástico

Um processo estocástico pode ser definido como um conjunto de variáveis aleatórias definidas segundo uma lei de probabilidades (Barros, 2004). Uma série temporal é uma realização de um processo estocástico. Assim, uma série  $y_t$  pode ser definida como uma função  $y$  da variável independente  $t$ , que é gerada por um processo estocástico desconhecido (Souza e Camargo, 2004).

Tendo em vista que o processo estocástico gerador de dados não é descrito por funções determinísticas, seu comportamento futuro deve ser descrito probabilisticamente. O processo somente pode ser estatisticamente determinado quando todas as suas funções de distribuição de probabilidade são conhecidas até a T-ésima ordem (Barros, 2004).

Na prática, as funções de probabilidade até a T-ésima ordem não são conhecidas e tem-se acesso a uma única realização do processo estocástico, a partir do qual se deseja inferir sobre o mecanismo gerador da série. Devido a isto, consideram-se duas restrições: estacionariedade e ergodicidade.

Um processo é dito estacionário se não existem mudanças sistemáticas em suas características. A estacionariedade pode ser do tipo forte ou fraca.

Na estacionariedade forte ou estrita, a forma da distribuição conjunta do processo permanece invariante mediante uma translação temporal. Como na prática é difícil especificar a distribuição conjunta de um processo estocástico, há a versão mais fraca, na qual somente alguns momentos do processo permanecem inalterados no tempo (Medeiros, 2005).

Assim, um processo é dito fracamente estacionário ou estacionário de segunda ordem se sua média é constante e sua autocovariância depende apenas do *lag*, isto é:

$$\begin{aligned}
E[Z_t] &= \mu & \forall t \\
\text{Var}[Z_t] &= \sigma^2 & \forall t \\
E[(Z_t - \mu)(Z_{t-k} - \mu)] &= \gamma_k & \forall t
\end{aligned}
\tag{2.7}$$

Um processo é dito ergódico se apenas uma realização é suficiente para obter todas as suas propriedades estatísticas. Por definição, todo processo ergódico é também estacionário.

A série temporal e o processo estocástico relacionam-se de maneira semelhante a uma amostra e sua população de origem. Então, o objetivo da análise de uma série temporal é retirar uma amostra finita de dados equidistantes no tempo (série temporal) de uma realidade (processo estocástico) e identificar um modelo que seja capaz de inferir sobre o comportamento da realidade.

### 2.3. Previsão de Séries Temporais

Uma previsão é uma estimativa quantitativa de eventos futuros, baseada na informação atual e passada (Souza, 1983).

O horizonte de previsão é o intervalo de tempo, contado a partir da origem de previsão, para o qual as previsões devem ser realizadas. Além disso, denomina-se número de passos à frente da previsão o número de intervalos previstos a partir da origem (Souza, 1989).

Assim, a previsão denotada por  $\hat{Z}_t(k)$  é definida como a esperança condicional de  $Z_{t+k}$  dados os valores passados:

$$\hat{Z}_t(k) = E(Z_{t+k} | Z_t, Z_{t-1}, \dots) \tag{2.8}$$

onde  $Z_{t+k}$  são os valores a serem previstos para  $k=1,2,\dots$

A maioria dos métodos de previsão de séries temporais se baseia na suposição de que observações passadas contêm todas as informações sobre o padrão de comportamento da série temporal e de que esse padrão é recorrente no tempo.

Os métodos de previsão, classificados como métodos quantitativos, baseiam-se na extrapolação de características de observações passadas e no inter-relacionamento entre essas observações, fornecendo previsões acuradas se o futuro apresentar comportamento similar ao passado.

Segundo Souza (1989), uma previsão quantitativa elaborada a partir de métodos estatísticos é caracterizada por: sua origem, o número de passos à frente, seu valor pontual e por uma medida de incerteza a ela associada (a variância, por exemplo).

O propósito dos métodos de previsão consiste em distinguir o padrão de qualquer ruído que possa estar contido nas observações e, então, usar esse padrão para prever os valores futuros da série temporal. Assim, pela identificação dessa componente, a previsão para períodos de tempo subsequentes ao instante observado pode ser desenvolvida.

Apesar de a quase totalidade dos métodos de previsão de séries temporais estar fundamentada apenas na análise das observações da série de interesse para a especificação de algum modelo que descreva essas observações (modelos univariados), alguns procedimentos de previsão buscam explicar o comportamento de uma série temporal pela evolução dos fenômenos observacionais de outras séries (modelos multivariados).

## 2.4. Método Ingênuo (Naíve)

O Método Ingênuo, também conhecido como *Random Walk* (Passeio Aleatório), é o mais simples para realizar a previsão de uma variável (Souza, 2004). Sob esta perspectiva, o valor previsto é igual ao último valor observado, independente do número de passos à frente. O método pode ser representado pela equação:

$$Z_t = Z_{t-1} + \varepsilon_t \quad (2.9)$$

A previsão  $\tau$  passos à frente a partir do instante  $t$  é denotada por  $\hat{Z}_t(\tau)$ . Pode-se escrever a equação de previsão da seguinte forma:

$$\hat{Z}_t(\tau) = E[(Z_{t-1+\tau} + \varepsilon_{t+\tau}) | Z_t] = E(Z_{t-1+\tau}) \quad (2.10)$$

Por sua simplicidade, o Método Ingênuo é usado como critério de comparação (*benchmark*), isto é, o modelo a ser testado tem que ser sempre melhor do que este. Para tal, pode-se fazer uso de alguns critérios de comparação.

A métrica MAPE, *Mean Absolute Percentage Error*, indica o valor médio do erro percentual das previsões e é definido pela seguinte fórmula:

$$MAPE = \frac{\sum_{k=1}^N \left| \frac{a_k - y_k}{a_k} \right|}{N} \times 100 \quad (2.11)$$

onde  $N$  é o número de previsões realizadas e  $a_k$  e  $y_k$  os valores observado e previsto para o índice  $k$ , respectivamente.

O coeficiente U de Theil é outra métrica utilizada para avaliar o desempenho da previsão em relação ao Método Ingênuo. Este coeficiente analisa a qualidade da previsão baseado na equação:

$$U = \frac{\sqrt{\sum_{i=1}^N (a_k - y_k)^2}}{\sqrt{\sum_{i=1}^N (a_k - a_{k-1})^2}} \quad (2.12)$$

O coeficiente pode ser interpretado da seguinte forma:

$$\begin{cases} U > 1 & \text{O erro do modelo estimado é maior que o erro da previsão ingênua} \\ U < 1 & \text{O erro do modelo estimado é menor que o erro da previsão ingênua} \end{cases}$$

Quanto mais próximo de zero for o coeficiente  $U$ , melhor é a previsão do modelo estimado.

## 2.5. Metodologia de Box & Jenkins

A metodologia proposta por Box & Jenkins surgiu na década de 70 e provocou uma revolução na área de análise e previsão de séries temporais. Esta classe de modelos é largamente utilizada na prática e se firmou como uma das técnicas disponíveis mais eficientes.

Uma das características fundamentais desta metodologia é interpretar a série temporal como uma realização de um vetor aleatório multivariado cuja dimensão é a série disponível (Box & Jenkins, 1970).

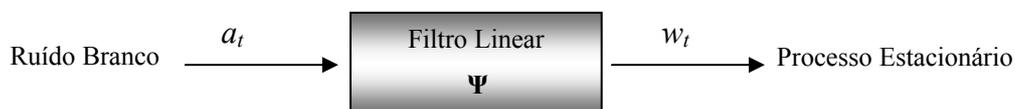
A partir de uma única realização do processo e com argumentos de estacionariedade e ergodicidade, procura-se detectar o sistema gerador, através das informações contidas na série.

Os modelos de Box & Jenkins são didaticamente divididos em:

- Modelos autorregressivos (AR), onde a série é explicada em função de suas próprias observações passadas;

- Modelos médias móveis (MA), que explicitam os valores futuros da série em função dos resíduos passados;
- Modelos autorregressivos de médias móveis (ARMA), que são uma combinação dos dois modelos anteriores;
- Modelos (ARIMA), que diferem dos ARMA pela aplicação de diferenças sucessivas na série temporal quando esta não respeita as condições de estacionariedade e ergodicidade, e
- Modelos SARIMA, para séries sazonais.

O modelo de Box & Jenkins tem como base a Teoria Geral de Sistemas Lineares, a qual supõe que a passagem de um ruído branco por um filtro linear de memória infinita gera um processo estacionário de segunda ordem (Souza e Camargo, 2004). Este mecanismo está representado na Figura 2.1.



**Figura 2.1** Mecanismo de geração de uma série temporal

O objetivo do modelo de Box & Jenkins é, então, encontrar o sistema inverso que, a partir de uma série temporal, seja capaz de gerar um ruído branco, como mostrado na Figura 2.2. Quando o ruído branco é obtido, pode-se dizer que a estrutura de dependência da série temporal foi capturada.



**Figura 2.2** Análise de uma série temporal

A metodologia pode ser aplicada a séries estacionárias ou não, e a condição necessária para a aplicação do modelo a séries não estacionárias, é que a estacionariedade possa ser obtida através de diferenças e/ou transformações não lineares da série original. A filosofia de tal modelagem é o princípio da parcimônia e a construção de modelos através de um ciclo iterativo.

Os operadores utilizados para a manipulação dos modelos são:

- Operador de translação para o passado, denotado por  $B$  e definido por:

$$\begin{aligned} BZ_t &= Z_{t-1} \\ B^m Z_t &= Z_{t-m} \end{aligned} \quad (2.13)$$

- Operador de translação para o futuro, denotado por  $F$  e definido por:

$$\begin{aligned} FZ_t &= Z_{t+1} \\ F^m Z_t &= Z_{t+m} \end{aligned} \quad (2.14)$$

- Operador de diferença, denotado por  $\Delta$  e definido por:

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t \quad (2.15)$$

sendo  $\Delta = 1 - B$ .

- Operador da soma, denotado por  $S$  e definido por:

$$SZ_t = \sum_{j=0}^{\infty} Z_{t-j} = Z_t + Z_{t-1} + \dots + (1 + B + B^2 + \dots)Z_t \quad (2.16)$$

sendo  $S = \Delta^{-1}$ .

O modelo de Box & Jenkins segue a formulação:

$$w_t = \Psi(B)a_t = \sum_{k=0}^{\infty} \psi_k a_{t-k} \quad (2.17)$$

O polinômio  $\Psi(B)$  possui infinitos parâmetros. Box *et al.* (1994) argumentam que, sob algumas restrições, todo polinômio infinito pode ser expresso pelo quociente de dois polinômios finitos. Então segue que:

$$\Psi(B) = \frac{\theta(B)}{\phi(B)} \quad (2.18)$$

onde:

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  é o polinômio de médias móveis MA(q) e

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  é o polinômio autorregressivo AR(p).

Generalizando, têm-se os modelos ARIMA(p,d,q):

$$\phi(B)\nabla^d Z_t = \theta(B)a_t \quad (2.19)$$

O melhor modelo deve ser parcimonioso, ou seja, deve conter o menor número de parâmetros possível para ajustar-se à série de dados observados.

### **2.5.1. Processo de modelagem dos modelos ARIMA**

Pode-se afirmar que a metodologia de Box & Jenkins segue quatro estágios principais:

### **Etapa 1 – Identificação**

Considerada como a fase crítica da metodologia, as principais ferramentas da identificação são a FAC (Função de Autocorrelação), a FACP (Função de Autocorrelação Parcial) e os correlogramas resultantes, que são as representações gráficas das FAC's e FACP's contra o tamanho da defasagem. De acordo com o conceito, a autocorrelação parcial é análoga ao conceito de coeficiente de regressão parcial; assim, esta mede a correlação entre observações (séries temporais) que sejam  $k$  períodos afastados.

Na identificação, o objetivo é determinar as ordens (p,d,q) para a componente não sazonal e/ou as ordens (P,D,Q), para a componente sazonal, além de estimativas preliminares dos parâmetros a serem usadas na fase seguinte.

### **Etapa 2 – Estimação**

Esta etapa consiste em estimar os parâmetros dos termos autorregressivo e de média móvel incluídos no modelo, através da maximização da função de verossimilhança condicional, equivalente à estimação de Mínimos Quadrados, supondo-se normalidade de  $a_t$ .

Na estimação dos parâmetros desconhecidos, procura-se minimizar a soma do quadrado dos erros, ou seja, no caso do modelo ARIMA(p,d,q), busca-se minimizar:

$$S(\xi) = \sum_{t=1}^T a_t^2, \text{ onde } a_t = \theta^{-1}(B) \cdot \phi(B) \cdot \nabla^d Z_t \quad (2.20)$$

Com a suposição da normalidade dos erros pode-se escrever a função de verossimilhança e proceder à estimação dos parâmetros por Máxima Verossimilhança.

### Etapa 3 – Verificação

Nesta fase é necessário verificar se o modelo representa adequadamente os dados. Existem diversos testes para tal procedimento e a grande maioria se baseia na análise dos resíduos dos modelos, ou seja, se os resíduos seguem um comportamento ruído branco.

Desta forma, o processo é iterativo e se o modelo selecionado não é adequado, retorna-se à etapa de identificação para a escolha de outro modelo. Um procedimento muitas vezes utilizado é o de identificar não só um único modelo, mas alguns modelos que serão então estimados e verificados. Se o propósito é realizar previsões, escolhe-se entre os modelos ajustados o melhor no sentido de, por exemplo, fornecer o menor Erro Quadrático Médio de previsão. É possível que vários pesquisadores identifiquem modelos diferentes para uma mesma série temporal (Morettin, 2006).

### Etapa 4 – Previsão

O objetivo nesta fase é prever um valor  $Z_{t+\tau}$ , com horizonte de previsão  $\tau \geq 1$ , supondo que temos observações  $\dots, Z_{t-2}, Z_{t-1}, Z_t$ , até o instante  $t$ , que é chamado origem das previsões. Esta previsão é de origem  $t$  e horizonte  $\tau$ , denotada por  $\hat{Z}_t(\tau)$ . A expressão de previsão pode ser escrita da seguinte maneira, utilizando a equação de diferenças (sendo  $\tau \geq 1$ ):

$$\hat{Z}_t = E[Z_{t+\tau} | Z_t] = \phi_1[Z_{t+\tau-1}] + \dots + \phi_p[Z_{t+\tau-p}] - \theta_1[a_{t+\tau-1}] - \dots - \theta_q[a_{t+\tau-q}] + [a_{t+\tau}] \quad (2.21)$$

Porém, os seguintes conceitos devem ser seguidos:

$$[Z_{t+k}] = \hat{Z}_t(k) \rightarrow k > 0$$

$$[Z_{t+k}] = Z_{t+k} \rightarrow k \leq 0$$

$$[a_{t+k}] = 0 \rightarrow k > 0$$

$$[a_{t+k}] = a_{t+k} \rightarrow k \leq 0$$

### 2.5.2. Testes de adequação do modelo

Existem diversos testes para verificar a adequação do modelo estimado. Box & Jenkins (1970) sugerem, entre outros, o Critério de Sobrefixação, o Teste dos Resíduos Gerados, que pode ser realizado através da verificação do Teste de Portmanteau e do Teste de Periodograma Acumulado.

O Critério de Sobrefixação consiste, basicamente, na elaboração de um modelo com um número de parâmetros superior ao do modelo fixado, que cubra as supostas direções de discrepâncias (Souza, 2004). Caso sejam encontrados parâmetros significativos, há um indicativo de que o modelo foi subidentificado.

Para testar os resíduos, o Teste de Portmanteau estima a autocorrelação residual e através da estatística:

$$Q = n \sum_{j=1}^k r_j^2(\hat{a}) \quad (2.22)$$

A hipótese nula é a de que os resíduos são aleatórios e, para que isso ocorra, o valor de Q deve ser menor do que o valor da abscissa da distribuição Qui-Quadrado com  $(k-p-q)$  graus de liberdade.

No teste do Periodograma acumulado, compara-se o Periodograma acumulado da série dos resíduos com o de um ruído branco, com o intuito de se

encontrarem componentes periódicos (Montgomery, 1976). A aplicação desse teste é feita, geralmente, em séries de carga semanal ou mensal, visto que o modelo comporta apenas a sazonalidade, não incluindo a modelagem dos ciclos.

## 2.6. Regressão Harmônica

A Análise Harmônica ou Análise de Fourier é, tradicionalmente, utilizada para resolver equações diferenciais parciais presentes na Física. Na análise de séries temporais, o objetivo básico é aproximar uma função do tempo por uma soma de componentes senoidais, cujos coeficientes são as transformadas de Fourier da série (Morettin, 2006).

A essência da análise de Fourier é a representação de um conjunto de dados em termos de funções senoidais. Neste ponto, isto parece apropriado para justificar a escolha destas funções, visto que muitas outras famílias de funções periódicas possuem as mesmas propriedades das senóides. Qualquer destas famílias poderia ser utilizada de forma similar (Artis *et al.*, 2007).

As oscilações nas séries podem ser descritas em termos senoidais pela Análise Espectral. Este é um método que descreve a tendência para as oscilações de uma dada frequência que aparece nos dados, em vez das próprias oscilações.

Segundo Bloomfield (1976), a propriedade mais básica das senóides que as torna geralmente adequadas à análise de séries temporais é o seu comportamento sob uma simples mudança de escala de tempo. Uma senóide de frequência  $\omega$  (em radianos por unidade de tempo) ou período  $2\pi/\omega$  pode ser escrita como:

$$f(t) = R \cos(\omega t + \phi) \quad (2.23)$$

onde  $R$  é a amplitude e  $\phi$  é a fase.

Outra característica útil das senóides é o seu comportamento sob amostragem (isto é, observando uma função da variável contínua  $t$  em um conjunto de valores igualmente espaçados  $t_0, t_1, \dots$ ), se seu intervalo amostral é  $\Delta$ , as senóides:

$$R \cos(\omega_1 t + \phi) \quad \text{e} \quad R \cos(\omega_2 t + \phi) \quad (2.24)$$

são imperceptíveis quando  $\omega_1 - \omega_2$  é um múltiplo de  $2\pi / \Delta$  (Bujosa *et al.*, 2007). Este fenômeno é conhecido como *aliasing*.

Muitas vezes o interesse da Análise Harmônica está em buscar periodicidades nos dados observados (Amaral, 1975). Há duas situações frequentes que podem ocorrer:

- As frequências são conhecidas e o objetivo é estimar amplitudes e fases;
- Há necessidade de estimar amplitudes, frequências e fases.

Quando o modelo possui uma única periodicidade, que é conhecida, pode-se representar os dados segundo a equação (Morettin, 2006):

$$Z_t = \mu + R \cos(\omega t + \phi) + \varepsilon_t \quad (2.25)$$

ou, de forma equivalente,

$$Z_t = \mu + A \cos(\omega t) + B \sin(\omega t) + \varepsilon_t \quad (2.26)$$

em que  $A = R \cos \phi$ ,  $B = -R \sin \phi$ ,  $R$  é a amplitude,  $\phi$  é o ângulo de fase,  $\omega$  é a frequência e  $\varepsilon_t$  é a componente aleatória.

De (2.25) e (2.26) tem-se que  $R^2 = A^2 + B^2$ , então:

$$\phi = \begin{cases} \arctan\left(-\frac{B}{A}\right), & A > 0, \\ \arctan\left(-\frac{B}{A}\right) - \pi, & A < 0, B > 0, \\ \arctan\left(-\frac{B}{A}\right) + \pi, & A < 0, B < 0, \\ -\pi/2, & A = 0, B > 0, \\ \pi/2, & A = 0, B < 0, \\ \text{arbitrário} & A = 0, B = 0. \end{cases} \quad (2.27)$$

O problema é, então, estimar  $A$ ,  $B$  e  $\mu$  para um  $\omega$  fixado, o que é feito utilizando o Método de Mínimos Quadrados. Os estimadores  $\hat{\mu}$ ,  $\hat{A}$  e  $\hat{B}$  são obtidos pela minimização da soma do quadrado dos erros dada pela equação:

$$S(\mu, A, B) = \sum_{t=1}^N (Z_t - \mu - A \cos(\omega t) - B \sin(\omega t))^2 \quad (2.28)$$

A estimação segundo a equação acima resulta em um sistema de equações cuja solução não é trivial. Assim, uma maneira alternativa para encontrar os estimadores é usar a notação matricial para reescrever a equação (2.26) (Bloomfield, 1976).

$$Z = W\theta + \varepsilon \quad (2.29)$$

em que:

$$Z = (Z_1, Z_2, \dots, Z_N)'$$

$$\theta = (\mu, A, B)$$

$$W = \begin{bmatrix} 1 & \cos(\omega) & \sin(\omega) \\ 1 & \cos(2\omega) & \sin(2\omega) \\ \vdots & \vdots & \vdots \\ 1 & \cos(N\omega) & \sin(N\omega) \end{bmatrix}$$

que fornece o estimador:

$$\hat{\theta} = (W'W)^{-1}W'Z \quad (2.30)$$

cuja solução é:

$$\hat{\mu} = \frac{\sum_{t=1}^N Z_t}{N} = \bar{Z}$$

$$\hat{R} = \hat{A}^2 + \hat{B}^2 \quad (2.31)$$

$$\hat{\phi} = \arctg\left(\frac{-\hat{B}}{\hat{A}}\right)$$

Se  $\omega \neq \pi$ :

$$\hat{A} = \frac{2}{N} \sum_{t=1}^N Z_t \cos(\omega t)$$

$$\hat{B} = \frac{2}{N} \sum_{t=1}^N Z_t \sin(\omega t) \quad (2.32)$$

Se  $\omega = \pi$ :

$$\hat{A} = \frac{1}{N} \sum_{t=1}^N Z_t (-1)^t$$

$$\hat{B} = 0 \quad (2.33)$$

Quando  $\omega \neq \frac{2\pi}{N}$  e não muito próximo de zero, uma solução aproximada é dada por (Morettin, 2006):

$$\begin{aligned}
\tilde{\mu} &= \hat{\mu} = \bar{Z} \\
\tilde{A} &= \frac{2}{N} \sum_{t=1}^N (Z_t - \bar{Z}) \cos(\omega t) \\
\tilde{B} &= \frac{2}{N} \sum_{t=1}^N (Z_t - \bar{Z}) \text{sen}(\omega t) \\
\tilde{R} &= \tilde{A} + \tilde{B} \\
\tilde{\phi} &= \text{arctg} \left( -\frac{\tilde{B}}{\tilde{A}} \right)
\end{aligned} \tag{2.34}$$

Quando um modelo tem periodicidades múltiplas ( $k$  periodicidades), a equação (2.26) pode ser reescrita da forma:

$$Z_t = \mu + A_1 \cos(\omega_1 t) + B_1 \text{sen}(\omega_1 t) + \dots + A_k \cos(\omega_k t) + B_k \text{sen}(\omega_k t) + \varepsilon_t \tag{2.35}$$

onde  $\omega_1, \dots, \omega_k$  são frequências conhecidas.

O modelo (2.35) é uma regressão linear, conhecida como Regressão Harmônica, que segue a equação (2.29) onde as matrizes que a compõem são agora:

$$\theta = (\mu, A_1, \dots, A_k, B_1, \dots, B_k)'$$

$$W = \begin{bmatrix} 1 & \cos(\omega_1) & \text{sen}(\omega_1) & \dots & \cos(\omega_k) & \text{sen}(\omega_k) \\ 1 & \cos(2\omega_1) & \text{sen}(2\omega_1) & \dots & \cos(2\omega_k) & \text{sen}(2\omega_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \cos(N\omega_1) & \text{sen}(N\omega_1) & \dots & \cos(N\omega_k) & \text{sen}(N\omega_k) \end{bmatrix}$$

Quando  $\omega_1 = \frac{2\pi i}{N}, \omega_2 = \frac{2\pi j}{N}, \dots, \omega_k = \frac{2\pi m}{N}$ , isto é, são frequências de

Fourier, a solução exata é dada por:

$$\hat{\mu} = \bar{Z} \tag{2.36}$$

$$\begin{aligned}\hat{A}_k &= \frac{2}{N} \sum_{t=1}^N Z_t \cos(\omega_k t) \quad k=1,2,\dots \\ \hat{B}_k &= \frac{2}{N} \sum_{t=1}^N Z_t \text{sen}(\omega_k t) \quad k=1,2,\dots\end{aligned}\tag{2.37}$$

Para  $\omega = \pi$  :

$$\begin{aligned}\hat{A}_k &= \frac{1}{N} \sum_{t=1}^N Z_t (-1)^t \\ \hat{B}_k &= 0\end{aligned}\tag{2.38}$$

### 2.6.1. Regressão Harmônica com Tendência Polinomial

Quando a série a ser modelada possui alguma tendência, o modelo (2.35) pode ser estendido para (Al-Madfai *et al.*, 2001):

$$\begin{aligned}Z_t &= \mu + at + bt^2 + \dots + ct^p + A_1 \cos(\omega_1 t) + B_1 \text{sen}(\omega_1 t) + \dots + \\ &+ A_k \cos(\omega_k t) + B_k \text{sen}(\omega_k t) + \varepsilon_t\end{aligned}\tag{2.39}$$

o que significa que o modelo segue a equação (2.29) cuja matriz  $Z$  permanece a mesma considerada para a Regressão Harmônica com múltiplas periodicidades e a matriz  $W$  é da forma:

$$W = \begin{bmatrix} 1 & t & t^2 & \dots & t^p & \cos(\omega_1) & \text{sen}(\omega_1) & \dots & \cos(\omega_k) & \text{sen}(\omega_k) \\ 1 & 2t & 2t^2 & \dots & 2t^p & \cos(2\omega_1) & \text{sen}(2\omega_1) & \dots & \cos(2\omega_k) & \text{sen}(2\omega_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & Nt & Nt^2 & \dots & Nt^p & \cos(N\omega_1) & \text{sen}(N\omega_1) & \dots & \cos(N\omega_k) & \text{sen}(N\omega_k) \end{bmatrix}$$

onde  $p$  é o grau do polinômio que modela a tendência.

O vetor  $\theta$  é agora:

$$\theta = (\mu, a, b, \dots, c, A_1, \dots, A_k, B_1, \dots, B_k)'$$

A estimação via Mínimos Quadrados Ordinários é dada pela equação (2.30).

### 2.6.2. Periodograma

Quando os  $\omega$  são frequências desconhecidas é necessário estender o método anterior de forma a incluir a estimação destas frequências. Assim, os estimadores de  $\mu$ ,  $A$  e  $B$  dependem de  $\omega$  e são, portanto, denotados por  $\hat{\mu}(\omega)$ ,  $\tilde{A}(\omega)$ ,  $\tilde{B}(\omega)$ , respectivamente (Young *et al.*, 1999).

O melhor valor para  $\omega$  é o que maximiza a quantidade:

$$\tilde{R}^2(\omega) = \tilde{A}^2(\omega) + \tilde{B}^2(\omega) \quad (2.40)$$

o que é equivalente a maximizar a quantidade abaixo, denominada Periodograma:

$$\begin{aligned} I(\omega) &= \frac{N}{8\pi} \tilde{R}^2(\omega) \\ &= \frac{1}{2\pi N} \left[ \left( \sum_{t=1}^N (Z_t - \bar{Z}) \cos(\omega t) \right)^2 + \left( \sum_{t=1}^N (Z_t - \bar{Z}) \text{sen}(\omega t) \right)^2 \right] \end{aligned} \quad (2.41)$$

O Periodograma é uma técnica utilizada para determinar as componentes de frequência contidas na série (Taylor *et al.*, 2007).

Uma notação mais compacta para o Periodograma é:

$$I(\omega) = \frac{2}{N} \left| \sum_{t=1}^N Z_t e^{-i\omega t} \right|^2 \quad (2.42)$$

Apesar de as equações (2.41) e (2.42) serem definidas para qualquer frequência  $\omega \in [-\pi, \pi]$ , o cálculo das ordenadas do Periodograma faz-se apenas para um conjunto finito de abscissas (Morettin, 2006). Espera-se que o Periodograma apresente um pico quando calculado para uma frequência que coincida exatamente com uma das frequências “verdadeiras”.

### 2.6.3. Teste de Fisher

Pode-se recorrer a testes de hipóteses para analisar a significância estatística dos máximos do Periodograma. Os principais são o Teste de Fisher e sua extensão, o Teste de Whittle, que são apresentados a seguir.

Segundo Morettin (1983), o Teste de Fisher é utilizado para testar o valor do maior pico no Periodograma. Considera-se o modelo:

$$X_t = Y_t + Z_t \quad (2.43)$$

onde  $X_t$ ,  $t=1,2,\dots,N$  são observações discretas de um processo estacionário com média zero e espectro misto.  $Y_t$  e  $Z_t$  são processos descorrelacionados, onde  $Y_t$  é a parte contínua do espectro e  $Z_t$ , um ruído branco.

A hipótese nula é:  $H_0 : A_j = 0$  para todo  $j$ , sob a condição de que  $X_t$  seja Normal, onde  $A_j$  é o coeficiente de ordem  $j$  da Regressão Harmônica. A Estatística de Fisher é:

$$g = \frac{\max[I_j(\omega)]}{\sum_{j=1}^m I_j(\omega)} \quad (2.44)$$

em que  $m = \frac{N}{2}$ .

Fisher (1929) mostrou que, sob  $H_0$ , a distribuição exata de  $g$  (equação 2.44) é:

$$P(g > z) = m(1-z)^{m-1} - \binom{m}{2}(1-2z)^{m-1} + \dots + (-1)^{k-1}(1-kz)^{m-1} \quad (2.45)$$

onde  $k = \frac{1}{z}$ .

Podem-se encontrar tabelas para a distribuição acima dada por Fisher (1929) e Shimshoni (1971).

Se  $H_0$  é rejeitada, pode-se concluir que há uma periodicidade em  $X_t$  cuja frequência corresponde ao  $\max[I_j(\omega)]$ . Se isto ocorre para  $j = j'$ , então esta frequência é  $\hat{\omega}_0 = \frac{2\pi j'}{N}$ .

Whittle (1952) sugere a possibilidade de testar o pico de 2ª ordem pela omissão do termo  $I_j(\omega)$  em (2.44) e ajustando-se o valor de  $m$  para  $m-1$ .

Se é conhecido que  $X_t$  tem exatamente  $r$  componentes periódicas, quando  $H_0$  é falsa, então o teste baseado em:

$$g^{(r)} = \frac{I^{(r)}(\omega)}{\sum_{j=1}^m I_j(\omega)} \quad (2.46)$$

pode ser usado, onde  $I^{(r)}(\omega)$  é a  $r$ -ésima maior coordenada.

## 2.7. Ajuste polinomial de curvas

Um ajuste de curvas é um mecanismo ou artifício que fornece uma relação funcional quando há uma relação estatística (Anton, 2001).

Quando algum fenômeno ou situação é analisado através de dados numéricos, há a necessidade de saber qual a relação funcional correspondente  $y = f(x)$ .

Uma das formas de se trabalhar com uma função definida por uma tabela de valores é a interpolação polinomial. Porém esta técnica não é adequada quando se quer extrapolar ou quando há erros inerentes ao experimento em questão, visto que estes são imprevisíveis.

O ajuste por Mínimos Quadrados é uma boa solução para este problema. Considerando um ajuste polinomial, ou seja,  $y = a + bx + cx^2 + \dots$ , pode-se proceder ao ajuste de Mínimos Quadrados para descobrir a função que define a curva utilizando a equação (Anton, 2001):

$$S = (X'X)^{-1} X'y \quad (2.47)$$

onde:

$$S = \begin{bmatrix} a \\ b \\ c \\ \vdots \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots \\ 1 & x_2 & x_2^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \\ 1 & x_n & x_n^2 & \dots \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Encontrada a equação que determina a dinâmica dos dados, pode-se utilizar alguns critérios para verificar qualidade do ajuste. O coeficiente de Correlação de Pearson é calculado como segue:

$$r^2 = 1 - \frac{N \sum (y - y_{aj})^2}{N \sum y^2 - (\sum y)^2} \quad (2.48)$$

onde  $y$  são os dados observados e  $y_{aj}$  são os valores ajustados pelo polinômio.

Quanto mais próximo de 1 for o  $r^2$ , melhor é o ajuste da curva. Outra forma de validar o ajuste é o MAPE (equação 2.11): quanto menor for este valor, mais adequada é a curva ajustada.

### **3**

## **Modelo Proposto**

A previsão de demanda de energia tem sido objeto de diversos estudos baseados em séries temporais, a maioria dos quais utilizam modelos de função de transferência para representar os dados.

A utilização de variáveis meteorológicas como fatores explicativos para estes modelos, por exemplo, é decorrente do efeito do clima sobre o comportamento do consumidor, o que influencia a demanda de eletricidade. Há diversos outros fatores que influenciam o comportamento do consumidor, dentre os quais podem ser citados a ocorrência de feriados, mudanças de cunho político, econômico, etc.

Estes fatores contribuem para a não estacionariedade do processo bem como para a multiplicidade de periodicidades e a presença de distúrbios repetidos, embora não-sistemáticos. Tais processos não podem se tornar estacionários pela simples aplicação das técnicas convencionais de diferenciação.

As observações exóticas não afetam os inter-relacionamentos que governam a dinâmica dos dados, porém desfiguram as ferramentas de identificação dos modelos como, por exemplo, as Funções de Autocorrelação e Autocorrelação Parcial.

Entre as abordagens clássicas de pré-processamento dos dados que lidam com este problema, pode-se citar a diferenciação linear e sazonal que apresenta resultados efetivos quando os dados são bem-comportados. Tal técnica, porém, não consegue tratar os distúrbios aperiódicos e as observações exóticas.

Outra técnica clássica, a análise de intervenção necessita de uma quantidade suficiente de observações ‘limpas’ para produzir bons resultados. O problema

inerente a este método é que pode haver super parametrização do modelo quando há muitos pontos exóticos e janelas de tempo a modelar.

O HPA, *Hierarchical Profiling Approach*, é uma técnica estatística de modelagem inovadora introduzida por Al-Madfai, Ameen & Ryley (2001), que proporciona uma estrutura generalizada de modelagem para eventos específicos. Além disso, o HPA proporciona um melhor entendimento do processo observado, pois modela as componentes da série explicitamente e é capaz de lidar com dados de múltiplas periodicidades.

A modelagem HPA é capaz de tratar observações atípicas únicas, periódicas ou não, bem como janelas de tempo atípicas enquanto, ao mesmo tempo, funciona como uma poderosa técnica de pré-branqueamento.

Diversos estudos já mostraram a habilidade preditiva do modelo. Al-Madfai, Ameen & Ryley (2001) mostram uma aplicação do HPA a dados diários de demanda de eletricidade do País de Gales. Outra aplicação em séries de energia pode ser encontrada em Al-Madfai, Ameen & Ryley (2004), onde os autores modelam, novamente, dados diários de demanda de energia e tratam as férias escolares e o feriado de Natal como observações exóticas.

Al-Madfai *et al.* (2009) aplicaram o HPA a dados de consumo de energia de meia em meia hora onde são estudados, separadamente, a Páscoa, o Natal, as férias escolares e o efeito das estações do ano, obtendo excelentes resultados.

Outras aplicações de sucesso do HPA estão relacionadas à previsão do número de crimes e assaltos no País de Gales. Entre elas, pode-se citar Al-Madfai (2005) e Al-Madfai *et al.* (2007). Nestes trabalhos, as previsões obtidas pela aplicação do modelo HPA são estudadas e distribuídas espacialmente.

Todos estes resultados apontam a adequabilidade desta nova técnica de modelagem, principalmente a dados de alta frequência, e servem como motivação para investigar o comportamento do modelo na modelagem de outras séries temporais.

Um dos principais conceitos da modelagem HPA é a definição de perfil. Segundo Al-Madfai, Ameen & Ryley (2001), um perfil pode ser definido como uma mudança típica do comportamento normal de uma série temporal, durante ou em resposta a um evento conhecido. Os perfis são hierárquicos e podem modelar eventos dentro do ano (eventos atípicos e janelas de tempo), ciclos anuais (sazonalidade) e ciclos maiores.

A criação dos perfis identifica e modela os eventos distintos periódicos e aperiódicos em séries de tempo e associa mudanças de nível nos dados a eventos salientes na série.

A metodologia HPA se baseia na criação e identificação de perfis pré-existentes nos dados para gerar previsões ou simplesmente explicar a dinâmica que rege seu comportamento. O HPA combina, hierarquicamente, diferentes níveis de perfis em sua aplicação.

A metodologia decompõe a variabilidade da série temporal em componentes de caráter determinístico, estocástico e ruído para identificar os padrões e modelar os dados.

Os perfis identificados são tipicamente modelados utilizando funções contínuas de tempo cuidadosamente selecionadas e acumuladas aditivamente na componente determinística, porém outras abordagens alternativas podem ser utilizadas.

A componente estocástica, resultado do ajuste da componente determinística, pode ser modelada utilizando uma abordagem adequada, geralmente um modelo ARIMA. O ruído é suposto aleatório.

A forma geral do modelo HPA é:

$$y_t = f(t) + Z_t \quad (3.1)$$

onde  $y_t$  é a série observada,  $f(t)$  é a componente determinística que modela os perfis e  $Z_t$  é a componente estocástica a ser modelada.

A componente determinística visa, via procedimentos iterativos, perfilar os eventos atípicos em conjunto com as outras componentes, incluindo tendência, sazonalidade e ciclos que a série possa conter. Tal componente contém os perfis da série e pode ser identificada através de análises quantitativas e qualitativas ou através da opinião de especialistas.

Os perfis identificados são acumulados aditivamente de forma a capturar o máximo de informação determinística proveniente da série. Há diversas metodologias que podem modelar a componente determinística; a principal delas é a modelagem via Mínimos Quadrados Ordinários.

Após a modelagem da componente determinística, a componente estocástica é criada iterativamente a cada correção da série, ou seja,  $Z_t$  é o perfil corrigido remanescente a cada nível.

O HPA combina, assim, diferentes níveis de perfis.

- O primeiro nível modela e corrige as variações dentro do ano;
- O segundo nível trata os comportamentos atípicos, tais como feriados e mudanças provocadas pelas estações do ano;
- O nível seguinte trata o comportamento geral da série durante o ano, incluindo tendência, sazonalidade e ciclos, etc.

O último perfil pode ser modelado através da aplicação de Regressão Harmônica com tendência polinomial conforme a equação (2.39). Dado o ajuste do perfil final, os perfis podem ser combinados.

O perfil geral é, então, composto pela agregação de todos os subperfis mencionados anteriormente:

$$f(t) = \sum f_i(t) \quad (3.2)$$

onde  $f_i(t)$  corresponde a cada perfil ajustado nas etapas anteriores.

Os desvios entre a série original e o perfil geral são dados por:

$$y_t - f(t) = Z_t \quad (3.3)$$

onde  $Z_t$  é a componente estocástica resultante da modelagem dos perfis.

O último passo do HPA tem por objetivo a construção de um modelo de Box & Jenkins (1976) que consiste em ajustar modelos auto-regressivos integrados de médias móveis, ARIMA (p,d,q). Além do modelo ARIMA, outras abordagens como Modelo de Espaço de Estados, Redes Neurais Artificiais, Método Holt, Método Holt-Winters, Amortecimento Exponencial e Passeio Aleatório podem ser usados para modelar a componente estocástica.

Cabe salientar que na modelagem HPA tanto a quantidade de perfis, quanto a metodologia que trata cada perfil variam de acordo com a análise do profissional.

Somente uma análise profunda e cuidadosa dos dados pode definir a forma de tratar a componente determinística do modelo. Após a modelagem desta componente é possível proceder à estimação de um modelo de previsão para a componente estocástica e, finalmente, para a série como um todo.

Dadas as previsões para a componente estocástica, as previsões para a série original são obtidas pela agregação destas com as previsões da componente determinística construída a partir dos perfis. A equação de previsão  $k$  passos a frente é:

$$y_t(k) = f(t+k) + Z_t(k) \quad (3.4)$$

A Figura 3.1 apresenta o fluxograma que resume a modelagem HPA.

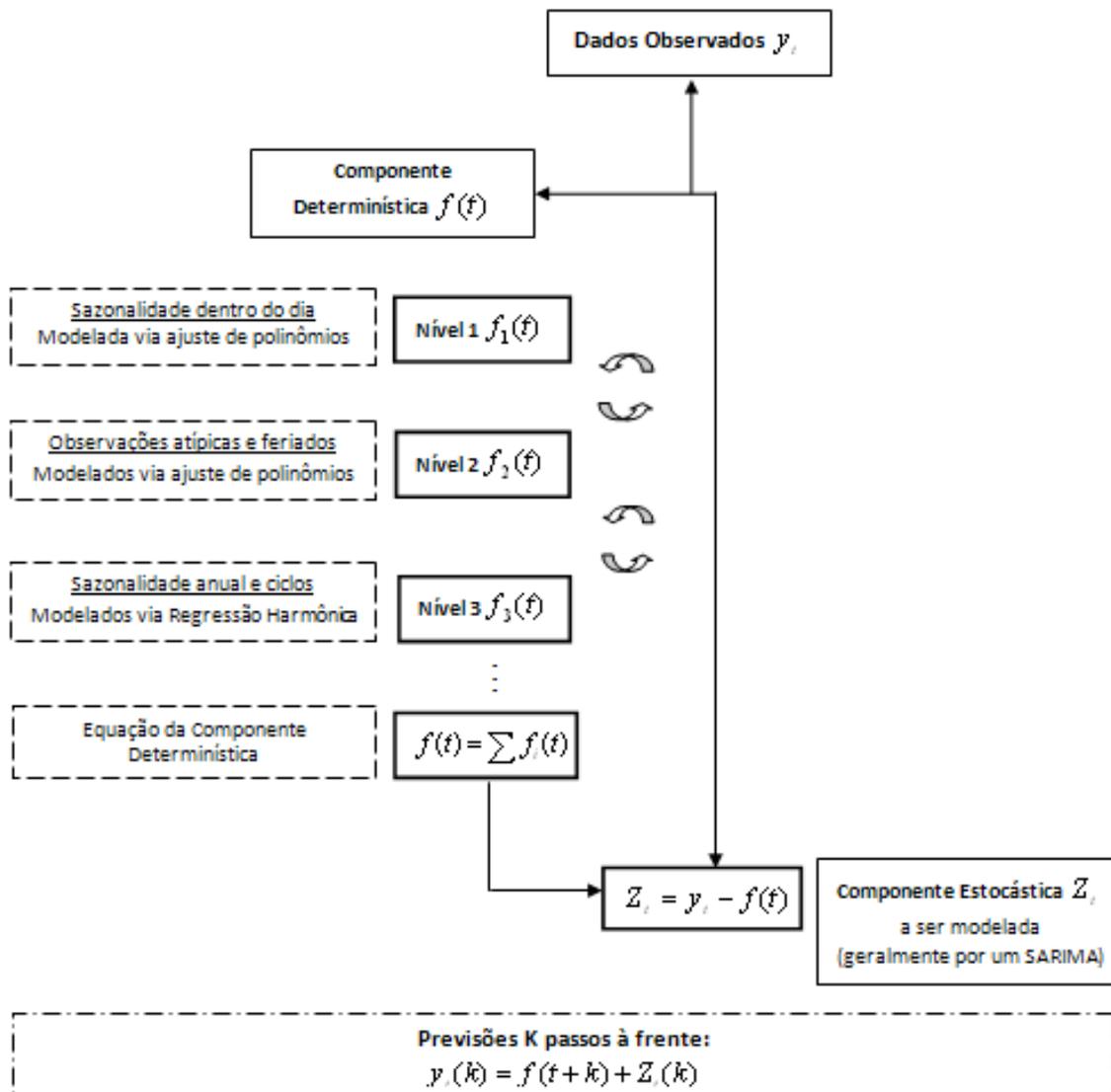


Figura 3.1 Fluxograma da modelagem HPA

A seguir, todas as etapas da modelagem serão mais detalhadas e o processo de modelagem será descrito conforme a aplicação em um conjunto de dados reais.

## 4 Aplicação do Modelo

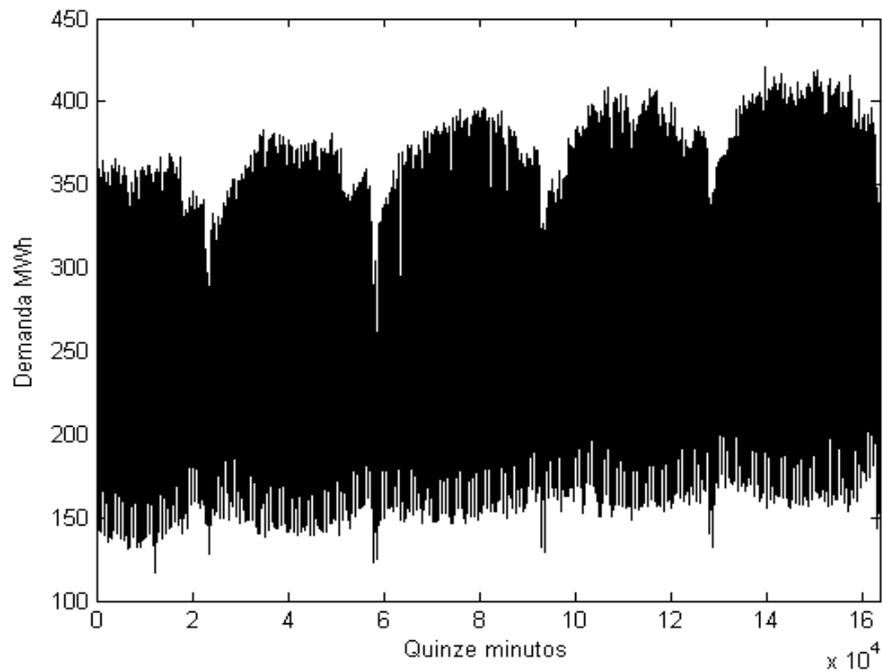
É possível encontrar na literatura diversas aplicações que utilizam técnicas estatísticas e de computação intensiva para realizar previsões de curto prazo na área de energia elétrica. Entre elas podem-se citar: Esteves (2003), que modela dados com múltiplos ciclos através do Método de Amortecimento Direto e Holt-Winters com múltiplos ciclos e Lourenço (1998) que faz uso de Redes Neurais e Lógica Fuzzy para modelar dados horários. Miranda (2007) apresenta uma extensão do método Holt-Winters com múltiplos ciclos que considera a temperatura como variável exógena, enquanto Rizzo (2001) aplica Amortecimento Direto e o Modelo de Gupta a dados horários. Sobral (1999) combina Redes Neurais, Lógica Fuzzy, modelos estatísticos e dados de temperatura para modelar dados de alta frequência.

Uma gama variada de outras técnicas e modelos pode ser empregada no estudo de previsão a curto prazo. O HPA é um desses métodos inovadores que tem obtidos bons resultados (Ivaha *et al.*, 2007).

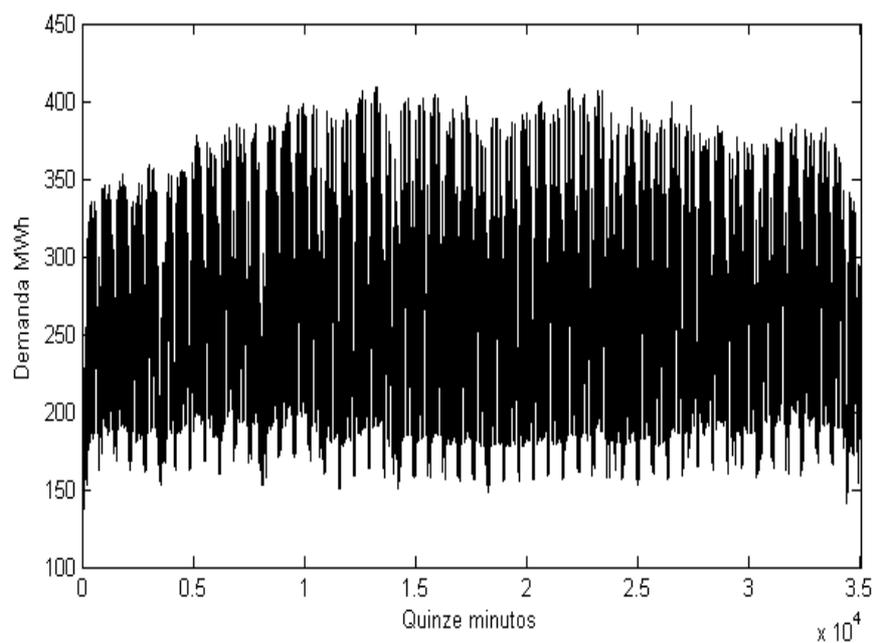
Neste capítulo são apresentados os resultados obtidos pela aplicação da metodologia *Hierarchical Profiling Approach* a dados de demanda brasileiros. O objetivo do modelo é realizar previsões a cada 15 minutos para sete dias à frente, ou seja, prever 672 passos.

Para tanto, o modelo foi aplicado à base de dados de uma grande concessionária de energia elétrica do Brasil, localizada na região Sudeste. Os dados referem-se à demanda de energia elétrica em MWh no período compreendido entre maio de 2002 e dezembro de 2006. As informações foram verificadas a cada 15 minutos, fornecendo 96 observações diárias e 163.676 no total.

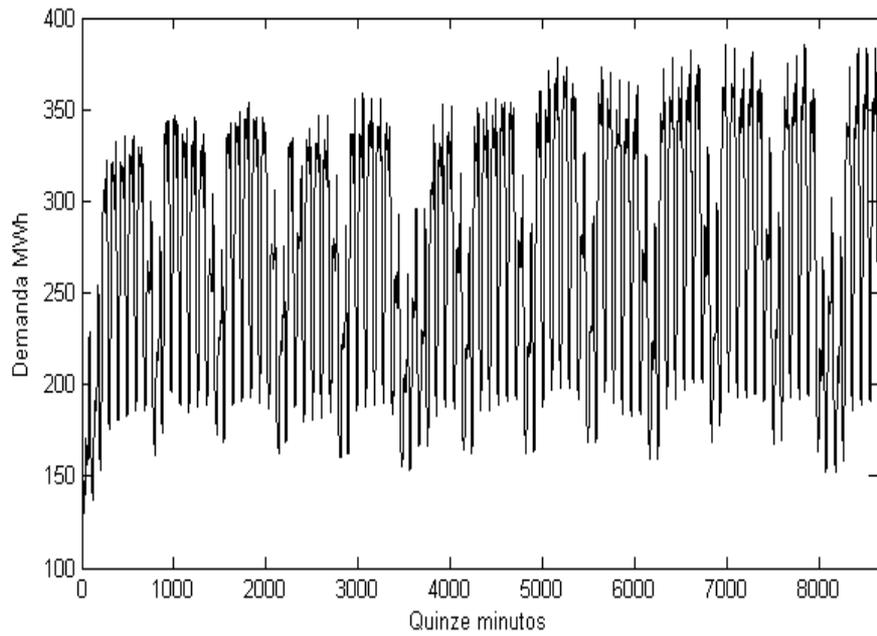
As características da série de carga estão ilustradas nas Figuras 4.1 a 4.6, nas quais se podem observar as sazonalidades existentes.



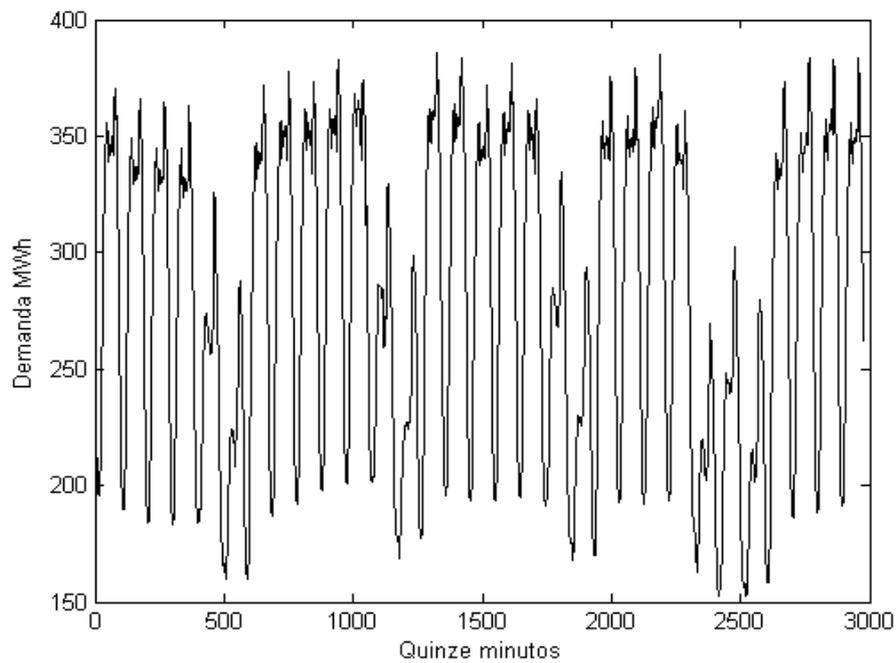
**Figura 4.1** Série de demanda de energia em todo o histórico



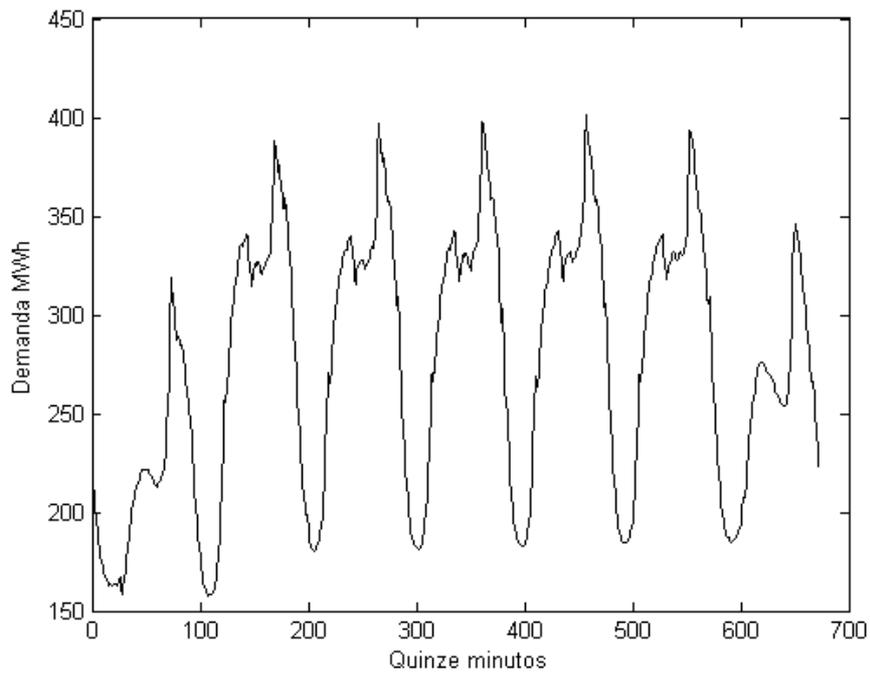
**Figura 4.2** Demanda de energia a cada quinze minutos de um ano



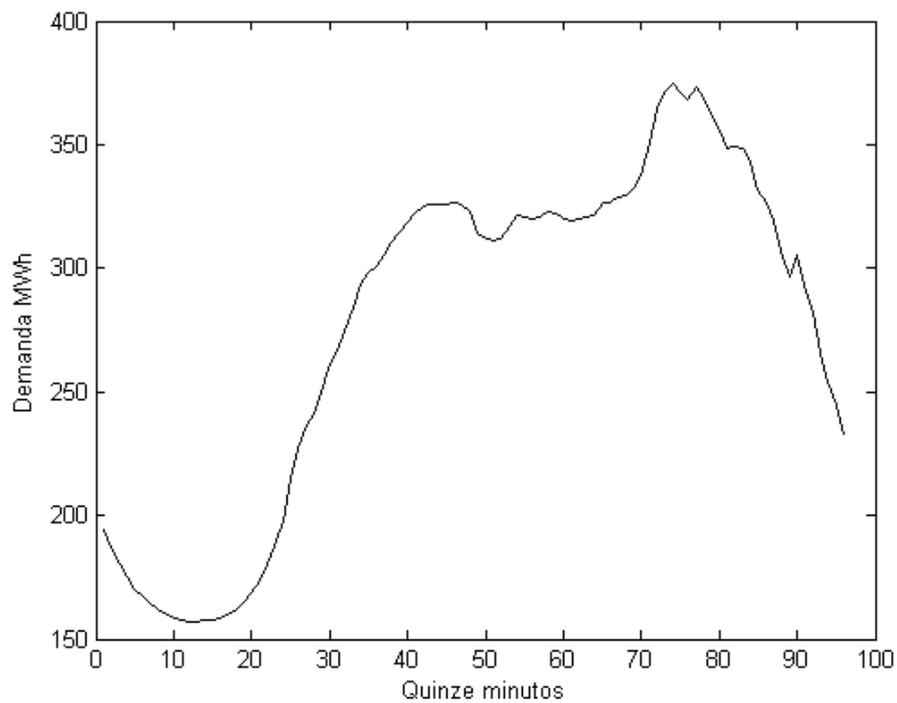
**Figura 4.3** Demanda de energia a cada quinze minutos de um trimestre



**Figura 4.4** Demanda de energia a cada quinze minutos de um mês



**Figura 4.5** Demanda de energia a cada quinze minutos de uma semana



**Figura 4.6** Demanda de energia a cada quinze minutos de um dia

Na Figura 4.1 nota-se que a periodicidade anual é bastante clara, visto que é possível distinguir facilmente cada ano observado. Nota-se, ainda, que há uma tendência crescente para os quatro anos de dados. Quando o período analisado é restrito a um ano de dados, tal tendência não é percebida.

Quando a análise é feita no âmbito mensal, Figura 4.4, pode-se perceber claramente cada divisão equivalente a uma semana, o que é um indício da presença de sazonalidade semanal. Esta análise pode ser estendida para a Figura 4.5, que mostra uma semana de dados: os sete dias aparecem bastante destacados e pode-se notar que a demanda para o sábado e o domingo apresenta comportamento diferente dos demais dias da semana.

Analisando a Figura 4.6, pode-se notar que a demanda é bastante baixa nas horas iniciais do dia e aumenta significativamente no fim do mesmo. Este comportamento permanece praticamente inalterado para os dias úteis sofrendo alterações nos fins de semana e feriados.

Maiores informações sobre os dados serão apresentadas a cada fase da especificação do HPA. A cada perfil considerado será possível obter informações mais detalhadas sobre a dinâmica subjacente aos dados, além de modelar cada comportamento particular que influencia a previsão.

#### **4.1.**

##### **Perfil de Nível 1 – Padrões para os dias da semana**

O primeiro nível do modelo HPA trata da periodicidade mais geral observada nos dados, ou seja, o perfil de nível 1 modela a sazonalidade dentro do ano. Para isto, analisa-se, em primeiro plano, o comportamento da demanda segundo o dia da semana analisado.

Dado que podem existir até setes padrões diferentes para este perfil, foram identificados os dias da semana para o conjunto de dados em estudo. As

observações correspondentes a cada dia da semana foram agrupadas e calculou-se uma média para cada dia da semana. Assim, foram obtidos sete padrões de demanda que correspondem ao comportamento médio durante cada dia da semana para os dados observados. Cada dia da semana padrão é uma série composta por 96 observações, ou seja, uma série específica de um dia da semana onde cada observação corresponde à média de todos os dados observados para um determinado horário correspondente.

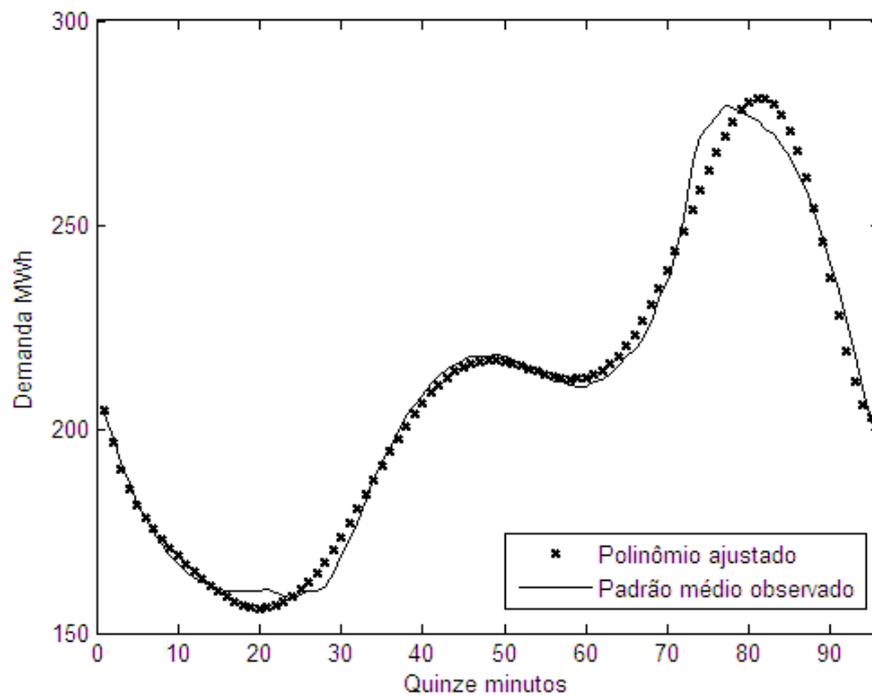
Uma análise mais cuidadosa revelou que não há diferenças significativas entre os padrões obtidos para as terças, quartas e quintas-feiras. Assim, optou-se por criar um único padrão para estes dias da semana. Para o nível 1, então, foram identificados cinco padrões de comportamento diários que são responsáveis por modelar a sazonalidade dentro do dia.

Segundo Al-Madfai (2005), há diversas abordagens possíveis para se modelarem os padrões obtidos. Optou-se, neste trabalho, por realizar um ajuste polinomial de curvas para estimar a equação que representa o comportamento de cada dia da semana. As curvas foram estimadas via Mínimos Quadrados Ordinários. Foram testados polinômios com graus variando de 1 a 20. A melhor curva para cada perfil foi escolhida com base em dois critérios: maximização do Coeficiente de Correlação de Pearson (equação 2.48) e minimização do Erro Médio Percentual Absoluto, MAPE (equação 2.11).

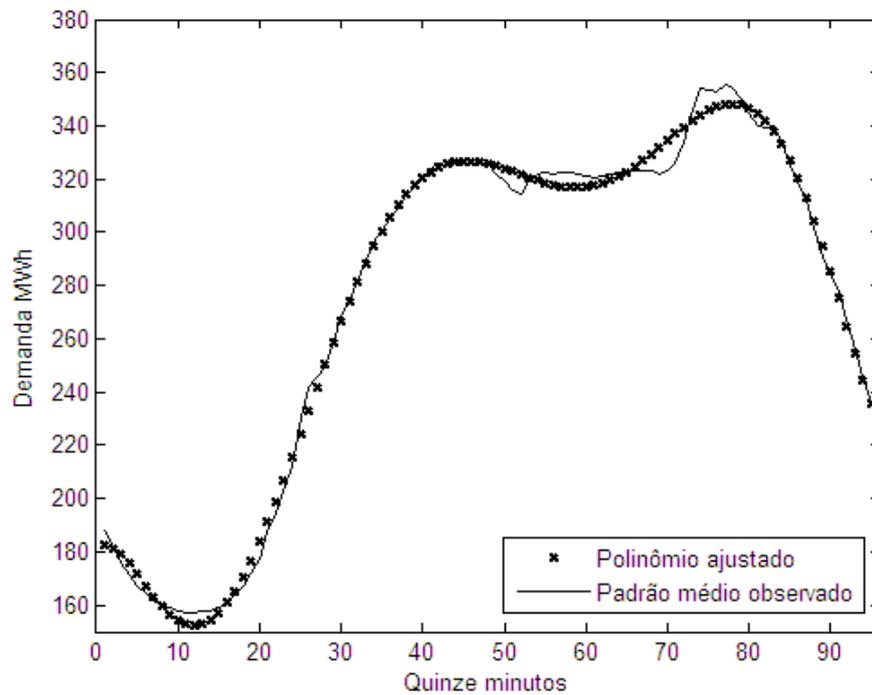
As Figuras 4.7 a 4.12 apresentam a comparação entre o perfil padrão observado e o perfil estimado para cada dia da semana, seguidos pelos respectivos polinômios estimados.

A equação do polinômio estimado para o padrão do domingo é:

$$\begin{aligned} \text{dompad} = & 215,13 - 11,95t + 1,56t^2 - 1,3t^3 + 6,02 \cdot 10^{-3}t^4 - 1,51 \cdot 10^{-4}t^5 + \\ & + 2,05 \cdot 10^{-6}t^6 - 1,42 \cdot 10^{-8}t^7 + 3,93 \cdot 10^{-11}t^8 \end{aligned} \quad (4.1)$$



**Figura 4.7** Padrão médio observado e polinômio estimado para o perfil de domingo



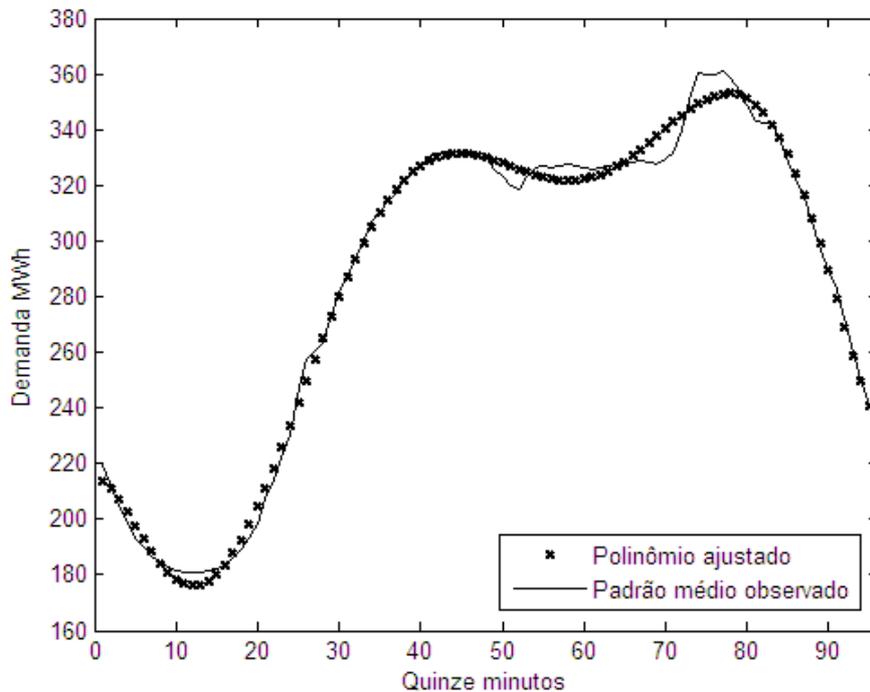
**Figura 4.8** Padrão observado e polinômio estimado para o perfil de segunda-feira

Para o padrão das segundas-feiras a equação do polinômio estimado é:

$$\begin{aligned} \text{segpad} = & 181,06 + 2,68t - 1,46t^2 + 0,13t^3 - 4,27 \cdot 10^{-3}t^4 + 6,72 \cdot 10^{-5}t^5 + \\ & + 5,08 \cdot 10^{-7}t^6 + 1,48 \cdot 10^{-9}t^7 \end{aligned} \quad (4.2)$$

O polinômio estimado para o perfil que combina as terças, quartas e quintas-feiras é mostrado a seguir. A comparação gráfica pode ser observada na Figura 4.9.

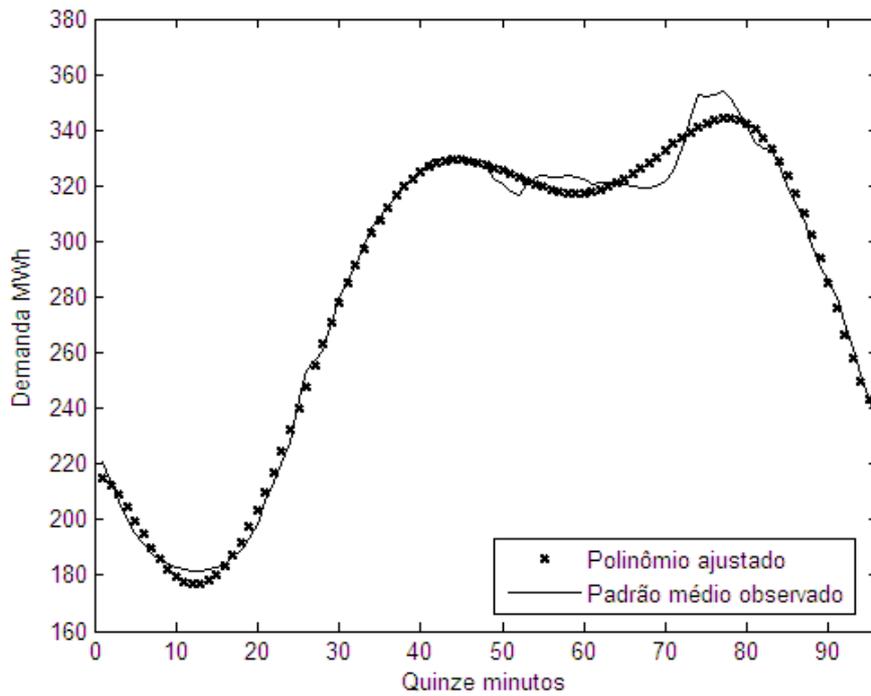
$$\begin{aligned} \text{tqqpad} = & 214,96 + 8,35 \cdot 10^{-2}t - 1,18t^2 + 0,11t^3 - 3,87 \cdot 10^{-3}t^4 + \\ & + 6,19 \cdot 10^{-5}t^5 - 4,73 \cdot 10^{-7}t^6 + 1,39 \cdot 10^{-9}t^7 \end{aligned} \quad (4.3)$$



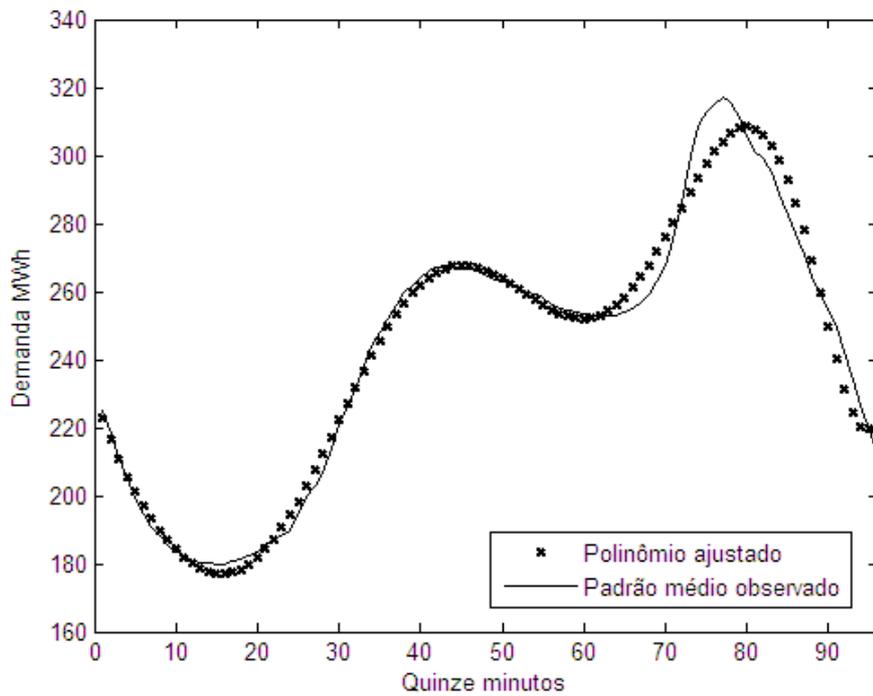
**Figura 4.9** Padrão observado e polinômio estimado para o perfil médio de terça, quarta e quinta-feira

Para o padrão das sextas-feiras a equação do polinômio estimado é:

$$\begin{aligned} \text{sexpad} = & 215,37 + 0,73t - 1,27t^2 + 0,12t^3 - 3,96 \cdot 10^{-3}t^4 + 6,29 \cdot 10^{-5}t^5 + \\ & - 4,79 \cdot 10^{-7}t^6 + 1,40 \cdot 10^{-9}t^7 \end{aligned} \quad (4.4)$$



**Figura 4.10** Padrão observado e polinômio estimado para o perfil de sexta-feira



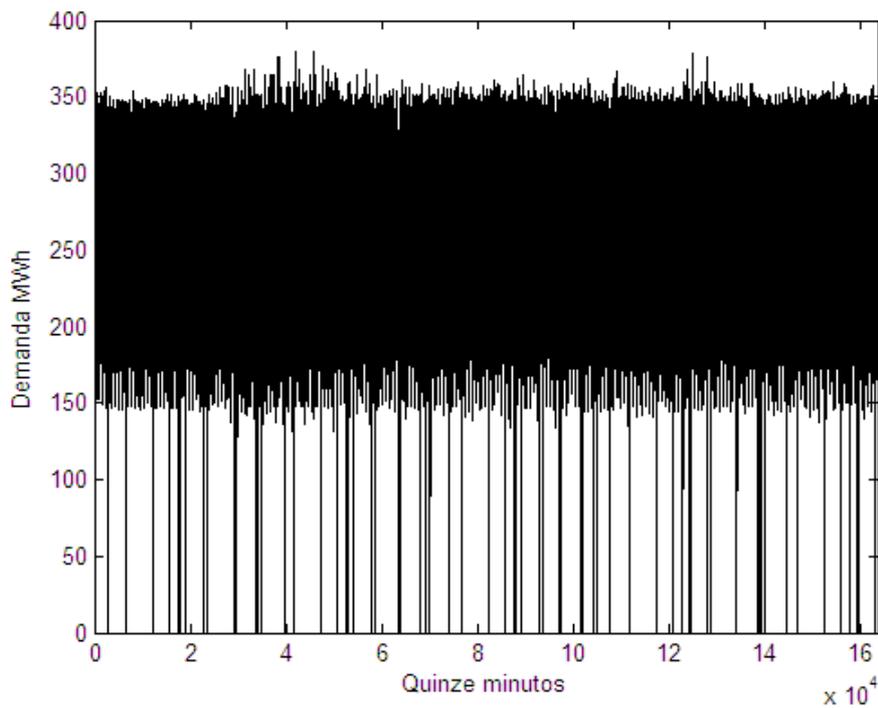
**Figura 4.11** Padrão observado e polinômio estimado para o perfil de sábado

O polinômio estimado para o perfil dos sábados é mostrado a seguir.

$$\begin{aligned} \text{sabpad} = & 230,6 - 8,18t + 0,67t^2 - 0,06t^3 + 3,31 \cdot 10^{-5}t^4 - 9,95 \cdot 10^{-5}t^5 + \\ & + 1,52 \cdot 10^{-6}t^6 - 1,14 \cdot 10^{-8}t^7 + 3,35 \cdot 10^{-11}t^8 \end{aligned} \quad (4.5)$$

Definidos os cinco padrões referentes ao perfil de nível 1, o próximo passo foi criar uma série temporal do mesmo tamanho que a série original, cujos valores são os padrões encontrados para os dias da semana.

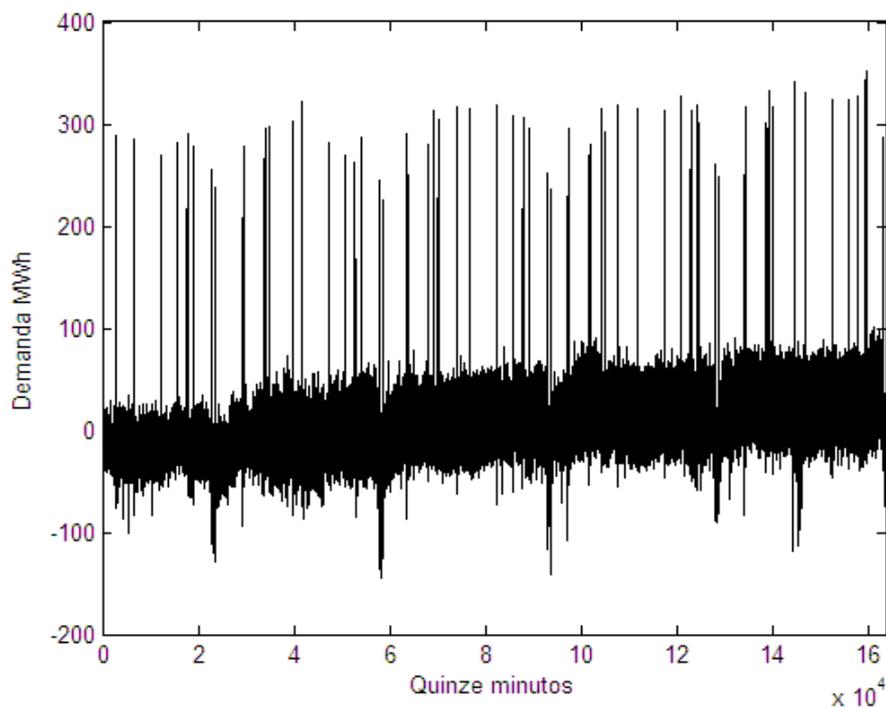
Os dias de domingo da nova série, por exemplo, são compostos pelos valores previstos após a aplicação do polinômio estimado para o padrão de domingo; o mesmo foi feito para os quatro padrões restantes encontrados. Cabe salientar que este procedimento foi realizado apenas para os dias em que não há ocorrência de feriados. Os dias de feriado receberam valor zero. Esta série, que pode ser vista na Figura 4.12, representa o perfil obtido no nível 1.



**Figura 4.12** Série obtida no nível 1

Para prosseguir a análise para os próximos níveis é necessário criar uma série ajustada sem o efeito da sazonalidade dentro do dia, modelada no nível 1. Para isso, utiliza-se o resíduo produzido pela diferença entre a série observada e a série obtida no nível 1.

A série remanescente após o ajuste de nível 1 é exposta na Figura 4.13. Nota-se que, desconsiderando os picos que correspondem aos feriados observados nos quatro anos de dados, a série apresenta-se dispersa em torno de zero. Este comportamento é esperado visto que, após a modelagem do último nível, a série remanescente deve ter o comportamento de um ruído aleatório, sem o efeito da sazonalidade, tendência ou outra função determinística.



**Figura 4.13** Série residual proveniente do ajuste de nível 1

## **4.2. Perfil de Nível 2 – Tratamento de feriados**

Retirado o efeito da sazonalidade dentro do dia da semana para os dias úteis, a próxima etapa consiste em realizar o mesmo procedimento para os dias de feriado.

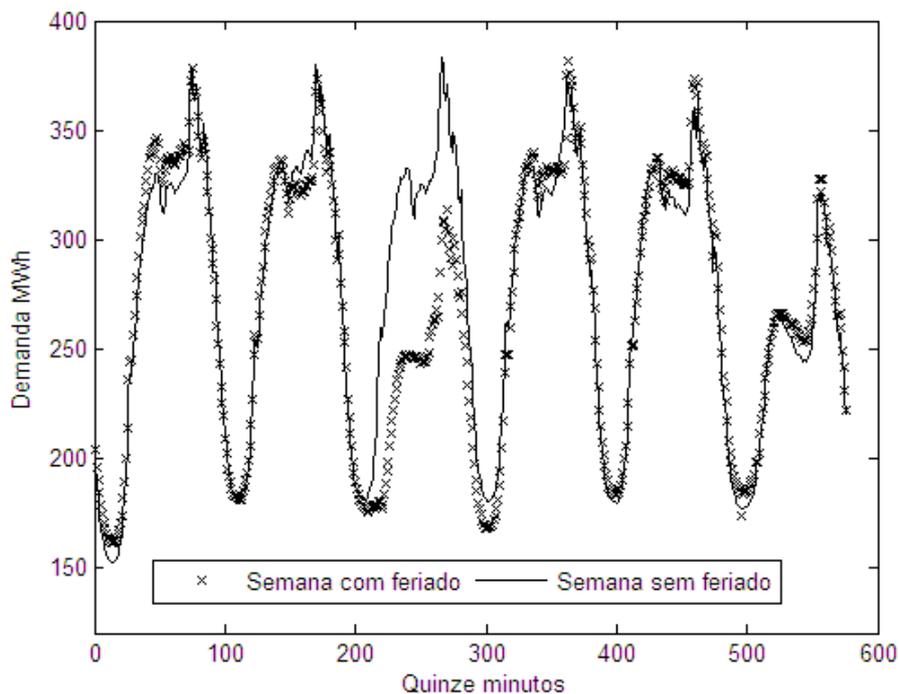
O nível 2 do HPA consiste na modelagem de observações atípicas, como os feriados, e outros eventos discrepantes que possam prejudicar as previsões. Cabe lembrar que a série a ser tratada neste nível não é mais a série original e sim, a série de resíduos gerada após a modelagem dos perfis de nível 1.

Com a finalidade de retirar o efeito dos dias atípicos, a primeira etapa consiste em identificar o dia da semana de ocorrência do feriado e agrupar os dados segundo este critério. Feito isso, o procedimento de modelagem é idêntico ao adotado no tratamento dos dias úteis do nível 1 do modelo.

Assim como em outros países, o Brasil possui uma gama de feriados que geram um perfil de demanda de carga elétrica diferente do comportamento dos dias úteis. Além de causar impacto na previsão, os feriados também distorcem a identificação e estimação do modelo (Vaage, 2000).

No meio acadêmico é possível ignorar os feriados ao verificar o poder preditivo de um modelo. Tal possibilidade não é verificada no mercado, visto que as concessionárias de energia elétrica necessitam de previsões acuradas para estes dias a fim de evitar sobrecargas no sistema ou mesmo subutilização da carga gerada.

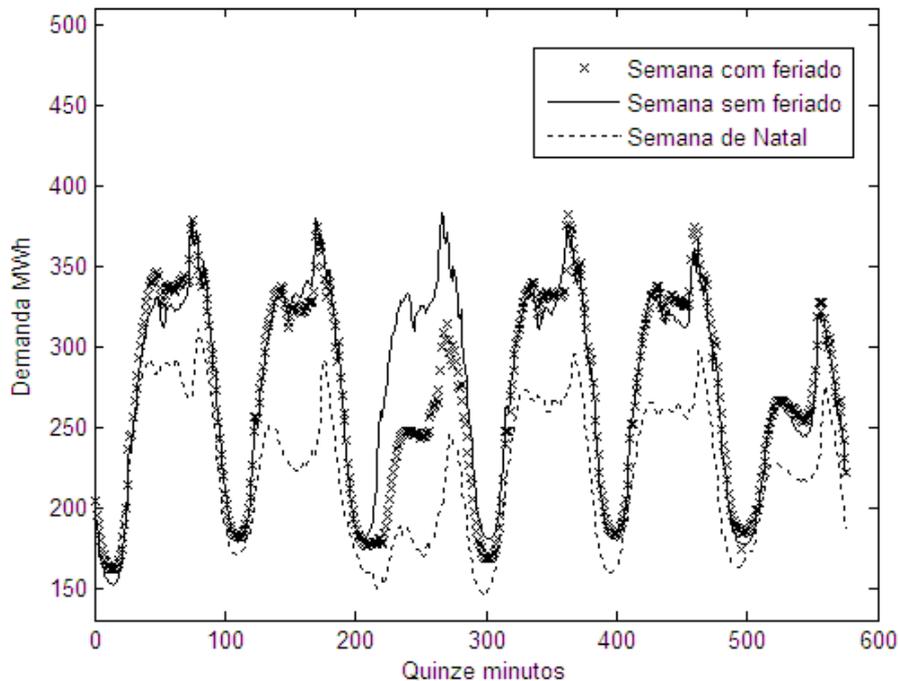
A diferença entre uma semana comum e uma semana onde há um feriado pode ser observada claramente na Figura 4.14. A ocorrência do feriado não parece influenciar significativamente a demanda dos dias próximos a este evento.



**Figura 4.14** Demanda de carga elétrica em uma semana que ocorreu um feriado e em uma semana onde não houve feriado

Miranda (2007) observa o problema da influência de feriados na demanda de energia elétrica e utiliza fatores de correção baseados no comportamento das semanas anterior e posterior ao feriado para minimizar este efeito. Tendo em vista que o feriado gera comportamentos distintos em dias da semana diferentes, este problema foi solucionado pela autora com a criação de regras para cada dia da semana. Tais regras foram ajustadas no mesmo período para o qual a previsão foi realizada, a fim de proporcionar uma melhor correção.

Os fatores de correção não serão aplicados nos feriados comuns deste trabalho, porém os feriados de Natal e Ano Novo receberão tratamento especial. Observou-se, para este conjunto de dados, que os feriados de Natal e Ano Novo comportam-se de maneira peculiar se comparados aos dias úteis e demais feriados. A demanda de carga elétrica para estes dias é bem inferior à dos demais. Este comportamento pode ser observado na Figura 4.15.



**Figura 4.15** Demanda em uma semana de feriado comum, uma semana de ocorrência de Natal e em uma semana onde não houve feriado

Tendo em vista estas discrepâncias, optou-se pelo desenvolvimento de um banco de regras que foi incorporado ao modelo de previsão de forma exógena. Este banco de regras contém fatores de redução/aumento que serão incorporados à previsão da energia, sendo assim capaz de ajustar melhor o modelo, reduzindo o erro de previsão (Bunn & Farmer, 1985).

No caso estudado nessa dissertação, havia disponibilidade de quatro anos de dados, o que não é suficiente para análise de feriados, porém, é possível realizar algumas análises capazes de reduzir os erros do modelo de previsão gerados para esses dias.

O primeiro passo desta análise consiste na identificação dos dias de ocorrência de feriados. Com isso, pode-se calcular a diferença percentual entre o dia do feriado e o mesmo dia da semana, observado nas semanas anterior e posterior (visto que, caso não ocorra nenhum fato extraordinário, o montante de carga do dia observado será semelhante ao das semanas antecedente e subsequente).

O objetivo deste procedimento é analisar a variação de carga elétrica de um dia normal e de um feriado. Se na semana anterior/posterior a um feriado que está sendo estudado, houve ocorrência de outro feriado, este dia será descartado da análise. Neste caso, o procedimento é retroceder/avançar uma semana a mais para realizar a análise.

Após o cálculo destes fatores, os feriados de Natal e Ano Novo podem ser corrigidos. A aplicação das regras faz com que estes feriados “especiais” apresentem, agora, comportamento semelhante aos demais feriados. Assim, as análises posteriores referentes a feriados poderão incluir mais estas duas datas. Os feriados tratados neste trabalho são mostrados nas Tabelas 4.1 e 4.2.

**Tabela 4.1** Lista de feriados fixos

| <b>Feriados Fixos</b>      |
|----------------------------|
| Confraternização Universal |
| Tiradentes                 |
| Dia do Trabalho            |
| Independência do Brasil    |
| Nossa Senhora Aparecida    |
| Finados                    |
| Proclamação da República   |
| Consciência Negra          |
| Natal                      |

**Tabela 4.2** Lista de feriados móveis

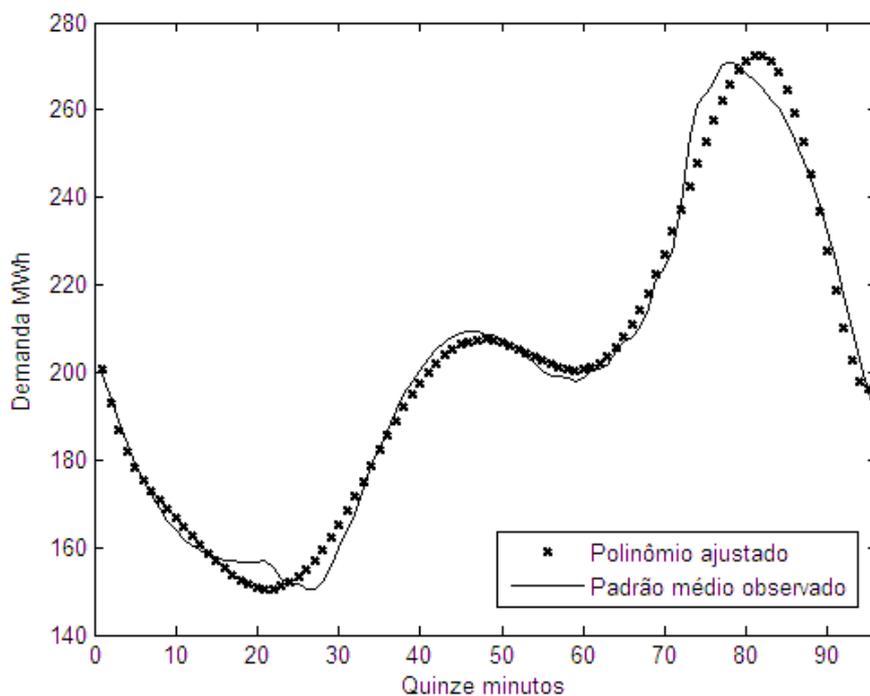
| <b>Feriados Móveis</b> |
|------------------------|
| Carnaval               |
| Sexta-feira da Paixão  |
| Páscoa                 |
| Corpus Christi         |

Cabe ressaltar que o feriado da Consciência Negra só é considerado a partir do ano 2006, quando foi instituído.

Cada dia da semana com feriado teve seu comportamento médio representado por uma série composta por 96 observações. Cada instante é obtido pela média do período correspondente aos feriados ocorridos no mesmo dia da semana considerado.

Foram, então, estimados sete polinômios correspondentes a cada dia da semana de ocorrência de feriado, que são responsáveis por definir o comportamento padrão da demanda de carga elétrica nos dias de feriado.

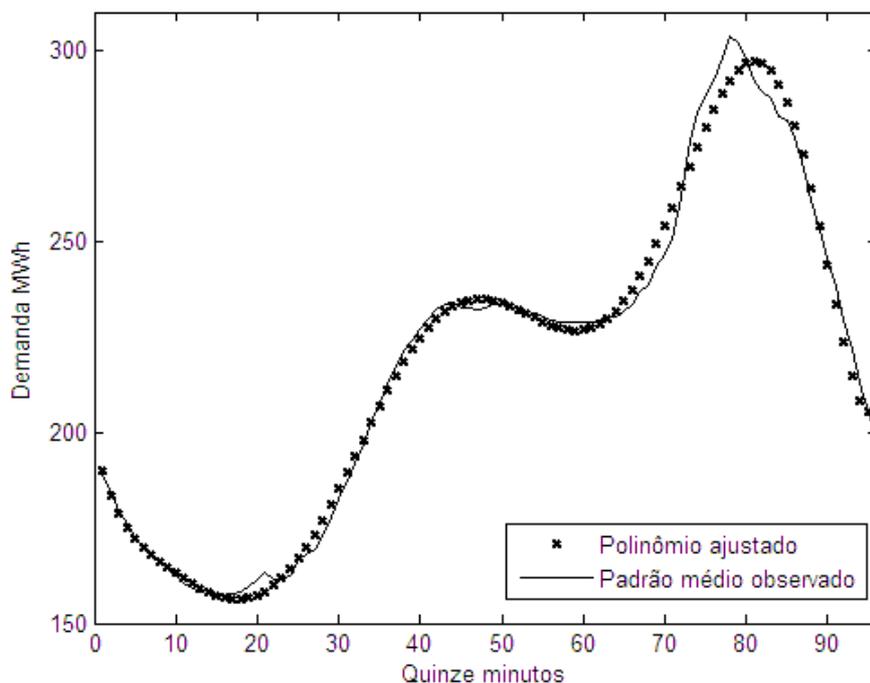
Os ajustes e equações para os sete padrões são apresentados a seguir. Os critérios de seleção dos polinômios foram os mesmos utilizados para os dias úteis.



**Figura 4.16** Padrão observado e polinômio estimado para o perfil de domingo de feriado

A equação que define o polinômio ajustado para o padrão médio observado para os feriados que ocorrem aos domingos é:

$$\begin{aligned} \text{dompadfer} = & 211,67 - 12,3t + 1,72t^2 - 0,15t^3 + 6,81 \cdot 10^{-3}t^4 - 1,70 \cdot 10^{-4}t^5 + \\ & + 2,30 \cdot 10^{-6}t^6 - 1,58 \cdot 10^{-8}t^7 + 4,36 \cdot 10^{-11}t^8 \end{aligned} \quad (4.6)$$



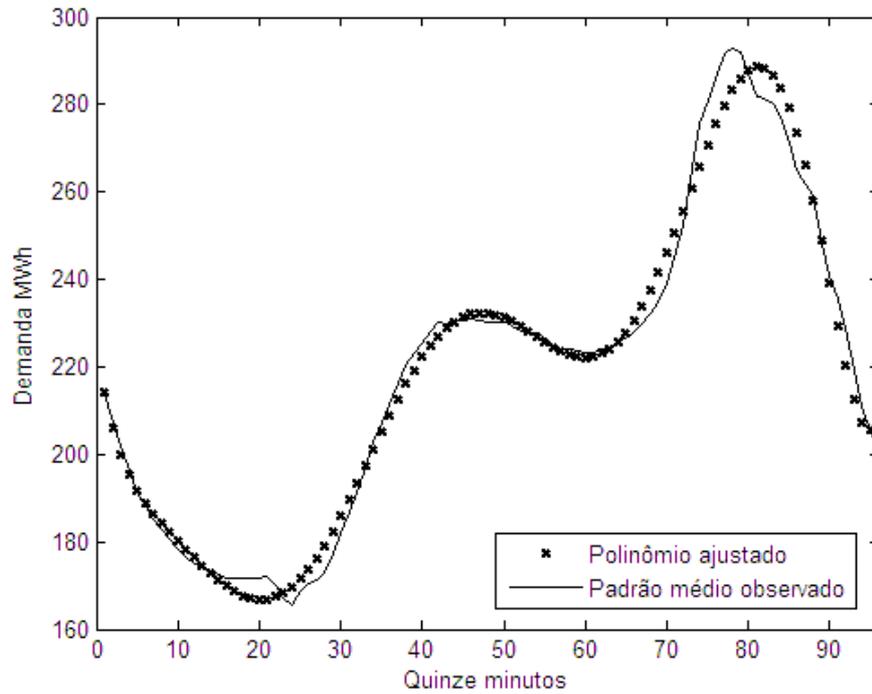
**Figura 4.17** Padrão observado e polinômio estimado para o perfil dos feriados às segundas-feiras

A equação que define o comportamento médio da demanda em um feriado que ocorre numa segunda-feira é:

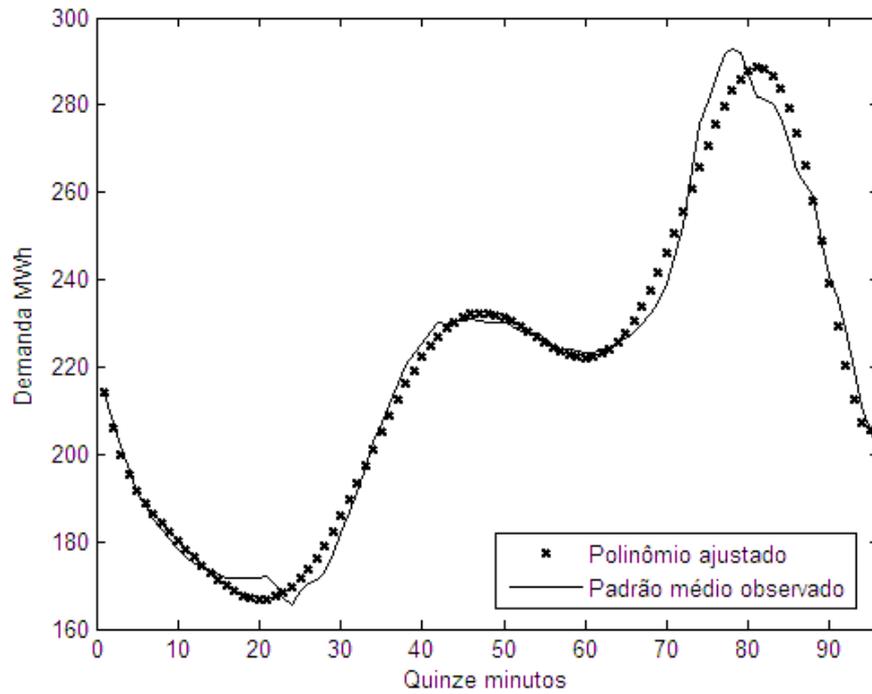
$$\begin{aligned} \text{segpadfer} = & 198,12 - 9,54t + 1,34t^2 - 0,12t^3 + 5,82 \cdot 10^{-3}t^4 - 0,02 \cdot 10^{-3}t^5 + \\ & + 2,11 \cdot 10^{-6}t^6 - 14,88 \cdot 10^{-9}t^7 + 41,64 \cdot 10^{-12}t^8 \end{aligned} \quad (4.7)$$

De maneira análoga, o perfil padrão para um feriado que ocorre numa terça-feira é definido pela equação:

$$\begin{aligned} \text{terpadfer} = & 216,38 - 13,27t + 1,88t^2 - 0,15t^3 + 7,04 \cdot 10^{-3}t^4 - 173,91 \cdot 10^{-6}t^5 + \\ & + 2,34 \cdot 10^{-6}t^6 - 16,09 \cdot 10^{-9}t^7 + 44,27 \cdot 10^{-12}t^8 \end{aligned} \quad (4.8)$$



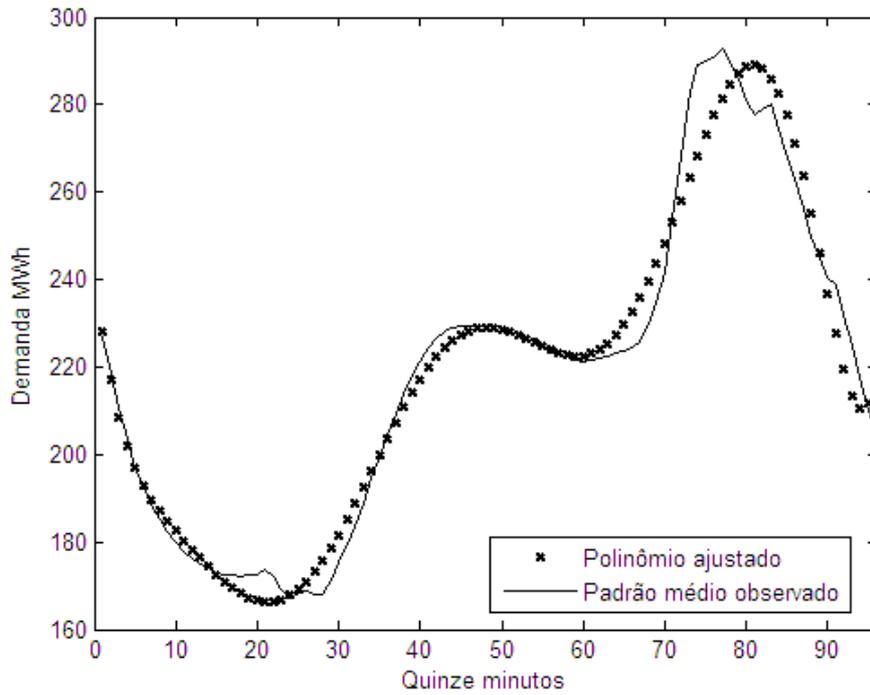
**Figura 4.18** Padrão observado e polinômio estimado para o perfil dos feriados às terças-feiras



**Figura 4.19** Padrão observado e polinômio estimado para o perfil dos feriados às quartas-feiras

O polinômio que define o padrão para um feriado que ocorre numa quarta-feira é:

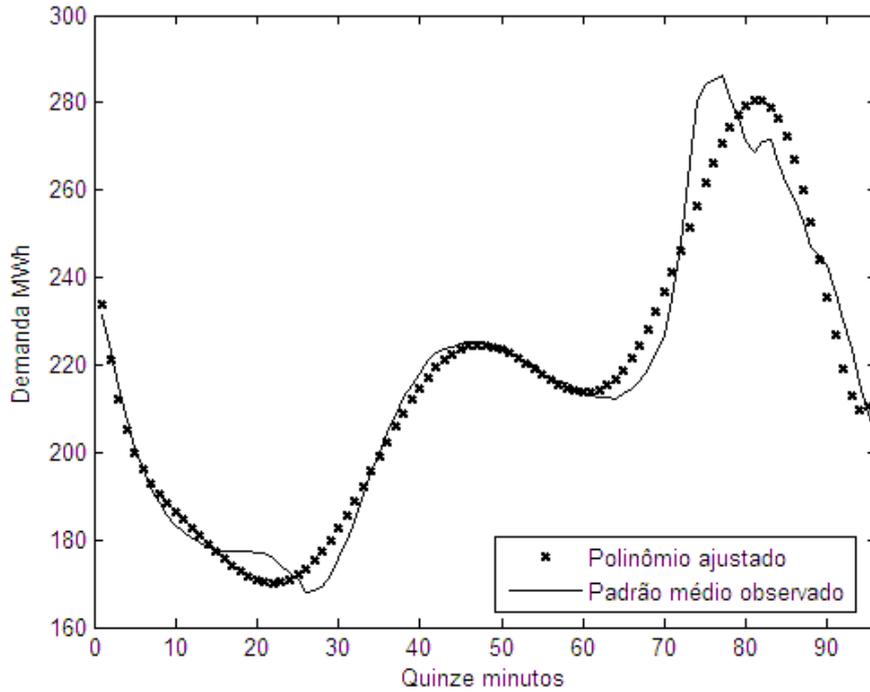
$$\begin{aligned} \text{quapadfer} = & 225,16 - 12,42t + 1,74t^2 - 0,15t^3 + 6,97 \cdot 10^{-3}t^4 - 175,01 \cdot 10^{-6}t^5 + \\ & + 2,37 \cdot 10^{-6}t^6 - 16,44 \cdot 10^{-9}t^7 + 45,41 \cdot 10^{-12}t^8 \end{aligned} \quad (4.9)$$



**Figura 4.20** Padrão observado e polinômio estimado para o perfil dos feriados às quintas-feiras

O padrão que representa o comportamento médio que ocorre às quintas-feiras pode ser definido por:

$$\begin{aligned} \text{quipadfer} = & 243,28 - 17,11t + 2,31t^2 - 0,18t^3 + 6,97 \cdot 10^{-3}t^4 - 199,25 \cdot 10^{-6}t^5 + \\ & + 2,65 \cdot 10^{-6}t^6 - 18,15 \cdot 10^{-9}t^7 + 49,80 \cdot 10^{-12}t^8 \end{aligned} \quad (4.10)$$



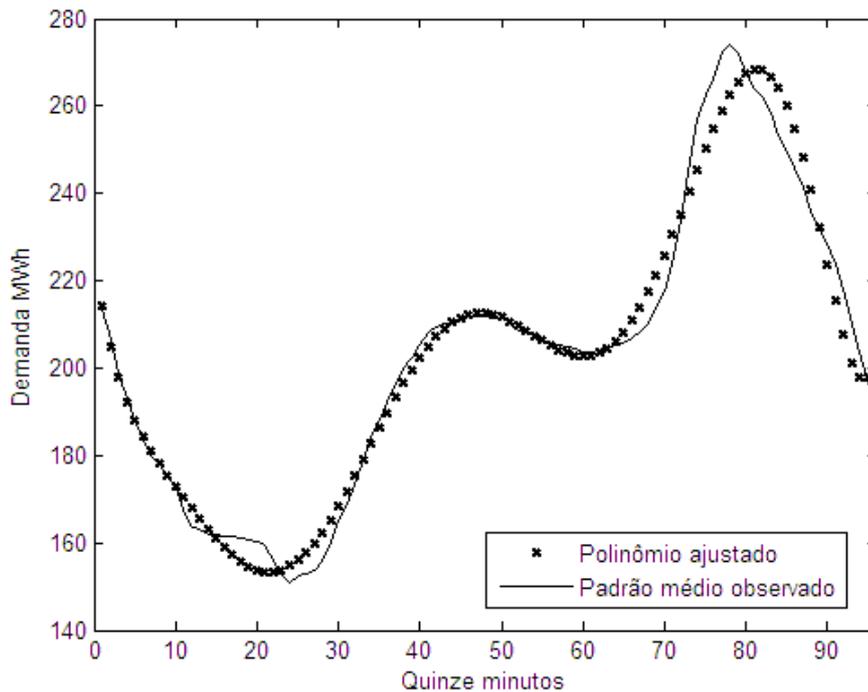
**Figura 4.21** Padrão observado e polinômio estimado para o perfil dos feriados às sextas-feiras

O polinômio que representa o padrão para os feriados às sextas-feiras é:

$$\begin{aligned} \text{sexpadfer} = & 250,13 - 18,80t + 2,58t^2 - 0,20t^3 + 8,79 \cdot 10^{-3}t^4 - 210,01 \cdot 10^{-6}t^5 + \\ & + 2,76 \cdot 10^{-6}t^6 - 18,68 \cdot 10^{-9}t^7 + 50,85 \cdot 10^{-12}t^8 \end{aligned} \quad (4.11)$$

Por fim, definiu-se o padrão que representa o comportamento dos feriados que ocorrem aos sábados. A equação que define este padrão é:

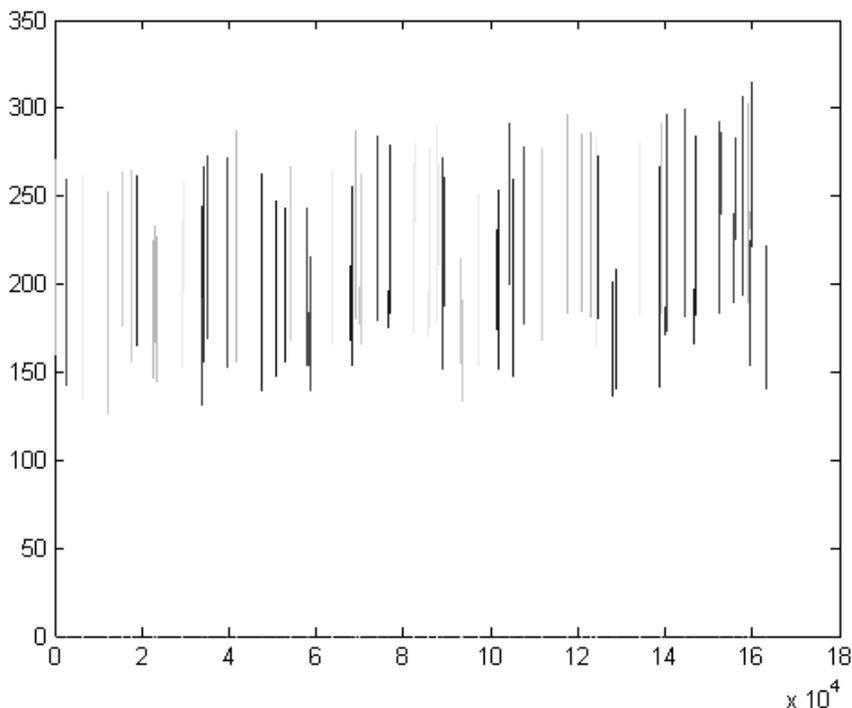
$$\begin{aligned} \text{sabpadfer} = & 226,26 - 13,68t + 1,81t^2 - 0,15t^3 + 7,09 \cdot 10^{-3}t^4 - 176,93 \cdot 10^{-6}t^5 + \\ & + 2,39 \cdot 10^{-6}t^6 - 16,50 \cdot 10^{-9}t^7 + 45,47 \cdot 10^{-12}t^8 \end{aligned} \quad (4.12)$$



**Figura 4.22** Padrão observado e polinômio estimado para o perfil dos feriados aos sábados

Após a estimação dos sete padrões que compõem o perfil de nível 2, o passo seguinte consiste em criar uma série temporal composta pelos padrões encontrados para os dias de feriado. A série apresenta valor nulo para as observações correspondentes aos dias úteis.

Os polinômios estimados foram alocados nos dias de feriado em toda a extensão da série. Assim, é possível determinar matematicamente o comportamento da demanda de carga de energia elétrica em qualquer dia de feriado deste conjunto de dados, sabendo apenas o dia da semana de ocorrência. A série criada para o nível 2 pode ser vista na Figura 4.23.

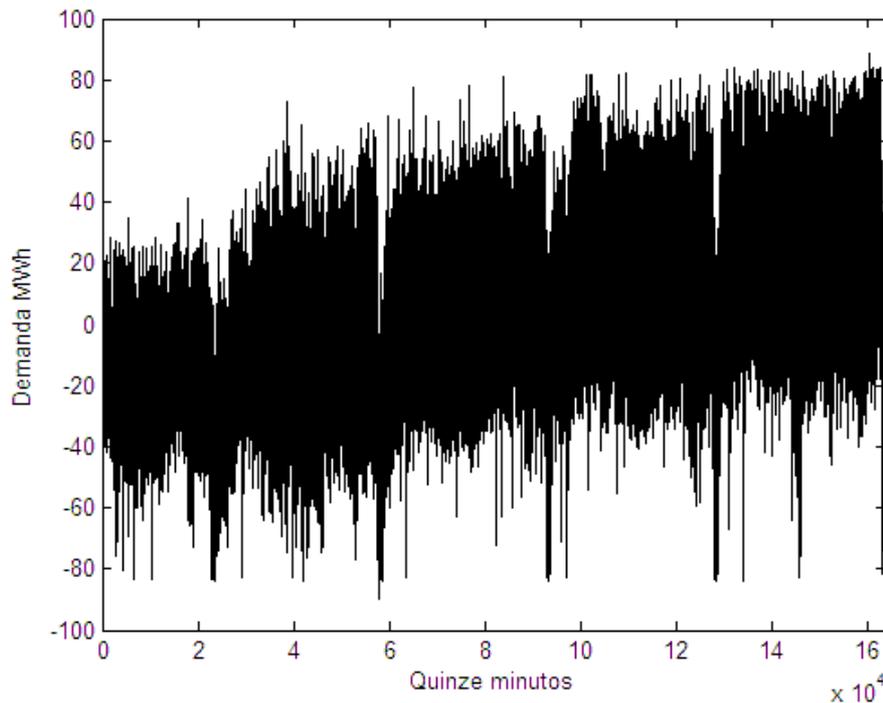


**Figura 4.23** Série de polinômios para os feriados obtida no nível 2

Para a modelagem dos próximos níveis, é necessário criar uma série ajustada que esteja sem o efeito da sazonalidade dentro do dia (tratada no nível 1) e, também, sem o efeito provocado pela ocorrência dos feriados (nível 2). Para isso, calculam-se os resíduos produzidos pela diferença entre a série observada, a série obtida no nível 1 e a série de ajuste de polinômios estimados para os feriados, obtida no nível 2.

A série remanescente após o ajuste dos níveis 1 e 2 é exposta na Figura 4.24. Pode-se observar, numa comparação com a série de dados original, que a tendência de crescimento é mais explícita na série ajustada. Nesta última, a média da série é aproximadamente nula.

O objetivo para os próximos níveis é retirar o efeito da componente determinística remanescente, a fim de obter um resíduo semelhante a um ruído branco, que será modelado com vistas a obter as previsões para os dados. Para isso, é necessário modelar a componente sazonal dos dados. O próximo nível tem a função de identificar e tratar tal componente.



**Figura 4.24** Série de resíduos obtida após o ajuste dos níveis 1 e 2

### 4.3.

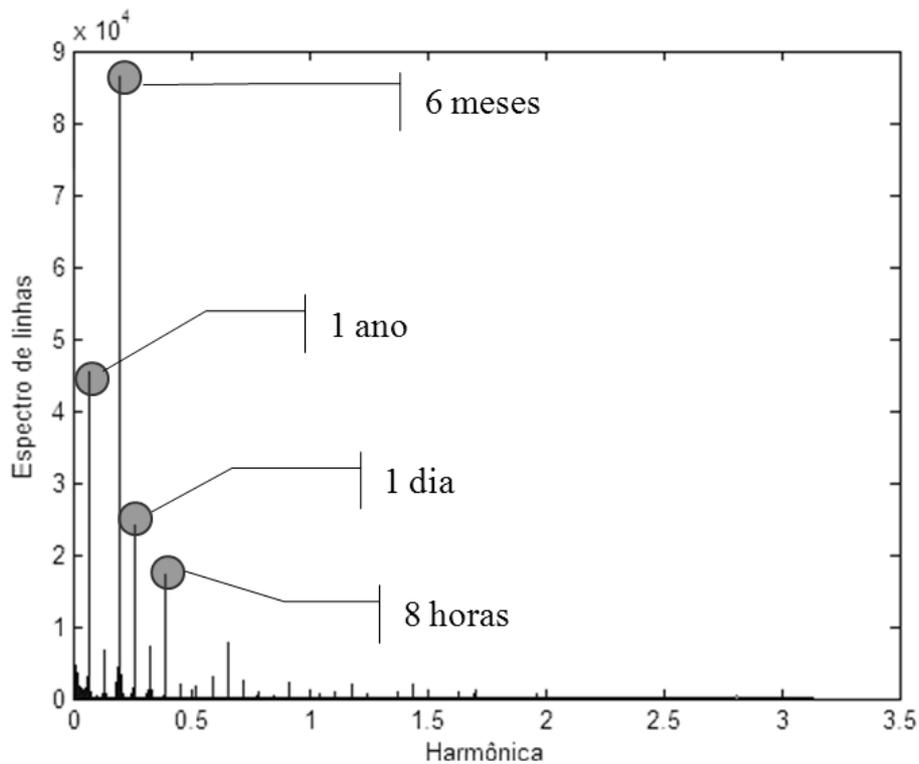
#### **Perfil de Nível 3 – Identificação e modelagem do padrão sazonal**

Segundo Al-Madfai *et al.* (2007), o perfil de nível 3 tem a função de modelar periodicidades anuais ou maiores, ou seja, modela a forma geral da série em torno do ano, incluindo tendência, sazonalidade e distúrbios periódicos.

Para isso é necessário investigar a série temporal com a finalidade de identificar a periodicidade presente nos dados. Tendo em vista que a sazonalidade dentro dia e o efeito proveniente da ocorrência de feriados foram retirados da série, a periodicidade não é mais notada de forma intuitiva como na série original.

Uma ferramenta útil para identificar periodicidades desconhecidas em um conjunto de dados é o Periodograma (seção 2.6.2). Aplicou-se esta técnica à série resultante do ajuste dos níveis 1 e 2 e identificaram-se periodicidades correspondentes a oito horas, um dia, seis meses e um ano.

O Periodograma pode ser observado na Figura 4.25. Os picos maiores representam as periodicidades mais significativas contidas nos dados.



**Figura 4.25** Periodograma da série de demanda de carga de energia elétrica

Os quatro maiores picos encontrados foram submetidos aos testes de Fisher e Whittle e as periodicidades ( $\omega$ ) foram consideradas estatisticamente significativas para este conjunto de dados. Outras periodicidades foram testadas, mas concluiu-se que elas são estatisticamente irrelevantes.

Conhecidas as frequências presentes nos dados é possível modelar a parcela final da componente determinística: a sazonalidade. Tendo em vista que tratam-se de periodicidades múltiplas há algumas abordagens específicas para tratar estes dados. Neste trabalho, optou-se por fazer uso da Regressão Harmônica com frequências conhecidas.

Como o objetivo é obter a componente aleatória após este ajuste, é necessário que os resíduos se apresentem dispersos aleatoriamente em torno de zero. Para isso, além de simplesmente modelar a componente sazonal, é preciso

tratar, também, a tendência dos dados. Assim, a técnica correta a ser empregada é a Regressão Harmônica com tendência polinomial para frequências conhecidas.

A equação da regressão é:

$$\begin{aligned} X_t = & \mu + at + bt^2 + \dots + c^n + A_1 \cos(\omega_1 t) + B_1 \text{sen}(\omega_1 t) + \\ & + A_2 \cos(\omega_2 t) + B_2 \text{sen}(\omega_2 t) + A_3 \cos(\omega_3 t) + B_3 \text{sen}(\omega_3 t) + \\ & + A_4 \cos(\omega_4 t) + B_4 \text{sen}(\omega_4 t) + \varepsilon_t \end{aligned} \quad (4.13)$$

onde:

$$\begin{aligned} \omega_1 &= 0,0001793146 \\ \omega_2 &= 0,0003586293 \\ \omega_3 &= 0,0656291616 \\ \omega_4 &= 0,1963495408 \end{aligned}$$

e  $\mu, a, b, \dots, c, A_1, B_1, A_2, B_2, A_3, B_3, A_4$  e  $B_4$  são os parâmetros a estimar.

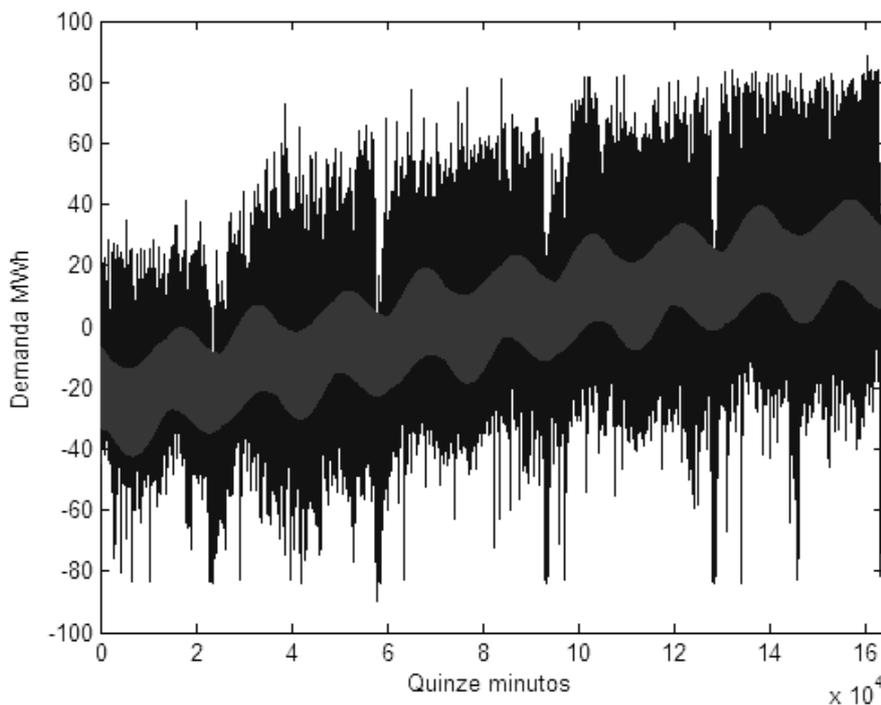
O grau do polinômio que define a tendência foi testado iterativamente até chegar ao modelo com menor MAPE e todos os parâmetros significativos. Foram estimados inúmeros modelos com variadas configurações até se chegar ao modelo mais adequado sob o ponto de vista estatístico.

Assim, a tendência é definida por um polinômio de grau 3 e os parâmetros estimados para a regressão que modela a sazonalidade são:

$$\begin{aligned} \mu &= -24,3784 \\ a &= 335,35 \cdot 10^{-6} \\ b &= 337,31 \cdot 10^{-12} \\ c &= -3,44 \cdot 10^{-15} \\ A_1 &= 0,3149 \\ B_1 &= -0,4667 \end{aligned}$$

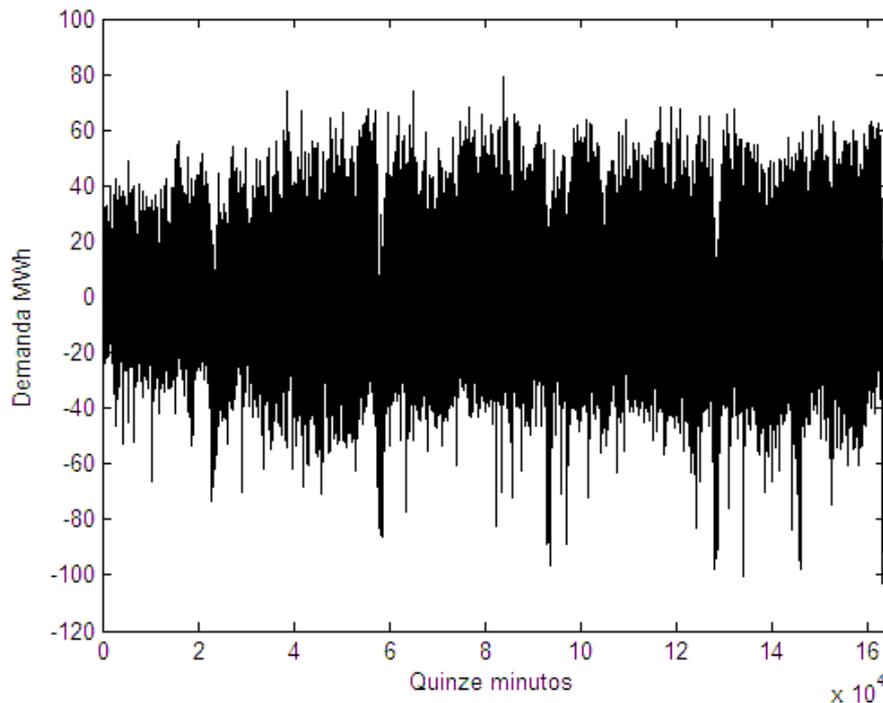
$$\begin{aligned}A_2 &= 4,2393 \\B_2 &= -3,5282 \\A_3 &= 5,9748 \\B_3 &= -0,1392 \\A_4 &= -2,8741 \\B_4 &= 7,8447\end{aligned}$$

A Figura 4.26 mostra a comparação entre valores observados e ajustados segundo a Regressão Harmônica. Nota-se que a dinâmica dos dados não é completamente entendida pelo modelo, porém espera-se que os resíduos apresentem o comportamento aleatório para prosseguir às próximas etapas do HPA.



**Figura 4.26** Ajuste da série resultantes dos níveis 1 e 2 via Regressão Harmônica

Pode-se perceber, no ajuste realizado, que o modelo conseguiu captar a tendência contida nos dados. Dado que a regressão também modelou a sazonalidade, a diferença entre a série inicial (sem o efeito dos dois primeiros níveis) e série resultante (sem o efeito dos três níveis) é um ruído aleatório, conforme mostra a Figura 4.27.



**Figura 4.27** Série de resíduos obtidos após o ajuste dos níveis 1, 2 e 3

#### 4.4 Criação da Função Determinística

Após a identificação dos perfis e o processo inicial de modelagem, o “modelo fundamental” da série foi obtido.

Os dados foram examinados para perfis de níveis superiores a 3 e nada foi encontrado. Isto significa que toda a componente determinística foi capturada pelos ajustes dos níveis discutidos anteriormente.

Ajustados os perfis, a função determinística  $f(t)$  pode ser obtida através da soma dos perfis obtidos nos níveis 1, 2 e 3. Cabe ressaltar que não há uma equação simples que caracterize esta função.

Pode-se considerar que, para este conjunto de dados, cada dia da semana observado tem uma equação definida que depende do ajuste nos perfis iniciais.

Para uma segunda-feira, por exemplo, a função determinística que define o comportamento da demanda de carga de energia elétrica é:

$$f_{seg}(t) = (1-i)secpad(t) + i.secpadfer(t) + X_t(t) \quad (4.14)$$

onde  $secpad(t)$  é o polinômio definido para o padrão de uma segunda-feira de não feriado,  $secpadfer(t)$  é o polinômio que define o padrão para uma segunda-feira de feriado,  $X_t$  é a equação estimada pela Regressão Harmônica que define o padrão sazonal e  $i$  é um indicador da forma:

$$i = \begin{cases} 1 & \text{se é feriado} \\ 0 & \text{caso contrário} \end{cases}$$

que pode ser reescrita da forma:

$$\begin{aligned} f_{seg}(t) = & (1-i).181,06 + 2,68t - 1,46t^2 + 0,13t^3 - 4,27.10^{-3}t^4 + 6,72.10^{-5}t^5 + \\ & + 5,08.10^{-7}t^6 + 1,48.10^{-9}t^7 + i.[226,26 - 13,68t + 1,81t^2 - 0,15t^3 + \\ & + 7,09.10^{-3}t^4 - 176,93.10^{-6}t^5 + 2,39.10^{-6}t^6 - 16,50.10^{-9}t^7 + \\ & + 45,47.10^{-12}t^8] - 24,4 + 335,35.10^{-6}t + 337,31.10^{-12}t^2 - 3,44.10^{-15}t^3 + \\ & + 0,31\cos(1,79.10^{-4}t) - 0,47\sen(1,79.10^{-4}t) + 4,24\cos(2,15.10^{-3}t) - \\ & + 3,53\sen\cos(2,15.10^{-3}t) + 5,97\cos(9,32.10^{-3}t) - 0,14\sen(9,32.10^{-3}t) + \\ & - 2,87\cos(0,07t) + 7,84\sen(0,07t) \end{aligned}$$

As equações que definem a componente determinística para os demais dias da semana são criadas de forma análoga à equação para as segundas-feiras.

## 4.5 Modelagem da Componente Estocástica

Conforme exposto na equação (3.1), a série pode ser decomposta em duas componentes: estocástica e determinística. Dado que a última foi modelada, a série resultante da diferença entre a série original e os ajustes dos níveis contém apenas a componente estocástica.

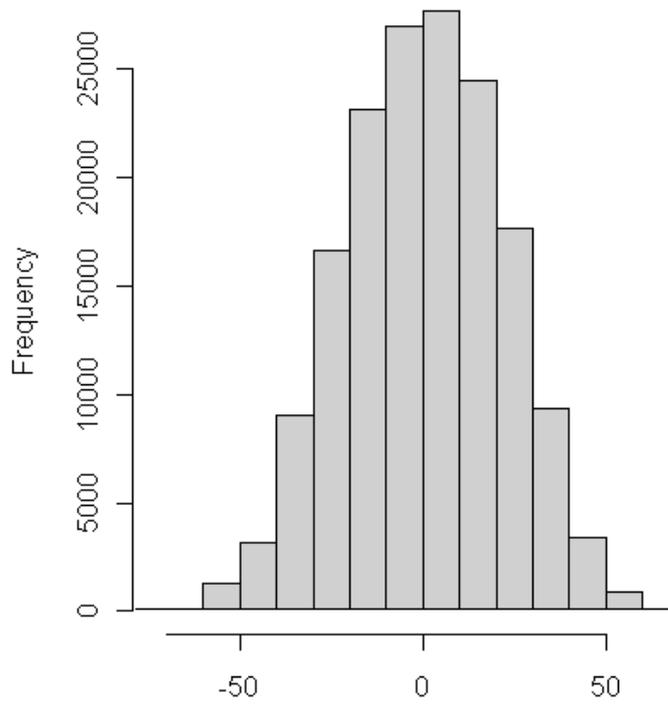
O objetivo final da modelagem antes de fazer a previsão é, então, modelar a componente estocástica. Visto que se trata de um ruído sem presença de tendência ou componente sazonal, qualquer abordagem simples pode ser utilizada para realizar tal tarefa.

Neste trabalho, optou-se por fazer uso da conhecida metodologia de Box & Jenkins. Dadas as características da série, será utilizado um modelo do tipo ARMA(p,q).

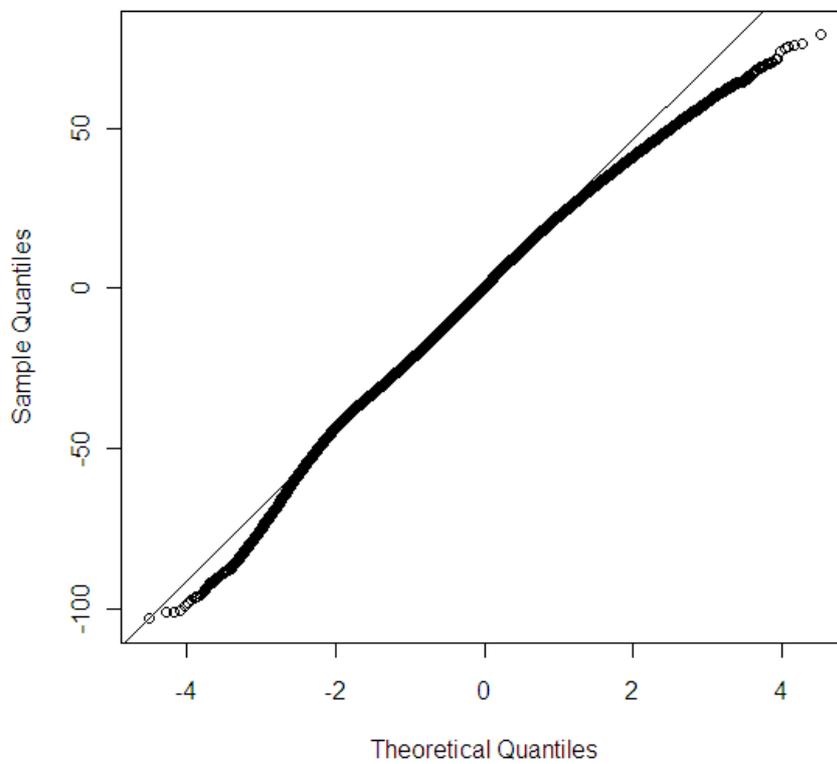
Antes de proceder à modelagem é necessário verificar os pressupostos de normalidade, estacionariedade e homocedasticidade.

A análise da série resultante do ajuste dos três níveis considerados (Figura 4.27) revela que os dados são estacionários, visto que não é possível notar a presença da componente de tendência na série.

Para verificar a hipótese de Normalidade, podemos fazer uso de alguns testes estatísticos, como o Jarque-Bera, além das análises do histograma e do gráfico Normal Q-Q Plot (Figuras 4.28 e 4.29).



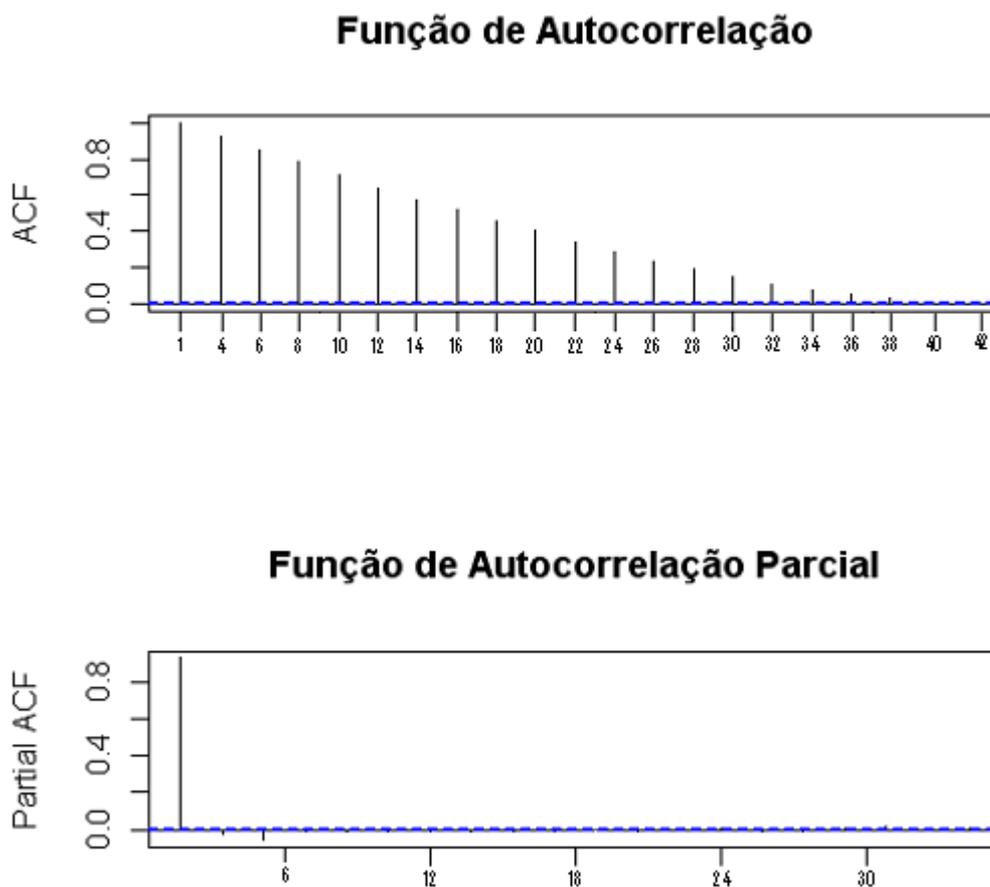
**Figura 4.28** Histograma dos dados



**Figura 4.29** Normal Q-Q Plot dos dados de demanda de carga elétrica

As figuras anteriores revelam que os dados se distribuem normalmente. O histograma apresenta simetria e a forma de sino característica da distribuição Normal. No gráfico Normal Q-Q não há desvios significantes da reta Normal. Esta análise é comprovada pela aplicação do teste de Jarque-Bera, que indica normalidade ao nível de 5% de significância.

Cumpridos os pressupostos, deve-se recorrer à análise do correlograma e autocorrelograma para identificar as ordens  $p$  e  $q$  do modelo. Estes gráficos podem ser observados na Figura 4.30.



**Figura 4.30** FAC e FACP da série de demanda de carga elétrica

Pode-se afirmar que a FAC apresenta decaimento exponencial lento e a FACP apresenta alguns *lags* significantes. Este é um indicativo de que trata-se de um modelo ARMA.

Partiu-se da estimação e análise residual de um modelo ARMA(1,1) e, via Critério de Sobrefixação, chegou-se ao modelo ARMA(2,2), cuja equação é:

$$Z_t = -0,0001 + 0,8084Z_{t-1} - 0,8240Z_{t-2} + a_t + 0,8647a_{t-1} - 0,0242a_{t-2} \quad (4.15)$$

onde  $Z_{t-1}$  e  $Z_{t-2}$  são as observações defasadas em 1 e 2 *lags*, respectivamente, e  $a_t$ ,  $a_{t-1}$  e  $a_{t-2}$ , o ruído.

Os resíduos gerados pelo modelo foram analisados segundo os testes de Portmanteau e Jarque-Bera, onde concluiu-se que os mesmos são descorrelatados e normais. Pode-se afirmar, então, que o modelo é bem especificado e permite realizar previsões.

Finalizada a modelagem da componente estocástica, pode-se concluir que a equação do modelo final é dada pela soma da componente determinística (que contém os perfis) com a componente estocástica.

O modelo HPA para a demanda de carga elétrica de 15 em 15 minutos é:

$$y_t = f(t) + Z_t \quad (4.16)$$

onde  $y_t$  são os dados observados,  $f(t)$  a componente determinística especificada para cada dia da semana e  $Z_t$  é a componente estocástica, dada pela equação (4.15).

A seguir, podem ser obtidas as previsões 672 passos à frente, objetivo do trabalho.

## 5 Resultados

Obtida a equação fundamental do modelo HPA, é possível escrever a equação de previsão para  $k$  passos à frente:

$$y_t(k) = f(t+k) + Z_t(k) \quad (5.1)$$

onde  $y_t(k)$  são os valores previstos  $k$  passos à frente partindo da origem no tempo  $t$ ,  $f(t+k)$  são as previsões  $k$  passos à frente para a componente determinística e  $Z_t(k)$  é a previsão para a componente estocástica.

O objetivo deste trabalho é realizar previsões para uma semana, ou seja, 672 passos à frente. Para isso, escolheu-se a semana de dados observados entre 24 e 30 de dezembro de 2006, visto que ela compreende 6 dias comuns e 1 feriado, a fim de testar a adequação do modelo para estes dias, além do ajuste dos fatores de correção para o feriado de Natal.

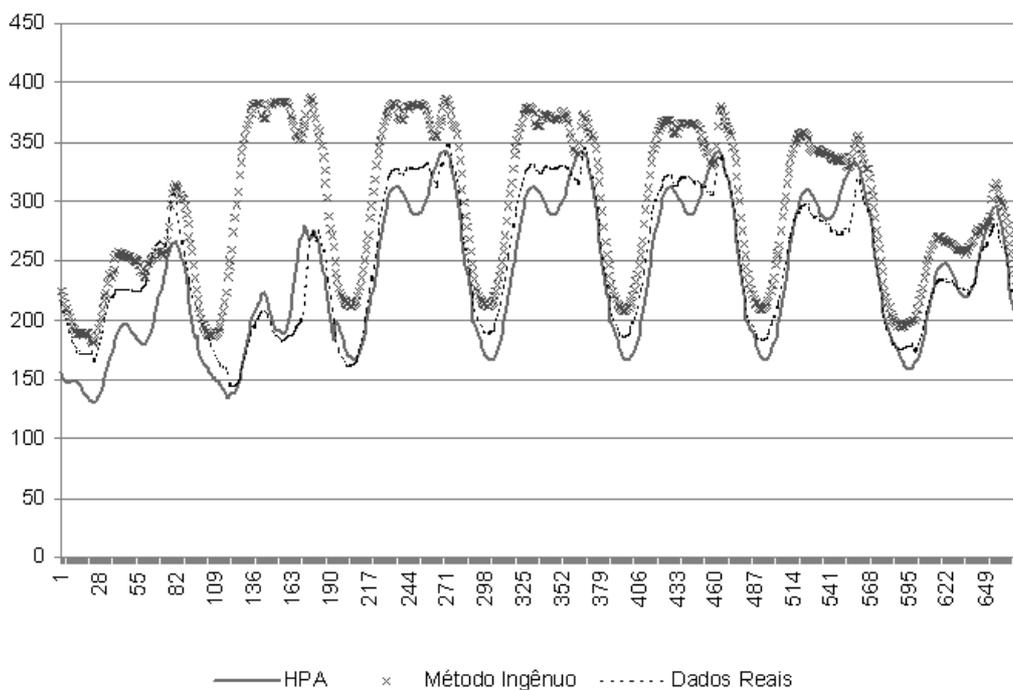
Para gerar as previsões do HPA, os dias da semana considerados foram separados e, para cada um deles, à exceção do dia 25 de dezembro, estimou-se as previsões para os dias úteis utilizando os polinômios obtidos no nível.

Para o Natal, estimaram-se as previsões geradas para os feriados que ocorrem às segundas-feiras e, em seguida, aplicaram-se os fatores de correção obtidos no nível 2.

Para concluir a parcela da previsão proveniente da componente determinística, foram estimadas as previsões para a tendência e sazonalidade via Regressão Harmônica. A adição destes três fatores compõe a previsão da componente determinística. A esta, foram incorporadas as previsões para a componente estocástica, gerada pela aplicação do modelo ARMA(2,2).

A fim de verificar o poder preditivo do modelo estimado, foi feita uma comparação com o Método Ingênuo de previsão via critério MAPE e Coeficiente U de Theil. Neste último método, a previsão para uma semana a frente é exatamente igual à última semana observada, ou seja, no Método Ingênuo as previsões correspondem aos valores observados para a semana compreendida entre 17 e 23 de dezembro de 2006.

A comparação entre os dados reais e as previsões geradas pelos dois modelos pode ser observada graficamente na Figura 5.1.



**Figura 5.1** Comparação entre as previsões geradas pelo HPA e pelo Método Ingênuo

Analisando a comparação, pode-se notar que, mesmo nos dias de não-feriado, onde não há um comportamento tão específico quanto no dia de feriado, as previsões geradas pelo HPA se mostraram mais aderentes do que as previsões geradas pelo Método Ingênuo.

O HPA detectou um padrão de queda de demanda no meio do dia para todas as datas observadas. Tal comportamento não é verificado, na realidade, para as terças, quartas e quintas-feiras. O modelo capturou perfeitamente o comportamento do feriado de Natal, acompanhando toda a dinâmica da demanda de carga para este dia.

Um modelo só pode ser utilizado para realizar previsões se, no mínimo, ele fornece erros menores que o Modelo Naíve. Calculando-se o MAPE para os dois modelos considerados tem-se, para o HPA, o valor de 5,46% e, para o Método Ingênuo, 15,08%.

O coeficiente U de Theil também foi calculado para o modelo estudado, fornecendo um valor de 0,3616 o qual indica que o erro de previsão do HPA é menor do que o do Método Ingênuo.

Estes resultados indicam que o HPA pode ser empregado com o intuito de gerar previsões para este conjunto de dados e que suas previsões se apresentam aderentes, inclusive nos feriados, que tendem a distorcer as previsões por apresentarem um padrão de demanda diferente dos dias comuns.

## 6 Conclusões

Tendo em vista o modelo vigente no setor elétrico brasileiro, a previsão de demanda de carga elétrica para dados de alta frequência tem um papel fundamental na distribuição de energia, sendo, por isso, necessários aprimoramentos para tornar tais previsões cada vez mais precisas.

Neste trabalho, foi utilizada uma técnica inovadora para modelar dados de alta frequência: o modelo *Hierarchical Profiling Approach* (HPA). Com este, calculou-se a previsão para uma semana, ou seja, 672 passos à frente, para os dados de demanda de energia elétrica, observados a cada 15 minutos, provenientes de uma concessionária de energia do Sudeste brasileiro.

Em um primeiro momento, o HPA tratou a componente determinística da série, modelando perfis que capturaram a sazonalidade dentro do dia (perfil nível 1), o comportamento de observações atípicas como os feriados (perfil nível 2) e as demais periodicidades contidas na série (perfil nível 3). Através da aplicação do Periodograma, foram identificadas periodicidades correspondentes a oito horas, um dia, seis meses e um ano. Estas componentes foram modeladas por uma Regressão Harmônica com tendência polinomial. A componente determinística foi criada pela adição dos perfis encontrados.

O resíduo gerado pela diferença entre a série original e a componente determinística gerou a componente estocástica que foi modelada utilizando um modelo ARMA(2,2).

As previsões foram compostas pela agregação das previsões das componentes determinística e estocástica e comparadas às previsões do Modelo Naíve. A série prevista pelo HPA gerou um MAPE de 5,46% e, o Método Ingênuo, 15,08%.

O coeficiente U de Theil também foi calculado para verificar a qualidade das previsões fornecidas pelo modelo estudado. Como resultado, obteve-se o valor 0,3616 o qual indica que o erro de previsão do HPA é menor do que o do Método Ingênuo de previsão.

Com base nos resultados apresentados, pode-se afirmar que o modelo HPA fornece uma análise lúcida e poderosa da série estudada. A aplicação do modelo pode ser estendida para uma variedade de dados da vida real onde há dificuldade em obter a estacionariedade utilizando os métodos estatísticos convencionais.

A aplicação do HPA proporciona um melhor entendimento da dinâmica subjacente aos dados, visto que as componentes da série são estudadas a fundo e, ao mesmo tempo, serve como uma técnica de pré-branqueamento dado que separa as componentes até obter a parcela estocástica presente na série.

Esta metodologia pode, ainda, ser usada para prever em massa e tem a vantagem de poder ser implementada em softwares já existentes. O modelo pode, ainda, ser visto como uma abordagem quantitativa para incorporar a opinião de especialistas na análise de séries temporais estocásticas.

### **6.1. Sugestões**

A inclusão de variáveis explicativas no modelo pode contribuir para captar padrões que não foram detectados pelo modelo estimado. Poder-se-ia estudar a inclusão de variáveis como temperatura, luminosidade, chuvas, população atendida, etc, além do efeito das estações do ano. O grande problema para tal melhoria consiste na obtenção destas variáveis. Por isto, caso o acesso a esses dados torne-se viável, pode ser feita uma nova análise, aprimorando a que foi aqui apresentada.

A inclusão de opiniões mais detalhadas provenientes de especialistas da área energética também poderia trazer ganhos significativos para o poder explicativo do modelo.

Para a base de feriados, caso também esteja disponível um histórico maior, pode-se analisar o seu impacto de acordo com a condição meteorológica do dia a ser previsto. Poderiam ser analisados separadamente os feriados de verão e os de inverno, por exemplo, o que não pôde ser feito aqui devido à pouca quantidade de dados. Assim, poder-se-ia classificar um dia de feriado a ser previsto como dia quente, frio ou intermediário e analisar o comportamento da demanda de energia.

Com relação ao modelo proposto, uma sugestão é testar técnicas diferentes para modelar a componente estocástica, como o Método Holt, Holt-Winters, Redes Neurais Artificiais, etc. e verificar qual técnica produz o modelo mais aderente. Além disso, pode-se estudar a possibilidade de gerar intervalos de confiança para as previsões.

## 7

### Referências Bibliográficas

AL-MADFAI, H.; AMEEN, J.; RYLEY, A.; **Daily electricity demand forecasting: A hierarchical profiling approach.** ETK/NTTS, 2001.

AL-MADFAI, H.; **Hierarchical Profiling of daily Crime Time Series Data as a Precursor to Modelling.** International Symposium of Forecasting, San Antonio, Texas, 2005.

AL-MADFAI, H.; AMEEN, J.; RYLEY, A.; **The hierarchical profiling approach to STLF of multi-year daily electricity demand in South Wales.** University of Glamorgan, 2004.

AL-MADFAI, H.; IVAHA, C.; HIGGS, G.; WARE, A.; CORCORAN, A.; **The Simple Spatial Disaggregation Approach to spatio-temporal crime forecasting.** International Journal of Innovative Computing, Information and Control, Volume 3, Number 3, June 2007.

AL-MADFAI, H.; AKINWALE, A.; LAKIN, S.; THOMAS, S.; LLOYD, S.; **Using Hierarchical Profiling Approach (HPA) Forecasts of Multi-Year Half Hourly Electricity Consumption Data as a Tool in Energy Management.** 29<sup>th</sup> International Symposium of Forecasting, 2009.

AMARAL, E.; **Análise harmônica e previsão a longo prazo.** Reunião Internacional de Biometria, Piracicaba. Anais. Piracicaba: Fundação Cargill, 1975. p. 232-271

ANEEL; **Perguntas e respostas sobre tarifas das distribuidoras de energia elétrica/ Agência Nacional de Energia Elétrica.** Brasília, 2007.

ANTON, H.; **Elementary Linear Algebra with Applications**. Editora Bookman, 2001.

ARTIS, M.; CLAVEL J. G.; HOFFMAN, M.; NACHANE, D.; **Harmonic Regression Models: A Comparative Review With Applications**. Working Paper Series, No 333, ISSN 1424-0459, 2007.

BARROS, M.; **Processos Estocásticos**. Rio de Janeiro. Papel Virtual, 2004.

BLOOMFIELD, P.; **Fourier analysis of time series: an introduction**. New York: J. Wiley, 1976.

BOX, G. E. P.; JENKINS, G. M.; **Time Series Analysis, Forecasting and control**. San Francisco; Holden-Day, Inc., 1970.

BOX, G. E. P.; JENKINS, G. M.; REINSEL; **Time Series Analysis: forecasting and control**. New Jersey, Prentice Hall, 1994.

BUJOSA, M.; GARCÍA-FERRER, A.; YOUNG, P. C.; **Linear dynamic harmonic regression**. Computational Statistics & Data Analysis, 52, p. 999-1024, 2007.

BUNN, D. W.; FARMER, E. D.; **Comparative models for electrical load forecasting**. Wiley, Belfast, 1985.

CASELLA, G; BERGER, R.L.; **Statistical Inference**. Duxbury, 2<sup>nd</sup> ed., 2001.

ESTEVES, G. R.T.; **Modelos de Previsão de Carga de Curto Prazo**. Dissertação de mestrado, DEE, PUC - Rio, abril 2003.

FISHER R. A.; **Tests of significance in harmonic analysis**. Proc. Roy. Soc. A, 125, p. 54-59, 1929.

IVAHA, C.; AL-MADFAI, H.; HIGGS, G.; WARE, A.; **The Dynamic Spatial Disaggregation Approach: A Spatio-Temporal Modelling of Crime**. Proceedings of the World Congress on Engineering 2007 Vol II. WCE 2007, July 2 - 4, London, U.K., 2007.

LOURENÇO, P. M.; **Um Modelo de Previsão de Curto Prazo de Carga Elétrica Combinando Métodos Estatísticos e Inteligência Computacional**. PUC - Rio, Tese de Doutorado, 1998.

MEDEIROS, M. C.; **Técnicas de Pesquisa em Economia**. Notas de Aula, 2005.

MIRANDA, C. V.; SOUZA, R.C.; **Previsão de dados de alta frequência para carga elétrica usando Holt-Winters com dois ciclos**. Dissertação de mestrado, DEE, PUC-Rio, 2007.

MONTGOMERY, D.C., JOHNSON, L. A.; **Forecasting and Time Series Analysis**. New York, McGraw-Hill Book Co., 2 ed., 1990.

MORETTIN, P. A.; TOLOI, C. M. C.; **Análise de Séries Temporais**. São Paulo: Edgard Blücher, 2006.

MORETTIN, P. A.; MESQUITA, A.R.; ROCHA, J.G.C.; **Rainfall At Fortaleza In Brazil Revisited**. In: Time Series Analysis and Forecasting, 1985, Toronto. Amsterdam : North Holland, 1983, v. 6, p. 67-85.

QUADRELLI, G.; **Modelos Comparativos de Previsão de Carga Elétrica de Curto Prazo**. Dissertação de mestrado, DEE, PUC - Rio, abril 1998.

RIZZO, G. M., **Previsão de Carga de Curtíssimo Prazo no Novo Cenário Elétrico Brasileiro**. Dissertação de mestrado, DEE, PUC - Rio, fevereiro 2001.

SHIMSHONI, M.; **Fisher's test of significance in harmonic analysis**. Geophys. J. R. Astron Soc., 23, p. 373-377, 1971.

SOBRAL, A. P. B.; **Modelo de Previsão Horária de Carga Elétrica para Light**. Dissertação de mestrado, DEE, PUC - Rio, março 1999.

SOUZA, R.C., CAMARGO, M. E.; **Análise e previsão de séries temporais: os modelos ARIMA**. Ijuí, SEDIGRAF, 2004.

SOUZA, R. C.; **Métodos Automáticos de Amortecimento Exponencial para Previsão de Séries Temporais**. Monografia GSM-10/83, maio 1983.

SOUZA, R. C.; **Modelos Estruturais para Previsão de Séries Temporais: Abordagens Clássica e Bayesiana**. 17º Colóquio Brasileiro de Matemática, IMPA, 1989.

TAYLOR, J. C.; PEDREGAL, D. J.; YOUNG, P. C.; TYCH, W.; **Environmental Time Series analysis and forecasting with the Captain Toolbox**. Environmental Modelling & Software 22, p. 797-814, 2007.

VAAGE, K.; **Detection of Outliers and Level Shifts in Time Series: An Evaluation of Two Alternative Procedures**. Journal of Forecasting, Nº 19, p. 23-37, 2000.

WHITTLE, P.; **The simultaneous estimation of a time series harmonic components and covariance structure**. Trabajos Estad., 3, p. 43-57, 1952.

YOUNG, P. C.; PEDREGAL, D. J.; TYCH, W.; **Dynamic Harmonic Regression**. Journal of Forecasting, 18, p.369-934, 1999.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)