

INTERVALO DE CONFIANÇA PARA PREVISÃO DE CARGA A CURTO-PRAZO

GILZA SANTOS SIMÃO FERREIRA

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO -

UENF

CAMPOS DOS GOYTACAZES – RJ

AGOSTO/2013

INTERVALO DE CONFIANÇA PARA PREVISÃO DE CARGA A CURTO-PRAZO

GILZA SANTOS SIMÃO FERREIRA

Dissertação apresentada ao Centro de Ciências e Tecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Engenharia de Produção.

Orientador: Professor Lacir Jorge Soares

CAMPOS DOS GOYTACAZES – RJ
AGOSTO/2013

INTERVALO DE CONFIANÇA PARA PREVISÃO DE CARGA A CURTO-PRAZO

GILZA SANTOS SIMÃO FERREIRA

Dissertação apresentada ao Centro de Ciências e Tecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Mestre em Engenharia de Produção.

1º de agosto de 2013.

Comissão Examinadora:

Prof. Ricardo Linden (Doutor em Engenharia Elétrica) –
CEPEL/FSMA

Prof. Rogério Figueiredo Daher (Doutor em Produção Vegetal) – UENF/CCTA

Prof. José Ramón Arica Chávez (Doutor em Engenharia de Sistemas e Computação) –
UENF/LEPROD

Prof. Lacir Jorge Soares (Doutor em Engenharia Elétrica) –
UENF/LESCE (Orientador)

Carinhosamente, dedico esta dissertação a meus pais, Gerson Ferreira e Maria de Fátima Ferreira, por terem confiado em mim e me apoiado sempre.

Agradecimentos

Aos meus pais, grandes encorajadores e intercessores em mais esta caminhada. Quantas foram as vezes, que mesmo distantes fisicamente, estiveram ao meu lado, dizendo para eu não desistir.

Imensa gratidão a meu orientador Professor Lacir Soares. Aproximadamente 2 anos de uma convivência, acompanhamento e orientação acadêmica. Obrigada por toda paciência, apoio, confiança e principalmente conselhos.

As minhas irmãs Gisiane Ferreira e Fani Ferreira, parentes, amigos (as) em especial a Keilla e Luiz Matias e todos, que de alguma forma, se fizeram presentes e me apoiaram, o meu eterno agradecimento!

RESUMO

Ferreira, G. S. S.; Soares, L. J. **INTERVALO DE CONFIANÇA PARA PREVISÃO DE CARGA A CURTO-PRAZO**. Campos dos Goytacazes, 2013. 64p. Dissertação de Mestrado - Laboratório de Engenharia de Produção, Universidade Estadual do Norte Fluminense Darcy Ribeiro.

Este trabalho apresenta uma metodologia estatística para construir intervalos de confiança para previsões de carga a curto prazo, por hora (de um à sete dias à frente). A construção de intervalos de confiança para os valores previstos, como um quantificador da confiabilidade da previsão, é um instrumento auxiliar à tomada de decisão dos operadores do sistema elétrico permitindo uma diminuição nos riscos de despacho e operação. Os intervalos de confiança (IC) são construídos baseado no cálculo dos quantis empíricos dos erros de previsão, observados no passado. A técnica simulação *bootstrap* é utilizada para identificar a distribuição densidade de probabilidade dos erros evitando a identificação por meio de integração numérica. A técnica *bootstrap* baseia-se na construção de distribuições amostrais por (re)amostragem com reposição. São utilizadas 1000 replicações para identificação do parâmetro de interesse, média. Os intervalos *bootstrap* são construídos com 95% de confiança. A volatilidade da série é cuidadosamente considerada, de tal modo que os robustos intervalos de confiança sejam computados. Todo o ciclo de modelagem é baseado em argumentos estatísticos rigorosos. Nossa proposta é colocar a prova, o conjunto de dados de uma concessionária de eletricidade do sudeste do Brasil.

Palavras-chave

Intervalos de Confiança, Bootstrap, Modelagem Estatística.

ABSTRACT

Ferreira, G. S. S.; Soares, L. J. **CONFIDENCE INTERVAL FOR SHORT TERM LOAD FORECAST** . Campos dos Goytacazes, 2013. 64p. MSc. Thesis - Engineering Production Laboratory, Universidade Estadual do Norte Fluminense Darcy Ribeiro.

The goal of this work is to present a statistical methodology to compute confidence intervals for short-term load forecasting. The paper is focused solely on hourly short-term forecasts (from one- to seven-days-ahead), but the developments can be directly extended to data observed at higher frequencies (half-hourly, for example). Two approaches are considered to obtain confidence intervals (CI) for load forecast. The first one is based on the calculation of empirical quantiles of relative forecast error observed in the past and the second one uses simulation or by block bootstrapping the residuals. Numerical integration can be avoided either by approximating the integral by simulation or by block bootstrapping the residuals. In this paper we adopt the block bootstrap approach. Confidence intervals provide a way of quantifying the uncertainty of the forecast. They could be used to assess as precisely as possible the operating margins. All the forecasts are computed using of the block-bootstrap procedure with 1000 replications (paths). The length of the block is set to 60 observations. The final forecast is the average over the paths. As the expected frequency is 5%, there is still space for improvement in the construction of the confidence intervals. The volatility of the series is also carefully considered, such that robust confidence intervals are computed. All the modeling cycle is based on rigorous statistical arguments. Our proposal is put to proof the dataset from electricity utility in Southeastern Brazil.

Keywords

Confidence Interval, Bootstrap, Statistical Model building.

Sumário

1	Introdução	1
1.1	Previsão de carga e Intervalos de Confiança	1
1.2	Estrutura deste trabalho	2
2	Modelagem de Previsão de Carga	3
2.1	Introdução	3
2.2	Modelagem Estatística	3
2.2.1	Metodologia Box & Jenkins	3
2.2.2	Modelos sazonal autorregressivo integrado médias móveis - SARIMA (Seasonal Integrated AutoRegressive Moving Average)	10
2.2.3	Modelos em duas Etapas	11
2.3	Modelagem com Inteligência Artificial	12
2.4	Resumo	13
3	Pesquisa Bibliográfica	14
3.1	Introdução	14
3.2	Métodos Estatísticos	14
3.3	Inteligência artificial	15
3.4	Intervalos de Confiança	17
3.5	Resumo	20
4	Intervalos de Confiança	21
4.1	Introdução	21
4.2	(Re)amostragem	22
4.2.1	Método <i>Bootstrap</i>	23
4.2.2	Método Monte Carlo	26

4.3	Resumo	27
5	Experimento	28
5.1	Os dados	28
5.2	Características das Curvas de carga	28
5.3	Classificação dos dias	29
5.4	Carga determinística	29
5.4.1	Tendência	29
5.4.2	Sazonalidade	31
5.5	Preparação dos dados	33
5.5.1	Teste Jarque-Bera	34
5.5.2	Teste Kolmogorov-Smirnov	34
5.5.3	Teste qui-quadrado	35
5.6	Intervalos de Confiança	37
5.7	Resumo	37
6	Resultados	39
7	Conclusão	45
7.1	Sugestões para trabalhos futuros	46
	Referências Bibliográficas	47
A	Tabelas - 2000 - 2005	51

Lista de Tabelas

5.1	Classificação dos dias	30
6.1	Parâmetros do modelo AR	40
6.2	MAPE por estações (22/12/2003-21/12/2004)	41
6.3	Pontos fora do intervalo de confiança (Erro)	42
A.1	MAPE Total- Modelagem AR	52
A.2	TESTES DE NORMALIDADE	53
A.3	MAPE - Modelagem AR	54

Lista de Figuras

2.1	Função de autocorrelação	9
5.1	Série observada e série com a aplicação de uma diferença	31
5.2	Cargas médias diárias (01/01/2002-30/04/2005)	32
5.3	Perfil de carga típico de uma semana de verão (19/01/2004 a 25/01/2004) e inverno (12/07/2004 a 18/07/2004)	33
5.4	Curva de carga de quatro quarta-feiras nas diferentes estações período referente aos anos de 2002 e 2003.	33
6.1	MAPE(Erro Médios Absolutos Percentuais) do ano 2004(claro) e MAPE do ano 2005(escuro)	41
6.2	Intervalo de confiança para dias úteis e um dia de feriado-21 de abril	43
6.3	Intervalo de confiança construído para uma quarta-feira de cada estação	44

Capítulo 1

Introdução

1.1 Previsão de carga e Intervalos de Confiança

A construção de intervalos de confiança para previsão é um tópico de interesse na literatura de carga a curto prazo pois apresenta-se como um quantificador da confiabilidade da previsão. Na maioria dos trabalhos publicados os intervalos de confiança são construídos com a suposição de normalidade da distribuição dos resíduos de previsão [Guan et al., 2010]; [Papadopoulos et al., 2001]; [Rivals & Personnaz, 2000], mas tal suposição nem sempre é confirmada como mostra [Bartkiewicz, 2000].

Uma das alternativas encontradas na literatura para solucionar este problema é obter intervalos de confiança utilizando procedimentos *bootstrap*, os quais não requerem a suposição de normalidade da distribuição [Fan. & Yao, 2005]. O método *bootstrap*, proposto por [Efron, 1979], começou a ser utilizado na literatura de séries temporais a partir de 1984, quando a metodologia foi aplicada num contexto econométrico para previsão de demanda de energia no mercado americano.

Em séries temporais o *bootstrap* pode ser aplicado de duas maneiras: o *bootstrap* nos resíduos do modelo e o *moving blocks* [Souza & Camargo,2004]. No primeiro caso, satisfeita a condição de independência dos resíduos, ajusta-se um modelo probabilístico aos resíduos para que possam gerar novas séries (re)amostradas da amostra original. Portanto, têm-se B séries temporais *bootstrap*. O método *moving blocks* consiste em construir blocos de tamanho M a partir da série original e sortear esses blocos amostrados com reposição agregando-os para formar a amostra *bootstrap*. Repete-se esse passo B vezes, gerando B novas séries *bootstrap*.

O objetivo desta dissertação é, partindo da distribuição densidade de probabilidade dos erros de previsão, através do método de *bootstrap* residual, identificar a distribuição quantílica destes erros e construir os intervalos de confiança sem suposição a cerca da distribuição dos erros.

Os erros de previsão resultam em perdas financeiras para empresas do setor. Como ilustração das consequências do erro de previsão, foi estimado que um incremento de 1% no erro de previsão pode aumentar em 10 milhões de libras os custos de operação, por ano, para uma empresa de eletricidade do Reino Unido [Bunn & Farmer, 1985]. Neste contexto, a previsão de carga, a curto prazo tem um papel fundamental no planejamento de operações confiáveis e seguras.

A motivação para utilizar uma metodologia puramente estatística para previsão e construção dos intervalos de confiança se deve a confiabilidade na interpretação dos parâmetros do modelo estatístico.

1.2 Estrutura deste trabalho

No Capítulo 2 são apresentados diferentes tipos de modelagens de carga, desde os métodos estatísticos clássicos até métodos que utilizam inteligência artificial. No Capítulo 3 são apresentadas as referências relacionadas a previsão com estimativa de intervalos de confiança. No Capítulo 4 são apresentadas as técnicas de (re)amostragem: Simulação Monte Carlo e o *bootstrap* e os conceitos básicos de intervalos de confiança usando *bootstrap*. No Capítulo 5 são apresentadas as análises feitas sobre a série carga, a teoria dos teste utilizados (Jarque Bera, Teste Kolmogorov-Smirnov e Teste Teste qui-quadrado). No Capítulo 6 são apresentados os resultados obtidos. A conclusão do estudo, algumas observações e comentários finais encerram este trabalho, no Capítulo 7.

Capítulo 2

Modelagem de Previsão de Carga

2.1 Introdução

Na literatura de previsão de carga a curto prazo encontram-se propostas baseadas em modelos de regressão, séries temporais, redes neurais artificiais (RNA), sistemas especialistas, lógica *fuzzy* e máquinas de vetores de suporte.

A seguir são apresentados os diferentes tipos de modelagem desde os métodos estatísticos clássicos até métodos que utilizam inteligência artificial.

2.2 Modelagem Estatística

2.2.1 Metodologia Box & Jenkins

A metodologia Box & Jenkins consiste em um modelo univariado, ou seja, que se baseia em uma única série temporal e que parte do princípio de que os modelos podem ser construídos a partir da informação contida nos próprios dados. Assim, será o próprio comportamento da variável que responderá pela sua dinâmica futura. A seguir são apresentados, resumidamente, os modelos.

2.2.1.1 Modelo autorregressivo - AR (*AutoRegressive*)

O modelo autorregressivo é um processo onde o valor de uma variável no instante (t) depende de seu valor no instante anterior ($t - 1$), ($t - 2$) até o instante ($t - p$) mais um ruído aleatório

$a(t)$. Assim tem-se:

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \dots + \phi_p y(t-p) + a(t)$$

onde:

ϕ_p são as constantes da função autorregressiva;

p números de instantes anteriores usados na regressão; e

$a(t)$ é ruído aleatório.

A carga no instante t depende das cargas passadas até o instante $t-p$. Assim, o processo autorregressivo é dito de ordem p .

Definindo-se o operador de retardo, B , tal que:

$$y(t-1) = By(t)$$

e conseqüentemente:

$$y(t-m) = B^m y(t).$$

A equação da carga pode ser dada por:

$$\phi(B)y(t) = a(t),$$

onde:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.1)$$

2.2.1.2 Modelos médias móveis - MA (*Moving Average*)

No processo médias móveis, o valor da carga é uma função linear dos valores atuais e passados de uma série de ruídos brancos $a_t, a_{t-1}, \dots, a_{t-q}$, definido pelos erros de previsão ou resíduos

das observações disponíveis. Assim, tem-se:

$$y(t) = a(t) - \theta_1 a(t-1) - \theta_2 a(t-2) - \dots - \theta_q a(t-q)$$

E, similar ao método autorregressivo pelo operador B , tem-se:

$$y(t) = \theta(B)a(t),$$

onde:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \quad (2.2)$$

sendo θ são os parâmetros do modelo MA de ordem q .

2.2.1.3 Modelos autorregressivo médias móveis - ARMA (*AutoRegressive Moving Average*)

O modelo autorregressivo médias móveis tem ordem (p, q) com p termos autorregressivos, que representam a ordem de defasagens dos termos autorregressivos; e q termos médias móveis, sendo que este representa a ordem de defasagens do termo de média móveis. Assim tem-se:

$$y(t) = \phi_1 y(t-1) + \dots + \phi_p y(t-p) - a(t) - \theta_1 a(t-1) - \theta_2 a(t-2) - \dots - \theta_q a(t-q),$$

e pode-se escrever:

$$\phi(B)y(t) = \theta(B)a(t).$$

2.2.1.4 Modelos autorregressivo integrado médias móveis - ARIMA (*autoregressive integrated moving average*)

A teoria para séries temporais lida com processos estocásticos estacionários, que são séries temporais onde não há uma mudança sistemática na sua média (sem tendências), não há mudanças sistemática na sua variância e sem variações periódicas.

Um processo estacionário, assim definido, satisfaz as seguintes condições:

$$\begin{aligned} E[Z_t] &= E[Z_{t+k}] = \mu \\ \text{Var}[Z_t] &= E[(Z_t - \mu)^2] = \sigma^2 \\ \text{Cov}[Z_t, Z_{t+k}] &= E[(Z_t - \mu).(Z_{t+k} - \mu)] \end{aligned}$$

As duas primeiras condições indicam que a média e a variância de Z_t não variam com o tempo, e a terceira indica que as autocovariâncias não dependem do tempo e sim da distância k que separa as observações.

Um processo estacionário, que se caracteriza por possuir um comportamento geral de sua estrutura probabilística, invariante no tempo, pressupõe que uma realização amostral $x_t, x_{t+1}, \dots, x_{t+n}$, para qualquer t , embora aleatória, apresente a mesma forma geral de suas observações, o que possibilita fazer previsões.

Se o processo estocástico não for estacionário, este pode se tornar estacionário por meio de sucessivas diferenciações da série original. Para isso, diferencia-se a série temporal através de um operador ∇ tal que:

$$\nabla y(t) = y(t) - y(t-1) = (1 - B)y(t)$$

e

$$\nabla^d = (1 - B)^d y(t).$$

Que fornece o modelo:

$$\phi(B)\nabla^d y(t) = \theta(B)a(t),$$

onde:

$$y(t) = \phi_1 y(t-1) + \dots + \phi_p y(t-p) + a(t) - \theta_1 a(t-1) - \theta_2 a(t-2) - \dots - \theta_q a(t-q),$$

O processo acima é um modelo ARIMA com ordem (p, d, q) .

A construção do modelo ARIMA parte da concepção de que as séries temporais envolvidas

na análise são geradas por um processo estocástico estacionário.

2.2.1.5 Etapas da Metodologia Box & Jenkins

O objetivo da metodologia Box & Jenkins é determinar os três componentes que configuram qualquer estrutura a saber: p parâmetros autorregressivos, d processos de diferenciação (integração) e q parâmetros de médias móveis. O processo de construção destes modelos propostos por Box & Jenkins é baseado em um ciclo iterativo, composto dos seguintes estágios: identificação, estimação, análise do modelo e previsão.

2.2.1.6 Identificação do Modelo

A identificação é a etapa onde se estabelecem as características básicas da série, tais como o tipo do processo gerador (estacionário ou não-estacionário); a existência ou não de componentes sazonal e tendencial; e as estimativas iniciais para os valores dos parâmetros envolvidos.

O processo de identificação do modelo ARIMA deve se iniciar com o parâmetro d . Sua definição é muito simples, e pode ser feita pela ótica da estacionariedade, onde o parâmetro d nada mais é do que o grau de diferenciação necessário para tornar os dados estacionários. A definição dos parâmetros deve ser feita através da análise das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) dos dados históricos. A função autocorrelação deriva da função de autocovariância definida por:

$$\gamma_k = cov[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)].$$

Se o processo for estacionário, existirá a igualdade $E[z_t] = E[z_{t+k}] = \mu$. Então, o coeficiente de autocorrelação será dado por:

$$\rho_k = cov[z_t, z_{t+k}]/Var(z_t) = \gamma_k/\gamma_0$$

Sendo assim, ρ_k é uma medida padrão de dependência entre os intervalos de tempo (defasagens), onde, para todo k , $|\rho_k| \leq 1$.

A ordem do modelo $AR(p)$ indica o número de termos da função autocorrelação parcial,

assim:

$$AR(1) = z_t = \hat{\phi}_1 z_{t-1} + a_t \quad (2.3)$$

$$AR(2) = z_t = \hat{\phi}_1 z_{t-1} + \hat{\phi}_2 z_{t-2} + a_t \quad (2.4)$$

$$AR(3) = z_t = \hat{\phi}_1 z_{t-1} + \hat{\phi}_2 z_{t-2} + \hat{\phi}_3 z_{t-3} + a_t \quad (2.5)$$

$$AR(p) = z_t = \hat{\phi}_1 z_{t-1} + \hat{\phi}_2 z_{t-2} + \hat{\phi}_3 z_{t-3} + \dots + \hat{\phi}_p z_{t-p} + a_t \quad (2.6)$$

Logo os p coeficientes de autocorrelação parcial serão dados por $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \dots, \hat{\phi}_p$. Esses coeficientes podem ser calculados a partir das equações acima (2.3),(2.4),(2.5) e (2.6). No entanto, esse calculo não é fácil e tomaria um tempo considerável. Alternativamente, esses valores são calculados utilizando-se os próprios coeficientes de autocorrelação ρ_k . Onde:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \phi_3 \rho_{k-3} + \dots + \phi_p \rho_{k-p}$$

2.2.1.7 Estimação dos parâmetros

Identificada a ordem do modelo ARIMA (p, d, q) , a próxima etapa é estimar os parâmetros ϕ 's e θ 's da equação de regressão. Para estimar os parâmetros autorregressivos (ϕ), que são lineares, pode-se usar mínimos quadrados ordinários. E os parâmetros não lineares de médias móveis (θ) podem ser estimados pelo método da máxima verossimilhança. Resumidamente, o método de mínimos quadrados, consiste na obtenção de parâmetros $(\phi_1, \phi_2, \phi_3, \dots, \phi_p)$ e $(\theta_1, \theta_2, \theta_3, \dots, \theta_n)$ que minimize a soma das diferenças entre os pontos observados na amostra e os pontos estimados pela equação de regressão com os respectivos parâmetros ao quadrado, ou seja, que o somatório dos resíduos (a_t) seja mínimo:

$$S(\phi, \theta) = \min(\sum(a_t)^2) .$$

O método de verossimilhança consiste em selecionar aqueles estimadores que maximizam a probabilidade de estimar valores, a partir da regressão, iguais aos valores realmente obser-

vados na amostra, ou simplificada, encontrar estimadores que gerem valores os mais assemelhados aos valores da amostra.

2.2.1.8 Validação do Modelo

O modelo é definido por exame visual dos gráficos da função autocorrelação e da função autocorrelação parcial da série temporal. Os correlogramas podem sugerir a existência de algumas imperfeições nos dados e também auxiliam na verificação da existência de estacionariedade ou não da série. Examinando os gráficos da FAC e da FACP, seja da própria série ou da série diferenciada, é possível identificar os parâmetros p e q do modelo ARIMA (p, d, q) . A figura (2.1) mostra o correlograma da função autocorrelação da série de carga com uma diferença. O correlograma indica um AR de ordem 1.

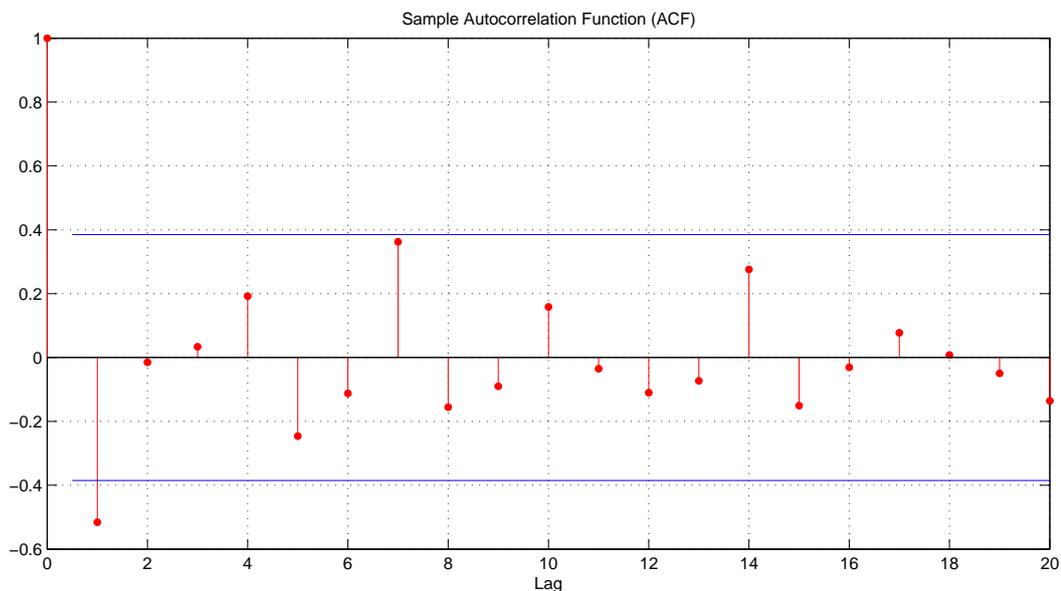


Figura 2.1: Função de autocorrelação

Segundo [Morettin, 2006] uma ideia fundamental no modelo Box & Jenkins é o princípio da parcimônia. De acordo com este princípio quanto menos parâmetros puderem ser adicionados ao modelo, melhor será a qualidade da regressão e, conseqüentemente, das estimativas. Pressupõem que ao se aumentar a quantidade de parâmetros, se aumenta também a possibilidade de erros.

2.2.2 Modelos sazonal autorregressivo integrado médias móveis - SARIMA (Seasonal Integrated AutoRegressive Moving Average)

Os modelos Box & Jenkins analisam a correlação existente entre as observações da série em um determinado instante do tempo e seus antecessores, mas existem observações que apresentam correlação com observações de defasagem superior a 1. Por exemplo, a curva de carga de uma quarta-feira pode esta relacionada com o que aconteceu na observação do dia anterior, mas ela também pode estar relacionada a quarta-feira da semana anterior, pois o mesmo dia da semana apresenta um padrão no perfil de carga. Os modelos que contemplam as séries que apresentam autocorrelação sazonal são conhecidos como SARIMA.

O modelo SARIMA $(p, d, q)(P, D, Q)$ é composto por uma parte não sazonal (ARIMA) com parâmetros (p, d, q) e uma parte sazonal com parâmetros (P, D, Q) , denotado pela equação:

$$\phi(B)\Phi(B^m)\nabla^d\nabla_m^D Z_t = \theta(B)\Theta(B^m)a_t,$$

onde:

$\Phi(B^m) = 1 - \Phi_1(B_1^m) - 1 - \Phi_2(B_2^{2m}) - \dots - 1 - \Phi_p(B_p^{pm})$ é o operador autorregressivo sazonal de ordem p .

$\Theta(B^m) = 1 - \Theta_1(B_1^m) - 1 - \Theta_2(B_2^{2m}) - \dots - 1 - \Theta_q(B_q^{qm})$ é o operador de médias móveis sazonal de ordem Q .

O operador $\nabla_m^D Z_t = Z_t - Z_{t-m}$ onde:

$\nabla_m^D = (1 - B^m)^D$, onde:

D = número de diferenças sazonais;

m = período sazonal.

Para identificação, verificação e estimação de modelos sazonais não há, em princípio, nenhuma dificuldade adicional. A diferença é que temos que distinguir a série em relação ∇ e ∇_m a fim de produzir estacionariedade. Com isso obtemos os valores de d e D , que na maioria das vezes assumem valores iguais no máximo, 2 [Morettin, 2006].

A obtenção das ordens p, P, q e Q é feita, como no modelo não sazonal, através das funções de autocorrelação e autocorrelação parcial da série diferenciada.

2.2.3 Modelos em duas Etapas

Existem diferentes maneiras de realizar modelagem de carga em dois níveis. [Ramanathan et al., 1997] propõem um modelo em duas etapas e decompõe a carga em quatro tipo de variáveis: determinística, temperatura, carga básica e erros. As horas são modeladas separadamente para remover o ciclo diário.

No modelo TLSAR (sigla para *Two-Level Seasonal AutoRegressive*) proposto em [Soares, 2003], o primeiro componente representa as tendências, o ciclo anual e os efeitos dos diferentes tipos de dia. O segundo componente captura as possíveis dependências de horas próximas e descreve a dinâmica de curto prazo. Um resumo do modelo TLSAR é apresentado como base para a posterior construção de intervalos de confiança, usando os erros de previsão.

2.2.3.1 Modelo Matemático TLSAR

Seja um modelo sazonal autorregressivo em dois níveis, definido como:

$$L_{h,d} = L_{h,d}^P + L_{h,d}^I, \quad (2.7)$$

onde a “carga potencial”, $L_{h,d}^P$, da hora h do dia d é definida como uma soma da tendência e da sazonalidade presente nas séries de curvas de cargas, dada por:

$$L_{h,d}^P = T_{h,d} + S_{h,d}, \quad (2.8)$$

onde:

$T_{h,d}$ é a componente de tendência, determinística ou estocástica;

$S_{h,d}$ é o um termo sazonal que representa os efeitos dos dias da semana e estações do ano.

O segundo componente, “carga irregular”, é definido por um processo autorregressivo que pode ser linear ou não linear com a seguinte equação:

$$L_{h,d}^I = f(L_{h,d-1}^I + \dots + L_{h,d-p}^I + u_{h,d}; \Psi), \quad (2.9)$$

onde:

$f(\cdot; \Psi)$ é uma função autorregressiva indexada por um vetor de parâmetros Ψ ;

$u_{h,d}$ é o termo de erro.

A carga irregular é modelada por um modelo autorregressivo (*AR*). São construídos 24 modelos de previsão, um modelo para cada hora do dia.

As defasagens do modelo *AR* e o número de harmônicas, que captura os efeitos sazonais, são determinados pelo Critério de Informação Bayesiana de Schwarz-SBIC [Schwarz, 1978], definido por:

$$SBIC = \ln(\hat{\sigma}^2) + p \frac{\ln(n)}{n} \quad (2.10)$$

onde :

$\hat{\sigma}^2$ é o estimador de máxima verossimilhança da variância residual;

p é o número de parâmetros estimados no modelo, incluindo o termo constante; e

n é o número de observações da série.

Outro critério largamente utilizado para seleção é o Critério de informação de Akaike, mas como este critério não apresentou diferença significativa que alterassem o processo de seleção dos parâmetros, somente o SBIC foi usado para a seleção das defasagens do modelo **AR**. O número de harmônicas, representada por uma soma de senos e cossenos que capturam os efeito sazonais e outros relativos às estações do ano e comportamentos sociais.

Para avaliar a previsão dos modelos são analisados o MAPE (erro médio absoluto percentuai), dado por :

$$MAPE = \frac{1}{N} \sum_1^N \left[\frac{L_{h,d} - \hat{L}_{h,d}}{L_{h,d}} \right] \times 100\% \quad (2.11)$$

onde:

N é o número de valores previstos;

$L_{h,d}$ é o valor da carga da hora h do dia d ;

$\hat{L}_{h,d}$ é o valor previsto da carga da hora h do dia d .

2.3 Modelagem com Inteligência Artificial

As técnicas de inteligência artificial tem evoluído nas últimas décadas e assim tem surgido diversas técnicas de construção de modelos não lineares a partir de dados reais. A seguir

serão abordados, resumidamente, três destas técnicas: as redes neurais, sistemas especialistas e lógica *nebulosa*.

As redes neurais são modelos computacionais inspirados no sistema nervoso dos seres vivos e tem como principal característica simular o seu funcionamento. São compostas por estruturas interligadas denominadas neurônios. No processo de previsão a rede neural é treinada para um problema específico, essa etapa é conhecida como aprendizagem e então é capaz de prever valores futuros com base em informações passadas.

Os sistemas especialistas são sistemas que tentam imitar o raciocínio de um especialista. A ideia é traduzir as previsões intuitivas em pensamento analógico para definir os passos da lógica formal [Rahman & Bhatnagar,1998]. Uma dificuldade no seu uso é que se necessita de um especialista capaz de prever a carga e assim converter o seu conhecimento em regras matemáticas, processo que pode ser extremamente difícil.

A aplicação da lógica *nebulosa* à previsão de carga se dá por esta ser capaz de aproximar uma função não linear com exatidão e de encontrar padrões em grandes conjuntos de dados.

2.4 Resumo

São encontradas na literatura diferentes abordagens de previsão de carga a curto prazo como: modelos de regressão, séries temporais, redes neurais artificiais, sistemas especialistas, lógica *nebulosa* e máquinas de vetores de suporte.

Neste trabalho é utilizada a modelagem apresentada em [Soares, 2003]. O modelo é composto por dois componentes: o primeiro componente, “carga potencial” e o segundo componente “carga irregular”. O componente de “carga potencial” é modelado como a soma da sazonalidade e das tendências presentes na carga. A “carga irregular” é modelada por um modelo autorregressivo (**AR**) de ordem 7. A estratégia de modelagem considera os valores observados para cada hora, tratando cada série separadamente, resultando num total de 24 modelos. Para capturar a sazonalidade anual usa-se uma soma de senos e cossenos para modelar a sazonalidade semanal (efeito dos dias da semana e de dias especiais, como os feriados) são usadas variáveis *dummy*. O segundo componente, “carga irregular” é definido pelos fatores que representam as variáveis de influência lenta.

Capítulo 3

Pesquisa Bibliográfica

3.1 Introdução

Neste capítulo, a literatura existente sobre a modelagem para previsão e construção de intervalos de confiança utilizando estatística clássica e inteligência artificial entre outros métodos é analisada e comentada. As referências relacionadas a previsão com estimativa de intervalos de confiança são aqui revistas e resumidas. Essa literatura oferece antecedentes para o restante do trabalho.

3.2 Métodos Estatísticos

Os métodos estatísticos recorrem a um modelo matemático que represente a carga elétrica como uma função de diferentes fatores, tais como: tempo, condições meteorológicas, em especial a temperatura e classes de consumidores. Entre os métodos estatísticos os mais usados são os métodos de regressão e as séries temporais.

[Ramanathan et al., 1997] propõem um modelo para cada hora, totalizando, 24 modelos, variante no tempo. Nesta modelagem, também, são incluídos os efeitos dos dias atípicos como finais de semana e feriados.

[Soares, 2003] propõe uma metodologia de previsão de carga baseada na construção de um modelo em dois níveis, estimando parâmetros para cada hora do dia. A estrutura do modelo é composta por séries formadas por observações horárias de carga, como soma de dois componentes. O primeiro componente, “carga regular”, é composto por variáveis determinísticas

como: hora do dia, dia da semana, etc. O segundo componente, “carga irregular”, modela as variáveis de influência lenta. Como é adotada a modelagem hora por hora, sendo tratadas separadamente as previsões para cada hora, são construídos 24 modelos de previsão. O diferencial dessa modelagem é que os erros de observações passadas não são utilizados para gerar futuras previsões. [Cancelo et al.,2008] propõem uma adaptação deste modelo incorporando informações climáticas.

[Engle et al., 1992] apresentaram vários modelos de regressão para a previsão de um dia à frente do pico de carga. Seus modelos incorporam influências determinísticas: feriados e influências estocásticas: carga média e influência exógena: condições meteorológicas.

[Fan & Donald, 1994] descrevem a implementação dos modelos ARIMAX (autorregressivo integrado de média móvel com entradas exógenas) para previsão de carga.

[Petiau, 2009] adota a modelagem em dois níveis para previsão onde o primeiro componente captura a sensibilidade climática (mudanças bruscas de temperatura) por uma função de transferência não linear e não estacionária. O segundo componente é modelado através de um modelo SARIMA que reflete fatores endógenos como a sazonalidade diária e semanal. “

3.3 Inteligência artificial

O uso de técnicas de inteligência artificial tem sido largamente utilizado para previsão de carga a curto prazo. Os trabalhos usando redes neurais artificiais tem tido ampla aplicação desde a década de 1990.

[Andrade & Silva, 2009], realizaram uma análise comparativa de previsões entre seguintes modelos: ARIMA, redes neurais artificiais e o sistema de inferência neuro-fuzzy adaptativo (ANFIS). A série de demanda de carga é composta por dados medidos em intervalos de cinco em cinco minutos, durante sete dias da semana em subestações localizadas em Andradina, Ubatuba e Votuporanga. Para seleção dos modelos ARIMA, inicialmente foi feita a estacionarização das séries com aplicação de duas diferenciações. Os modelos foram selecionados através da função de autocorrelação, função de autocorrelação parcial e testes de *Ljung Box* e *Box-teste de Pirce*. São geradas previsões doze passos à frente utilizando um ambiente de desenvolvimento baseado na linguagem de programação *R*. O modelo de redes neurais

artificiais foi desenvolvido, com arquitetura MLP (*Multilayer Perceptron*) formada por cinco neurônios de entrada, três neurônios na camada oculta e um neurônio de saída. Foi utilizado um sistema adaptativo *neuro-fuzzy* com cinco funções de entrada e três funções de associação e geradas as previsões para esses modelos. O MAPE das previsões, geradas pelas redes neurais artificiais, apresentou-se melhor em todas as três subestações. O modelo ANFIS, demonstrou um bom resultado nas primeiras etapas mas o crescimento foi expressivo a partir da terceira previsão.

[Chyssolouris et al., 1996] desenvolveram um método para estimar intervalos de confiança que considera uma distribuição normal para os erros em lugar da covariância para as saídas. Este método difere de outros métodos existentes, devido ao fato de não serem necessárias informações sobre as segundas derivadas da saída da rede neural. Os autores utilizaram um conjunto de 279 observações que foram divididas em duas partes. A primeira parte é formada por 225 observações utilizadas para o treinamento e as 54 restantes para teste. A rede é constituída por um algoritmo de retro propagação (*backpropagation*) com 5 neurônios na camada de entrada, 8 na camada oculta e 1 na camada de saída. Para um nível de confiança de 80%, oito dos 54 valores previstos estão fora do intervalo de confiança, logo 85% dos valores estão contidos no intervalo. A rede neural 5-8-1 também foi treinada para adicionar aos dados de saída o viés de 0.0183, estando 78% dos valores, neste caso, contidos no intervalo. A rede neural treinada com ou sem o ruído nos dados, para um nível de 80% de confiança, não inclui previsões abaixo desse intervalo.

[Townsend, 1999] analisa o problema de precisão de estimativa das saídas da rede neural através de um modelo de perturbação. As fontes de ruído, modelado inicialmente, estão no vetor de entrada e nos pesos. O modelo de perturbação é aplicado a redes de bases de funções radiais. A rede neural RBF (redes de funções de bases radiais) usada, possui 100 neurônios na camada oculta e 10000 neurônios uniformemente distribuídos por todo espaço de entrada. O mapa auto-organizável de *Kohonen* é utilizado para o treinamento, onde são definidas as posições centrais e seguidas por inversão de matriz para determinação da segunda camada. São usados 10000 pares de dados (entrada e saída) para o treinamento e para o teste.

3.4 Intervalos de Confiança

[Petiau, 2009] emprega os quantis empíricos para construir intervalos de confiança, utilizando os erros relativos e compara com intervalos de confiança construídos por um método básico em que a suposição de gaussianidade na distribuição dos erros é admitida. São consideradas, também, as classificações dos dias: dias normais, dias de férias, dias que antecedem feriados, feriados e dia após um feriado.

Os limites superior e inferior do intervalo, construídos por quantis, tem a seguinte forma:

$$\begin{cases} L_{d_0, h | (d_0-1)}^{sup, \alpha} &= \widehat{L}_{d_0, h | (d_0-1)} (1 + q^{1-\frac{\alpha}{2}} (E\{d_0\})) \\ L_{d_0, h | (d_0-1)}^{inf, \alpha} &= \widehat{L}_{d_0, h | (d_0-1)} (1 + q^{\frac{\alpha}{2}} (E\{d_0\})) \end{cases}$$

Sendo o procedimento independente para cada hora do dia, o tamanho do intervalo de confiança depende dos valores previstos. No método de referência para cada hora h supõe que o erro absoluto é independente e identicamente distribuído. Quando comparada as características (cobertura, largura e capacidade de diferenciar situações) dos intervalos de confiança dos dois métodos a um nível de confiança **95%** o método proposto obteve os melhores resultados. As coberturas dos métodos encontram-se entre **79%** e **86%** com p-valores diferentes para o modelo básico e o modelo construído pelo método adaptativo. O autor ressalta que em muitos casos, coberturas empíricas são próximas do nível de confiança esperado para intervalos de confiança construídos pelo método adaptativo e valores altos do p-valor permitem não rejeitar a hipótese nula.

[Silva & Moulin, 2000] utilizam três técnicas para calcular de intervalos de confiança. As técnicas são: saída de erro, (re)amostragem dos erros e regressão multilinear adaptada para redes neurais. O problema analisado é a previsão de cargas a curto prazo. A previsão das cargas é obtida com o auxílio de redes de *perceptrons* de múltiplas camadas. A rede neural adaptada para a regressão multilinear encontra intervalos de confiança relativo de 0.46, os intervalos gerados para previsões obtidas através do erro de saída são de 0.28 e 0.29 são os intervalos calculados para previsões geradas por (re)amostragem .

[Zang et al., 2003] estudam intervalos de confiança para a previsão gerada por uma rede neural em cascata utilizando uma base de inferência bayesiana. Este método considera os

ruídos dos pesos, os ruídos dos dados de entrada e os ruídos gerados no processo de previsão. Neste método a distribuição de saída é aproximada para uma distribuição gaussiana. A variância da saída pode ser calculada pelo método *Quasi-Newton* sem memória. O método *Quasi-Newton* sem memória é rápido e com boas características computacionais. As estimativas de previsão e os intervalos de confiança foram implementados com C++ em um *Pentium* III 500 MHz. Os dados são do operador sistema inglês, referentes ao período de 1º de maio de 1999 ao final de novembro de 2000. Os resultados do erro médio percentual absoluto das previsões são comparados com os resultados do operador sistema inglês. São obtidos pelos autores um erro médio percentual absoluto (MAPE) de 8,8% e pelo operador sistema inglês um MAPE de 9,73%. Os autores também encontram intervalos de confiança com uma cobertura de 66,6% que é cerca de 68% da cobertura gaussiana, já que é assumida uma distribuição gaussiana para a saída.

[Rivals & Personnaz, 2000] apresentam resultados teóricos da construção de intervalos de confiança para uma regressão não linear, baseada na estimação por mínimos quadrados ordinários e utilizando a expansão linear de Taylor da saída correspondente ao modelo não linear. A metodologia desenvolvida é aplicada a um modelo de rede neural. Os autores mostram, para um problema real analisado e simulado, que a expansão linear de Taylor além de fornecer um intervalo de confiança em qualquer ponto, fornece também, uma ferramenta para detectar super ajuste (*overfitting*). A metodologia foi aplicada a um conjunto de dados de 69 valores onde a suposição de homocedasticidade é considerada. O treinamento foi feito com redes neurais com um neurônio de saída e uma camada oculta. O número ideal de neurônios da camada oculta é selecionado com base no MSPE (erro quadrado médio de desempenho). O fato da proporção correspondente ao quociente do MSPE (erro quadrado médio de desempenho)/MSTE (erro médio de formação) está próximo de 1, indica que n é suficientemente grande, então a família de redes neurais selecionada contém uma boa aproximação da regressão. Os candidatos com mais de três neurônios na camada oculta não foram aprovados pois apresentaram MSPE (erro quadrado médio) igual ou superior 1.7×10^{-1} . E quociente MSPE (erro quadrado médio de desempenho)/MSTE (erro médio de formação) igual ou superior a 1.1×10 . Sendo MSPE ideal para o conjunto de dados (N=69 valores) de 2.1×10^{-2} e com dois neurônios na camada oculta. A proporção correspondente ao quociente MSPE)/MSTE é igual a 1,3 indicando que $n=2$ é grande o suficiente, para a família

de redes neurais selecionada. O intervalo de confiança com um nível de significância de 95% é calculado usando a suposição que a distribuição segue uma distribuição *t-Student*.

[Bartkiewicz, 2000] propõem o cálculo de intervalo de confiança para previsões de carga a curto prazo. As previsões horárias para dois dias a frente são obtidas com rede neural. O modelo é formado por 24 equações, um modelo para cada hora do dia. A rede é constituída por uma arquitetura MLP (*Multilayer Perceptron*) com 13 neurônios na camada de entrada 10 neurônios na camada oculta e 1 neurônio na camada de saída. Para o treinamento da rede neural são usadas observações do ano de 1996 e o teste o período das observações é relativo a 1º de janeiro de 1997 à 1º de janeiro de 1998. Os feriados e os domingos foram descartados no treinamento, sendo utilizados apenas os dias de semana, incluindo o sábado. Por exemplo as previsões feitas para as 13 horas e para as 15 tiveram erros elevados devido a flutuações inesperadas da carga. O autor diz que este fato é relativamente raro e consequência da aleatoriedade da demanda de energia desta empresa de previsão. Ressalta que esse é um problema comum na Polônia onde ocorreu alterações inesperadas nas condições climáticas, na homogeneidade do dia e na instabilidade dos consumidores.

[Guan et al., 2010] apresentam uma proposta para previsão e cálculos de intervalos de confiança utilizando redes neurais, baseado em ondaletas (*wavelet*) treinadas por algoritmos híbridos de *Kalman*. São utilizados dados em intervalos de cinco em cinco minutos para produzir previsões para a próxima hora ao mesmo tempo que são estimados os intervalos de confiança. Para capturar características das componentes da carga, das componentes de baixa frequência e das componentes de alta frequência, a rede neural foi treinada de duas formas: utilizando filtros de *Kalman* estendido EKFNN (*Extended Kalman Filter Neural Network*) para as previsões das componentes de baixa frequência e filtro de *kalman un-scented* UKFNN (*Unscented Kalman Filter Neural Network*) para as componentes de alta frequência, pois as componentes de baixa frequência são aproximadamente lineares e as componentes de alta frequências são não-lineares. Os erros têm distribuição normal com média $\mu = 2 \times 10^{-3}$ e variância 10^{-6} respectivamente. Os intervalos de confiança são obtidos derivando a variação do desvio padrão. Os autores utilizam um conjunto de dados referentes ao período compreendido entre janeiro e dezembro de 2008 para analisar os resultados obtidos pelo método MWNNHK (*Multilevel Neural Networks Hybrid kalman*). O período de outubro de 2006 a dezembro de 2007 é utilizado para o treinamento da rede neural.

3.5 Resumo

Nesse capítulo são apresentados alguns métodos de previsão de carga, são eles: métodos estatísticos e inteligência artificial. Entre os métodos estatísticos os mais utilizados são os métodos de regressão e os métodos baseados em séries temporais. Os modelos estatísticos tem a vantagem de serem técnicas matemáticas já bem desenvolvidas e estudadas.

Os métodos de previsão baseados em redes neurais, frequentemente utilizam uma arquitetura MLP com algoritmo de aprendizagem. Os intervalos de confiança são construídos assumindo uma distribuição gaussiana para os erros.

Capítulo 4

Intervalos de Confiança

4.1 Introdução

Os intervalos de confiança se apresentam como um quantificador da confiabilidade da previsão. A construção de intervalos de confiança sintetiza as incertezas decorrentes da estimação de parâmetros a partir de amostras finitas. A relevância da análise de algumas das particularidades da construção de intervalos de confiança se dá pela importância das informações que podem ser extraídas na abordagem de aspectos práticos da previsão de carga, minimizando os riscos de despacho na operação e auxiliando a tomada de decisão pelo operador do sistema.

Segundo [Portela et al., 2011] a estimação de intervalos para um quantil, \mathbf{X}_F , associado a uma probabilidade avaliada de acordo com um dado modelo com um número k de parâmetros, $\mathbf{F}_x(\mathbf{x}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k)$, em que $\boldsymbol{\theta}_k$ designam os parâmetros, consiste em identificar os limites inferior, λ_i , e superior, λ_s , tais que tornam válida a equação:

$$P(\lambda_1 \leq \mathbf{X}_F \leq \lambda_2) = 1 - \alpha \quad (4.1)$$

no qual $(1 - \alpha)$ denota o nível de confiança.

Como \mathbf{X}_F é uma grandeza populacional e não uma variável aleatória, a interpretação correta da equação (4.1) é: a probabilidade do intervalo $[\lambda_i, \lambda_s]$ conter o verdadeiro, embora desconhecido, valor de \mathbf{X}_F é $100(1 - \alpha)\%$. A equação reflete as incertezas decorrentes da estimação dos parâmetros $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k$ e, conseqüentemente, de \mathbf{X}_F , frente às flutu-

ações amostrais, não incluindo as incertezas associadas à seleção da lei de probabilidades $F_x(\mathbf{x}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k)$. A princípio, supomos ser verdadeira a função de distribuição de probabilidades que governa o comportamento da variável aleatória \mathbf{X} .

As previsões das séries de cargas geradas para sete passos à frente são calculadas de acordo com a proposta apresentada por [Teräsvirta et. al.,2005]. De acordo com a equação(2.9) a previsão da componente irregular um passo à frente denotada por $L_{h,d+1}^I$ será dada por:

$$\widehat{L}_{h,d+1|d}^I = f(L_{h,d}^I, \dots, L_{h,d-p+1}^I; \widehat{\Psi}_d) \quad (4.2)$$

onde ψ_d indica que as estimativas dos parâmetros são obtidas usando observações que antecedem o dia d . Para dois passos a frente a previsão, denotada por $L_{h,d+2}^I$, será obtida pela equação:

$$\widehat{L}_{h,d+2|d}^I = \int_{-\infty}^{\infty} f(\widehat{L}_{h,d+1}^I | d + u_{h,d+1}, L_{h,d}^I, \dots, L_{h,d+2}^I; \widehat{\Psi}_d) du_{h,d+1}. \quad (4.3)$$

Para encontrar a previsão pontual para horizontes mais distantes seria necessária a resolução de uma integral multidimensional. No entanto a integração numérica da equação (4.3) pode ser evitada pela aproximação da integral por simulação ou utilizando técnica de *bootstrapping* nos resíduos. A última alternativa requer supor que os erros da equação (2.9) sejam independentes. Foram produzidas as previsões das cargas $L_{h,d+1}^I, L_{h,d+2}^I, \dots, L_{h,d+7}^I$. A descrição completa e detalhada do processo pode ser encontrada em [Soares & Medeiros,2005].

4.2 (Re)amostragem

A técnica de(re)amostragem surgiu em meados de 1935 e consiste em sortear com reposição dados pertencentes a uma amostra, de modo a formar novas amostras. As técnicas de (re)amostragem são úteis em especial quando o cálculo de estimadores por métodos analíticos é complicado. Por exemplo, quando desejamos construir intervalos de confiança para um determinado parâmetro da população, em que a normalidade não é verificada. Será necessário, então, encontrar a distribuição empírica dos dados, mediante (re)amostragem. Os intervalos de confiança para quaisquer parâmetros da população são construídos com base nas distribuições amostrais, que por sua vez dependem da forma da distribuição da

população.

Existem diferentes técnicas de (re)amostragem sendo mais largamente usada na literatura a Simulação Monte Carlo e o *bootstrap*. Neste trabalho será utilizado o *bootstrap* para estimar a distribuição de quantis.

4.2.1 Método *Bootstrap*

O bootstrap é um método de (re)amostragem por simulação de dados introduzido por [Efron, 1979]. Este método de simulação baseia-se na construção de distribuições amostrais por (re)amostragem com reposição. Essa técnica visa à obtenção de estimativas pontuais ou intervalares empírica para os estimadores dos parâmetros de interesse, através da (re)amostragem do conjunto de dados original [Carrasco, 2002].

Quando se tem alguma suposição sobre a distribuição dos dados usa-se o método conhecido como *bootstrap* paramétrico. No caso de não se conhecer a distribuição usa-se o método conhecido como *bootstrap* não-paramétrico.

A ideia básica da técnica *bootstrap* é tratar a amostra original como se fosse a população de interesse e retirar várias amostras com reposição, ou seja, (re)amostrar a amostra original com reposição e para cada (re)amostra, calcular a estimativa de interesse. Ao final da (re)amostragem tem-se várias estimativas distintas para cada (re)amostra, porque se faz amostragem com reposição em cada caso, criando assim, uma aproximação da população .

4.2.1.1 Intervalo de Confiança *Bootstrap*

4.2.1.2 Intervalo de Confiança *Bootstrap* padrão

O intervalo *Bootstrap* padrão para a média μ de uma distribuição normal com σ conhecido, onde θ é o estimador de \bar{X} .

Como $z_{\frac{\alpha}{2}}\sigma\sqrt{n}$ é o $100(1 - \frac{\alpha}{2})$ percentil da distribuição $\bar{X} - \mu$ e $-z_{\frac{\alpha}{2}}\sigma\sqrt{n}$ é o $100(\frac{\alpha}{2})$ percentil dessa distribuição. Logo, o intervalo de confiança de $100(1 - \alpha)$ será dado por:

$$P(100(\frac{\alpha}{2})\text{percentil} \leq \bar{X} - \mu \leq 100(1 - \frac{\alpha}{2})\text{percentil}) = 1 - \alpha \quad (4.4)$$

O intervalo de confiança *bootstrap* com cobertura de $100(1 - \alpha)$ é dado por:

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{n} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{n} \quad (4.5)$$

4.2.1.3 Intervalo de Confiança *Bootstrap* percentil

O intervalo de confiança percentil é o mais simples e o mais difundido entre os intervalos de confiança *bootstrap*. O uso desse intervalo baseia-se na tentativa de aproximar os percentis da distribuição de um estimador usando percentis gerados por *bootstrap*.

Realizando as replicações \mathbf{x}^{*b} de $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ e estimando a estatística *bootstrap* de interesse $\hat{\boldsymbol{\theta}}^*$, o intervalo construído pelo método percentil com $(1 - \alpha)$ de probabilidade de cobertura é obtido pelos percentis (α) e $(1 - \alpha)$ de \hat{F} , a função distribuição acumulada de $\hat{\boldsymbol{\theta}}^*(\mathbf{b})$. A equação para esse intervalos de confiança é dada por:

$$(L_I(\mathbf{x}^*); L_S(\mathbf{x}^*)) = (\hat{F}_\alpha^{-1}; \hat{F}_{1-\alpha}^{-1}) \quad (4.6)$$

Como $\hat{F}^{-1}(\alpha) = \hat{\boldsymbol{\theta}}_\alpha^*(\mathbf{b})$ o percentil $100 - \alpha$ de $\hat{\boldsymbol{\theta}}_\alpha^*(\mathbf{b})$ a equação pode ser escrita como:

$$(L_I(\mathbf{x}^*); L_S(\mathbf{x}^*)) = (\hat{\boldsymbol{\theta}}_\alpha^*; \hat{\boldsymbol{\theta}}_{1-\alpha}^*) \quad (4.7)$$

Assim, o intervalo consiste na porção central de tamanho $(1 - \alpha)$ da distribuição de $\boldsymbol{\theta}^*(\mathbf{b})$. Segundo [Efron & Tibshirani, 1993] a equação refere-se à situação ideal em que as replicações *bootstrap* são infinitas. Como na prática usa-se um números finitos de replicações, 1000 replicações proporcionam bons resultados para o cálculo das estimativas *bootstrap*.

4.2.1.4 Intervalo de Confiança *t-Bootstrap*

O método *t-Bootstrap* segundo [Efron & Tibshirani, 1993] é uma generalização do usual método *t* de *Student*, aplicável, particularmente em estatísticas como a média amostral, a mediana ou percentis amostrais.

Este método constrói os intervalos sem que se faça nenhuma consideração sobre normalidade da distribuição. O procedimento estima a distribuição estatística diretamente dos dados. São geradas B amostras $\mathbf{x}_1^*, \mathbf{x}_2^* \dots \mathbf{x}_B^*$, independentes, estima-se a estatística teste $\boldsymbol{\theta}^*(\mathbf{b})$,

$b = 1, 2, \dots, B$, e encontra-se:

$$z^{*b} = \frac{\theta^*(b)}{\widehat{ep}(\theta^*(b))} \quad (4.8)$$

Em que $\widehat{ep}(\theta^*(b))$, $b = 1, 2, \dots, B$. O α -ésimo percentil de z^{*b} é estimado por $t_{(\alpha)}$, tal que:

$$\sum_{b=1}^B \frac{z^*(b) \leq \widehat{t}_\alpha}{B} = \alpha \quad (4.9)$$

Quando o estimador do desvio-padrão não estiver disponível, então realiza-se (re)amostragem para estimar este erro padrão, para cada uma das B amostras *bootstrap*. Assim, o número de amostras *bootstrap* necessárias para a construção do intervalo de confiança *t-bootstrap* é dado pelo produto de B pelo número de novas amostras *bootstrap* requerida para cada estimação do $\widehat{ep}(\theta^*(b))$. Para a aplicação do algoritmo de *bootstrap* neste trabalho usou-se a base estrutural do *t-bootstrap* mas sem supor que a distribuição dos resíduos segue uma distribuição *t-Student*. Os intervalos de confiança são construídos de acordo com as seguintes etapas abaixo :

1. dada uma amostra aleatória, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, calcula-se $\widehat{\theta}$;
2. amostre com reposição a amostra *bootstrap* $\mathbf{x}^{*b} = (\mathbf{x}_1^{*b}, \mathbf{x}_2^{*b}, \dots, \mathbf{x}_w^{*b})$.
3. calcula-se a mesma estatística usando as amostras geradas no passo 2 para $\widehat{\theta}^{*b}$.
4. use a amostras *bootstrap* \mathbf{x}^* para calcular o $\widehat{SE}\theta^{*b}$. Isso pode ser calculado usando a fórmula ou estimando por *bootstrap* ;
5. calcula-se o z^{*b} usando as informações contidas nos passos 3 e 4;
6. repete-se as etapas de 2 a 5, B vezes, onde $B \geq 1000$.
7. ordena-se os z^{*b} valores em ordem crescente. Estima-se os quantis $\widehat{t}^{(1-\alpha/2)}$ e $\widehat{t}^{(1-\alpha/2)}$
8. estima-se o erro padrão $\widehat{SE}_{\widehat{\theta}}$ usando as B replicações de $\widehat{\theta}^*$ (do passo 3).
9. os intervalos de confiança são construídos,

$$(\widehat{\theta} - z^{(1-\frac{\alpha}{2})} \mu_{\widehat{\theta}}, \widehat{\theta} - z^{(\frac{\alpha}{2})} \mu_{\widehat{\theta}}) \quad (4.10)$$

O método aqui descrito para gerar intervalos de confiança considera parâmetros vindos de populações de tamanho infinito ou grandes o suficiente para que possam ser tratadas como de tamanho infinito.

4.2.2 Método Monte Carlo

Uma alternativa para associar intervalos de confiança a quantis é o método de simulação de Monte Carlo. Esse método é apropriado para resolver problemas de grande escala e muitas vezes, é usado para calcular o valor esperado de uma variável que é função de várias variáveis estocásticas, que não podem ser tratadas analiticamente.

O método em si já era conhecido há séculos, mas passou a ser efetivamente utilizado somente nas últimas décadas. Pelo fato das simulações por esse método estarem baseadas em eventos que ocorrem aleatoriamente e, ainda, por terem uma similaridade com jogos de azar, durante o Projeto Manhattan na Segunda Guerra Mundial, Ulam e Von Neumann denominaram "Monte Carlo", fazendo referência à famosa cidade de Mônaco conhecida mundialmente como a capital dos jogos de azar.

O método de Monte Carlo pode ser usado como uma alternativa para construção de intervalos de confiança para quantis. A partir do conjunto de estimativas dos parâmetros de $F_X(x|\theta_1, \theta_2, \dots, \theta_k)$, obtém-se M (com M muito grande) amostras com a mesma dimensão da amostra original. Essas amostras serão ordenadas para uma dada probabilidade de não excedência de, F , e utilizadas no cálculo das M curvas de quantis, permitindo a construção do correspondente intervalo de confiança $100(1 - \alpha)\%$ para F_X .

1. determina-se a pseudo-população ou modelo que representa a verdadeira população de interesse;
2. (re)amostra-se a pseudo-população;
3. calcula-se a estatística de interesse;
4. repete-se os procedimento 2 e 3 M vezes;
5. usa-se os M valores encontrados no passo (4) para estudar a distribuição da estatística e depois de ordenadas para uma dada probabilidade, permitem a construção do correspondente intervalo de confiança para a distribuição original, até então desconhecida.

4.3 Resumo

Neste capítulo são abordadas as técnicas de (re)amostragem para o cálculo dos intervalos de confiança. São elas: *bootstrap* e Monte Carlo. (Re)amostrar consiste em sortear com reposição, dados pertencentes a uma amostra, de modo a formar uma nova amostra. São úteis quando o cálculo de estimadores por métodos analíticos for complicado ou até mesmo impossível.

O *bootstrap* é um método em que são realizadas (re)amostragens (com reposição) dos dados de uma amostra de tamanho n finito, tentando aproximar a distribuição das (re)amostragens a distribuição empírica dos dados. A estimativa numa amostra *bootstrap* se aproxima do valor real quando o número de (re)amostragem tende a infinito, ou seja, quando as (re)amostragens são realizadas várias vezes, sendo que o valor de B é escolhido de acordo com a finalidade para a qual o método *bootstrap* está sendo usado.

O método de simulação de Monte Carlo é apropriado para resolver problemas de grande escala e muitas vezes, é usado para calcular o valor esperado de uma variável que é função de várias variáveis estocásticas, que não podem ser tratadas analiticamente. Além de representar uma alternativa simples e de fácil implementação, tendo em vista os recursos computacionais atualmente disponíveis, a construção de intervalos de confiança utilizando a simulação de Monte Carlo não presume que a distribuição de amostragem seja normal. De fato, a técnica permite a construção da distribuição empírica dos quantis a partir de amostras das séries obtidas por simulação.

Capítulo 5

Experimento

5.1 Os dados

São utilizados dados de uma série de observações horárias de uma concessionária de distribuição de energia elétrica do sudeste brasileiro. As observações são referentes ao período de 1 de setembro de 2000 a 30 de abril de 2005. O conjunto de dados, então, é separado em 24 séries, cada uma contendo a carga de uma hora específica do dia.

Os dados são divididos em dois períodos: *in-sample*, observações entre 1º de setembro de 2001 a 31 de dezembro de 2005 e *out-of-sample*, observações realizadas entre 1º de janeiro de 2005 a 30 de abril de 2005. Utilizando os dados do *in-sample*, são estimados os parâmetros determinísticos e seu resíduo é usado como entrada para estimação dos parâmetros da “carga irregular” e, posteriormente, ajustar o melhor modelo **AR** selecionado pelas defasagens significativas.

5.2 Características das Curvas de carga

As curvas de carga representam a demanda ou o consumo de energia em função do tempo em um determinado período. Assim para caracterizar as curvas de carga é necessário conhecer a forma como os consumidores utilizam a energia, posto que a carga é uma variável dependente de seu consumidor, seja ele residencial, comercial ou industrial.

O perfil de carga dos consumidores residenciais apresenta um crescimento constante (vide figura 5.2), sensível às condições meteorológicas e as mudanças econômicas e tarifárias, sendo

difícil defini-lo, pois cada pessoa comporta-se de acordo com as suas necessidades individuais.

A carga comercial apresenta um perfil sensível à sazonalidade resultante de eventos como datas comemorativas, condições meteorológicas, mudanças econômicas e qualidade de energia entregue pelas empresas da energia elétrica.

O perfil de carga do consumo industrial é bem específico, definido pelo número de turnos de operação. Geralmente, tem curvas de cargas mais regulares, indicando pouca variabilidade sazonal e até mesmo horária, para os consumidores de maior porte. O consumo industrial pode adicionar incertezas às previsões, pela ocorrência de eventos inesperados, como paralisações ou greves, que afetam o nível da carga demandada [?].

Outros fatores que caracterizam o perfil de carga são os serviços públicos (iluminação pública, metrô, tráfego ferroviário, etc) e os fatores climáticos, como: temperatura, umidade relativa do ar, o nível de luminosidade e a velocidade do vento.

5.3 Classificação dos dias

Na modelagem da sazonalidade semanal o perfil de curva de carga dos diferentes tipos de dia foi modelado como uma variável *dummy*. Na literatura, perfil da curva de carga pode ser previsto pelas curvas de carga de dias similares no histórico. Consequentemente, os dias são agrupados em classes de diferentes tipos de dia. Essa distinção é importante pois em dias úteis as cargas são mais elevadas que em outros dias. Portanto informar ao modelo o tipo de dia é fundamental. Os dias são aqui classificados em 10 classes.

5.4 Carga determinística

5.4.1 Tendência

A tendência de uma série indica o seu comportamento de “longo prazo”, isto é, se ela cresce, decresce ou permanece estável. Há dois tipos de tendência: determinística e estocástica. No caso de tendência determinística, regressão linear e mínimos quadrados podem ser usados para estimar e modelar a componente de tendência como em [Ramanathan et al., 1997].

código	descrição
1	domingo
2	segunda-feira
3	terça-feira
4	quarta-feira
5	quinta-feira
6	sexta-feira
7	sábado
8	feriado oficial ou religioso
9	dia após feriado
10	dia anterior ao feriado

Tabela 5.1: Classificação dos dias

Quando a série apresenta tendência estocástica é necessário uma transformação para torná-la estacionária. Uma série é estacionária se ela se desenvolve no tempo, aleatoriamente, ao redor de uma média constante, refletindo alguma forma de equilíbrio. Segundo [Morettin, 2006], a maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias. Mas quando elas não apresentam esta característica, é necessário diferenciá-las para obter a estacionariedade.

O processo consiste em tomar diferenças sucessivas da série original até obter uma série estacionária. Se uma série temporal for diferenciada uma vez e a série resultante for estacionária, diz-se que a série é integrável de ordem 1.

Existem testes que identificam a natureza da tendência, testes para raízes unitárias como: o teste Aumentado de *Dickey-Fuller* (ADF) e o teste de *Phillips-Perron* (PP), cujo detalhamento teórico foge ao escopo deste trabalho.

Neste trabalho, foi aplicado uma diferenciação a série de carga, com o objetivo de testar se este procedimento produziria uma série estacionária. A figura (5.1) apresenta a série observada referente ao período compreendido entre 01 de setembro de 2000 e 31 de dezembro de 2004 e a série diferenciada. A análise visual permite verificar que a série diferenciada aparenta ser estacionária. As séries originais apresentam uma não estacionariedade homogênea que não é notada na série com uma diferenciação.

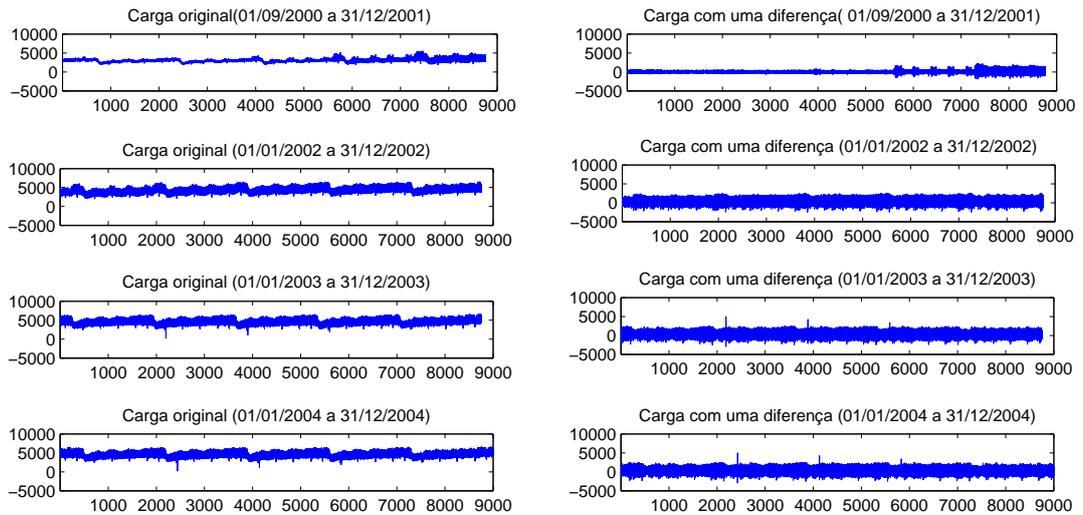


Figura 5.1: Série observada e série com a aplicação de uma diferença

5.4.2 Sazonalidade

A sazonalidade pode ser definida como a propriedade de determinadas séries de apresentar comportamentos cíclicos ou periódicos. As séries de cargas elétricas apresentam sazonalidade diária, mensal e anual.

Na literatura são encontradas diferentes propostas para modelar a sazonalidade. [Cottet & Smith, 2003], por exemplo, usam funções trigonométricas em suas abordagens.

Neste trabalho o objetivo é descrever o comportamento periódico fixo. Para capturar a sazonalidade é usada a decomposição de [Soares & Medeiros, 2008], onde a componente sazonal é dada por:

$$S_{h,d} = \alpha_0 + \sum_{r=1}^H [\alpha_{1r} \cos(\omega rd) + \alpha_{2r} \sin(\omega rd)] + \sum_{k=1}^K \mu_k \delta_k, \quad (5.1)$$

onde: $\omega = \frac{2\pi}{365}$ rd/s é a frequência fundamental; $\alpha_0, \rho, \alpha_{1r}, \alpha_{2r}, r = 1, \dots, H, \mu_k, k = 1, \dots, K$ são parâmetros e $\delta_i, i = 1, \dots, k$, são variáveis *dummy* identificando os dias da semana, feriados, dias especiais (dia após feriados e dia anterior a feriado), totalizando 10 tipos de dias tabela (5.1).

A sazonalidade anual é modelada como uma soma de senos e cossenos, esse comportamento

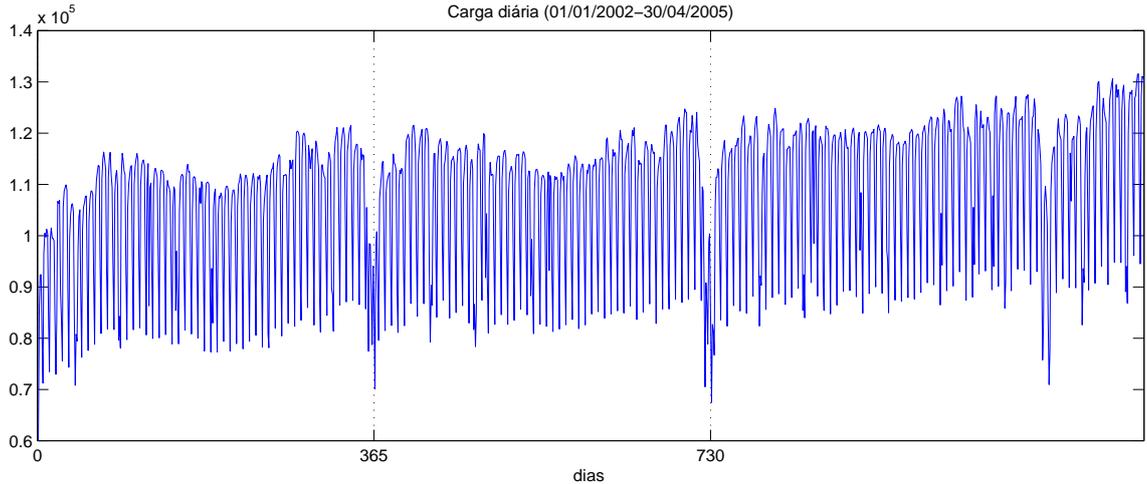


Figura 5.2: Cargas médias diárias (01/01/2002-30/04/2005)

senoidal da curva de carga pode ser observado por inspeção visual do gráfico (5.2). Observa-se, também, que as cargas são mais elevadas no verão e na primavera, e mais baixas no outono e inverno. Este padrão se explica pelo uso intenso de aparelhos de condicionamento de ar e também pela intensificação das atividades do setor industrial. Segundo [Oliveira et al., 2000], o setor industrial brasileiro aumenta o seu consumo de eletricidade na primavera (setembro, outubro e novembro), como mecanismo de preparação dos estoques necessários para o período natalino, contribuindo para o aumento na curva de carga.

Além do ciclo sazonal anual temos o ciclo sazonal semanal (efeito dos dias da semana e de dias especiais, como os feriados). A modelagem da sazonalidade semanal é feita através de variáveis *dummy*, tabela (5.1).

Inspecionando a figura (5.3), observa-se que em uma semana os dias úteis (segunda-feira a sexta-feira) têm perfis de carga semelhantes, enquanto os perfis de carga dos fins de semana apresentam um comportamento diferente.

A figura (5.4) mostra os perfis de carga nas quatro estações do ano. Para ilustrar melhor o comportamento da curva de carga selecionamos quatro quartas-feiras nas diferentes estações do ano. Nota-se que o perfil de carga das quarta-feiras de inverno e outono são similares e diferentes do verão e primavera. A “carga potencial” é modelada como:

$$L_{h,d}^P = \alpha_0 + \rho d + \sum_{r=1}^H [\alpha_{1r} \cos(\omega r d) + \alpha_{2r} \sin(\omega r d)] + \sum_{k=1}^K \mu_k \delta_k, \quad (5.2)$$

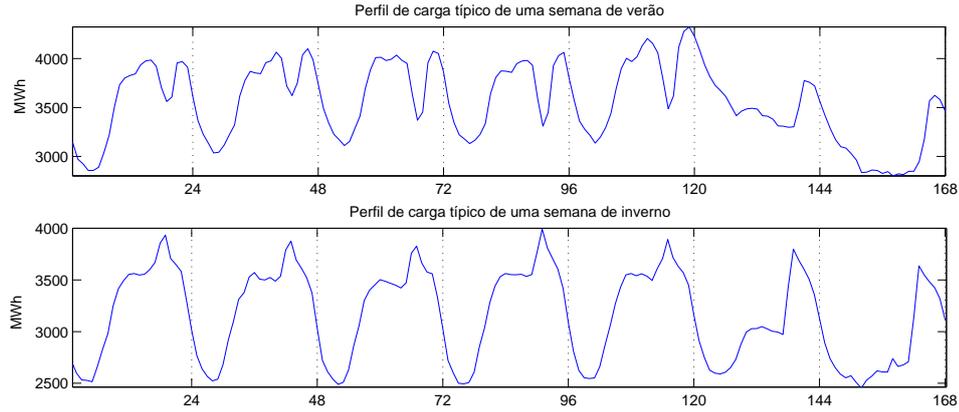


Figura 5.3: Perfil de carga típico de uma semana de verão (19/01/2004 a 25/01/2004) e inverno (12/07/2004 a 18/07/2004)

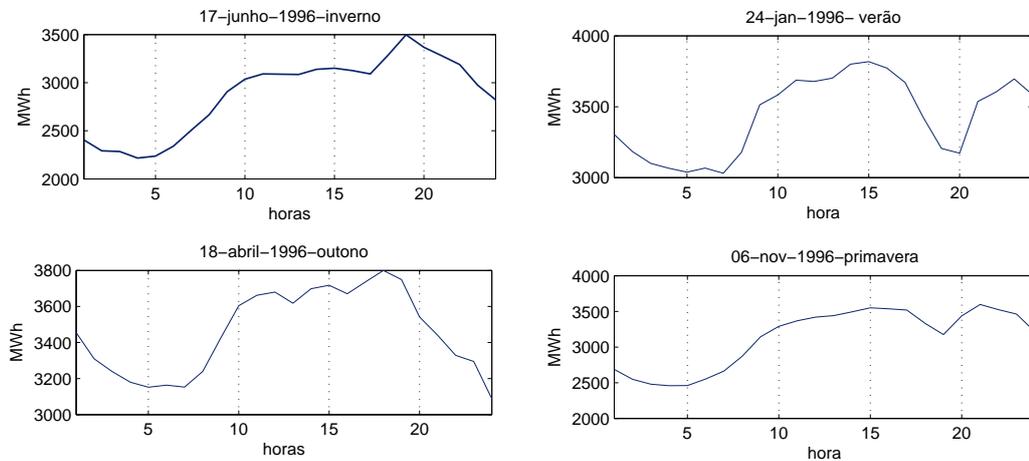


Figura 5.4: Curva de carga de quatro quarta-feiras nas diferentes estações período referente aos anos de 2002 e 2003.

onde: $\omega = \frac{2\pi}{365}$ rd/s é a frequência fundamental, ρ é a inclinação da tendência linear, d é o dia; $\alpha_0, \alpha_{1r}, \alpha_{2r}$, com $r = 1, \dots, H$; $\mu_k, k = 1, \dots, k$ são parâmetros a serem estimados; $\alpha_{1r} \cos(\omega rd) + \alpha_{2r} \sin(\omega rd)$ é a harmônica de ordem r , com $r = 1, \dots, H$; $\delta, i = 1, \dots, D$ são variáveis binárias identificando os diferentes tipos de dia.

5.5 Preparação dos dados

Para realizar as previsões adota-se a modelagem em dois níveis em que o “componente irregular” da carga é modelado por modelo **AR**. Inicialmente aplicaram testes para verificar a presença ou não de normalidade e autocorrelação dos resíduos. São apresentados a seguir a formulação teórica dos teste utilizados. Os resultados obtidos encontram-se na tabela(A.2).

5.5.1 Teste Jarque-Bera

O teste *Jarque-Bera* calcula o grau de adequação da distribuição de uma série à distribuição normal, para grandes amostras. Ele baseia-se nas diferenças entre assimetria e curtose da distribuição da série em relação à distribuição normal. Seu cálculo é feito através da fórmula abaixo:

$$JB = \frac{N - K}{6} \left(A^2 + \frac{(C - 3)^2}{4} \right) \quad (5.3)$$

onde:

N é o número de observações;

A é a assimetria;

C é a curtose;

k o número estimado de coeficientes usados.

Segundo [Gujarati, 2000], em uma distribuição normal, o valor da assimetria é 0 e o da curtose é 3, portanto $(C - 3)$ representa o excesso de curtose.

Testamos, então, a hipótese nula de que a amostra foi extraída de uma distribuição normal contra a hipótese alternativa de que ela não segue uma distribuição normal. Sob a hipótese nula de normalidade, a estatística *Jarque-Bera* segue distribuição qui-quadrado, χ^2 , com 2 graus de liberdade. Essa estatística *Jarque-Bera* resulta na probabilidade observada, que deve ser maior que o nível de confiança proposto para que a hipótese nula não seja rejeitada. Caso contrário, se a probabilidade calculada for menor que o valor observado, a hipótese de normalidade deve ser rejeitada.

5.5.2 Teste Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov (K-S) proposto pelos probabilistas russos Kolmogorov e Smirnov, baseia-se na diferença entre a função distribuição empírica, F_n , e função de distribuição proposta, F_0 . A função distribuição empírica da amostra defini-se como a proporção das observações da amostra que são menores ou iguais a x para todos os valores reais x como uma medida da discrepância entre a função de distribuição empírica F_n e a função de distribuição proposta F_0 .

Assim, se H_0 for verdadeira, a distância vertical máxima entre as imagens das duas dis-

tribuições não deve de ser muito grande, e logo se espera que D_n tome um valor pequeno. Conhecendo a distribuição de D_n é possível avaliar o seu valor observado d_n , e decidir se este é suficientemente pequeno para que não se rejeite H_0 a um dado nível de significância α .

Para uma variável aleatória qualquer x , o teste K-S baseia-se na análise do ajustamento entre a função de distribuição, F_0 , admitida na hipótese nula H_0 e a função de distribuição empírica F_n . Assim, sendo F_n a função de distribuição (desconhecida) da população, as hipóteses a testar serão:

$$H_0 : F(x) = F_0, -\infty < x < \infty$$

$$H_1 : F(x) \neq F_0(x)$$

onde F_0 é a função de distribuição proposta, contínua e completamente especificada. No teste de Kolmogorov-Smirnov considera-se a estatística:

$$D_n = \sup_{-\infty < x < +\infty} |F_n(x) - F_0(x)|$$

como uma medida da discrepância entre a função de distribuição empírica F_n e a função de distribuição proposta F_0 .

Assim, se H_0 for verdadeira, a distância vertical máxima entre as imagens das duas distribuições não deve de ser muito grande, e logo se espera que D_n tome um valor pequeno.

Conhecendo a distribuição de D_n é possível avaliar o seu valor observado d_n , e decidir se este é suficientemente pequeno para que não se rejeite H_0 a um dado nível de significância α .

5.5.3 Teste qui-quadrado

O teste qui-quadrado é baseado na soma dos erros absolutos das frequências observadas comparadas com as frequências esperadas. Esse teste é específico para dados agrupados, em que as classes que possuírem valores menores que três ou cinco devem ser agrupadas em outras classes, sendo um fator limitante para uso em dados com poucas classes.

Considere-se uma amostra aleatória de n elementos, extraída de uma população com distribuição desconhecida, sobre os quais se observa uma característica (qualitativa ou quantitativa). Os valores possíveis da característica em estudo são, num primeiro passo, repartidas por m classes mutuamente exclusivas, $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$ (serão intervalos da reta real se a característica é quantitativa e contínua).

Então, o problema que se põe é o de testar a hipótese nula, que a probabilidade desconhecida, p_i , de obter uma observação em uma classe \mathbf{A}_i é igual à probabilidade de obter uma observação na classe \mathbf{A}_i assumindo que a observação foi extraída de uma população com a distribuição especificada. Contra a hipótese alternativa que a observação não foi extraída de uma população com a distribuição especificada.

Onde:

- O_i o número de observações ou frequência absoluta observada da classe \mathbf{A}_i ;
- p_i a probabilidade desconhecida de obter uma observação na classe \mathbf{A}_i ;
- p_{0i} a probabilidade de obter uma observação na classe \mathbf{A}_i ; assumindo que a observação foi extraída de uma população com a distribuição especificada em \mathbf{H}_0 ;

As hipóteses a testar serão:

$$\mathbf{H}_0 : p_i = p_{0i}, i = 1, \dots, m$$

$$\mathbf{H}_1 : p_i \neq p_{0i}$$

Assim, a frequência esperada da classe \mathbf{A}_i , quando \mathbf{H}_0 é verdadeira, é dada por:

$$e_i = n \times p_{0i} \tag{5.4}$$

A estatística de teste, do teste Qui-quadrado, é dada por:

$$Q_{obs} = \sum_{i=1}^m \frac{(O_i - e_i)^2}{e_i} \tag{5.5}$$

que, sendo verdadeira a hipótese nula, tem distribuição assintótica do Qui-quadrado com $m - k - 1$ graus de liberdade (χ_{m-k-1}^2), onde k é o número de parâmetros desconhecidos

da distribuição proposta em H_0 , estimados a partir da amostra. Se a função (densidade) de probabilidade proposta em H_0 , estiver completamente especificada, e se não há parâmetros desconhecidos, então $Q \sim \chi_{m-k-1}^2$.

Se a hipótese nula for de fato verdadeira, a diferença entre cada valor observado e o respectivo valor esperado não será grande, e conseqüentemente a estatística teste terá um valor observado pequeno. De modo intuitivo, quanto maior for o valor observado de Q , menos plausível é a hipótese nula. Assim, para um nível de significância α , a hipótese nula é rejeitada se $Q > \chi_{1-\alpha}^2$, com $\chi_{1-\alpha}^2$, o quantil de probabilidade $1 - \alpha$ da distribuição χ_{m-k-1}^2 .

5.6 Intervalos de Confiança

Finalmente os intervalos de confiança para as previsões geradas pelo modelo TLSAR, de um à sete passos a frente, podem ser construídos.

A distribuição dos erros é identificada através do método *bootstrap*, descrito no capítulo (4) seção (4.2.1.4). Foram realizadas 1000 replicações *bootstrap* para estimar o parâmetro de interesse. Considerando não ser conhecido o verdadeiro valor do desvio padrão populacional são realizadas para cada replicação *bootstrap*, 50 replicações *bootstrap* para estimar o desvio padrão da estatística de interesse. Os intervalos de confiança de $100(1 - \alpha)\%$ são dados pela equação 4.10 ao nível de nível de significância de ($\alpha = 0.05$).

5.7 Resumo

Neste capítulo apresenta-se a teoria do modelo utilizado no decorrer deste trabalho. As previsões são feitas para uma série de observações horárias de uma concessionária de distribuição de energia elétrica do sudeste brasileiro. São utilizadas observações do período de 1 de setembro de 2000 a 30 de abril de 2005. Os dados são separados em séries horárias, totalizando 24 séries, cada uma contendo a carga de uma hora específica do dia. As observações entre 1º de setembro de 2001 a 31 de dezembro de 2005 são usadas para estimar o modelo e as observações entre 1º de janeiro de 2005 a 30 de abril de 2005 para validar o modelo.

Para caracterizar as curvas de carga é necessário conhecer como os consumidores utilizam a energia, uma vez que a carga é dependente de seu consumidor. Os consumidores são classi-

ficados como: residencial, comercial ou industrial. O perfil de carga do consumo residencial é crescente e constante, o perfil de carga do consumo comercial é sensível à sazonalidade e o consumo industrial tem perfil de carga com curvas mais regulares.

Na modelagem adotada, modelo em dois níveis, a primeira etapa consiste em modelar as variáveis determinísticas que compõem a “carga potencial”: sazonalidade e tendência.

Como na literatura o perfil da curva de carga pode ser previsto pelas curvas de carga de dias similares no histórico, a classificação aqui adotada é apresentada na tabela (5.1).

O teste de *Jarque-Bera* é utilizado para verificar se os resíduos, gerados ao final da modelagem das variáveis determinísticas, se distribuem normalmente. Para identificar a distribuição dos erros são usadas 1000 replicações *bootstrap* e os intervalos são construídos com **95%** de confiança.

Capítulo 6

Resultados

Este capítulo apresenta os resultados do experimento descrito no capítulo 5 onde foi detalhada a modelagem para obtenção das previsões e a estratégia para construir os intervalos de confiança.

As observações referentes ao período entre 1º de julho de 2001 a 31 de dezembro de 2004 foram usadas para estimar os modelos. A amostra de validação (ou seja, o período utilizado para testar a precisão dos modelos) refere-se a 1º de janeiro de 2005 a 30 de abril de 2005.

A primeira etapa da modelagem consiste em calcular as variáveis determinísticas ou os componentes da “carga potencial” que capturam a influência de componentes sazonais e tendências presentes na carga. Para capturar a sazonalidade usa-se uma soma de senos e cossenos. O número de harmônicas para cada hora do dia é selecionado pelo critério SBIC, assim como, as defasagens significativas do modelo AR , (tabela 6.1).

Os resíduos gerados pela primeira etapa da modelagem são testados, através do teste de Jarque-Bera, Kolmogorov-Smirnov e Qui-quadrado para verificar se as suas distribuições são gaussianas. O resultado dos testes rejeitam a hipótese de normalidade, tabela A.2.

Os erros médios absolutos percentuais (MAPE), para as séries correspondentes às 24 horas, são apresentados para horizontes de previsão de um a sete passos a frente. O MAPE da previsão dos períodos correspondentes aos anos de 2004 e 2005 variam entre **1.98%** e **5.79%**, de um a sete passos a frente. A figura(6.1) mostra o MAPE encontrado para os anos de 2004 e 2005. O menor MAPE é obtido em previsões realizadas um passo a frente, pois os erros de previsão aumentam a medida que o horizonte de previsão cresce. Os dois anos apresentam

Tabela 6.1: Parâmetros do modelo AR

hora	harmônicas	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\phi}_7$
1	8	0.6025	-	-	0.103	-	-	0.1386
2	8	0.6347	-	-	0.1027	-	-	0.1197
3	8	0.7065	-0.093	-	0.126	-	-	0.1146
4	8	0.6726	-0.0881	-	0.1166	-	-	0.1368
5	8	0.6024	-0.1162	-	0.0512	-	-	0.2573
6	8	0.4773	-0.1036	-	-0.0064	-	-	0.3379
7	8	0.3696	-0.0728	-	0.062	-	-	0.182
8	8	0.3336	-0.0466	-	0.0824	-	-	0.1617
9	7	0.3188	-0.0225	-	0.1025	-	-	0.1575
10	8	0.3144	-0.0196	-	0.1007	-	-	0.1775
11	8	0.3243	-0.0232	-	0.0898	-	-	0.1995
12	7	0.3364	-0.0089	-	0.1002	-	-	0.1968
13	8	0.3513	-0.0227	-	0.0669	-	-	0.1998
14	8	0.2986	-0.001	-	0.0613	-	-	0.1653
15	8	0.3124	-0.0021	-	0.0688	-	-	0.1563
16	8	0.3275	-0.0105	-	0.0664	-	-	0.1684
17	8	0.3109	0.0039	-	0.0752	-	-	0.1903
18	8	0.3189	0.0335	-	0.0962	-	-	0.2055
19	9	0.3887	0.0126	-	0.1337	-	-	0.2068
20	8	0.4364	-0.0095	-	0.1588	-	-	0.1777
21	8	0.4735	-0.0328	-	0.1488	-	-	0.209
22	7	0.5294	-0.0536	-	0.1278	-	-	0.1878
23	8	0.5519	-0.0417	-	0.1237	-	-	0.1795
24	8	0.5543	0.0081	-	0.1103	-	-	0.1548

erros médios de previsão crescentes, mas este crescimento não é explosivo, por exemplo: o MAPE calculado para previsões realizadas um passo à frente para o ano de 2004 é de **2.82%** e para sete passos à frente é de **3.32%**. No ano de 2005 o MAPE para um passo à frente é de **3.40%** e sete passos à frente **4.49%**. Ao comparar o MAPE encontrado para os sete passos de previsão do ano de 2004 com o MAPE encontrado para as previsões de 2005, verifica-se que o MAPE do ano de 2005 é maior para todos os horizontes de previsão. Neste ano além dos padrões sazonais citados no capítulo 5 acrescenta-se uma forte tendência de crescimento econômico no país.

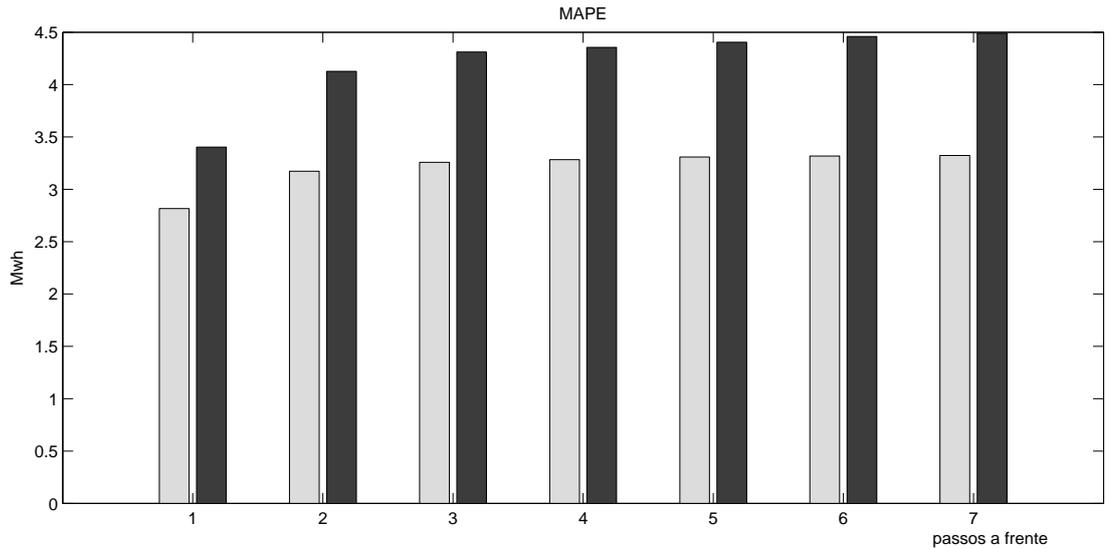


Figura 6.1: MAPE(Erro Médios Absolutos Percentuais) do ano 2004(claro) e MAPE do ano 2005(escuro)

O MAPE calculado para as quatro estações do ano, de um a sete passos à frente, são mostrados na tabela (6.2). Os maiores registros são apresentados no verão e primavera e os menores no inverno. Segundo [Hippert et al., 2001] as curvas de carga de dias do inverno tendem a ser mais estáveis que as curvas de carga do verão, uma vez que a carga nos meses frios é menos sensível a alterações climáticas.

Tabela 6.2: MAPE por estações (22/12/2003-21/12/2004)

estações	total						
	passos à frente						
	1	2	3	4	5	6	7
verão	3.39	3.97	4.16	4.21	4.25	4.29	4.29
outono	2.71	2.85	2.85	2.85	2.84	2.86	2.89
inverno	1.85	2.04	2.06	2.06	2.03	2.02	2.02
primavera	3.46	3.92	4.03	4.08	4.16	4.15	4.14

A figura (6.2) mostra intervalos com **95%** de confiança construídos para o período compreendido entre 17 de abril de 2005 a 30 de abril de 2005. Verifica-se que os intervalos construídos seguem a distribuição da curva de carga das observações, pois nenhuma suposição sobre sua distribuição foi adotada e sim realizada a identificação a partir dos resíduos. No entanto são encontrados pontos da curva de carga fora do intervalo de confiança construído, conforme ilustrado pela figura (6.3). A tabela (6.3) mostra o número e o percentual de observações

que estão fora dos intervalos, período referente a 1º de janeiro de 2005 a 30 de abril de 2005. Esse fato ocorre em dias de feriados, dias considerados difíceis em matéria de previsões de

Tabela 6.3: Pontos fora do intervalo de confiança (Erro)

passos a frente	pontos	%
1	538	20%
2	353	12%
3	367	12.7%
4	362	12.5%
5	395	13.7%
6	398	13.8%
7	423	14.6%

carga, pois ocorrem em menor número no conjunto de observações. Por exemplo, entre as 33600 observações foram encontradas 1584 observações referentes a feriados. Tal feito não é verificado nos demais dias. A figura (6.3) mostra intervalos de confiança construídos para quatro quartas nas quatro estações do ano de 2005.

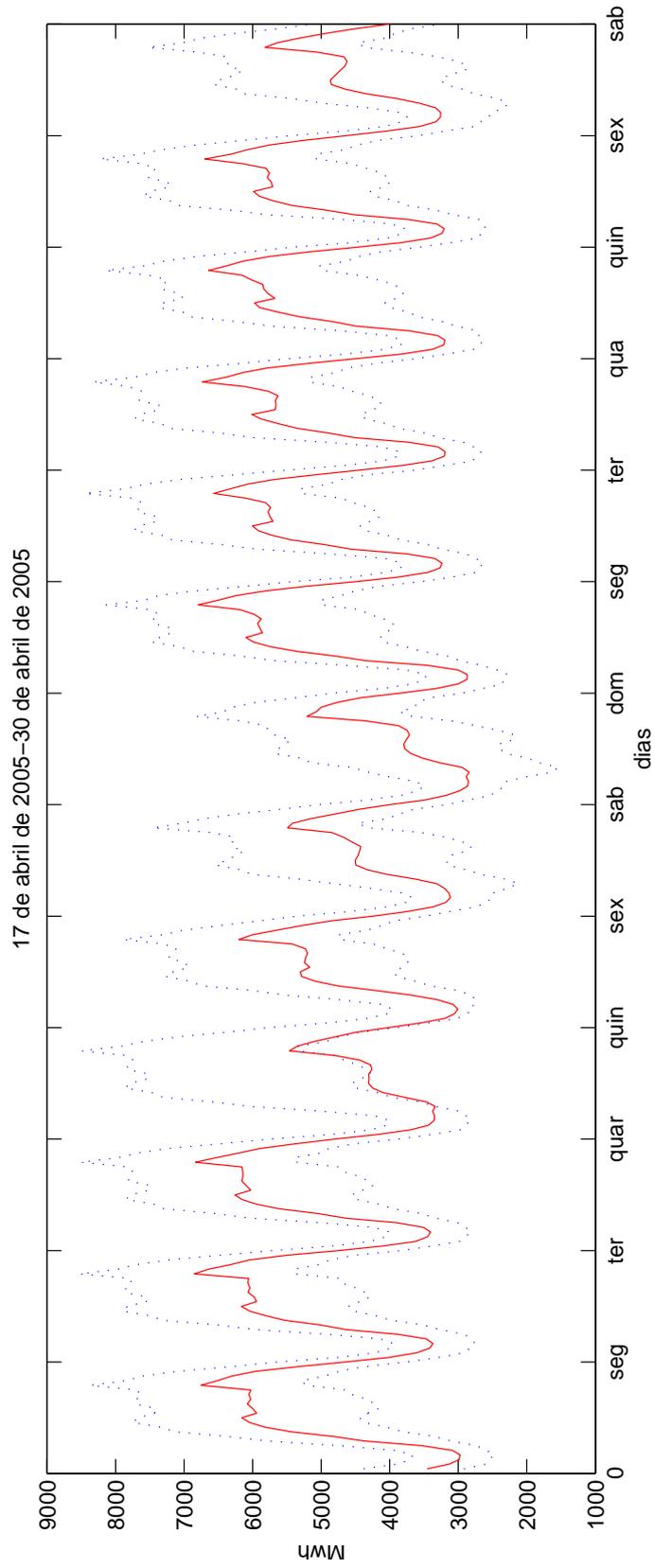


Figura 6.2: Intervalo de confiança para dias úteis e um dia de feriado-21 de abril

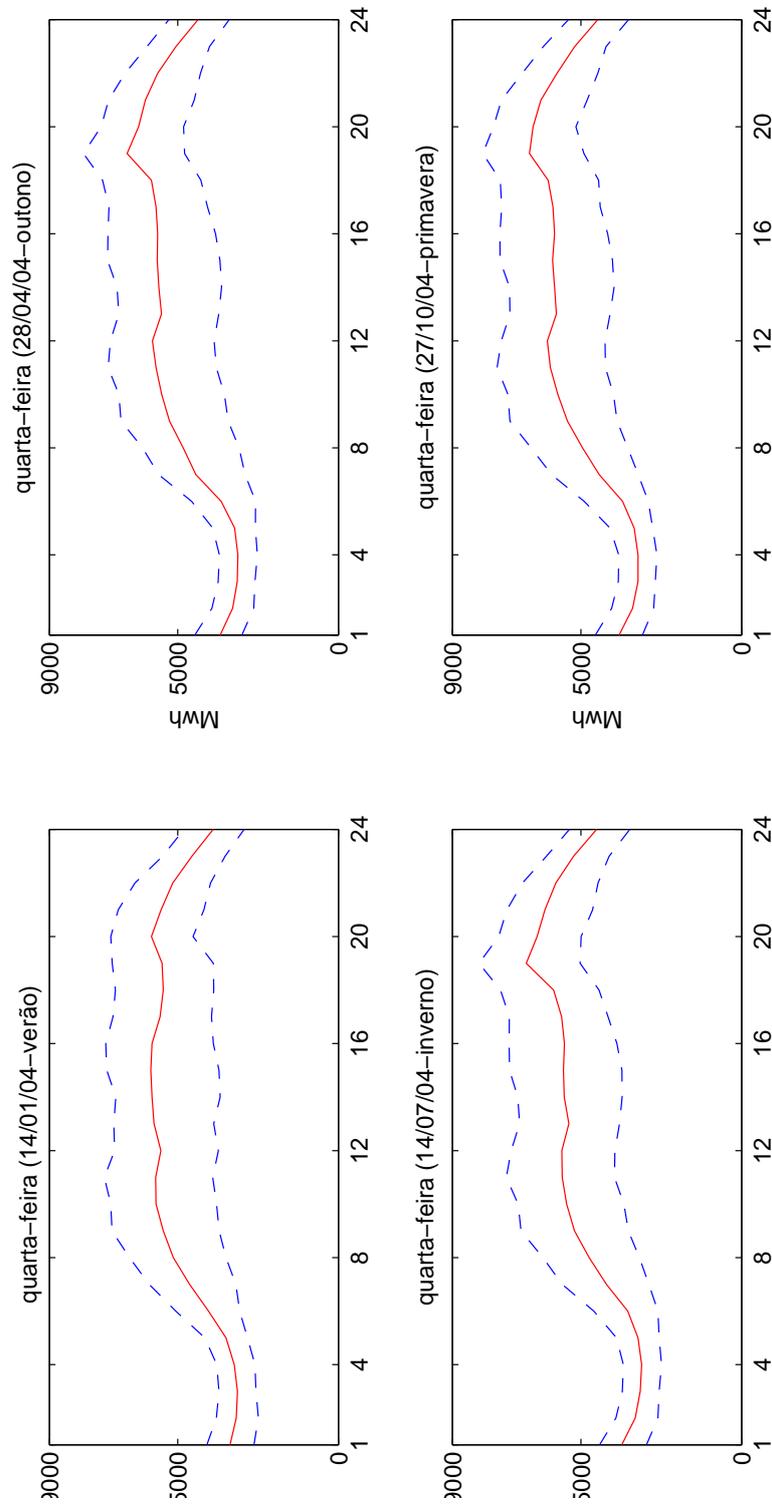


Figura 6.3: Intervalo de confiança construído para uma quarta-feira de cada estação

Capítulo 7

Conclusão

Este trabalho propõe um método alternativo de construção intervalo de confiança para a demanda horária de energia. Foram estudados alguns modelos para a previsão de carga a curto prazo como: modelos de regressão, séries temporais e redes neurais artificiais. Isto foi considerado necessário para gerar as previsões e posteriormente construir os intervalos de confiança.

A metodologia aqui adotada é apresentada em [Soares, 2003]. Essa estratégia apoia-se somente em argumentos estatísticos e o tratamento longitudinal dos modelos o que não é comum na literatura. Esta modelagem é atrativa pois permite a previsão de hora em hora.

O método adotado para construção dos intervalos de confiança, *bootstrap*, possibilita a obtenção dos intervalos sem a necessidade de pressupostos sobre a distribuição do estimador. Vários tipos de intervalos de confiança *bootstrap* podem ser considerados, com variado custo computacional.

Os métodos aqui descritos para construir os intervalos de confiança consideram os parâmetros vindos de populações de tamanho infinito ou grandes o suficiente para que possam ser tratadas como de tamanho infinito. Métodos *bootstrap* adequados à populações finitas são apresentados por Booth et al.(1994).

O problema de determinar quantas replicações B são necessárias para a obtenção de boas estimativas dos limites inferior e superior de intervalos de confiança construídos pelo método *bootstrap* é discutido por [Efron & Tibshirani, 1993]. O trabalho de Hall (1986) explora a questão da determinação ótima de B com maior profundidade. Uma forma simples de de-

terminar o número necessário de replicações B , consiste em conduzir um estudo empírico, replicando a aplicação do método *bootstrap* para diversos valores de B .

Os intervalos *bootstrap*, construídos com **95%** de confiança, apresentaram resultados melhores para os dias úteis quando comparados com os feriados. Na literatura, muitas vezes os feriados são classificados na mesma categoria dos domingos [Hsu and Yang.,1995], mas sem nenhuma dúvida, dias especiais diferentes têm perfis de carga diferentes.

Este modelo pode ser aplicado a outros dados de concessionárias que apresentam padrões sazonais semelhantes, como muitos no Brasil.

7.1 Sugestões para trabalhos futuros

Para trabalhos futuros, algumas propostas podem ser feitas, por exemplo: aperfeiçoar os resultados obtidos acrescentando outros fatores que influenciam o comportamento da carga como variáveis climáticas; comparar intervalos de confiança em que a distribuição dos erros é identificada utilizando a técnica *bootstrap* com os intervalos em que a distribuição é identificada por simulação Monte Carlo.

Referências Bibliográficas

- [Andrade & Silva, 2009] L. Andrade and E. Silva. **Very short-term load forecasting based on arima model and intelligent systems.** *IEEE*, 2009.
- [Efron, 1979] B. Efron. **Bootstrap methods: Another look at the jackknife.** *The Annals of Statistics*, 1979.
- [Bartkiewicz, 2000] W. Bartkiewicz. **Confidence intervals prediction for the short-term electrical load neural forecasting models.** *W Elektrotechnik and Informations technik*, pages 8–12, 2000.
- [Bunn & Farmer, 1985] B. D. W. and F. E. D. **Economic and operational context of electric load prediction,**in d. w. bunn and e. d. farmer(eds). *Comparative Models for Electrical Load Forecasting*, pages 3–11, 1985.
- [Carrasco, 2002] C. C. G. **Comparação de procedimentos para construção de intervalos de confiança para os parâmetros do modelo poli-log-logístico na presença de censuras aleatórias.** Dissertação de mestrado, Departamento de Estatística da UFSCar, 2002.
- [Chyssolouris et al., 1996] M. Chyssolouris G. and R. A. **Confidence interval prediction for neural network models.** *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 7:229–232, 1996.
- [Cancelo et al.,2008] A. E. José Ramón Cancelo and R. Grafe. **Forecasting the electricity load from one day to one week ahead for the spanish system operator.** *Internacional of forecasting*, pages 588–602, 2008.
- [Cottet & Smith, 2003] C. R. and S. M. **Bayesian modeling and forecasting of intraday electricity load.**

- [Efron & Tibshirani, 1993] E. B. and T. R. **An introduction to the bootstrap**. Chapman & Hall, 1993.
- [Engle et al., 1992] M. C. Engle R.F. and R. J. **Modeling peak electricity demand**. *Journal of Forecasting*, pages 241–251, 1992.
- [Fan. & Yao, 2005] J. e. Y. Fan. **Nonlinear time series: Nonparametrics and parametric methods**. Springer, 2005.
- [Fan & Donald, 1994] F. J.Y. and M. J.D. **A real-time implementation of short-term load forecasting for distribution power systems**. *IEEE*, pages 988–994, 1994.
- [Guan et al., 2010] M. L. C. M. Che Guan, Luh P.B. and F. P.B. **Hybrid kalman algorithms for very short-term load forecasting and confidence interval estimation**. *IEEE*, pages 1–8, 2010.
- [Gujarati, 2000] D. N. Gujarati. **Econometria básica**. Makron Books, 2000.
- [Hippert et al., 2001] P. C. Hippert H. S. and S. R.C. **Neural networks for short-term load forecasting: A review end evaluation**. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, pages 44–55, 2001.
- [Hsu and Yang.,1995] Y. Hsu and C. Yang. **Electrical load forecasting**. *Applications of Neural Networks*, pages 157–189, 1995.
- [Morettin, 2006] C. M. C. T. Morettin Pedro A. **Análise de séries temporais**. Edgard Blücher LTDA, São Paulo, 2006.
- [Oliveira et al., 2000] G. B. d. S. Adilson de Oliveira and J. de Medeiros Braga. **Diversidade sazonal do consumo de energia elétrica no brasil**. *Pesquisa e Planejamento economico*, 2000.
- [Petiau, 2009] B. Petiau. **Confidence interval estimation for short-term load forecasting**. *IEEE Bucharest Power Tech Conference*, pages 1–6, 2009.
- [Papadopoulos et al., 2001] P. E. G. Papadopoulos and A. Murray. **Confidence estimation methods for neural networks: A practical comparison**. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, pages 1278–1287, 2001.

- [Portela et al., 2011] M. N. Artur Tiago Silva and M. M. Portela. **Sobre a estimação de intervalos de confiança para os quantis de variáveis aleatórias hidrológicas.** *Associação Portuguesa dos Recursos Hídricos*, pages 63–73, 2011.
- [Rahman & Bhatnagar,1998] S. Rahman and R. Bhatnagar. **An expert system based algorithm for short term load forecast.** *IEEE*, pages 392–399, 1998.
- [Rivals & Personnaz, 2000] I. Rivals and L. Personnaz. **Construction of confidence intervals for neural networks based on least squares estimation.** *Pergamon Neural Networks*, pages 463–484, 2000.
- [Ramanathan et al., 1997] G. C. W. J. V.-A. F. Ramanathan R., Engle R. and B. C. **Short-run forecasts of electricity loads and peaks.** *International Journal of Forecasting*, pages 161–174, 1997.
- [Schwarz, 1978] G. Schwarz. **Estimating the dimension of a model.** *Annals of Statistics*, 1978.
- [Silva & Moulin, 2000] L. S. Moulin and A. P. A. da Silva. **Confidence intervals for neural network based short-term load forecasting.** *IEEE Transactions on Power Systems*, 15:1191–1196, 2000.
- [Soares & Medeiros,2005] L. J. Soares and M. C. Medeiros. **Modeling and forecasting short-term electric load demand: A two step methodology.** *Textos para Discussão* 495.
- [Soares, 2003] L. J. Soares. **Ensaio em previsão a curto prazo.** Tese de doutorado, Pontifícia Universidade Católica do Rio de Janeiro, 2003.
- [Soares & Medeiros, 2008] L. J. Soares and M. C. Medeiros. **Modeling and forecasting short-term electricity load: A comparison of methods with an application to brazilian data.** *International Journal of Forecasting*, 2008.
- [Souza & Camargo,2004] R. Souza and M. Camargo. **Análise e previsão de séries temporais: Os modelos arima.** Ijuí:SEDIGRAF, 2004.
- [Teräsvirta et. al.,2005] T. Teräsvirta, D. van Dijk and M. C. Medeiros. **Linear models, smooth transition autoregressions, and neural networks for forecasting**

macroeconomic time series: A re-examination. *International Journal of Forecasting*, 21(4):755 – 774, 2005.

[Townsend, 1999] T. N. . T. L. **Estimations of error bounds for neural-network function approximators.** *IEEE TRANSACTIONS ON NEURAL NETWORKS*, pages 217–230, 1999.

[Zang et al., 2003] L. P. Li Zang and K. K. **Energy clearing price prediction and confidence interval estimation with cascaded neural networks.** *IEEE TRANSACTIONS ON POWER SYSTEMS*, pages 99–105, 2003.

Apêndice A

Tabelas - 2000 - 2005

Tabelas de MAPE (Modelagem AR) - 2004 - 2005

Tabela A.1: MAPE Total- Modelagem AR

hora	passos à frente						
	1	2	3	4	5	6	7
1	2.47	2.98	3.18	3.26	3.34	3.40	3.43
2	2.12	2.66	2.86	2.93	3.01	3.08	3.11
3	1.98	2.56	2.75	2.82	2.90	2.97	3.00
4	2.06	2.64	2.80	2.86	2.92	2.98	3.01
5	2.59	3.24	3.34	3.37	3.40	3.42	3.42
6	3.59	4.31	4.36	4.37	4.36	4.36	4.36
7	3.63	4.08	4.12	4.12	4.15	4.17	4.18
8	3.59	3.97	4.02	4.02	4.06	4.09	4.10
9	3.52	3.88	3.94	3.95	3.99	4.03	4.05
10	3.49	3.84	3.90	3.91	3.95	3.98	3.99
11	3.47	3.82	3.87	3.88	3.92	3.94	3.95
12	3.54	3.90	3.96	3.98	4.02	4.04	4.05
13	3.57	4.00	4.05	4.07	4.10	4.11	4.12
14	5.24	5.65	5.71	5.73	5.76	5.79	5.79
15	4.11	4.51	4.57	4.59	4.62	4.65	4.66
16	3.95	4.36	4.41	4.43	4.45	4.47	4.48
17	3.82	4.12	4.17	4.20	4.22	4.24	4.24
18	3.42	3.69	3.74	3.76	3.78	3.80	3.81
19	2.71	3.01	3.06	3.09	3.13	3.16	3.18
20	2.23	2.50	2.57	2.60	2.64	2.67	2.70
21	2.24	2.56	2.63	2.65	2.70	2.72	2.75
22	2.35	2.75	2.83	2.86	2.90	2.93	2.96
23	2.54	2.99	3.09	3.14	3.17	3.22	3.25
24	2.58	3.07	3.22	3.29	3.35	3.40	3.44
min	1.98	2.50	2.57	2.60	2.64	2.67	2.70
med	3.45	3.76	3.80	3.82	3.85	3.87	3.88
max	5.24	5.65	5.71	5.73	5.76	5.79	5.79

Tabela A.2: TESTES DE NORMALIDADE

hora	Jarque-Bera		Kolmogorov-Smirnov	Qui-quadrado
	p-valor	estatística-t	estatística-t	estatística-t
1	0.001	89.51435	70.3228	0.5496
2	0.001	111.45388	86.9217	0.5431
3	0.001	119.89868	84.9463	0.548
4	0.001	134.05139	75.677	0.0388
5	0.001	191.16876	39.1935	0.5478
6	0.001	550.5133	20.9355	0.5193
7	0.001	3628.4694	45.7258	0.529
8	0.001	4943.9746	32.0167	0.534
9	0.001	4325.4783	42.409	0.5436
10	0.001	3686.6631	80.5134	0.539
11	0.001	3451.8045	58.2543	0.5399
12	0.001	2164.4298	47.474	0.5357
13	0.001	3217.7954	29.0254	0.5486
14	0.001	77026.028	45.4435	0.5431
15	0.001	25393.426	14.4237	0.5447
16	0.001	8376.9578	42.8915	0.5415
17	0.001	2322.2214	27.6039	0.5467
18	0.001	702.64542	53.744	0.5471
19	0.001	639.87808	56.7236	0.5346
20	0.001	933.07703	85.2156	0.0388
21	0.001	488.88915	110.3604	0.5449
22	0.001	263.14141	76.4725	0.5552
23	0.001	152.0634	30.6273	0.5529
24	0.001	88.72324	20.839	0.5366

Tabela A.3: MAPE - Modelagem AR

hora	Total							2004							2005						
	passos à frente							passos à frente							passos à frente						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	2.47	2.98	3.18	3.26	3.34	3.40	3.43	2.35	2.67	2.85	2.91	2.94	2.95	2.94	2.76	3.61	3.95	4.11	4.22	4.36	4.46
2	2.12	2.66	2.86	2.93	3.01	3.08	3.11	1.99	2.41	2.58	2.62	2.66	2.65	2.63	2.42	3.27	3.59	3.80	3.92	4.06	4.16
3	1.98	2.56	2.75	2.82	2.90	2.97	3.00	1.86	2.35	2.49	2.53	2.57	2.58	2.56	2.28	3.20	3.52	3.69	3.80	3.91	4.02
4	2.06	2.64	2.80	2.86	2.92	2.98	3.01	1.93	2.40	2.52	2.57	2.61	2.62	2.60	2.35	3.30	3.63	3.73	3.84	3.94	4.02
5	2.59	3.24	3.34	3.37	3.40	3.42	3.42	2.47	3.02	3.08	3.11	3.15	3.15	3.15	2.76	3.85	4.17	4.22	4.27	4.32	4.33
6	3.59	4.31	4.36	4.37	4.36	4.36	4.36	3.49	4.18	4.22	4.22	4.22	4.22	4.22	3.87	4.97	5.16	5.16	5.15	5.14	5.14
7	3.63	4.08	4.12	4.12	4.15	4.17	4.18	3.48	3.88	3.93	3.92	3.95	3.96	3.97	3.83	4.67	4.77	4.76	4.81	4.86	4.86
8	3.59	3.97	4.02	4.02	4.06	4.09	4.10	3.38	3.71	3.77	3.78	3.79	3.81	3.82	4.00	4.72	4.82	4.82	4.89	4.95	4.96
9	3.52	3.88	3.94	3.95	3.99	4.03	4.05	3.21	3.54	3.60	3.62	3.63	3.66	3.67	3.97	4.57	4.66	4.67	4.71	4.77	4.79
10	3.49	3.84	3.90	3.91	3.95	3.98	3.99	3.14	3.46	3.53	3.55	3.57	3.59	3.59	4.17	4.69	4.80	4.80	4.84	4.88	4.88
11	3.47	3.82	3.87	3.88	3.92	3.94	3.95	3.13	3.44	3.50	3.52	3.55	3.55	3.56	4.18	4.74	4.85	4.85	4.88	4.90	4.91
12	3.54	3.90	3.96	3.98	4.02	4.04	4.05	3.17	3.51	3.58	3.61	3.64	3.65	3.66	4.30	4.91	5.00	5.01	5.01	5.03	5.05
13	3.57	4.00	4.05	4.07	4.10	4.11	4.12	3.22	3.59	3.67	3.69	3.73	3.73	3.74	4.13	4.77	4.88	4.88	4.90	4.91	4.91
14	5.24	5.65	5.71	5.73	5.76	5.79	5.79	3.37	3.72	3.79	3.81	3.85	3.86	3.87	4.24	4.75	4.90	4.90	4.94	4.95	4.96
15	4.11	4.51	4.57	4.59	4.62	4.65	4.66	3.44	3.81	3.88	3.91	3.94	3.96	3.96	4.41	5.02	5.15	5.16	5.17	5.18	5.19
16	3.95	4.36	4.41	4.43	4.45	4.47	4.48	3.46	3.89	3.95	3.98	4.01	4.02	4.02	4.34	4.96	5.08	5.09	5.10	5.10	5.11
17	3.82	4.12	4.17	4.20	4.22	4.24	4.24	3.47	3.73	3.80	3.83	3.84	3.86	3.86	4.24	4.88	4.98	4.99	4.96	4.96	4.96
18	3.42	3.69	3.74	3.76	3.78	3.80	3.81	3.19	3.44	3.52	3.54	3.54	3.55	3.55	3.81	4.42	4.52	4.53	4.50	4.50	4.50
19	2.71	3.01	3.06	3.09	3.13	3.16	3.18	2.57	2.81	2.88	2.91	2.93	2.95	2.98	3.26	3.92	4.04	4.07	4.12	4.16	4.19
20	2.23	2.50	2.57	2.60	2.64	2.67	2.70	2.14	2.35	2.43	2.47	2.49	2.48	2.49	2.43	3.09	3.26	3.29	3.33	3.37	3.41
21	2.24	2.56	2.63	2.65	2.70	2.72	2.75	2.09	2.36	2.42	2.44	2.46	2.46	2.47	2.42	2.93	3.12	3.15	3.23	3.29	3.34
22	2.35	2.75	2.83	2.86	2.90	2.93	2.96	2.23	2.46	2.52	2.54	2.54	2.54	2.58	2.38	3.05	3.29	3.32	3.37	3.45	3.49
23	2.54	2.99	3.09	3.14	3.17	3.22	3.25	2.39	2.66	2.76	2.79	2.80	2.80	2.83	2.53	3.28	3.52	3.59	3.68	3.82	3.88
24	2.58	3.07	3.22	3.29	3.35	3.40	3.44	2.45	2.78	2.90	2.95	3.01	3.01	3.04	2.60	3.43	3.83	3.95	4.06	4.21	4.25
min	1.98	2.50	2.57	2.60	2.64	2.67	2.70	1.86	2.35	2.42	2.44	2.46	2.46	2.47	2.28	2.93	3.12	3.15	3.23	3.29	3.34
med	3.45	3.76	3.80	3.82	3.85	3.87	3.88	3.13	3.44	3.51	3.53	3.55	3.55	3.55	3.82	4.50	4.59	4.60	4.61	4.64	4.64
max	5.24	5.65	5.71	5.73	5.76	5.79	5.79	3.49	4.18	4.22	4.22	4.22	4.22	4.22	4.41	5.02	5.16	5.16	5.17	5.18	5.19

..