

ALEX VIEIRA FALKENBERG

**PREVISÃO DE CONSUMO URBANO DE ÁGUA EM
CURTO PRAZO**

Dissertação apresentada como requisito parcial
à obtenção do grau de Mestre no Curso de
Pós-Graduação em Métodos Numéricos em
Engenharia, Universidade Federal do Paraná.

Orientadora: Prof^ª Dra. Eng. Andrea Sell
Dyminski

Co-orientador: Prof. Dr. Eng. Eduardo Parente
Ribeiro

CURITIBA

2005

TERMO DE APROVAÇÃO

ALEX VIEIRA FALKENBERG

PREVISÃO DE CONSUMO URBANO DE ÁGUA EM CURTO PRAZO

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Curso de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientadora: Prof.^a Dr.^a Eng. Andrea Sell Dyminski

Departamento de Construção Civil, UFPR

Co-orientador: Prof. Dr. Eng. Eduardo Parente Ribeiro

Departamento de Eletricidade, UFPR

Prof.^a Dr.^a Maria Teresinha Arns Steiner

Departamento de Matemática, UFPR

Prof. Dr. Eng. Leandro dos Santos Coelho

Departamento de Engenharia de Produção e Sistemas, PUCPR

Curitiba, 1 de abril de 2005.

À memória de Rudolfo Falkenberg.

Agradecimentos

Aos professores Andrea Sell Dyminki e Eduardo Parente Ribeiro pela orientação, e paciência.

A meus pais, pelo apoio sempre presente.

Mutti, Miguel e Helga, pelo eterno carinho.

A Leonardo Romildo Früchting, pela amizade e ensinamentos.

Ao pessoal da USCO, pela oportunidade e companheirismo.

À Sanepar por tornar este trabalho possível.

Aos meus amigos dos bons tempos de universidade, que sempre conseguiram atrapalhar meus estudos da melhor maneira possível.

RESUMO

Nos últimos anos, as companhias de saneamento têm investido largamente na automação de seus sistemas de abastecimento de água (SAA), proporcionando dados em tempo real de vazão e de pressão das regiões de consumo. Após a etapa da automação e de posse das informações em tempo real e possibilidade de atuação automática sobre os elementos do SAA, é possível otimizar a operação para um melhoramento contínuo dos serviços de captação, tratamento e distribuição de água e redução de custos relacionados, tais como as perdas físicas, a energia e o uso de produtos químicos.

A correta tomada de decisões operacionais depende do conhecimento prévio do perfil de demanda ao longo do dia das regiões de consumo. O presente estudo objetiva apresentar diferentes modelos de previsão de consumo a curto prazo utilizando diferentes técnicas, tais como redes neurais artificiais, regressão linear múltipla e modelos do tipo Box e Jenkins. Estas previsões baseadas em dados históricos coletados através de um sistema SCADA (*Supervisory, Control and Data Acquisition*), para prever o consumo urbano de água a curto prazo, particularmente para as próximas 24 horas. Os modelos foram aplicados para três bairros com perfis de consumo distintos da cidade de Ponta Grossa: Uvaranas (residencial), Centro Histórico (comercial) e Distrito Industrial (misto).

Inicialmente, os dados provenientes de uma base de dados de um ano são coletados, filtrados e validados de forma a tornar possível a obtenção dos modelos de previsão. Para cada uma das regiões de consumo foram propostos e comparados segundo critérios de erro RMSE (*Root Mean Square Error*) e MAPE (*Mean absolute Percentage Error*), 6 modelos de redes neurais artificiais, 3 modelos de regressão linear múltipla, 7 modelos do tipo Box e Jenkins e dois modelos híbridos, objetivando a obtenção de possíveis alternativas para a previsão do consumo nas próximas 24 horas. Os resultados alcançados foram satisfatórios para os objetivos propostos, mesmo com a limitação dos dados e com a ausência de informações meteorológicas, fatores que influenciam diretamente no consumo.

Palavras-chave: previsão de consumo de água; sistemas de distribuição de água; redes neurais artificiais.

ABSTRACT

In the last years, the sanitation companies in Brazil has been broadly investing in their water supply systems automation, providing flow and pressure real time data from the consumption areas. After the automation stage, with all the information available in real time, and full control of the water supply system elements, it turns possible to optimize the operation on the lookout for a continuous improvement in quality of water treatment, distribution services and costs reduction, such as physical losses, energy and chemical products.

The correct decisions upon operational services depend on the previous knowledge of the water consumption along the day in the local distribution areas. The present study aims at introducing several forecasting models using different techniques such as artificial neural networks, multiple linear regression Box and Jenkins models, based on historical data collected by a SCADA (Supervisory, Control and Data Acquisition) system to forecast short terms urban water consumption, particularly for the next 24 hours. The models were applied for three urban consumption areas with distinctive demand profiles from Ponta Grossa city: Uvaranas (residential), Centro Histórico (commercial) and Distrito Industrial (mixed).

Firstly, the data provided by a one year database is collected, filtered and validated in order to turn possible obtaining reliable forecast models. For each one of the consumption areas, it was proposed and compared 6 artificial neural networks models, 3 multiple linear regression models, seven Box and Jenkins models and two hybrid models, intending to acquire a better alternative at forecasting the next 24 hours water consumption. The results were satisfactory considering the proposed objectives, even with the database limitation and the absence of meteorological information, factors influence directly in the water consumption.

Key-words: water demand forecasting; water distribution systems; artificial neural networks.

É preciso cavar o poço antes de ter sede

(Ditado popular)

*A água chia no púcaro que elevo à boca.
«É um som fresco» diz-me quem me dá a bebê-la.
Sorrio. O som é só um som de chiar.
Bebo a água sem ouvir nada com a minha garganta.*

(Fernando Pessoa)

LISTA DE FIGURAS

Figura 1 – Volume não faturado sobre volume produzido (percentual) – dados de 1995, (Silva e Conejo, 1998).....	5
Figura 2 – Possível perda física por extensão de rede ($m^3/Km/dia$) – dados de 1995, (Silva e Conejo, 1998).....	5
Figura 3 – Topologia de rede genérica de um sistema de supervisão e controle de distribuição de água.....	13
Figura 4 – Modelo de um neurônio artificial (Haykin, 2001).....	19
Figura 5 – Modelo de arquitetura de RNA do tipo <i>MLP</i>	21
Figura 6 – Topologia de comunicação do sistema de abastecimento de água de Ponta Grossa – primeira etapa.....	28
Figura 7 – Redução dos dados em relação ao número de consumos de horas passadas para o Centro Histórico.....	31
Figura 8 – Perfil de consumo de uma semana de Uvaranas para mês quente (vermelho) e frio (azul).....	32
Figura 9 – Perfil de consumo de uma semana de Centro Histórico para mês quente (vermelho) e frio (azul).....	33
Figura 10 – Consumo médio ao longo do dia de Uvaranas, Centro Histórico e Distrito Industrial.....	33
Figura 11 – Função de autocorrelação da série de Uvaranas.....	40
Figura 12 – Função de autocorrelação da série do Distrito Industrial.....	40
Figura 13 – Função de autocorrelação da série do Centro Histórico.....	41
Figura 14 – Função de autocorrelação parcial da série de Uvaranas.....	42
Figura 15 – Função de autocorrelação parcial da série do Distrito Industrial.....	42
Figura 16 – Função de autocorrelação parcial da série do Centro Histórico.....	43
Figura 17 – Periodograma acumulado para série de resíduos do AR(25) de Uvaranas.....	45

Figura 18 – Periodograma acumulado para série de resíduos do AR(27) do Centro Histórico.....	45
Figura 19 – Periodograma acumulado para série de resíduos do AR(29) do Distrito Industrial.....	46
Figura 20 – Previsão de consumo da próxima hora para Uvaranas – RNh tangente hiperbólica com 5 neurônios na camada oculta.	51
Figura 21 – Previsão de consumo da próxima hora para Centro Histórico – RLMd2.	52
Figura 22 – RMSE por dia da semana por hora para o Centro Histórico, RNd2 com função de ativação tangente hiperbólica – previsão horária.	53
Figura 23 – Previsão de consumo da próxima hora para Uvaranas – AR(25).	54
Figura 24 – RMSE por dia da semana por hora para o Centro Histórico, AR(29) – previsão horária.	56
Figura 25 – Previsão de consumo das próximas 24 horas – Uvaranas.	58
Figura 26 – Previsão de consumo das próximas 24 horas – Uvaranas.	59
Figura 27 – Previsão de consumo das próximas 24 horas – Uvaranas.	61
Figura 28 – Histogramas dos resultados da previsão de consumo das próximas 24 horas – Uvaranas	64
Figura 29 – Histogramas dos resultados da previsão de consumo das próximas 24 horas – Distrito Industrial.....	65
Figura 30 – Histogramas dos resultados da previsão de consumo das próximas 24 horas – Centro Histórico.....	66
Figura 31 – Histogramas dos resultados da previsão de consumo das próximas 12 horas – Uvaranas.....	66
Figura 32 – Variação do erro de previsão de consumo das próximas 24 horas em relação ao número de semanas para cálculo da parcela de ajuste – Uvaranas	68
Figura 33 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Uvaranas.....	78

Figura 34 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Uvaranas.....	78
Figura 35 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Distrito Industrial.	79
Figura 36 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Distrito Industrial.	80
Figura 37 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Centro Histórico.	81
Figura 38 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Centro Histórico.	81
Figura 39 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Uvaranas.....	85
Figura 40 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Uvaranas.....	85
Figura 41 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Distrito Industrial.	86
Figura 42 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Distrito Industrial.	87
Figura 43 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Centro Histórico.	88
Figura 44 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Centro Histórico.	88

LISTA DE TABELAS

Tabela 1 – Comparação de erros RMSE e MAPE para as redes neurais e regressão linear múltipla –Uvaranas.....	49
Tabela 2 – Comparação de erros RMSE e MAPE para as redes neurais e regressão linear múltipla – Distrito Industrial.	50
Tabela 3 – Comparação de erros RMSE e MAPE para as redes neurais e regressão linear múltipla –Centro Histórico.	50
Tabela 4 – Comparação de erros RMSE e MAPE para modelos ARMA(p,q) – bairro Uvaranas.....	54
Tabela 5 – Comparação de erros RMSE e MAPE para modelos ARMA(p,q) – Distrito Industrial.....	55
Tabela 6 – Comparação de erros RMSE e MAPE para modelos ARMA(p,q) – bairro Centro Histórico.	55
Tabela 7 – Erros RMSE e MAPE para previsões das próximas 24 horas com modelos de redes neurais, regressão linear múltipla e híbrido para previsão a partir de 9:00 de 23/12/2003 – Uvaranas.....	58
Tabela 8 – Erros RMSE e MAPE para previsões das próximas 24 horas com modelos de redes neurais, regressão linear múltipla e híbrido previsão a partir de 9:00 de 23/12/2003 – Uvaranas.....	60
Tabela 9 – Erros RMSE e MAPE para previsões das próximas 24 horas com modelos de Box e Jenkins a partir de 9:00 de 23/12/2003 – Uvaranas.	61
Tabela 10 – Índices dos modelos do tipo RNA e RLM.....	63
Tabela 11 – Índices dos modelos tipo RNA e RLM.....	75
Tabela 12 – Matriz de modelos tipo RNA e RLM escolhidos segundo critério de erro RMSE para a previsão horária do bairro de Uvaranas.....	75
Tabela 13 – Matriz de modelos tipo RNA e RLM escolhidos segundo critério de erro RMSE para a previsão horária do Centro Histórico.	76

Tabela 14 – Matriz de modelos tipo RNA e RLM escolhidos segundo critério de erro RMSE para a previsão horária do Distrito Industrial.	77
Tabela 15 – Modelos tipo RNA e RLM escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Uvaranas segundo critério de erro RMSE.....	77
Tabela 16 – Modelos tipo RNA e RLM escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Distrito Industrial segundo critério de erro RMSE.	79
Tabela 17 – Modelos tipo RNA e RLM escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Centro Histórico segundo critério de erro RMSE.	80
Tabela 18 – Índices dos modelos tipo Box e Jenkins	82
Tabela 19 – Matriz de modelos tipo Box e Jenkins escolhidos segundo critério de erro RMSE para a previsão horária do bairro de Uvaranas.	82
Tabela 20 – Matriz de modelos tipo Box e Jenkins escolhidos segundo critério de erro RMSE para a previsão horária do Centro Histórico.	83
Tabela 21 – Matriz de modelos tipo Box e Jenkins escolhidos segundo critério de erro RMSE para a previsão horária do Distrito Industrial.	84
Tabela 22 – Modelos tipo Box e Jenkins escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Uvaranas segundo critério de erro RMSE.....	84
Tabela 23 – Modelos tipo Box e Jenkins escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Distrito Industrial segundo critério de erro RMSE.	86
Tabela 24 – Modelos tipo Box e Jenkins escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Centro Histórico segundo critério de erro RMSE.	87

LISTA DE ABREVIATURAS E SIGLAS

ARIMA	- <i>Auto Regressive Integrated Moving Average</i>
ARMA	- <i>Auto Regressive Moving Average</i>
CLP	- Controlador Lógico Programável
EEB	- Estação elevatória de água bruta
EET	- Estação elevatória de água tratada
ETA	- Estação de tratamento de água
Fac	- função de autocorrelação
facp	- função de autocorrelação parcial
FT	- <i>Flow transmitter</i> (Transmissor de vazão)
LT	- <i>Level transmitter</i> (Transmissor de nível)
MAPE	- <i>Mean absolute Percentage Error</i> (erro percentual absoluto médio)
MLP	- <i>Multilayer perceptron</i>
MSE	- <i>Mean Square Error</i> (Erro quadrático médio)
PCV	- <i>Pressure control valve</i> (Válvula controladora de pressão)
PT	- <i>Pressure transmitter</i> (Transmissor de pressão)
RAP	- Reservatório apoiado
RMSE	- <i>Root Mean Square Error</i> (Raiz do erro quadrático médio)
RNA	- Redes Neurais Artificiais
SAA	- Sistema de abastecimento de água
SC	- <i>Speed controller</i> (Variador de frequência)
SCADA	- <i>Supervisory, Control and Data Acquisition</i> (Supervisão, Controle e Aquisição de Dados)
SQL	- <i>Structured Query Language</i>

LISTA DE SÍMBOLOS

$a_{j(ds,h)}$	- fator de ajuste para previsão de modelos auto-regressivos segundo hora e dia da semana
a_t	- termos aleatórios (modelo ARIMA)
d_1	- primeira diferença entre consumos
d_2	- segunda diferença entre consumos
ed_1	- entrada senoidal para dia da semana
ed_2	- entrada senoidal para dia da semana
eh_1	- entrada senoidal para hora
eh_2	- entrada senoidal para hora
$I(a)$	- periodograma da série temporal
k_1	- coeficiente do dia de maior consumo
k_2	- coeficiente da hora de maior consumo
$pa(f)$	- espectro da série temporal
$Pa(f)$	- função espectral acumulada da série temporal
qm	- vazão média diária anual
$RLMd1$	- estrutura de dados para regressão linear múltipla com data, consumo da última hora e diferença entre dois últimos consumos
$RLMd2$	- estrutura de dados para regressão linear múltipla com data, consumo da última hora, diferença entre dois últimos consumos e diferença entre duas últimas diferenças
$RLMh$	- estrutura de dados para regressão linear múltipla com data e consumo da última hora
$RNd1$	- estrutura de dados para redes neurais com data, consumo da última hora e diferença entre dois últimos consumos.
$RNd2$	- estrutura de dados para redes neurais com data, consumo da última hora, diferença entre dois últimos consumos e diferença entre duas últimas diferenças.
RNh	- estrutura de dados para redes neurais com data e consumo da última hora
θ_j	- termos referentes às médias móveis (modelo ARIMA)
φ_i	- termos auto-regressivos (modelo ARIMA)

SUMÁRIO

RESUMO	v
ABSTRACT	vi
LISTA DE FIGURAS	viii
LISTA DE TABELAS	xi
LISTA DE ABREVIATURAS E SIGLAS	xiii
LISTA DE SÍMBOLOS	xiv
1 INTRODUÇÃO	1
1.1 ESCOPO GERAL DO PROBLEMA	1
1.2 OBJETIVOS	1
1.3 JUSTIFICATIVA	2
1.4 ESTRUTURA DO TRABALHO	3
2 ANÁLISE DO PROBLEMA	4
2.1 CONSUMO DE ÁGUA	4
2.2 FATORES QUE INFLUENCIAM O CONSUMO	6
2.3 VARIAÇÕES DE CONSUMO	8
2.4 PREVISÃO DE CONSUMO DE ÁGUA	9
2.5 AUTOMAÇÃO EM SISTEMAS DE ABASTECIMENTO DE ÁGUA	10
2.6 A PREVISÃO DE CONSUMO DE ÁGUA EM SISTEMAS DE ABASTECIMENTO	15
2.7 A DETERMINAÇÃO DE UM MODELO DE PREVISÃO DE CONSUMO DE ÁGUA	17
2.7.1 PRINCIPAIS TÉCNICAS EMPREGADAS	17
2.7.2 REGRESSÃO LINEAR MÚLTIPLA	18
2.7.3 MODELOS TIPO BOX E JENKINS	18
2.7.4 REDES NEURAS ARTIFICIAIS	19
2.8 ESTUDOS PREVIAMENTE REALIZADOS	22
3 METODOLOGIA	26
3.1 COLETA E FILTRAGEM DOS DADOS	27
3.2 VARIÁVEIS UTILIZADAS PARA OS MODELOS DE PREVISÃO	30
3.3 CARACTERIZAÇÃO DO PERFIL DE CONSUMO	31
3.4 APLICAÇÃO DOS MODELOS DE PREVISÃO	34
3.4.1 PREVISÃO HORÁRIA COM REDES NEURAS ARTIFICIAIS	35
3.4.2 PREVISÃO HORÁRIA COM REGRESSÃO LINEAR MÚLTIPLA	37
3.4.3 PREVISÃO COM MODELOS BOX E JENKINS	38
3.5 MODELOS HÍBRIDOS	47
4 RESULTADOS	49
4.1 RESULTADOS PARA PREVISÃO HORÁRIA COM REDES NEURAS E REGRESSÃO LINEAR MÚLTIPLA	49
4.2 RESULTADOS PARA PREVISÃO HORÁRIA COMO MODELOS BOX E JENKINS	54
4.3 PREVISÃO DAS PRÓXIMAS 24 HORAS	57
4.3.1 Modelos tipo RNA e RLM	57
4.3.2 Modelos tipo Box e Jenkins	60

5	CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS	69
	REFERÊNCIAS	72
	APÊNDICE A – TESTES COM MODELOS DE RNA E RLM	75
	APÊNDICE B – TESTES COM MODELOS TIPO BOX E JENKINS	82

1 INTRODUÇÃO

1.1 ESCOPO GERAL DO PROBLEMA

O recente investimento em automação nos sistemas de abastecimento de água (SAA) pelas companhias de saneamento tem proporcionado uma gama de informações em tempo real de vazão e de pressão das cidades, além da possibilidade de atuação automática sobre os elementos do SAA, tais como bombas, inversores de frequência, válvulas de bloqueio, válvulas controladoras de pressão e vazão, dosadoras de produtos químicos, agitadores, etc. Alcançada a flexibilidade operacional, é possível otimizar a operação dos SAA para um melhoramento dos serviços de captação, tratamento e distribuição de água e redução de custos relacionados, tais como energia, produtos químicos e perdas físicas de água.

A correta tomada de decisões operacionais depende do conhecimento prévio do perfil de demanda ao longo do dia das regiões de consumo de água. Este tipo de previsão é classificado como previsão a curto prazo.

Diversas são as vantagens de se obter um modelo de previsão de consumo de água para otimização da distribuição de água, tais como a identificação imediata de vazamentos, ganho na qualidade da água distribuída, redução de custos de energia e planejamento otimizado da operação da estação de tratamento de água de modo a obter os melhores pontos de operação dos sistemas que compõem esta planta, principalmente daqueles relacionados a aplicações de produtos químicos.

Diversas técnicas de previsão baseadas em dados históricos podem ser aplicadas para este fim, dentre as quais foram utilizadas as técnicas de redes neurais artificiais, regressão linear múltipla e modelos do tipo Box e Jenkins .

1.2 OBJETIVOS

Este trabalho objetiva introduzir o contexto do problema de previsão de consumo urbano de água a curto prazo e sugerir diversas técnicas possíveis de previsão, análise e comparação de seus resultados de previsão de consumo de água.

Para aplicação das técnicas de previsão, foram utilizados os dados históricos de três regiões de consumo da cidade de Ponta Grossa (PR), com características de demanda distintas: Uvaranas (residencial), Distrito Industrial (misto) e Centro Histórico (comercial), cujos sistemas de abastecimento de água foram recentemente automatizados, e seus dados disponibilizados pela Companhia de Saneamento do Paraná – Sanepar.

Neste contexto, deseja-se alcançar um ou mais modelos de previsão suficientemente confiáveis a ponto de determinar ações de operação do sistema de abastecimento de água, ou indicar condições anormais em sua operação.

1.3 JUSTIFICATIVA

Considerando os modelos de previsão de consumo de água a curto prazo, o objetivo principal é aperfeiçoar a operação do sistema de abastecimento de água – SAA, de forma a oferecer um serviço / produto de melhor qualidade, mais confiável e a um menor custo.

De posse de um modelo de previsão de consumo confiável, a operação ótima do SAA, ao longo do dia, pode ser planejado objetivando finalidades de:

- a) Melhoria na qualidade do produto (água): será alcançada por intermédio da manutenção de estoques adequados nos reservatórios de distribuição. Por um lado, a potabilidade da água será favorecida eliminando-se a necessidade de superdosagem de produtos químicos na estação de tratamento de água (ETA) e, conseqüentemente, uma redução do custo de produção. Por outro lado, evita-se a perda de qualidade da água ocasionada pelo tempo alto de retenção entre produção e consumo.
- b) Melhoria na qualidade do serviço: é alcançada à medida que se evita a falta de água nas regiões mais críticas nos períodos de pico de consumo, além de uma ação eficaz na busca e detecção de vazamentos na rede.
- c) Menor custo: através de um planejamento correto da operação, reduz-se os custos com produtos químicos na ETA, custos de perda física na rede (vazamentos) e custos de energia, evitando-se operação de bombas em horários de picos de consumo de energia em detrimento de distribuição por

gravidade ou válvulas controladoras de pressão / vazão; contribuindo, portanto, com a atenuação de dois dos principais problemas ambientais da atualidade: desperdício de água e consumo elevado de energia.

1.4 ESTRUTURA DO TRABALHO

No presente capítulo há uma breve introdução ao trabalho com a descrição dos objetivos e justificativa.

O capítulo 2 aborda o problema do consumo de água de uma forma mais ampla e sua importância no contexto operacional dos sistemas de distribuição de água. Neste capítulo são apresentados os principais fatores determinantes do consumo de água, assim como os tipos de regiões de consumo e suas características. O tema da automação dos sistemas de abastecimento de água é abordado de forma a elucidar os possíveis problemas encontrados na coleta dos dados. Além disso, são apresentadas as técnicas de previsão empregadas neste estudo e uma revisão bibliográfica sobre o assunto da previsão de consumo de água.

No capítulo 3 é apresentada a metodologia adotada para obtenção dos modelos de consumo de água.

O capítulo 4 apresenta os resultados obtidos para previsão de consumo horário e nas próximas 24 horas das três regiões de consumo estudadas.

O capítulo 5 conclui o trabalho e sugere estudos futuros.

2 ANÁLISE DO PROBLEMA

2.1 CONSUMO DE ÁGUA

A água conduzida a uma cidade enquadra-se em uma das seguintes classes de consumo, ou de destino (Yassuda *et al.*, 1976):

a) Uso doméstico: água consumida nas habitações, compreendendo as parcelas destinadas a fins higiênicos, potáveis e alimentares, e à lavagem em geral. Esta classe pode ser dividida como de uso interno e externo, dado que a água para o uso externo destina-se à lavagem de calçadas, rega de jardins, lavagem de carros, etc., enquanto que a água para o uso interno destina-se ao consumo doméstico em geral. O consumo de uso doméstico varia com o nível sócio-econômico da população, sendo maior quanto mais elevado for este padrão;

b) Uso comercial ou industrial: água de uso comercial é destinada a bares, restaurantes, hotéis, postos de gasolina, garagens, entre outros, enquanto que a água de uso industrial é utilizada como matéria-prima para lavagem, refrigeração, alimentação de caldeiras e processos industriais em geral. Esta classe representa um consumo significativamente mais elevado que o de uso doméstico;

c) Uso público: parcela de água destinada à rega de jardins, lavagem de ruas e passeios, edifícios e sanitários de uso público, alimentação de fontes, esguichos e tanques fluxíveis de redes de esgoto;

d) Perdas: sob o ponto de vista operacional dos sistemas públicos de abastecimento, são consideradas perdas de água os volumes não contabilizados, dividindo-se em perdas físicas (parcela não consumida) e perdas não físicas (água consumida e não registrada). As perdas físicas originam-se de vazamentos do sistema, envolvendo captação, adução de água bruta, tratamento, reserva, adução de água tratada e distribuição de água, além de procedimentos operacionais. As perdas não físicas originam-se de ligações clandestinas ou não cadastradas, hidrômetros parados ou que submedem, fraudes em hidrômetros entre outros. A redução das perdas físicas permite

diminuir os custos de produção (energia, produtos químicos e outros), enquanto que a redução das perdas não físicas permite aumentar a renda tarifária, melhorando a eficiência dos serviços prestados e o desempenho financeiro do prestador de serviços (Silva e Conejo, 1998). As figuras 1 e 2 apresentam percentuais de perdas aproximados de diferentes estados brasileiros.

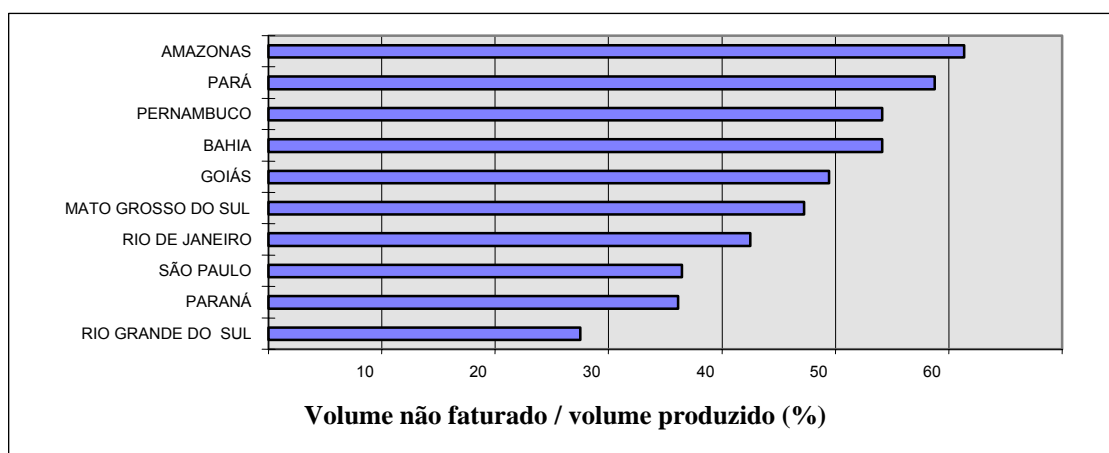


Figura 1 – Volume não faturado sobre volume produzido (percentual) – dados de 1995, (Silva e Conejo, 1998).

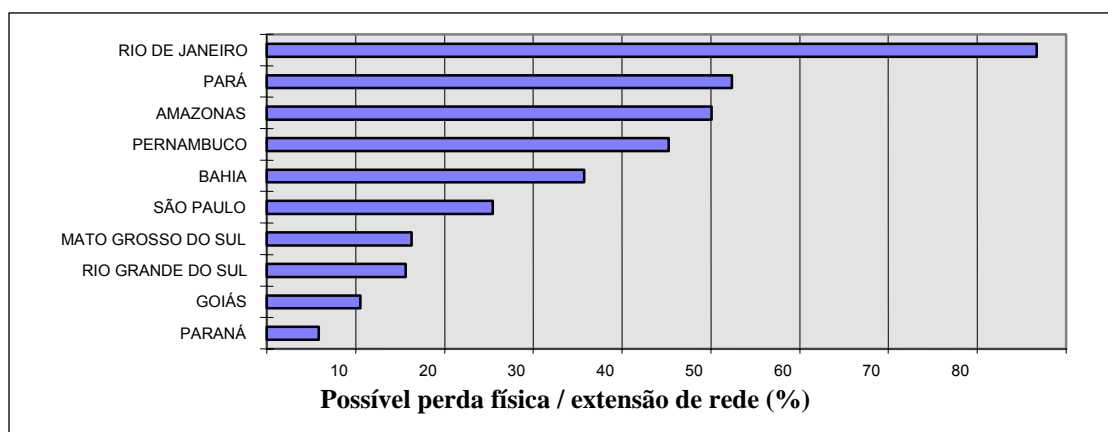


Figura 2 – Possível perda física por extensão de rede ($m^3/Km/dia$) – dados de 1995, (Silva e Conejo, 1998).

2.2 FATORES QUE INFLUENCIAM O CONSUMO

Diversos são os fatores que afetam o consumo de água de uma cidade (Yassuda *et al.*, 1976), ou, particularmente, de um bairro ou região de consumo, dentre os quais podem ser citados:

- a) Clima: o consumo aumenta conforme aumenta a temperatura. A umidade também exerce influência, dado que em regiões mais secas o consumo é maior. A presença de chuvas afeta diretamente o consumo, reduzindo-o drasticamente na sua ocorrência;
- b) Hábitos e nível de vida da população: os hábitos da população afetam o uso direto ou indireto da água, tais como banhos, lavagem de calçadas, lavagem de carros, rega de jardins, etc. Quanto melhor o nível de vida da população, maior é o consumo de água, devido a um maior número de possibilidades de emprego da água, tais como área maior de jardins, carros, máquinas de lavar roupa, piscinas, entre outras aplicações que visam o conforto e facilidades;
- c) Natureza da cidade ou bairro: bairros residenciais, comerciais e industriais apresentam um perfil de consumo diferenciado. As áreas industriais apresentam um consumo consideravelmente maior que os bairros tipicamente residenciais;
- d) Tamanho da cidade: a experiência tem demonstrado que o consumo de água tende a aumentar conforme o aumento da população da cidade. Entre os fatores que explicam esta tendência destaca-se a maior demanda industrial, comercial e pública, além do aumento do número de ligações de água e, proporcionalmente, as perdas físicas e não físicas da rede hidráulica;
- e) Localização geográfica da cidade: cidades litorâneas apresentam uma curva de consumo diferente das cidades localizadas em regiões serranas, principalmente pelos hábitos diferenciados da população e diferentes atividades econômicas;
- f) Turismo: cidades turísticas incorporam forte sazonalidade em certas épocas do ano;

g) Estação do ano: o consumo é diferenciado nas quatro estações do ano, basicamente pela interferência dos fatores climáticos;

h) Dias da semana e feriados: existem comportamentos variados de consumo às segundas-feiras em relação aos domingos, sábados e restante dos dias da semana. As pessoas adquirem certos hábitos que realçam esta diferenciação, tais como o fato de preferirem lavar roupas na segunda-feira ou nos sábados;

i) Regras de consumo: algumas regras de consumo são evidenciadas ao analisar o comportamento de uma população, por exemplo, após 2 ou 3 dias consecutivos de chuva, a presença de um dia ensolarado causa um alto consumo de água, visto que muitas pessoas aproveitam a oportunidade para lavarem suas roupas, carros e calçadas;

j) Medição de consumo: a ausência de medição efetiva e individual nas instalações prediais favorece o desperdício e a despreocupação com vazamentos;

k) Administração pública: a iniciativa política e uma preocupação maior com a manutenção e limpeza das áreas de uso público, tais como parques, praças e pavimentos edifícios públicos em geral;

l) Preço do serviço: o preço elevado das tarifas de água e esgoto promove a redução do consumo de água, interferindo diretamente nos hábitos da população;

m) Pressão na rede: quanto maior a pressão da rede hidráulica, maior o consumo devido à maior saída de água mesmo com pequena abertura das válvulas e torneiras, além do aumento de perdas. Segundo a norma técnica NBR 12218/1994, a pressão estática máxima das tubulações distribuidoras deve ser de 500 KPa (50 mca), e a pressão dinâmica mínima, de 100 KPa (10 mca). Valores fora desta faixa podem ser aceitos desde que justificados técnica e economicamente (Yoshimoto *et al.*, 1999). Portanto, as redes distribuidoras devem trabalhar com pressão mais reduzida possível, contanto que seja suficiente para atender às condições de pressão mínima definida pela norma.

Dado que o consumo de água é influenciado por diversos fatores tais como condições de tempo, variações sazonais, dia da semana e se um determinado dia é

ocasionalmente feriado, o consumo do sistema de distribuição de água urbano é uma série temporal variável no tempo, periódica, e não estacionária (An *et al.*, 1996).

2.3 VARIAÇÕES DE CONSUMO

Dado que o consumo apresenta variações ao longo do ano devido aos diversos fatores mencionados na seção 2.2, é possível determinar uma parcela proveniente da variação diária e uma parcela proveniente da variação horária de consumo. A variação diária é decorrente dos dias da semana e às diferentes épocas do ano, enquanto que a variação horária reflete os hábitos diários da população.

Dado o consumo total distribuído em um ano, define-se a vazão média diária anual conforme a seguinte equação.

$$q_m = \frac{\text{volume distribuído em um ano}}{365} \quad (1)$$

A relação entre o maior consumo diário verificado e a vazão média diária anual fornece o coeficiente do dia de maior consumo é dado por:

$$k_1 = \frac{\text{maior consumo diário no ano}}{\text{vazão média diária no ano}} \quad (2)$$

A relação entre a maior vazão horária observada num dia e a vazão média horária do mesmo dia fornece o coeficiente da hora de maior consumo, dado por:

$$k_2 = \frac{\text{maior vazão horária no dia}}{\text{vazão média horária no dia}} \quad (3)$$

O coeficiente k_1 é utilizado na determinação da vazão de dimensionamento de várias partes constitutivas de um sistema de abastecimento de água, tais como a captação, bombas, adutoras e estações de tratamento de água.

O coeficiente k_2 é utilizado no dimensionamento dos condutos de distribuição que partem dos reservatórios, pois permite conhecer as condições de maior solicitação das tubulações (Yassuda *et al.*, 1976).

Os reservatórios de distribuição permitem armazenar água nos períodos de menor consumo para suprir as necessidades dos períodos de maior consumo e demandas de emergência. O conhecimento prévio do perfil de consumo do dia fornece a vazão média do dia e as vazões das horas de maior consumo, possibilitando o planejamento de operação do SAA e limites de reserva necessários para o dia.

O planejamento inadequado da reserva pode tanto ocasionar desabastecimento nas horas de maior consumo, em caso de reserva deficiente, quanto ocasionar perda de qualidade da água, em caso de reserva muito superior à necessária.

2.4 PREVISÃO DE CONSUMO DE ÁGUA

Em geral, os modelos de previsão de demanda são baseados em dados históricos e podem ser classificados, segundo Jones *et al.*¹ *apud* (Silva e Rocha, 1999), como:

- a) Previsão: em um conceito mais amplo, abrange qualquer tipo de informação sobre o futuro;
- b) Estimativa: previsão condicional, baseada em pressupostos implícitos ou explícitos;
- c) Projeção: estimativa baseada em pressupostos que reconhecem, ao menos em parte, a continuação de uma ou mais tendências passadas;
- d) Extrapolação: uma estimativa baseada em pressupostos que se baseiam inteiramente na continuação de tendências passadas.

Basicamente pode-se dividir o tipo de previsão de consumo de água de acordo com suas finalidades em duas modalidades, (Silva e Rocha, 1999):

- a) Previsão a longo prazo: baseada na combinação de padrões de consumo com projeções demográficas tem como principal finalidade servir de parâmetro

¹ Jones, C.V.; Boland, J.J.; Crews, J.E.; et al. (1984) - Municipal water demand: statistical and management issues. Westview Press. Boulder.

para estimativa de capacidade dos sistemas produtores, solicitações máximas aos sistemas de distribuição e previsão de investimentos para futuras ampliações dos sistemas de abastecimentos de água;

b) Previsão a curto prazo: refere-se à previsão das próximas horas ou até dias. Este tipo de previsão fornece informações para uma resposta instantânea dos sistemas produtores e de distribuição a picos prolongados e flutuações acentuadas de demanda e elementos para tomada de decisões operacionais para otimizar a distribuição. Permite a avaliação de benefício líquido potencial de médias de conservação e uso racional da água voltada à redução de picos de consumo. Também identifica possíveis ocorrências de vazamentos na rede, agilizando a manutenção e reduzindo as perdas ocasionadas.

As previsões de curto prazo são mais complexas que as previsões de longo prazo, visto que estas são menos suscetíveis a variações localizadas de consumo e, quando necessário, são passíveis de revisão parcial ao longo de sua vigência. Já as de curto prazo envolvem elementos de operação e, quando erradas, são mais evidentes e menos passíveis de correção. (Silva e Rocha, 1999).

2.5 AUTOMAÇÃO EM SISTEMAS DE ABASTECIMENTO DE ÁGUA

O investimento em automatização dos sistemas de abastecimento de água justifica-se à medida que os sistemas tornam-se cada vez mais extensos e complexos. Basicamente, um sistema de abastecimento de água automatizado possui uma estação central de controle e operação onde deve estar o sistema SCADA (*Supervisory, Control and Data Acquisition*), recebendo informações de todas as áreas remotas do sistema, normalmente comandadas por controladores lógicos programáveis (CLPs), através de uma rede de comunicação de dados. Desta forma, na estação de controle é possível monitorar todas as variáveis importantes do sistema (vazão, pressão, nível, características físico-químicas da água, entre outros), assim como comandar e supervisionar o estado dos diversos equipamentos envolvidos na operação, tais como bombas, inversores de frequência, válvulas de bloqueio, válvulas controladoras de pressão e vazão, dosadoras de produtos químicos, agitadores, etc. Qualquer anormalidade operacional também é alarmada

no sistema de supervisão, a fim de que as resoluções necessárias sejam tomadas em tempo hábil.

Um SAA automatizado oferece uma gama de possibilidades adicionais em termos de melhorias de operação da planta, na tomada de decisões nas diversas situações operacionais, na redução de custos de energia, e na redução da pressão na rede, através de inversores de frequência, válvulas redutoras de pressão ou válvulas controladoras de pressão. Desta forma, o perfil de consumo tende a mudar depois de implantada a automatização do sistema de abastecimento de água. Por conseguinte, o uso de dados de consumo de origem anterior à implantação da automatização do sistema de abastecimento de água incutiria erros na previsão de consumo futuro pós-automação.

A topologia de rede de um sistema de abastecimento de água é composta, de forma genérica, pela camada mais baixa de CLPs remotos, que podem ler medidas de vazão locais através de rede ou de placas de entradas analógicas, ou então receber remotamente essas medições via modem ou rádio. Em um nível acima se encontram os CLPs mestres, que por sua vez também coletam dados dos medidores de vazão locais ou remotos. Recebendo os dados dos CLPs, estão as estações SCADA, permitindo a visualização e a operação de toda planta. A figura 3 apresenta uma topologia de rede genérica de um sistema de abastecimento de água automatizado, apresentando os principais tipos de ligações em rede dos medidores de vazão.

No canto inferior esquerdo da figura 3, a EEB01, ou estação elevatória de água bruta, verificada a qualidade do corpo de água através dos analisadores (AIT – *analytical indicator transmitter*), através de bombas com inversores de frequência (SC – *speed control*) recalcam uma vazão definida pela estação de tratamento de água (ETA), determinado de acordo com o volume de água necessário para tratar. Dois medidores de vazão (FT – *flow transmitter*), um na saída da captação e outro na chegada da ETA indicam o valor para controle do volume de água bruta estipulado e verificam a estanqueidade da linha, alarmando em caso de vazamentos. Após processos de tratamento, a água é reservada em um reservatório apoiado, neste caso (RAP01), de onde é distribuída para as regiões de consumo. Parte é distribuída para a região de consumo 1 via gravidade, e uma bomba *booster*

com inversor de frequência (EET02) controla a pressão jusante em um ponto remoto na região de consumo 2, via medida de pressão (PT – *pressure transmitter*) e valor de referência (*set point*) de pressão determinado. O controle de pressão normalmente ocorre através da instalação dos medidores de pressão remotos nos pontos mais altos da rede de distribuição, ou menos favorecidos, podendo haver mais de um medidor de pressão por elevatória de distribuição. Cada medidor de pressão, por sua vez, pode ter diferentes valores de referência de pressão (*set point*), tal que o *set point* escolhido para controle pela elevatória, bem como a variável medida, passará por uma lógica de seleção de medidores, tal que seja escolhido o medidor remoto com o maior desnível negativo ou menor desvio positivo de pressão em relação ao *set point* estipulado. Desta forma, garante-se o abastecimento nas regiões menos privilegiadas.

O nível do RAP01, indicando condições de operação servirá como segurança da EET01, que através de inversores de frequência controla a pressão na região de consumo 3, verificando entre os dois medidores de pressão remotos aquele que estiver em condições mais críticas. À medida que aumentar o consumo, os inversores de frequência tendem a acelerar, aumentando a vazão recalçada. Ao chegar a um limite de vazão estipulado, a EET01 passa a controlar vazão, a fim de gerenciar o volume de vazão destinado para cada região. Quando a vazão impulsionada pela EET01 não for suficiente para atender à região de consumo 3, a válvula controladora de pressão (PCV – *pressure control valve*) começa a atuar controlando a pressão da região de consumo 3, conforme seleção dos PT's. À medida que o consumo decresce, a PCV01 tende a fechar e a EET01 volta a controlar pressão na região de consumo 3. O nível baixo do RAP02 acionará a EET01 a alimenta-lo em um período do dia que não haja consumo elevado. Os medidores de vazão indicam o consumo de cada região, tornando possível verificar o perfil de consumo individual auxiliando no planejamento da operação.

A maleabilidade de programação do CLP facilita a operação de equipamentos envolvidos no SAA de uma forma mais segura. Os CLPs de cada área controlam os processos locais e enviam informações relevantes para CLPs de outras áreas. O sistema de supervisão monitora os estados de todas as áreas remotas através dos dados recebidos dos CLPs.

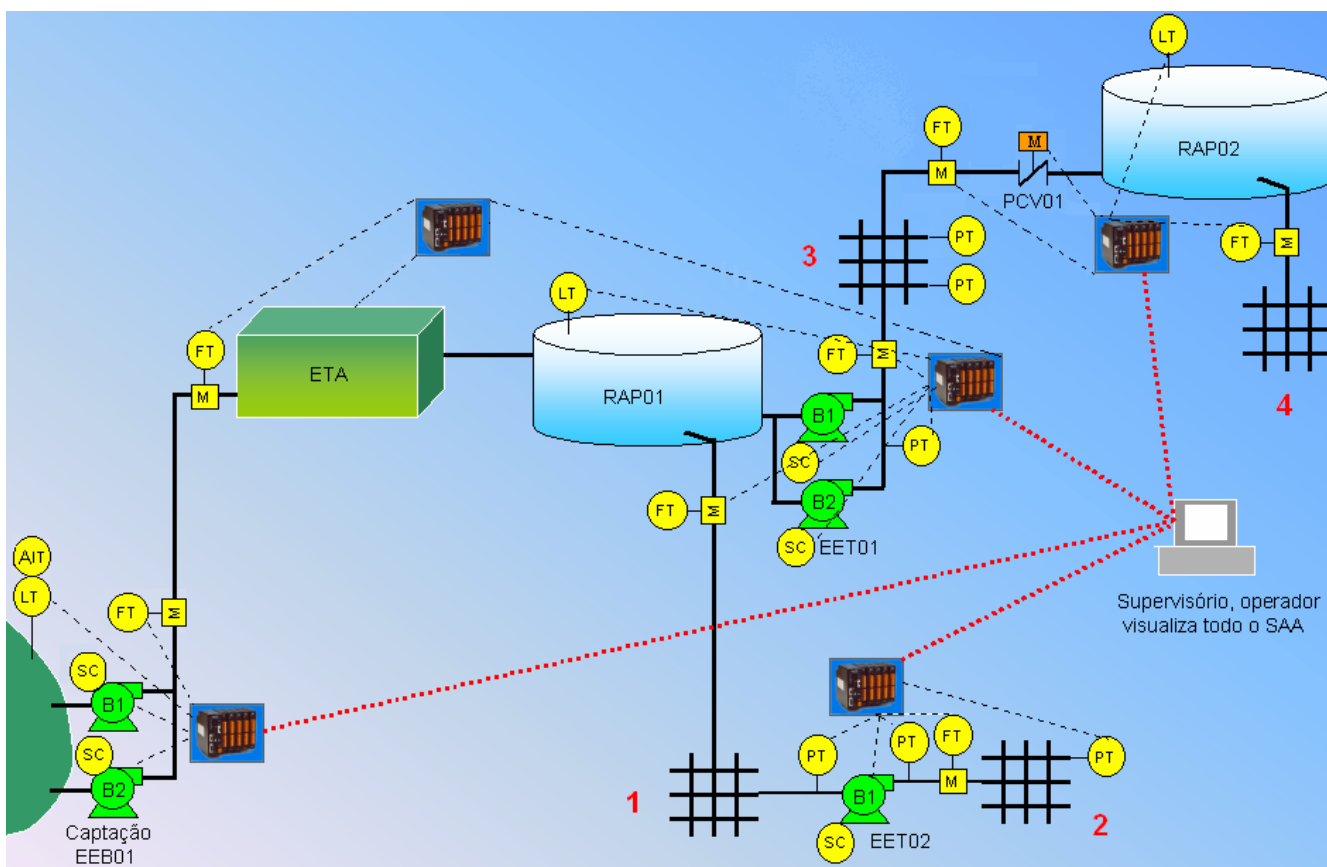


Figura 3 – Topologia de rede genérica de um sistema de supervisão e controle de distribuição de água.

Nos sistemas da Sanepar, o sistema de supervisão insere os dados de interesse em um banco de dados em intervalo de tempo estipulado, normalmente a cada 15 minutos. Os dados de interesse, no caso do problema aqui abordado, são os volumes de água que passam pelos medidores de vazão a cada intervalo de tempo. Caso os medidores de vazão estejam ligados em cartões de entradas analógicas nos CLPs, normalmente esta medição de volume deve ser configurada no medidor de vazão para indicar um pulso a cada medida de volume, normalmente um metro cúbico. Este pulso é lido pelo CLP através de uma entrada digital e incrementa um contador que é transmitido ao sistema de supervisão. O erro de leitura é, portanto, de $\pm 1 \text{ m}^3$ além do erro específico do medidor de vazão. Caso os medidores de vazão estejam em rede, a totalização é contínua, lida diretamente através de uma memória específica, sem necessidade de processamento indireto

pelo CLP, com apenas erro específico do medidor e, portanto, apresentado maior precisão.

A fim de que os dados de volume de água consumido coletados pelo banco de dados sejam confiáveis é necessário avaliar os tipos de falhas, interferências, ou medidas errôneas passíveis de ocorrência.

Os tipos de falha variam de acordo com a topologia de rede de cada sistema, mas de forma geral, podem-se citar algumas situações que causam a incerteza de uma leitura são:

- a) Falhas no medidor de vazão;
- b) Falha de comunicação em algum ponto da rede (entre o medidor de vazão e o CLP, entre CLPs ou entre CLPs e o sistema de supervisão): em geral, a falha de comunicação provoca ou congelamento dos dados ou falta de dados;
- c) Falta de energia prolongada: a falta de energia prolongada impede a operação adequada dos equipamentos envolvidos (bombas, inversores, válvulas, etc.) causando desabastecimento da rede hidráulica, demandando um tempo significativo no retorno da energia para que as condições de consumo voltem à normalidade;
- d) Falhas de equipamentos envolvidos: falha de bombas, inversores, válvulas, CLPs, etc;
- e) Vazamentos: o vazamento provoca consumo excessivo, fora dos padrões normais de consumo;
- f) Manutenção da rede hidráulica: a manutenção da rede hidráulica implica em operação de válvulas manuais, modificando a configuração original da rede a fim de abastecer as áreas envolvidas através de outras vias, o que torna o perfil de consumo equivocado durante este período. Para que esta situação seja considerada, deve haver uma intervenção manual do operador no sistema de supervisão.

A fim de que os dados de calibração e determinação dos parâmetros dos modelos de previsão de demanda sejam corretos, é necessária uma filtragem dos mesmos, conjugando os sinais de falha de comunicação e falhas de sensor com os

de leitura totalizada do volume de consumo. Da mesma forma, devem ser verificadas as datas onde ocorreram vazamentos e manutenções da rede hidráulica, a fim de que os dados coletados nestas ocasiões sejam ignorados. Esta última providência é mais complexa dado que requer uma influência manual para identificar no banco de dados o período dos registros a serem excluídos.

Portanto, ao se planejar um sistema de supervisão com treinamento automático de modelos de previsão, as possíveis falhas citadas anteriormente devem ser consideradas para validar os novos dados de treinamento.

2.6 A PREVISÃO DE CONSUMO DE ÁGUA EM SISTEMAS DE ABASTECIMENTO

De acordo com Petrella (2002) um dos motivos que explicam a crise mundial da água² seriam todos os fatores relacionados ao desperdício e ao mau gerenciamento dos recursos disponíveis. Os estudos sobre perdas nos sistemas de abastecimento de água de Silva *et al.* (1998) indicam que o índice de perdas no faturamento dos sistemas de abastecimento de água brasileiros (volume não faturado sobre volume produzido) varia entre aproximadamente 27% e 60%. As perdas de água podem ser divididas em perdas físicas, ocasionadas por vazamentos desde a captação até a distribuição, e perdas não físicas, ocasionadas por ligações clandestinas, defeitos em hidrômetros, fraudes, entre outras.

Segundo Yoshimoto *et al.* (1999), o efetivo controle de perdas físicas é feito através de quatro atividades complementares:

- a) gerenciamento de pressão;
- b) controle ativo de vazamentos;
- c) velocidade e qualidade dos reparos;
- d) gerenciamento da infra-estrutura.

² Existem mais de 1.4 bilhão de pessoas que não têm acesso à água potável, e mais de outros 2 bilhões de pessoas não têm qualquer sistema sanitário doméstico ou de purificação da água de esgoto. Se nada for feito para reverter a tendência atual, no ano 2025 o número de pessoas sem acesso à água potável alcançará mais de 4 bilhões, a metade da população mundial Petrella (2002).

A automação dos sistemas de abastecimento de água é, portanto, indispensável a fim de atingir tais objetivos, reduzindo a pressão na rede através de inversores de frequência e válvulas redutoras de pressão e permitindo um completo gerenciamento da rede hidráulica através de sistemas SCADA, assim como a coleta em tempo real de dados de pressão e vazão, e, conseqüentemente, permitir identificar o comportamento de consumo das diversas regiões de uma cidade.

O armazenamento em banco de dados destas variáveis fornece possibilidades de, através de análise de dados históricos de consumo e aplicação de técnicas diversas, obter mecanismos de previsão da demanda de água de uma determinada região de consumo.

A fim de que os dados históricos do perfil de consumo sejam consistentes, é necessária uma filtragem dos mesmos no banco de dados (Lertpalangsuntia *et al.* (1999), conjugando os sinais de falha de comunicação e falhas de sensor com os de leitura totalizada do volume de consumo. Da mesma forma, devem ser verificadas as datas onde ocorreram vazamentos e manutenções da rede hidráulica, a fim de que os dados coletados nestas ocasiões sejam ignorados.

O conhecimento prévio do perfil de consumo ao longo do dia (previsão a curto prazo) de uma dada região de consumo permite planejar a operação do sistema visando otimizar os custos de produção e transferência de água mantendo-se a qualidade exigida.

Várias são as vantagens da obtenção de um modelo de previsão de consumo a fim de otimizar a distribuição de água, dentre as quais podem ser citadas:

a) Identificação imediata de perda física no sistema: dispondo-se de um modelo confiável de previsão, este acusará em tempo hábil qualquer anormalidade (vazamentos);

b) Ganho da qualidade da água através de estoques adequados: limitando-se a transferir apenas o volume necessário para os reservatórios, evita-se a perda de qualidade da água não consumida em tempo adequado;

c) Planejamento da operação das elevatórias de água *in natura* e tratada segundo contratos de energia: conhecendo-se o consumo nos períodos de pico de

energia, pode-se planejar a reserva de água para este período e desligar os equipamentos, evitando-se as multas sobre as tarifas de energia elétrica;

d) Planejamento da operação diária da Estação de Tratamento de Água - ETA, de modo a obter os melhores pontos de operação de todos os sistemas que compõem esta planta;

e) Estimativa das variáveis de ajuste de controle de pressão e vazão ótimas ao longo do dia, bem como *set points* de nível de enchimento de reservatórios;

f) Logística de produtos químicos da ETA: através de uma previsão de longo tempo de consumo (diária e mensal), tem-se como determinar a logística dos produtos químicos utilizados, evitando-se perdas e atrasos.

Segundo Tucci (1998), um modelo é a representação de algum objeto ou sistema, numa linguagem ou forma de fácil acesso e uso, com o objetivo de entendê-lo e buscar suas respostas para diferentes entradas. À medida que os sistemas se tornam mais complexos, mais desafiadora e necessária é a obtenção de seus modelos. O modelo em si não deve ser visto como o objetivo final, mas sim uma ferramenta para obtenção de um ou mais objetivos, uma etapa para atingir uma meta.

2.7 A DETERMINAÇÃO DE UM MODELO DE PREVISÃO DE CONSUMO DE ÁGUA

2.7.1 PRINCIPAIS TÉCNICAS EMPREGADAS

A determinação do perfil de consumo não se trata, porém, de uma tarefa trivial, dado que diversos fatores contribuem para a variação do consumo de água em uma dada região, tais como fatores climáticos (temperatura, umidade, precipitação), características sócio-econômicas, época do ano, dia da semana, tipo do bairro (residencial, comercial, industrial ou agrícola), entre outros (An *et al.*, 1996), caracterizando a série temporal de consumo de água como dinâmica, multivariável e não linear

Dentre as técnicas mais utilizadas para previsão destacam-se a regressão linear múltipla (Johnson e Wichern, 1998), análise de séries temporais (Box &

Jenkins, 1994), redes neurais artificiais (Haykin, 2001) e sistemas nebulosos (Klir e Yuan, 1995).

2.7.2 REGRESSÃO LINEAR MÚLTIPLA

O objetivo da regressão linear múltipla é de estabelecer uma relação quantitativa entre um grupo de variáveis preditivas (X) e uma resposta (Y).

O modelo de regressão linear múltipla é descrito pela equação 4:

$$Y = X\beta + \varepsilon \quad (4)$$

onde:

Y é um vetor $n \times 1$ de observações.

X é uma matriz $n \times p$ de regressores.

β é um vetor $p \times 1$ de parâmetros.

ε é um vetor $n \times 1$ de ruídos.

A solução para o modelo é um vetor B que estima o vetor β tal que satisfaça a equação 5:

$$B = \hat{\beta} = (X^T X)^{-1} X^T Y \quad (5)$$

E o valor de resposta estimado será conforme a equação 6:

$$\hat{Y} = XB \quad (6)$$

2.7.3 MODELOS TIPO BOX E JENKINS

Para processos estacionários, Box e Jenkins (Box *et al.*, 1994) propuseram um modelo que considera termos regressivos da série, médias móveis e um ruído branco com distribuição normal, média zero e variância definida. Caso a série não seja estacionária, esta pode ser diferenciada d vezes até que se obtenha estacionariedade.

Um processo ARIMA(p,d,q), ou auto-regressivo-integrado-médias móveis de ordem p (número de parâmetros auto-regressivos), d (número de diferenciações aplicadas a série a fim de que ela se torne estacionária), e q (número de parâmetros médias móveis), é representado pela equação 7:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + a_t \quad (7)$$

onde ϕ_i são os termos auto-regressivos, θ_j são os termos referentes às médias móveis e a representa um termo aleatório com média zero e variância definida.

2.7.4 REDES NEURAIS ARTIFICIAIS

As redes neurais artificiais – RNAs, na sua forma mais geral, podem ser consideradas uma máquina que é projetada para modelar um problema específico de maneira inspirada na forma como o cérebro realiza uma tarefa particular ou função de interesse (Haykin, 2001). As RNAs utilizam unidades de processamento simples denominados neurônios (artificiais) interconectados através de sinapses, caracterizados por valores de pesos sinápticos, e ativados por uma função de ativação.

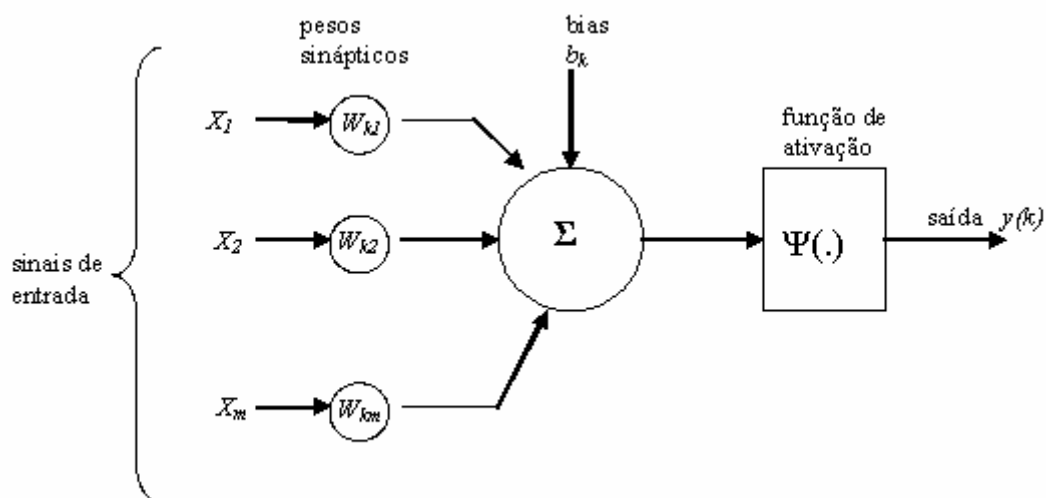


Figura 4 – Modelo de um neurônio artificial (Haykin, 2001).

A figura 4 apresenta o modelo de um neurônio artificial, onde:

x_j : entradas da rede;

w_{ki} : pesos sinápticos;

b_k : bias;

$\varphi(\cdot)$: função de ativação;

y_k : saída da rede.

Os três elementos básicos do modelo de um neurônio são: um conjunto de sinapses representadas pelos pesos sinápticos, que confere um grau de importância aos sinais de entrada. Outro elemento é um somador que realiza a soma dos sinais de entrada ponderados pelos pesos sinápticos, e, por fim, uma função de ativação, cujo papel é de restringir a amplitude de saída do neurônio, Haykin (2001).

Matematicamente, a saída da rede y_k é representada pela equação:

$$y_k = \varphi \left(\sum_{j=1}^m w_{kj} x_j + b_k \right) \quad (8)$$

As RNAs são capazes de adquirir conhecimento do meio através de aprendizagem e de generalizar, apresentando respostas a situações inéditas, podendo ser implementadas via *hardware* ou *software*.

Existem vários tipos de RNA, diferenciadas pela sua arquitetura e modos de aprendizagem. A figura 5 apresenta uma RNA tipo MLP (*Multi Layer Perceptron*) com uma camada de entrada com n entradas, uma camada oculta com m neurônios e uma única camada de saída com um neurônio.

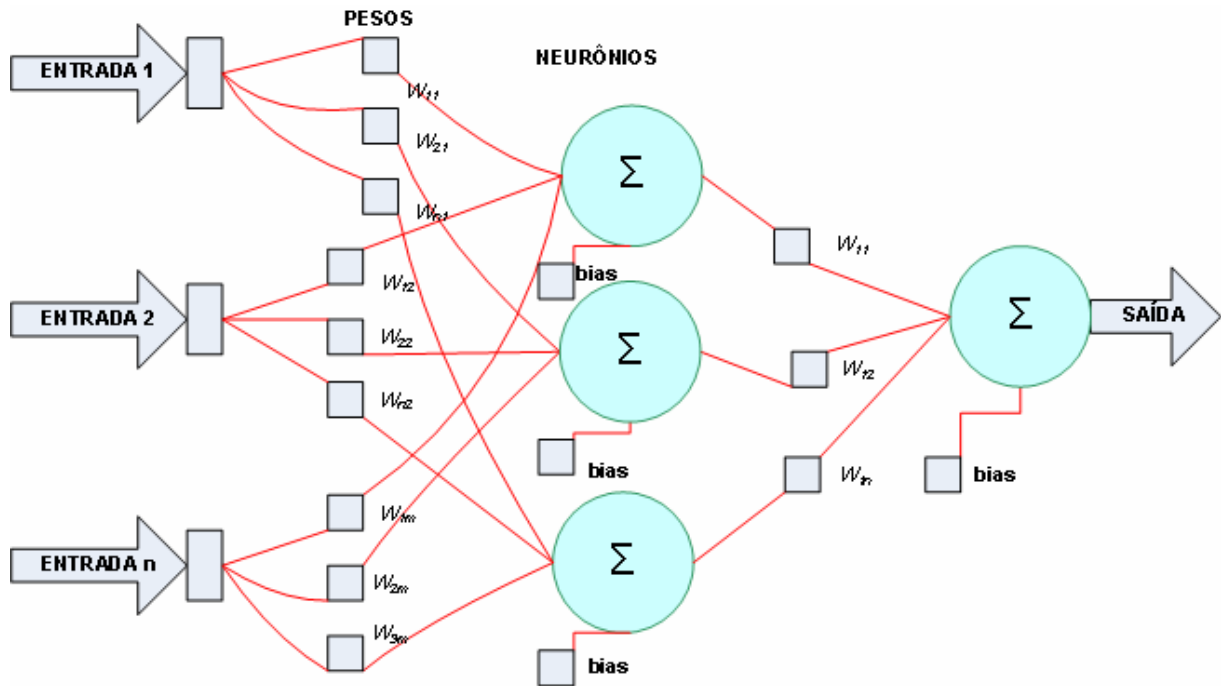


Figura 5 – Modelo de arquitetura de RNA do tipo *MLP*

A camada oculta tem como função intervir entre a camada externa da rede e a saída, conferindo à rede neural capacidade de extrair estatísticas de ordem elevada (Haikin, 2001).

O aprendizado da rede neural tipo *MLP* envolve a alteração dos pesos sinápticos de modo que o erro entre a saída desejada e a resposta fornecida pela rede neural seja minimizado. Este tipo de aprendizado (quando tem-se informações do resultado desejado para cada entrada) é denominado aprendizado supervisionado.

Em geral, as RNAs oferecem as seguintes propriedades e capacidades (Haykin, 2001):

- não-linearidade;
- mapeamento entrada-saída;
- adaptabilidade;
- tolerância à falhas;
- capacidade de generalização.

Em decorrência de sua versatilidade, as RNAs têm sido utilizadas para os mais diversos fins, tais como classificação de padrões, identificação de sistemas, controladores, previsão, entre outros (Haykin, 2001).

Para o caso em estudo, utilizaram-se as RNAs do tipo *MLP* com apenas uma camada de neurônios oculta respaldado pelo teorema da aproximação universal, que afirma que uma única camada oculta é suficiente para aproximar uniformemente qualquer função contínua com suporte em um hipercubo unitário, demonstrado por Cybenko³ *apud* (Haykin, 2001).

2.8 ESTUDOS PREVIAMENTE REALIZADOS

Diversos trabalhos tratando de previsão são encontrados na literatura aplicados às mais distintas áreas, dentre as quais, previsões de demanda de energia elétrica ocupam notável importância, com aplicações utilizando-se das mais variadas técnicas de previsão, (Chen *et al.*, (1995); Zhang e Dong (2001); Ozawa e Nimura (1999); Dash *et al.* (1995); Reis e Silva (2003); Rodrigues (2002)), muitos destes servindo como alento aos estudos de previsão de demanda de água pela similaridade dos problemas em discussão.

Estudos utilizando modelos de redes neurais aplicados a recursos hídricos e ciências ambientais tais como previsão de vazão de rios, podem ser encontrados em Galvão e Valença (1999) e Valença e Ludermir (2001).

Em se tratando de sistemas de saneamento, verificam-se estudos utilizando redes neurais artificiais para obtenção de melhor padrão de qualidade de água em estações de tratamento de água através da remoção da cor e turbidez no processo de coagulação, redução da dureza da água e busca de melhor desempenho de filtros (Baxter *et al.*, 2001).

Em se tratando de previsão de consumo de água, An *et al.* (1996) apresentam um estudo extensivo sobre regras de consumo extraídas a partir de dados históricos contendo informações de condições de operação passadas do sistema de

³ CYBENKO, G.1988. "Approximation by superpositions of a sigmoidal function", *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 303-314.

distribuição de água, associando a uma previsão de consumo e um fator de certeza. Florent *et al.* (1998) propõem dois modelos de previsão de demanda de água utilizando redes neurais artificiais, o primeiro objetivando prever o consumo da próxima hora e o segundo objetivando prever o consumo nas próximas doze horas. Para tal, utilizaram dados de consumo passado, assim como informações de dia da semana, hora e mês.

Liu *et al.* (1996) comparam sistemas nebulosos com redes neurais artificiais e modelos auto-regressivos para previsões a curto prazo e concluem que redes neurais e sistemas nebulosos são bons candidatos para este tipo de previsão, com performance superiores à do modelo auto-regressivo, visto que aqueles podem prever múltiplos pontos desde que haja quantidade suficiente de dados históricos disponíveis para teste. Por outro lado, os modelos auto-regressivos não necessitam de um pré-treinamento, porém podem prever apenas poucos dados adiante.

Lertapalangsunti *et al.* (1999) apresentam um *software* especialista ferramental para fazer previsões utilizando-se diversas técnicas, tais como redes neurais artificiais, sistemas nebulosos e regras de conhecimento. Advertem que previsões a partir de uma única técnica estão propensas a grandes erros de previsão, especialmente quando o domínio do problema inclui diferentes classes de padrões. É, portanto, aconselhável dividir o problema da previsão em subproblemas, aplicando diferentes técnicas para solucionar cada subproblema comparativamente. A maioria das abordagens estatísticas leva em consideração informações externas, tais como condições do tempo e distinções para dias de semana e feriados, dado que o uso exclusivo dos dados da série temporal pode ocasionar grandes erros de previsão.

Amaral (2000) procura explicar o consumo residencial médio e mensal da cidade de Piracicaba, SP, utilizando variáveis climáticas e sócio-econômicas, tais como o preço médio da água, preço marginal, consumo residencial total, consumo residencial médio, número de ligações e temperatura média mensal através de séries temporais.

Zhou *et al.* (2000) apresentam um modelo de previsão de consumo diário para a cidade de Melbourne, Austrália, baseado em um conjunto de equações representados por dois componentes: consumo básico de água – independente das

condições climáticas – e consumo sazonal de água – dependente das condições climáticas, este por sua vez subdividido em uma componente sazonal cíclica, uma componente sazonal climática e uma componente de persistência. A componente de consumo básico de água, não dependente das condições climáticas, baseou-se nos dados históricos dos meses mais frios, ou seja, de menor consumo. As variações desta parcela são determinadas por fatores sócio-econômicos, taxa do serviço e crescimento populacional, representando, portanto, pouca variação ao longo do tempo.

Baxter *et al.* (2001) ilustram a aplicação de redes neurais artificiais para previsão de consumo de água diário e horário para a cidade de Edmonton, Canadá, objetivando minimizar o custo de energia despendido em bombeamento, representando uma importante parcela do consumo de energia em sistema de distribuição e tratamento de água, através da otimização e planejamento de uso das bombas. A estimativa do consumo representa uma informação importante para a adequação operacional dos serviços de bombeamento e planejamento das reservas visando minimizar os custos de energia. Como entradas para o modelo de previsão horária, utilizaram as variáveis climáticas previstas disponíveis, tais como temperatura mínima e máxima do dia, precipitação do dia anterior, precipitação dos últimos cinco dias, informação de dia da semana e feriados, informação da estação do ano e consumo de água das 09h00min às 10h00min da manhã.

Zhou *et al.* (2001) utilizam uma série de consumo de um período de 142 anos para identificar o intervalo médio de recorrência dos dias de consumo excessivo de água, para um, dois, três e cinco dias consecutivos, para a cidade de Melbourne, Austrália, fundamentando-se no fato de que o comportamento climático ao longo do ano é essencial para explicar os padrões de consumo.

Zhou *et al.* (2002) apresentam um modelo de previsão de demanda para as próximas 24 horas para a cidade de Melbourne, Austrália, decompondo a demanda em consumo básico (não dependente das condições climáticas) – representado por 10% da média do consumo da hora separados em dia da semana e fins de semana – e consumo sazonal (dependente das condições climáticas). Para determinação da parcela de consumo sazonal, os perfis de demanda diurnos são estudados e definidos de acordo com o comportamento das horas de pico de consumo, entre

18h00min e 21h00min, variando conforme o dia da semana e mês referente. Através de regressão linear para cada hora e cada mês a partir do consumo, obtém-se a previsão do consumo e do perfil de consumo correspondente.

Trautwein Jr. (2003) compara técnicas de redes neurais artificiais com ajustamento exponencial para previsão a curtíssimo prazo, definido como o consumo nos próximos 15 minutos, utilizando dados históricos de uma semana para a cidade de Curitiba.

Liu *et al.* (2003) aplicam redes neurais artificiais para previsão de consumo residencial por habitante para a cidade de Weinan, China, utilizando como dados de entrada informações residenciais como número de moradores, consumo mensal da família, preço da tarifa de água, renda anual per capita dos moradores. Através dos resultados, relatam algumas conclusões:

- a) O consumo é diretamente proporcional à renda familiar;
- b) Para residências de mesma renda per capita, a residência cujo número de pessoas é maior possui menor consumo per capita, visto que há maior possibilidade de entre estas habitarem pessoas ou muito novas ou muito idosas, que consomem menos que adultos jovens;
- c) Para moradores com renda per capita muito alta, o consumo aumenta de forma insignificante com o número de habitantes, visto que estes preferem jantar fora a maioria das vezes ao invés de jantar em casa;
- d) O consumo é mais elástico para famílias de classe baixa.

3 METODOLOGIA

Obter modelos de previsão de séries temporais envolve ao menos as seguintes decisões:

- a) Seleção dos dados a serem incluídos no modelo e transformações de dados;
- b) Manipulação de dados inválidos: geralmente dados históricos contêm leituras inválidas ou fora de ordem. Deve-se definir um procedimento de como estas ocorrências podem ser detectadas e reajustadas ou eliminadas do banco de dados;
- c) Definição das variáveis externas a serem incluídas no modelo;
- d) Escolha das técnicas utilizadas para determinação dos modelos de previsão;
- e) Uma estratégia para determinar um modelo adequado – uma busca por todos os modelos possíveis é, em geral, bastante exaustiva e dispendiosa, e muitas vezes impraticável, portanto, o ideal é estabelecer uma estratégia para alcançar um modelo adequado de acordo com o tipo de técnica utilizada e da série temporal;
- f) Seleção da estrutura adequada do modelo selecionado;
- g) Definição do critério de avaliação do modelo e métricas de erro a serem utilizadas: RMSE, MAPE, AIC (Critério de Informação de Akaike), MSE (*Mean Square Error*), etc. Neste trabalho foram utilizados o RMSE – *root mean square error* (equação 9) e MAPE – *mean absolute percentage error* (equação 10);
- h) Manipulação de previsões ambíguas: deve-se testar o resultado das previsões de forma a garantir que a previsão se encontre dentro de uma faixa esperada factível.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (atual_i - previsto_i)^2} \quad (9)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{atual_i - previsto_i}{atual_i} \right| * 100 \quad (10)$$

3.1 COLETA E FILTRAGEM DOS DADOS

A obtenção de um bom modelo requer um banco de dados confiável e completo tal que seja possível identificar as principais características do perfil de consumo de uma determinada área de consumo.

Para este trabalho, os dados estão disponíveis em um banco de dados tipo MSAccess® cujos registros são inseridos por um sistema SCADA em funcionamento no centro de controle operacional da companhia de saneamento do Paraná – Sanepar, na cidade de Ponta Grossa, Paraná. Dispõe-se de um banco de dados de 30/09/2002 a 03/01/2004. Para este estudo, foram escolhidas três regiões de consumo do município de Ponta Grossa: Centro Histórico – consumo diário médio de 2.100 m³, Uvaranas – consumo diário médio de 11.000 m³ e Distrito Industrial, incluindo núcleo habitacional Santa Tereza, com consumo diário médio de 3.900 m³. A figura 6 ilustra a topologia de comunicação do sistema de abastecimento de água de Ponta Grossa. As estações de supervisão comunicam-se com os CLPs via rede de comunicação dedicada, que por sua vez comunicam-se com os CLPs das áreas remotas via modem ou rádio.

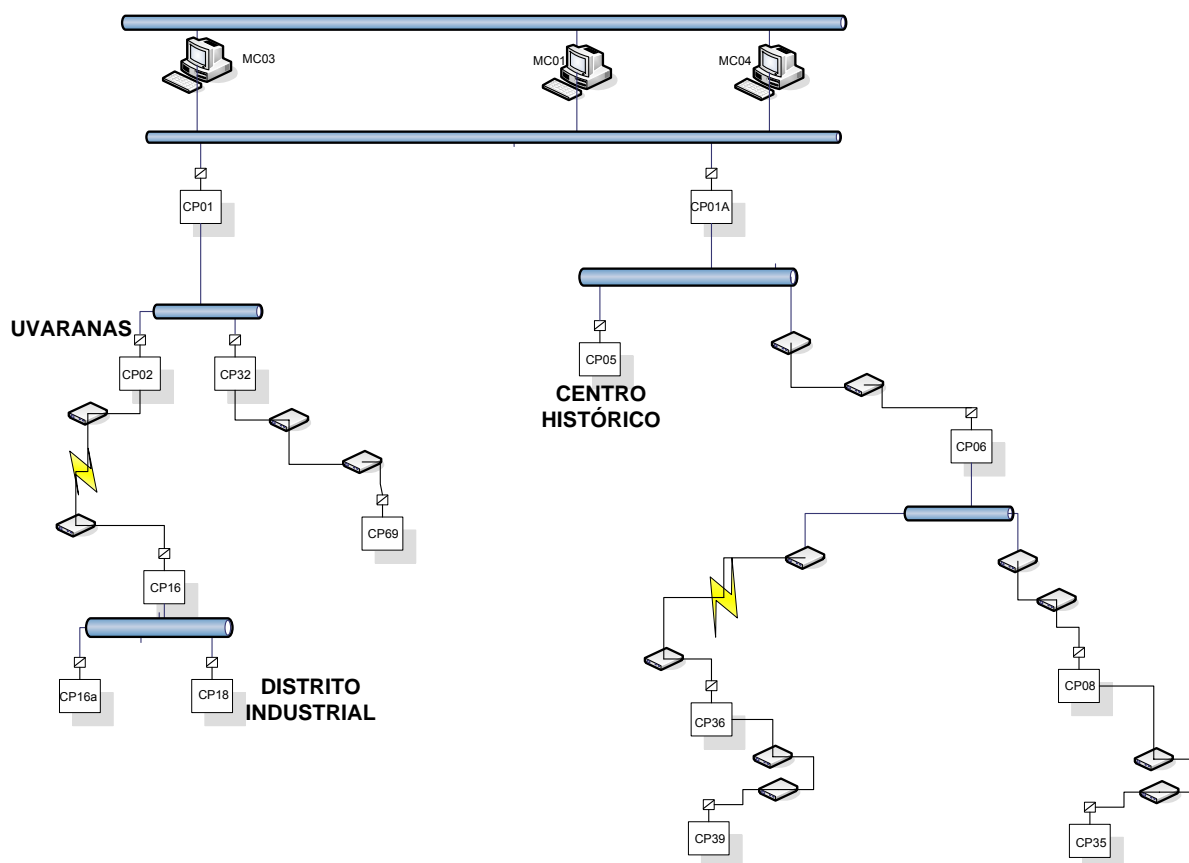


Figura 6 – Topologia de comunicação do sistema de abastecimento de água de Ponta Grossa – primeira etapa.

As tabelas de dados contêm informações de data, totalização dos medidores envolvidos na área remota, informações de falha de comunicação entre as remotas envolvidas no caminho para o centro de controle e outras informações diversas, tais como tempo de operação de bombas, medições de nível de reservatórios e demais informações pertinentes à área envolvida.

Os medidores de vazão eletromagnéticos analógicos fornecem um sinal de corrente de 4 a 20 mA, proporcional à vazão instantânea medida, a ser lido pelo CLP através de um cartão de entradas analógicas. Além disso, fornecem um sinal digital referente ao sentido de fluxo da vazão e outro sinal digital em forma de pulso, emitido quando a vazão instantânea, integralizada, atinge um valor definido de volume, (normalmente de um metro cúbico). Ambos os sinais digitais são lidos pelo CLP através de um cartão de entradas digitais. Ao ler o pulso de totalização, o CLP soma um metro cúbico à memória correspondente à totalização de vazão do medidor no sentido de fluxo de interesse. O valor de consumo de uma região pode

provir de um ou mais medidores da mesma área remota ou de áreas remotas diferentes, além de valores de nível de um reservatório de distribuição.

A informação da totalização de vazão é então enviada desde a área remota onde é coletada até o centro de controle operacional, onde está situado o sistema de supervisão (SCADA) que, por sua vez, através do disparo de um comando em SQL (*Structured Query Language*) em tempo definido, insere um registro com os dados de interesse, no caso data e totalização de vazão, no banco de dados. É válido observar que a data que o banco de dados recebe não é a data sincronizada ao valor de totalização, mas sim a data em que foi disparada a escrita dos dados do banco de dados do sistema de supervisão para o banco de dados tipo MSAccess®.

Inicialmente, um filtro das tabelas existentes do banco de dados foi feito de tal forma que ignore os registros nulos ou inseridos na ocorrência de falha de comunicação no caminho entre a área remota (fonte dos dados) e a área central coletora.

A partir do programa computacional Matlab®, uma conexão com o banco de dados MSAccess® é feita coletando os dados de totalização e data de cada região de consumo ordenada por data.

Visto que os dados são coletados a cada intervalo de 15 minutos, primeiramente uma rotina seleciona as linhas da tabela de dados cujas datas correspondem ao início de cada hora com intervalo limítrofe de cinco minutos, visando buscar os valores de consumo correspondentes ao intervalo de 60 ± 5 minutos. O erro de leitura, desta forma, pode ser decomposto em três parcelas.

- a) $\pm 1 \text{ m}^3$ referente ao incremento mínimo de volume;
- b) Erro de medição do medidor de vazão, interferindo na integralização da medida de volume;
- c) Volume referente a ± 5 minutos, folga cedida à coleta de dados, dependente da vazão instantânea da hora.

Em seguida, uma nova rotina efetua a diferença entre os dados de totalização entre duas leituras de horas adjacentes.

Como critério para filtragem das leituras inválidas e *outliers*, foi calculado a média e o desvio padrão de consumo para cada hora e dia da semana, e eliminados os consumos localizados fora da faixa $\mu \pm 3\sigma$, onde μ é a média e σ o desvio padrão. As linhas contendo valor de consumo igual a zero e a primeira linha subsequente também foram eliminadas, representando uma ausência de atualização das tabelas de comunicação entre áreas.

Os dados utilizados para as técnicas de redes neurais artificiais e regressão linear múltipla tiveram suas linhas de valores inválidos eliminadas. Visto que para identificação correta dos parâmetros dos modelos de séries temporais necessita-se preferencialmente da série temporal sem lacunas, os dados utilizados para modelos de análise de séries temporais tipo Box e Jenkins tiveram as lacunas de horas ausentes completadas com valores supostos de consumo. Escolheu-se para tal utilizar a média do consumo da hora. Visto que a base de dados possui uma lacuna entre os dias 10/12/2002 a 04/02/2003, optou-se por não utilizar os dados anteriores a esta lacuna para análise de séries temporais, a fim de evitar a necessidade de completar toda esta faixa com valores médios de consumo.

Dispondo destes dados, foi separado um conjunto de setenta por cento dos dados para treinamento, utilizado para obter os parâmetros do modelo, um conjunto de quinze por cento de dados para validação a fim de verificar a capacidade de generalização do modelo e um conjunto de quinze por cento de dados para teste, utilizado para verificar a confiabilidade do modelo através dos critérios de avaliação escolhidos.

3.2 VARIÁVEIS UTILIZADAS PARA OS MODELOS DE PREVISÃO

As variáveis utilizadas para os modelos de previsão serão basicamente as informações de data e de consumo das horas passadas. Embora como já comentado anteriormente, o consumo seja diretamente influenciado pelas condições climáticas, estas informações não serão utilizadas como entradas dos modelos, visto que não estão facilmente disponíveis de forma a serem acessadas em tempo real. Para tal, seria necessária uma conexão de fácil acesso do centro de controle operacional para um banco de informações meteorológicas. Supõe-se, portanto, que as influências climáticas sejam explicadas pelo próprio comportamento da série.

Com relação às informações de data, foram identificadas as informações de hora, dia da semana e mês. Com relação às informações de consumo, para os testes com redes neurais artificiais e regressão linear múltipla, foram selecionados para cada hora os consumos das últimas três horas, realizando desta forma um novo filtro nos dados de forma a incluir somente os registros cujas três últimas informações de consumo em horas adjacentes estejam disponíveis.

A figura 7 ilustra a redução dos dados históricos utilizáveis como entradas para os modelos de redes neurais e regressão linear múltipla relacionada ao número de consumos passados requeridos.

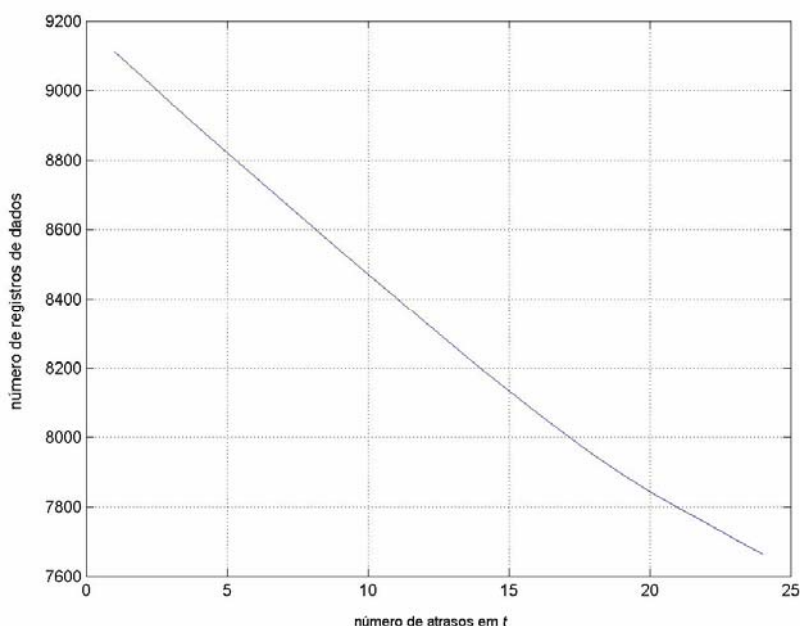


Figura 7 – Redução dos dados em relação ao número de consumos de horas passadas para o Centro Histórico.

3.3 CARACTERIZAÇÃO DO PERFIL DE CONSUMO

As regiões de consumo em estudo possuem comportamentos de consumo diferentes. A região de Uvaranas é tipicamente uma região residencial, o Centro Histórico, uma região tipicamente comercial e o Distrito Industrial, uma região mista entre residencial e industrial.

Como uma primeira análise destes perfis de consumo, as figuras 8 e 9 ilustram o perfil de consumo em dias de semana e fins de semana, de duas épocas distintas

do ano, um período no verão (de 23/02/2003 a 01/03/2003) e um período no inverno (de 10/08/2003 a 16/08/2003) para os bairros Uvaranas e Centro Histórico, respectivamente. Observa-se um maior consumo nos dias mais quentes, e um perfil de consumo diferenciado para dias de semana e fins de semana. Observa-se que para Uvaranas o consumo aumenta nos fins de semana, enquanto que no Centro Histórico o consumo diminui, caracterizando ser um bairro tipicamente comercial. Outra característica peculiar de Uvaranas é a grande similaridade entre as horas de baixo consumo entre os dois períodos, enquanto que para o Centro Histórico observa-se uma pequena elevação do consumo para os meses quentes.

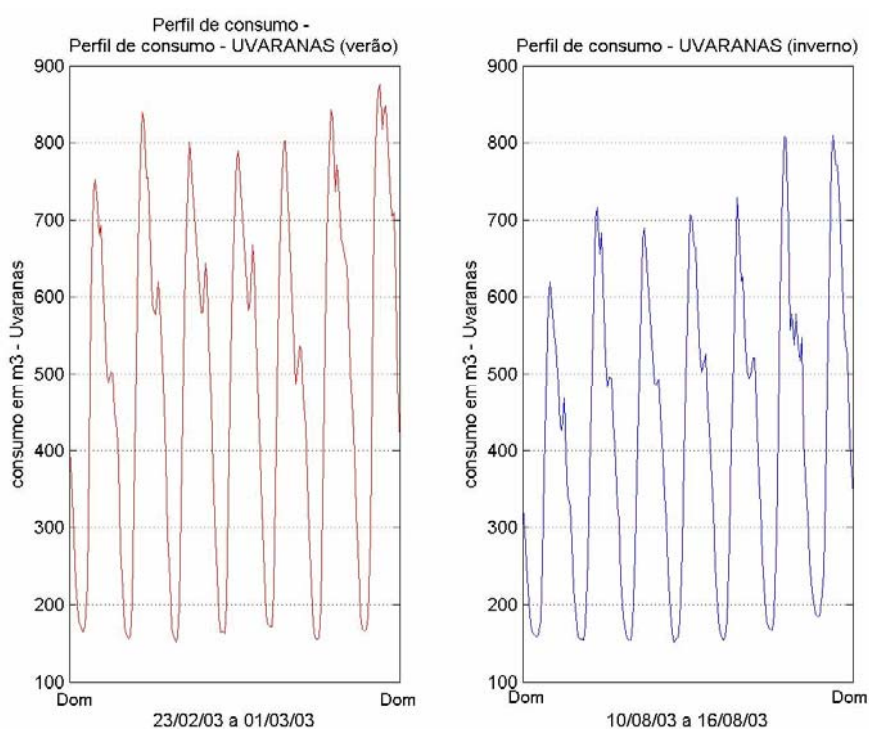


Figura 8 – Perfil de consumo de uma semana de Uvaranas para mês quente (vermelho) e frio (azul).

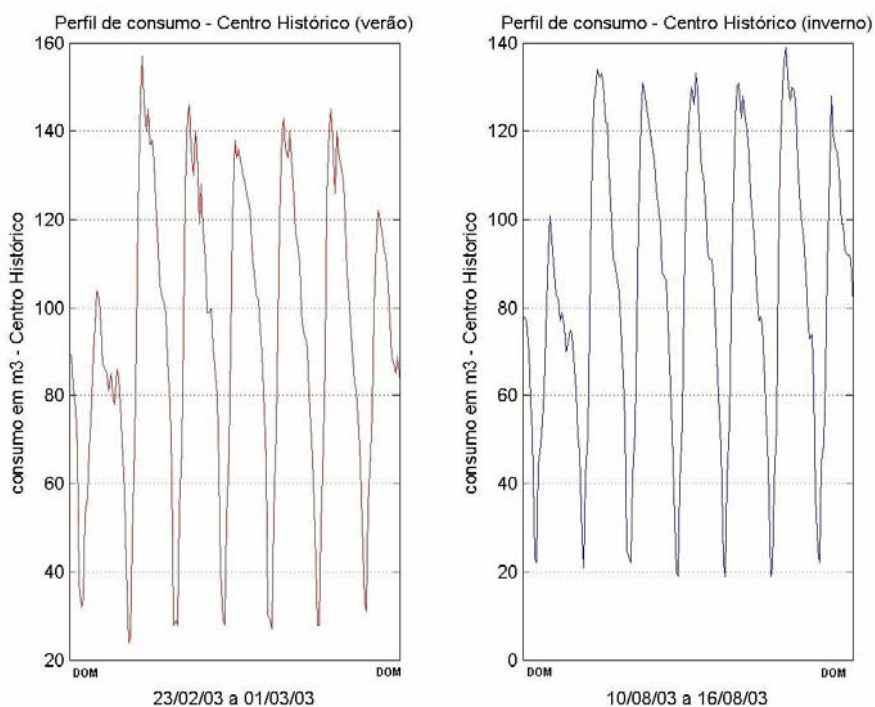


Figura 9 – Perfil de consumo de uma semana de Centro Histórico para mês quente (vermelho) e frio (azul).

A figura 10 apresenta a curva do consumo médio das horas obtido através de todo o conjunto de dados, dos bairros Uvaranas, Centro Histórico e Distrito Industrial.

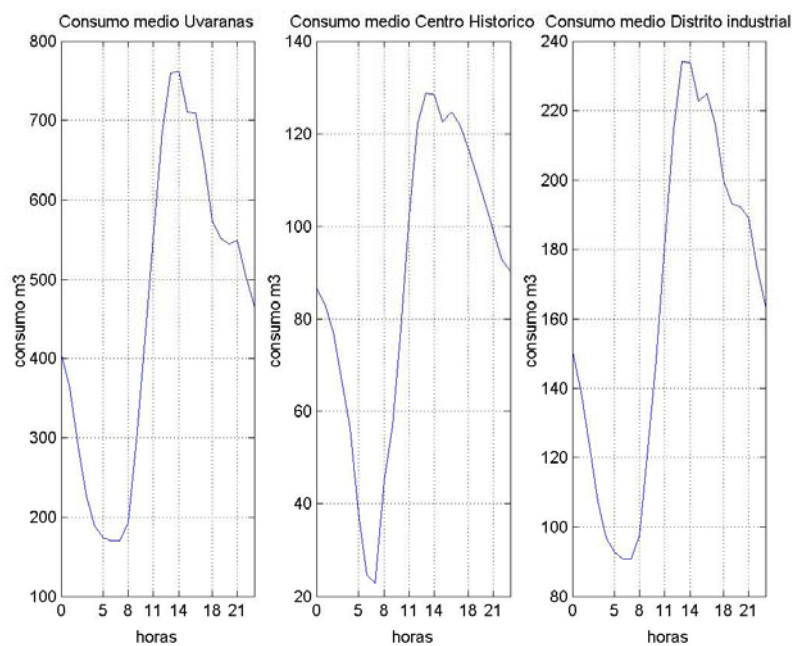


Figura 10 – Consumo médio ao longo do dia de Uvaranas, Centro Histórico e Distrito Industrial.

3.4 APLICAÇÃO DOS MODELOS DE PREVISÃO

Como primeiro passo para aplicar as técnicas de previsão, objetiva-se prever o consumo da próxima hora dado o consumo das horas passadas e as informações de dia da semana e hora.

Em se tratando de modelos com redes neurais artificiais e regressão linear múltipla, diversos testes foram realizados com o conjunto de dados representativos da data para definição da estrutura de entradas a ser utilizada para os modelos de previsão. A inclusão da informação do mês no conjunto de entradas ocasionou um decréscimo no desempenho dos modelos. Portanto, escolheu-se utilizar somente as informações de hora e dia da semana.

Da mesma forma, o embaralhamento dos dados de entrada com ou sem o uso da informação do mês ocasionou um decréscimo no desempenho dos modelos. Portanto, escolheu-se utilizar dados em ordem crescente de data para treinamento dos modelos.

Convencionou-se utilizar setenta por cento dos dados para treinamento, quinze por cento para validação e quinze por cento restantes dos dados para teste.

Para as técnicas de redes neurais artificiais e regressão linear múltipla utilizaram-se como entradas externas os dias da semana, a hora do consumo previsto, os consumos das horas passadas e relações entre estes consumos. Para as entradas de hora e dia da semana pode-se tanto utilizar entradas binárias, representando, portanto, sete entradas para os dias de semana e vinte e quatro entradas para a hora, ou então transformar as entradas em funções senoidais, como ilustrado nas equações 11 e 12, representando duas entradas para os dias de semana e duas entradas para a hora, reduzindo o número de entradas dos modelos e conseqüentemente o tempo de processamento. Verificou-se, porém, resultados sensivelmente melhores para as entradas binárias.

$$ed_1 = \sin\left(\frac{2\pi ds}{7}\right); ed_2 = \cos\left(\frac{2\pi ds}{7}\right) \quad (11)$$

Onde ds é o valor do dia da semana entre 1 e 7.

$$eh_1 = \sin\left(\frac{2\pi(h+1)}{24}\right); eh_2 = \cos\left(\frac{2\pi(h+1)}{24}\right) \quad (12)$$

Onde h é o valor da hora entre 0 e 23.

Para as entradas de consumos passados, variou-se o número de consumos comparando-se os erros de previsão. Também foram incluídos como entradas para RNA e RLM, relações entre parâmetros de consumo tais como a primeira diferença – d_1 – e a segunda diferença – d_2 – representadas pelas equações 13 e 14 respectivamente. A entrada d_1 representa a variação de consumo entre duas horas, enquanto que a entrada d_2 representa a variação da variação de consumo entre as horas.

$$d_{1_i} = t_{i-1} - t_{i-2} \quad (13)$$

Onde t_{i-1} é o consumo da hora passada e t_{i-2} é o consumo duas horas passadas.

$$d_{2_i} = (t_{i-1} - t_{i-2}) - (t_{i-2} - t_{i-3}) \quad (14)$$

Onde t_{i-3} é o consumo de 3 horas passadas.

3.4.1 PREVISÃO HORÁRIA COM REDES NEURAS ARTIFICIAIS

Foram utilizadas três tipos de estruturas de dados de entrada para modelos de rede neural, definidos aqui como RNh, RNd1 e RNd2. As três estruturas diferem entre si pelas entradas, da seguinte forma:

a) Entradas da RNh:

- Sete nós de entradas binárias para dia da semana;
- Vinte e quatro nós de entradas binárias para hora;
- Um nó de entrada do consumo da hora anterior em m^3 .

Saídas da RNh:

- Consumo na próxima hora em m^3 .

b) Entradas da RNd1:

- Sete nós de entradas binárias para dia da semana;
- Vinte e quatro nós de entradas binárias para hora;
- Um nó de entrada do consumo da hora anterior em m^3 ;

- Um nó de entrada da primeira diferença – d_1 – do consumo das duas últimas horas;

Saídas da da RNd1:

- Consumo na próxima hora em m^3 .

c) Entradas da RNd2:

- Sete nós de entradas binárias para dia da semana;
- Vinte e quatro nós de entradas binárias para hora;
- Um nó de entrada do consumo da hora anterior em m^3 ;
- Um nó de entrada da primeira diferença – d_1 – do consumo das duas últimas horas;
- Um nó de entrada da segunda diferença – d_2 – do consumo das três últimas horas;

Saídas da RNd2:

- Consumo na próxima hora em m^3 .

Para cada estrutura utilizaram-se duas redes neurais: uma rede neural artificial tipo *MLP* com uma camada oculta de neurônios com função de ativação linear, e uma rede neural tipo *MLP* com uma camada oculta de neurônios com função de ativação tangente hiperbólica e com neurônio de saída com função de ativação linear, com dados normalizados entre zero e um na entrada e na saída. Para as duas redes utilizou-se como algoritmo de treinamento, o de Levenberg-Marquardt, minimizando a função de soma do erro médio quadrático – MSE. O número de neurônios na camada oculta varia entre as redes neurais, tendo sido escolhido aquele que apresenta melhor resultados para a mesma estrutura de dados. Para escolha do número de neurônios na camada oculta, convencionou-se o seguinte procedimento:

1. Iniciou-se com um neurônio na camada oculta e verificam-se resultados;
2. Acrescentou-se um neurônio na camada oculta (limite de 5 neurônios) e verificam-se resultados;

3. Comparam-se resultados: caso a nova rede fosse melhor, voltou-se ao passo 2, caso contrário, manteve-se o número de neurônios da rede antiga.

Para a escolha do número de épocas de treinamento da rede neural, a fim de maximizar a capacidade de generalização da rede, evitando-se desta forma o *overfitting*, utilizou-se a função *early stopping* no Matlab® como critério de parada, que utiliza o método de validação cruzada.

O método da validação cruzada consiste em validar o modelo com um conjunto de dados diferente daquele utilizado para a estimação dos parâmetros. Utiliza-se o conjunto de treinamento para selecionar o modelo e o conjunto de validação para testar ou validar o modelo. Para evitar que haja um ajuste excessivo do modelo para com o conjunto de validação, acarretando desta forma uma perda na capacidade de generalização deste, utiliza-se um conjunto diferente do de validação, o conjunto de teste para verificação da capacidade de generalização da rede.

Neste caso, o erro sobre o conjunto de validação é monitorado durante o treinamento da rede, quando comumente diminui juntamente com o erro de treinamento, até que a partir de um determinado número de épocas de treinamento o erro de validação começa a aumentar, indicando que a rede está iniciando o processo de *overfitting* e, conseqüentemente, perdendo sua capacidade de generalização. Neste ponto o treinamento é interrompido e a rede mantém os pesos cujo erro de validação fora mínimo.

3.4.2 PREVISÃO HORÁRIA COM REGRESSÃO LINEAR MÚLTIPLA

Utilizou-se três tipos de estruturas para modelos de regressão linear múltipla, definidos aqui como RLMh, RLMd1 e RLMd2. As três estruturas diferem entre si pelas entradas, da seguinte forma:

a) Entradas da RLMh:

- Sete entradas binárias para dia da semana;
- Vinte e quatro entradas binárias para hora;
- Uma entrada do consumo da hora anterior em m³.

Saídas da RLMh:

- Consumo na próxima hora em m^3 .

b) Entradas da RLMd1:

- Sete entradas binárias para dia da semana;
- Vinte e quatro entradas binárias para hora;
- Uma entrada do consumo da hora anterior em m^3 ;
- Uma entrada da primeira diferença – d_1 – do consumo das duas últimas horas;

Saídas da RLMd1:

- Consumo na próxima hora em m^3 .

c) Entradas da RLMd2:

- Sete entradas binárias para dia da semana;
- Vinte e quatro entradas binárias para hora;
- Uma entrada do consumo da hora anterior em m^3 ;
- Uma entrada da primeira diferença – d_1 – do consumo das duas últimas horas;
- Uma entrada da segunda diferença – d_2 – do consumo das três últimas horas;

Saídas da RLMd2:

- Consumo na próxima hora em m^3 .

3.4.3 PREVISÃO COM MODELOS BOX E JENKINS

Para os modelos tipo Box e Jenkins, foi utilizado um conjunto de dados distinto do utilizado para os modelos de redes neurais e regressão linear múltipla, visto que houve necessidade do preenchimento das lacunas existentes a fim de que a série temporal estivesse completa. Desta forma, optou-se por comparar os resultados separadamente dos modelos de redes neurais artificiais e regressão linear múltipla. O conjunto de dados foi separado em três partes: setenta por cento destinado para estimação dos parâmetros dos modelos, quinze por cento destinado para avaliar a previsão de consumo horária e compor a matriz de erros dos modelos, e os quinze por cento restantes para avaliar a previsão nas próximas 24 horas.

Para este estudo, utilizou-se o software Matlab® pela possibilidade de conexão com o banco de dados, facilidade de manipulação dos dados na filtragem e formatação dos mesmos, por prover de todas as técnicas empregadas para previsão, e pela facilidade em adicionar rotinas complementares às simulações.

Uma desvantagem dos modelos Box e Jenkins é que sua utilização requer experiência e algum conhecimento específico por parte do profissional, além do uso automático de um pacote de computador, e mesmo assim, é possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal, Morettin e Toloí (2004).

A metodologia para obtenção de um modelo tipo Box e Jenkins é baseada em um ciclo iterativo definido pelos seguintes estágios:

- a) Consideração de uma classe geral de modelos (especificação);
- b) Identificação de um modelo baseado nas análises de autocorrelações, autocorrelações parciais e outros critérios;
- c) Estimação dos parâmetros do modelo identificado;
- d) Verificação do modelo ajustado através de uma análise de resíduos a fim de comprovar se este é adequado para os objetivos da previsão.

Como meio de identificar o modelo $ARMA(p,q)$ a ser utilizado, foram analisadas as funções de autocorrelação (fac) e de autocorrelação parcial ($facp$) dos processos, baseado nas seguintes características específicas Morettin e Toloí (2004):

1. Um processo $AR(p)$ tem fac que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;
2. Um processo $MA(q)$ tem fac finita, no sentido que ela apresenta um corte após o “lag” q ;
3. Um processo $ARMA(p,q)$ tem fac infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o “lag” $q-p$.

As figuras 11, 12 e 13 apresentam as funções de autocorrelação das séries de Uvaranas, Centro Histórico e Distrito Industrial, que segundo suas características, com decaimento de acordo com senóides amortecidas infinitas em extensão, indicam os modelos $AR(p)$ para todas as séries.

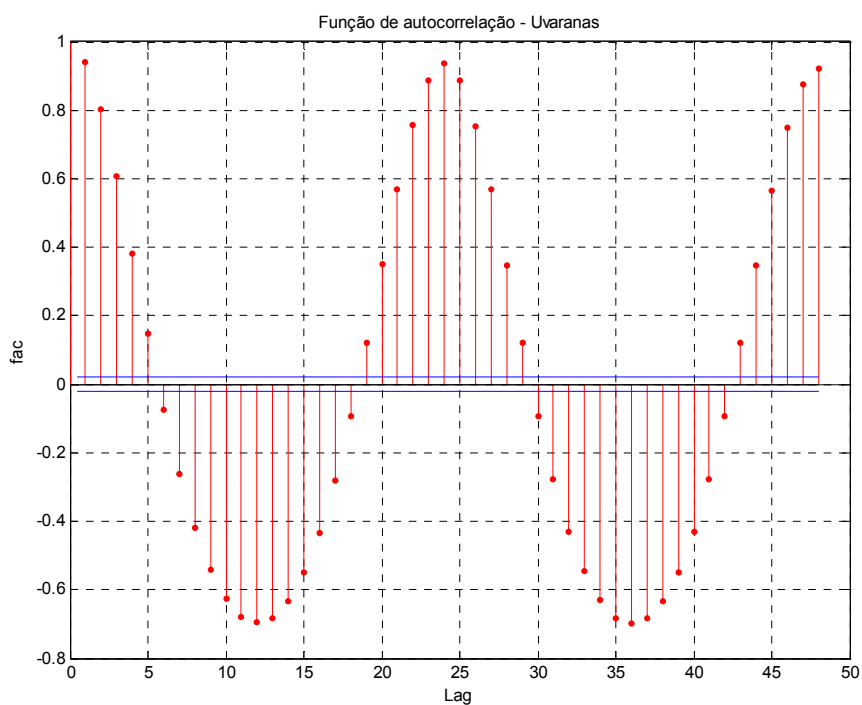


Figura 11 – Função de autocorrelação da série de Uvaranas.

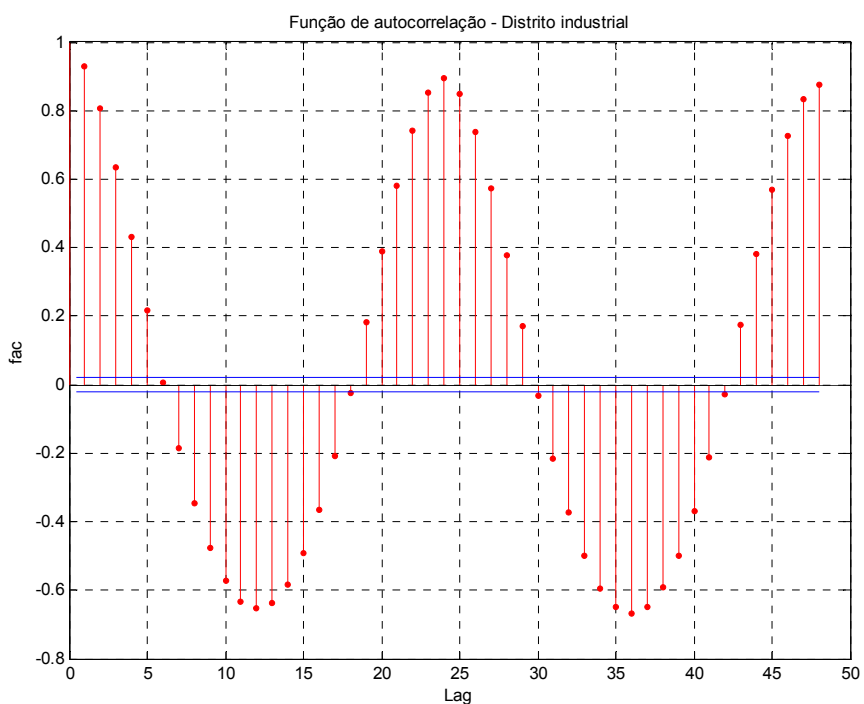


Figura 12 – Função de autocorrelação da série do Distrito Industrial.

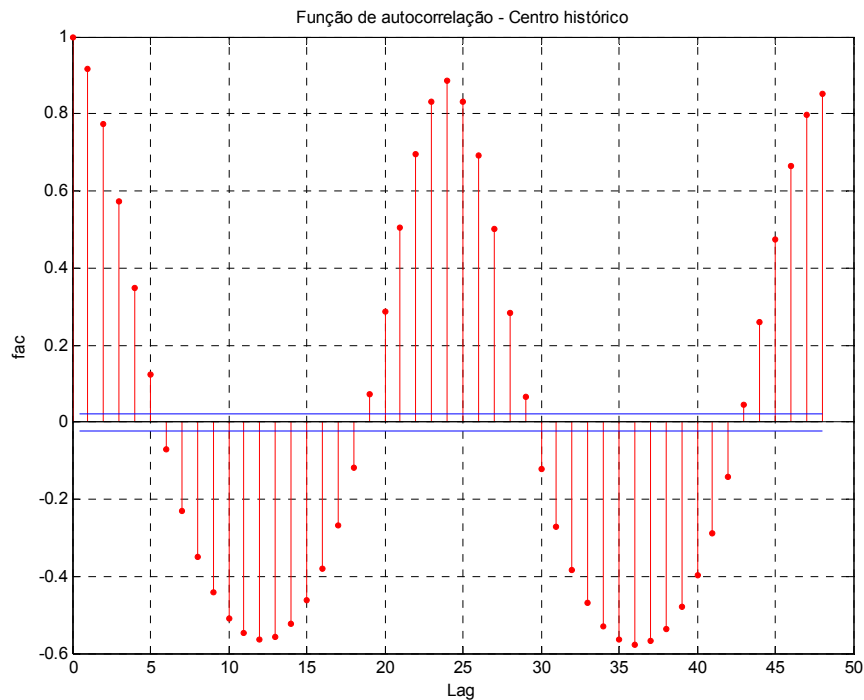


Figura 13 – Função de autocorrelação da série do Centro Histórico.

Box *et al.* (1994) propõem o uso da função de autocorrelação parcial (facp) e demonstram que:

1. Um processo $AR(p)$ tem $facp \Phi_{kk} \neq 0$ para $k \leq p$ e $facp \Phi_{kk} = 0$ para $k > p$;
2. Um processo $MA(q)$ tem $facp$ que se comporta de maneira similar à fac de um processo $AR(p)$: é denominada por exponenciais e/ou senóides amortecidas;
3. Um processo $ARMA(p,q)$ tem $facp$ que se comporta como a $facp$ de um processo MA puro.

As figuras 14, 15 e 16 apresentam as funções de autocorrelação parcial das séries de Uvaranas, Centro Histórico e Distrito Industrial e sugerem valores de $facp$ significativos até os *lags* 25, 27 e 29, respectivamente.

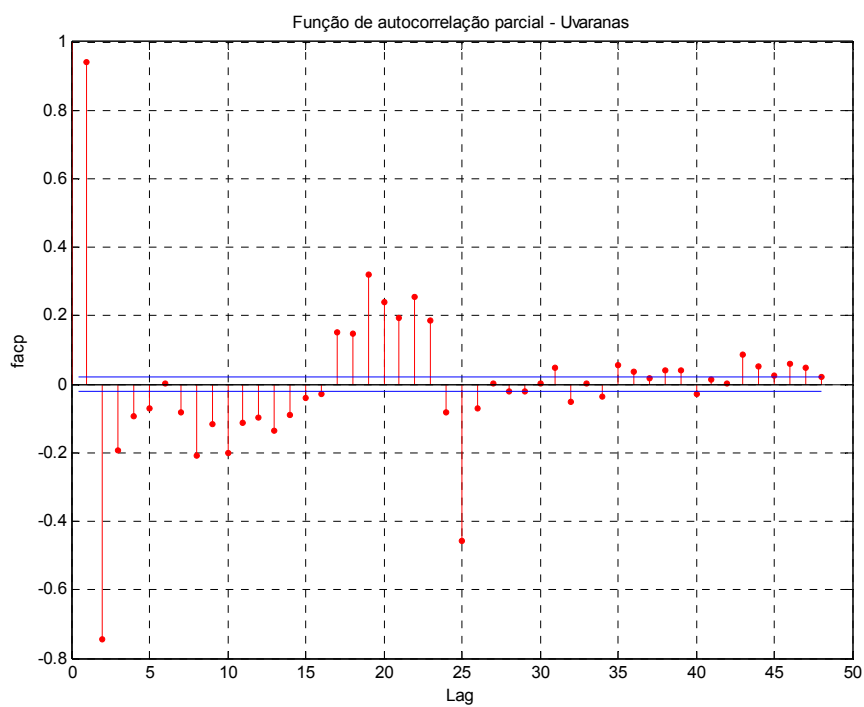


Figura 14 – Função de autocorrelação parcial da série de Uvaranas.

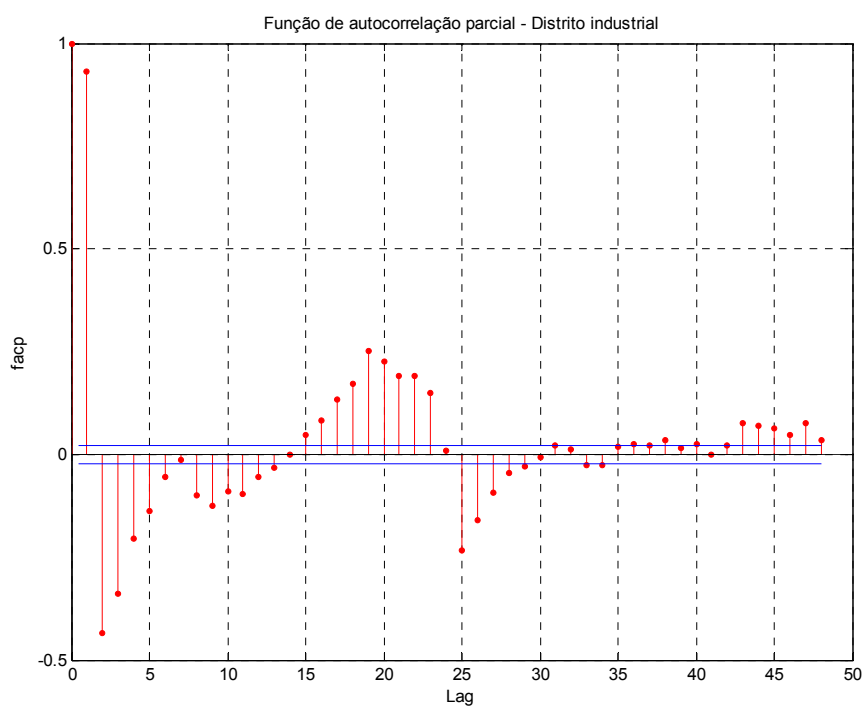


Figura 15 – Função de autocorrelação parcial da série do Distrito Industrial.

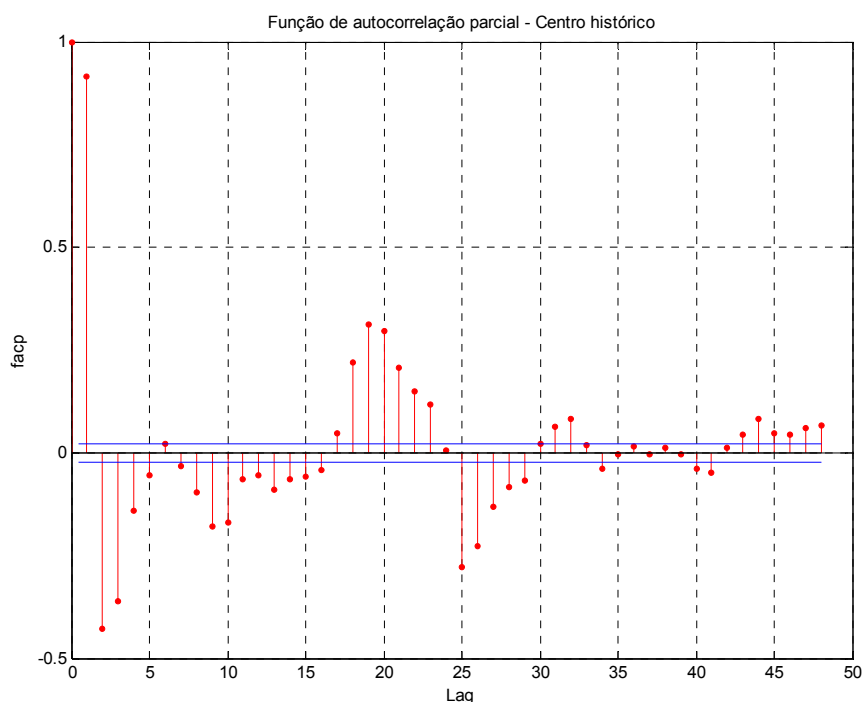


Figura 16 – Função de autocorrelação parcial da série do Centro Histórico.

Para estimação dos parâmetros dos modelos convencionou-se dos algoritmos disponíveis no *software* Matlab®.

Para cada região de consumo foram testados diferentes modelos baseados nas características das *fac* e *facp* das séries temporais. O melhor modelo identificado para Uvaranas foi o AR(25), para o Centro Histórico o AR(27), e para o Distrito Industrial o AR(29).

A análise dos resíduos tem como finalidade aplicar um conjunto de testes a fim de verificar se os resíduos são ou não aleatórios. A verificação de um ruído aleatório ou ruído branco informa que o modelo conseguiu explicar os dados corretamente, não restando nenhuma outra informação útil nos resíduos. Como análise dos resíduos aplicaram-se para as séries de resíduos das três regiões de consumo o teste do periodograma acumulado para os modelos AR(25) de Uvaranas, AR(27) do Centro Histórico e AR(29) do Distrito Industrial.

O periodograma foi um dos primeiros estimadores propostos nos estágios iniciais de análise espectral de séries temporais, com a finalidade de detectar periodicidade nos dados. Baseado na suposição de que uma série temporal é

constituída por ondas senoidais e cossenoidais com diferentes freqüências, seja um ruído branco a_t , $t=1, \dots, n$, sejam observações de um processo estocástico, o estimador $I(a)$, denominado periodograma, do espectro $p_a(f)$ do processo é dado por:

$$I(a) = \frac{1}{2\pi n} \left[\left(\sum_{t=1}^n a_t \cos \frac{2\pi i}{n} t \right)^2 + \left(\sum_{t=1}^n a_t \sin \frac{2\pi i}{n} t \right)^2 \right] \quad 0 < f_i < \frac{1}{2} \quad (15)$$

Um pico na freqüência $f_i=i/n$ indica uma periodicidade e período $1/f_i$. (Morettin e Toloi, 2004). Pode-se demonstrar que se a_t for ruído branco, seu espectro é constante e igual a $2\delta_a^2$ no intervalo $[0, 1/2]$, ou seja:

$$p(a) = 2\delta_a^2, \quad 0 < f_i < \frac{1}{2} \quad (16)$$

A função espectral acumulada para o ruído branco é determinada por:

$$P_a(f) = \int_0^f p_a(g) dg = \begin{cases} 0, & f < 0 \\ 2\delta_a^2 f, & 0 \leq f \leq \frac{1}{2} \\ \delta_a^2, & f > \frac{1}{2} \end{cases} \quad (17)$$

Assim, para um ruído branco a_t , o espectro acumulado varia linearmente com a freqüência. O periodograma acumulado normalizado $P(f)/\delta_a^2$ das séries dos resíduos dos modelos AR(25) de Uvaranas, AR(27) do Centro Histórico e AR(29) do Distrito Industrial estão representados nas figuras 17, 18 e 19 respectivamente.

Periodograma acumulado normalizado - AR(25) de Uvaranas

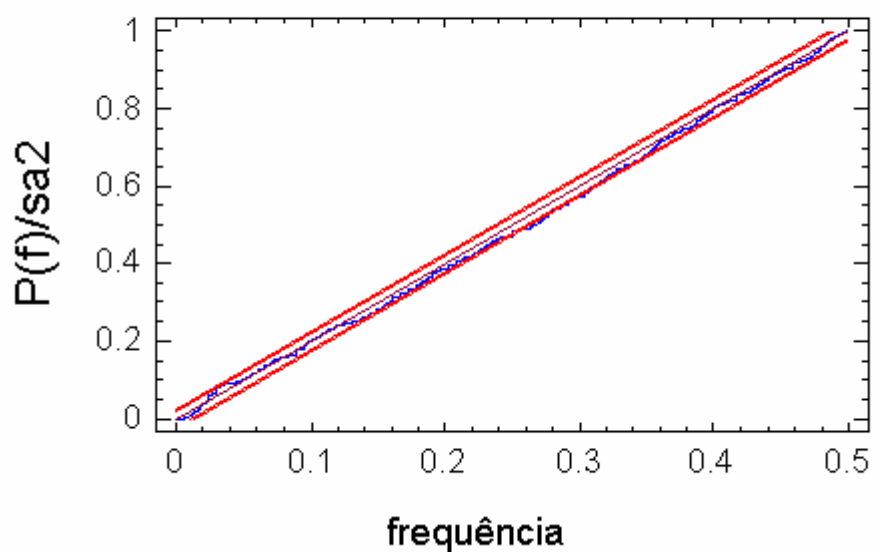


Figura 17 – Periodograma acumulado para série de resíduos do AR(25) de Uvaranas.

Periodograma acumulado normalizado - AR(27) Centro Histórico

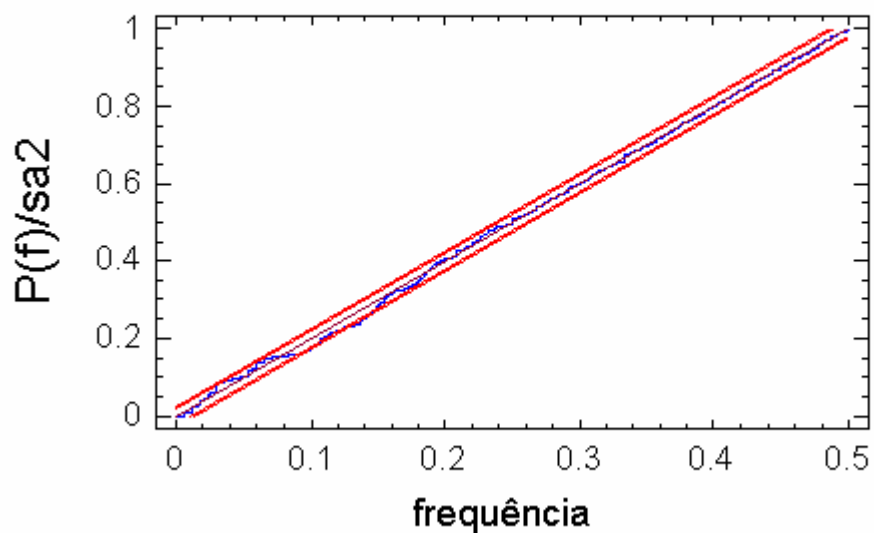


Figura 18 – Periodograma acumulado para série de resíduos do AR(27) do Centro Histórico.

Periodograma acumulado normalizado - AR(29) Distrito Industrial

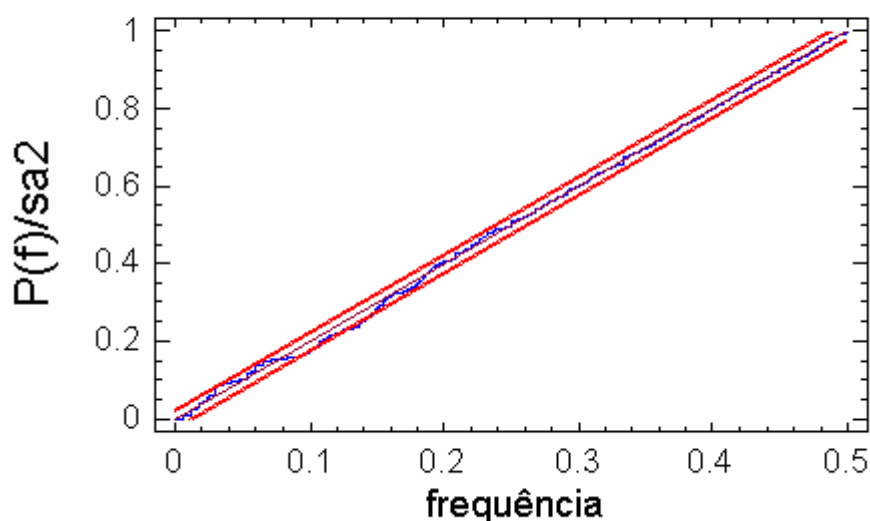


Figura 19 – Periodograma acumulado para série de resíduos do AR(29) do Distrito Industrial.

Os testes do periodograma acumulado revelam que os espectros acumulados das três séries de resíduos variam linearmente com a frequência indicando proximidade para um ruído branco.

Embora tenham sido identificados os modelos para as séries das três regiões de consumo, foram testados os modelos AR(1), AR(2), AR(5), AR(25), AR(27), AR(29) e o ARMA(2,1), para todas as séries a fim de comparar os resultados entre modelos para previsão horária. O objetivo principal desta comparação foi verificar a existência de algum modelo que se sobressaia sobre os modelos escolhidos para cada região de consumo (AR(25) para Uvaranas, AR(27) para Centro Histórico e AR(29) para Distrito Industrial) em dias da semana e horas específicas de previsão.

Verificou-se que ao simular os modelos tipo Box e Jenkins para previsões das próximas 24 horas no conjunto de teste, quanto mais distante a data de previsão do último registro de dados do conjunto utilizado para estimação dos parâmetros, maior o erro de previsão. A fim de tentar atenuar o erro dos modelos auto-regressivos quando aplicados em uma data mais afastada do conjunto de estimação dos parâmetros, experimentou-se adicionar uma parcela de ajuste ao modelo calculada a

partir da média dos erros de previsão horária do modelo auto-regressivo aplicado nos dados passados.

Desta forma, a cada previsão de 24 horas utilizando um modelo auto-regressivo, verifica-se a média dos erros da previsão horária do modelo de acordo com a hora de previsão e dia da semana, resultando numa matriz de ajustes de 7 x 24. Após obtida a matriz de ajustes, a cada novo passo de previsão, aplica-se o modelo e adiciona-se a parcela de ajuste segundo hora e dia da semana, de acordo com as equações 18 e 19, onde \hat{y}_t é o consumo previsto da próxima hora e $aj_{(ds,h)}$ o fator de ajuste de acordo com o dia da semana, ds e hora prevista, h .

$$\hat{y}_t = \sum_{i=1}^p \phi_i y_{t-i} + aj_{(ds,h)} \quad (18)$$

$$aj_{(ds,h)} = \frac{1}{n_{(ds,h)}} \sum_{t=1}^{n_{(ds,h)}} e_{t (ds,h)} \quad (19)$$

3.5 MODELOS HÍBRIDOS

Aplicando os diferentes modelos encontrados das diferentes técnicas de RNAs, RLM e tipo Box e Jenkins, verifica-se que muito raramente um modelo único é superior a todos os demais em todas as situações de dias da semana e hora para previsão horária de consumo. Na busca de um modelo resultante da mescla de vários modelos aplicáveis aproveitando-se os pontos fortes de cada modelo, foi proposto um modelo híbrido de previsão de consumo para as técnicas de RNA e RLM e um modelo híbrido de previsão de consumo para Box e Jenkins.

Para escolha de um modelo híbrido de previsão de consumo para cada bairro, primeiramente obteve-se para cada modelo de previsão duas matrizes 7 x 24 representando os dias da semana e as horas contendo os erros médios de previsão horária RMSE e MAPE para o conjunto de validação. Através destas n matrizes denominadas matrizes de erros dos n modelos, identifica-se o modelo que conduz ao menor erro de previsão horária para uma dada hora e dia da semana, obtendo-se a matriz de modelos escolhidos para RMSE e para o MAPE para cada região de

consumo. As matrizes dos modelos escolhidos para cada região de consumo estão representadas nos apêndices A (RNA e RLM) e B (Box e Jenkins).

4 RESULTADOS

4.1 RESULTADOS PARA PREVISÃO HORÁRIA COM REDES NEURAIS E REGRESSÃO LINEAR MÚLTIPLA

A seguir, as tabelas 1, 2 e 3 apresentam os erros RMSE e MAPE para previsão de consumo na próxima hora para o conjunto de dados de teste utilizando os modelos de redes neurais artificiais e regressão linear múltipla para os bairros de Uvaranas, Distrito Industrial e Centro Histórico, respectivamente.

Tabela 1 – Comparação de erros RMSE e MAPE para as redes neurais e regressão linear múltipla –Uvaranas.

Modelo \ Erro	RMSE (m³)	MAPE (%)
RNhl	29,771	4,402
RNht	27,263	3,760
RNd1I	28,531	4,220
RNd1t	28,175	4,394
RNd2I	28,778	4,281
RNd2t	27,580	4,102
RLMh	29,775	4,404
RLMd1	28,535	4,222
RLMd2	28,777	4,280

Tabela 2 – Comparação de erros RMSE e MAPE para as redes neurais e regressão linear múltipla – Distrito Industrial.

Modelo \ Erro	RMSE (m ³)	MAPE (%)
RNhI	12,169	5,822
RNht	12,446	5,257
RNd1I	12,121	5,831
RNd1t	10,440	5,372
RNd2I	12,110	5,823
RNd2t	12,137	5,697
RLMh	12,165	5,814
RLMd1	12,116	5,823
RLMd2	12,105	5,814

Tabela 3 – Comparação de erros RMSE e MAPE para as redes neurais e regressão linear múltipla –Centro Histórico.

Modelo \ Erro	RMSE (m ³)	MAPE (%)
RNhI	8,767	7,776
RNht	7,849	6,823
RNd1I	8,359	7,715
RNd1t	9,509	9,172
RNd2I	8,366	7,736
RNd2t	7,727	6,631
RLMh	8,766	7,764
RLMd1	8,358	7,704
RLMd2	8,365	7,725

A figura 20 mostra o gráfico da previsão da próxima hora utilizando a RNh com função de ativação tangente hiperbólica com 5 neurônios (camada oculta) para o bairro de Uvaranas, ilustrando a previsão horária de consumo utilizando-se apenas a informação de data e consumo na última hora.

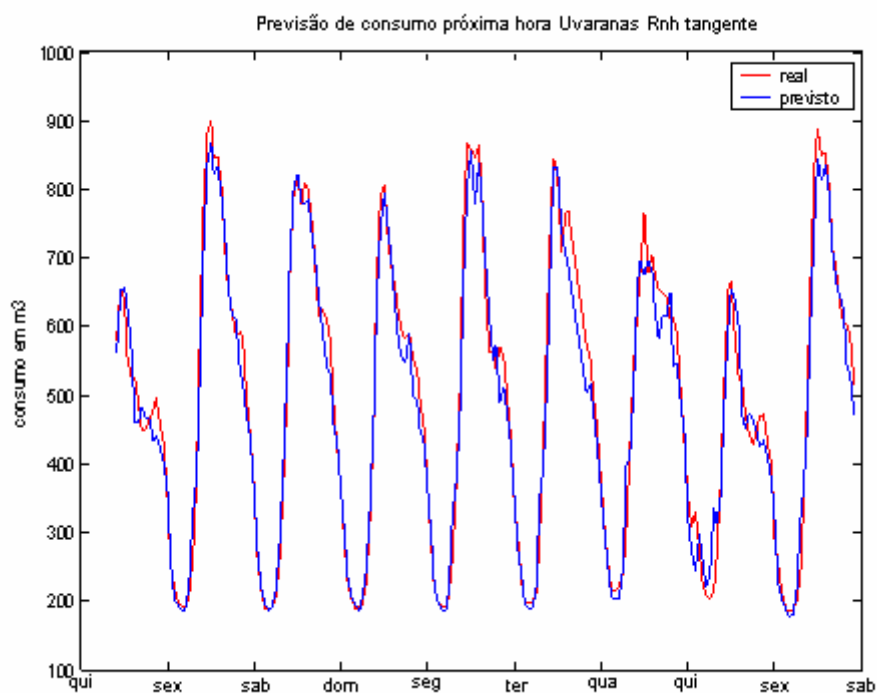


Figura 20 – Previsão de consumo da próxima hora para Uvaranas – RNh tangente hiperbólica com 5 neurônios na camada oculta.

A figura 21 mostra o gráfico da previsão da próxima hora utilizando regressão linear múltipla – RLMd1, indicado como melhor modelo de regressão linear múltipla para o bairro do Centro Histórico.

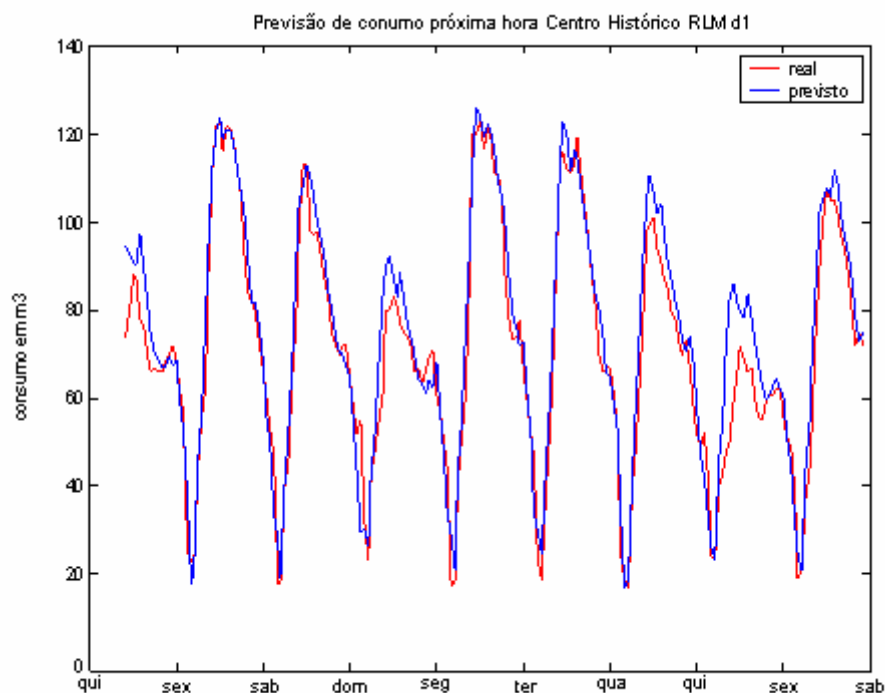


Figura 21 – Previsão de consumo da próxima hora para Centro Histórico – RLMd2.

A fim de analisar com maior profundidade as performances dos modelos, foram calculados os erros RMSE e MAPE de previsão da próxima hora para cada hora e dia da semana distintos. Para cada modelo de redes neurais e regressão linear múltipla calculou-se uma matriz de erros RMSE e MAPE de 7 linhas representando os dias de semana por 24 colunas representando as horas do dia, resultando em nove matrizes (seis modelos de RNA e três modelos de RLM), de RMSE e nove matrizes de MAPE para cada região de consumo. Como ilustração da variação do desempenho dos modelos em relação à hora do consumo previsto e ao dia da semana, a figura 22 apresenta a matriz de erros RMSE para a RNd2 tangente hiperbólica para o bairro Centro Histórico. Através desta matriz, observam-se os pontos fortes e fracos de cada modelo de acordo com a hora e o dia da semana. Em se tratando do RNd2 tangente para o Centro Histórico, verificam-se erros maiores de previsão para segundas-feiras das 10h00min às 15h00min, às quintas-feiras das 15h00min às 19h00min e às sextas-feiras das 14h00min às 18h00min. Tratando-se de um bairro comercial, estes períodos coincidem justamente com a elevação de consumo característicos nestes períodos de acordo com o perfil de consumo

ilustrado na figura 9. A elevação do consumo nas quintas e sextas-feiras podem ser explicadas pelo maior movimento nos bares.

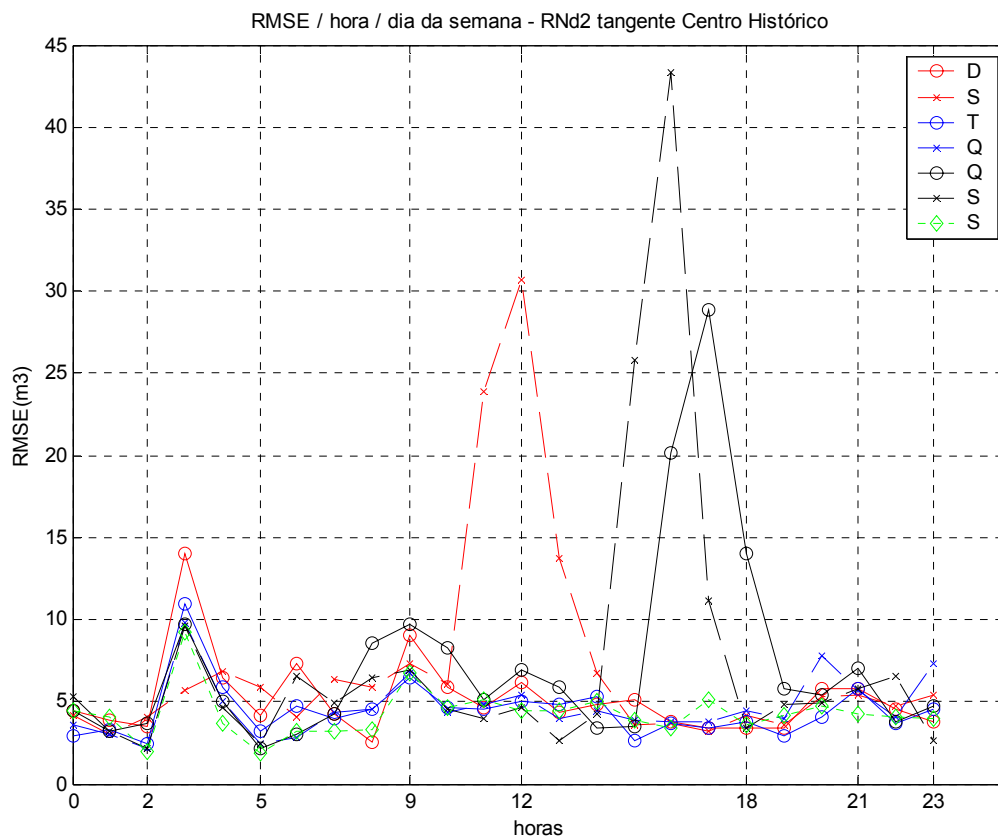


Figura 22 – RMSE por dia da semana por hora para o Centro Histórico, RNd2 com função de ativação tangente hiperbólica – previsão horária, onde D, S, T, Q, Q, S e S referem-se à: dom, seg, ter, qua, qui, sex, sab, nesta ordem.

A figura 23 ilustra a previsão horária para o bairro de Uvaranas com o uso do modelo AR(25), indicado como o melhor modelo.

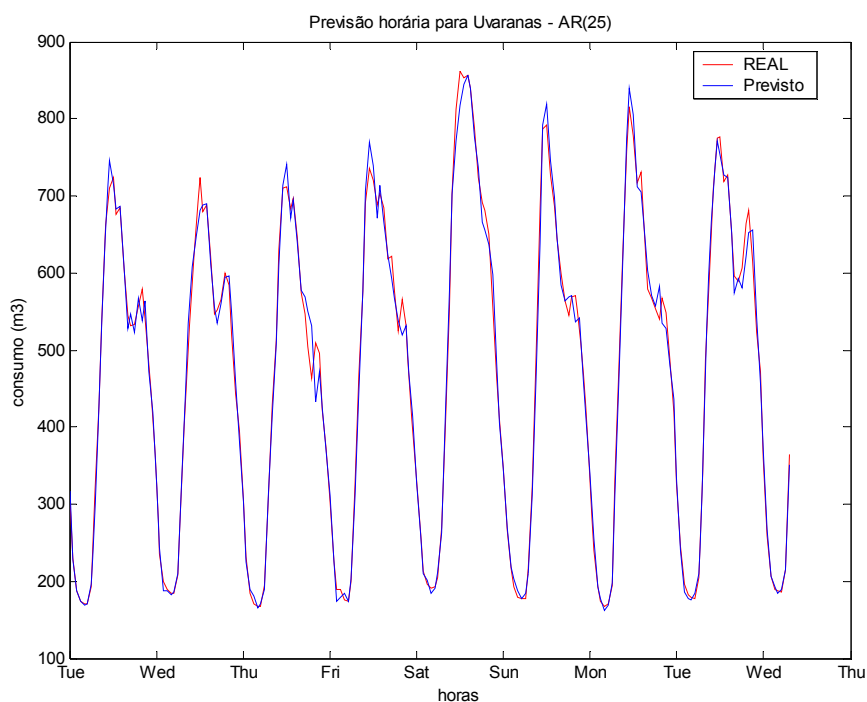


Figura 23 – Previsão de consumo da próxima hora para Uvaranas – AR(25).

4.2 RESULTADOS PARA PREVISÃO HORÁRIA COMO MODELOS BOX E JENKINS

A seguir, as tabelas 4, 5 e 6 apresentam os erros RMSE e MAPE para previsão de consumo da próxima hora para o conjunto de dados de teste utilizando os modelos de séries temporais para os bairros de Uvaranas, Distrito Industrial e Centro Histórico, respectivamente.

Tabela 4 – Comparação de erros RMSE e MAPE para modelos ARMA(p,q) – bairro Uvaranas.

Modelo \ Erro	RMSE (m ³)	MAPE (%)
AR(1)	33,573	4,454
AR(2)	40,859	4,579
AR(5)	40,858	4,668
AR(25)	33,801	4,129
AR(27)	33,602	4,133
AR(29)	33,658	4,148
ARMA(2,1)	41,806	4,977

Tabela 5 – Comparação de erros RMSE e MAPE para modelos ARMA(p,q) – Distrito Industrial.

Modelo \ Erro	RMSE (m ³)	MAPE (%)
AR(1)	17,087	5,591
AR(2)	21,027	6,159
AR(5)	19,861	6,157
AR(25)	16,967	5,733
AR(27)	16,556	5,680
AR(29)	16,512	5,675
ARMA(2,1)	24,400	7,101

Tabela 6 – Comparação de erros RMSE e MAPE para modelos ARMA(p,q) – bairro Centro Histórico.

Modelo \ Erro	RMSE (m ³)	MAPE (%)
AR(1)	10,764	7,987
AR(2)	13,864	9,720
AR(5)	12,814	9,446
AR(25)	10,635	8,390
AR(27)	10,046	8,176
AR(29)	10,025	8,100
ARMA(2,1)	15,916	11,45

Da mesma forma, para os modelos Box e Jenkins foram calculados os erros RMSE e MAPE de previsão da próxima hora para cada hora e dia da semana distintos, obtendo-se para cada modelo uma matriz de erros RMSE e MAPE de 7 linhas por 24 colunas, e a partir da comparação das matrizes de erros de todos os modelos obteve-se a matriz dos modelos escolhidos para cada região de consumo. A figura 24 ilustra o RMSE do modelo AR(29) para o bairro Centro Histórico. Verifica-se uma maior debilidade deste modelo para a previsão do consumo para às 12:00 de segunda-feira e para às 15:00 de sexta-feira.

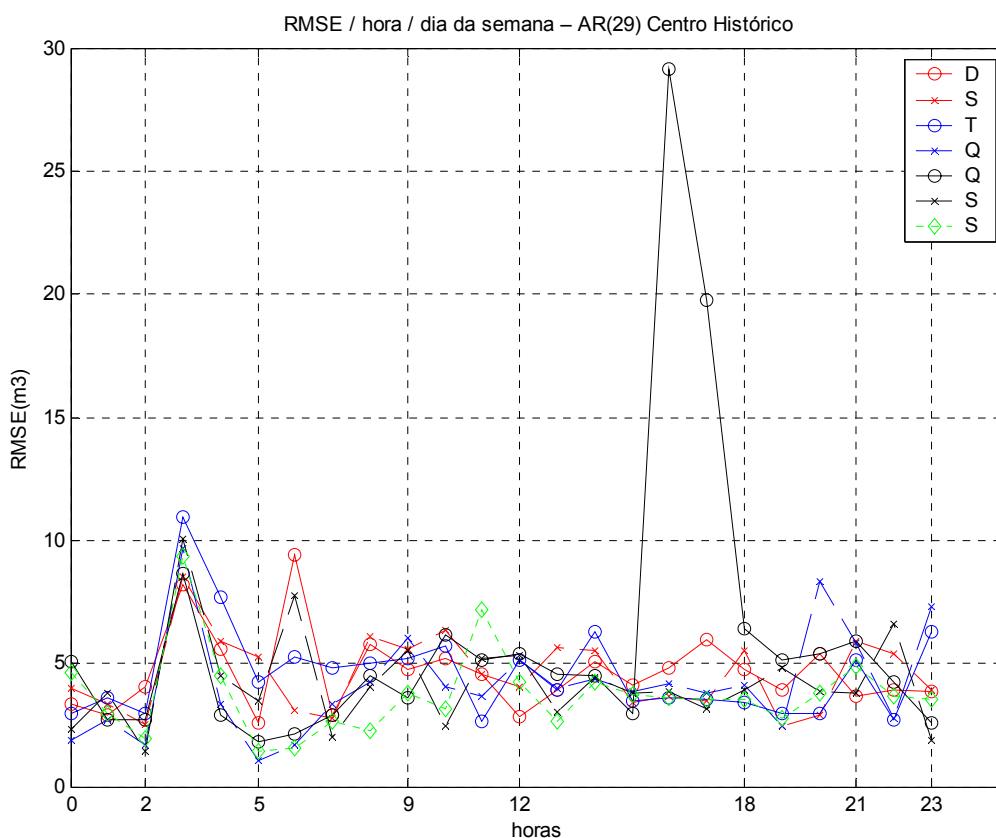


Figura 24 – RMSE por dia da semana por hora para o Centro Histórico, AR(29) – previsão horária, onde D, S, T, Q, Q, S e S referem-se à: dom, seg, ter, qua, qui, sex, sab, nesta ordem.

O apêndice A apresenta as tabelas de modelos escolhidos entre redes neurais e regressão linear múltipla para cada hora e dia da semana para os bairros de Uvaranas, Distrito Industrial e Centro Histórico utilizando-se o critério de erro RMSE. Observa-se que para o bairro de Uvaranas, tabela 12, que todos os modelos entre os de redes neurais e regressão linear múltipla são selecionados para alguma determinada hora e dia da semana. Particularmente para as 09h00min e 10h00min, justamente no período onde ocorre maior variação de consumo, observa-se uma predominância do modelo 6 (RNd2 tangente), dado que fornece a informação da variação do consumo nas duas últimas horas, bem como a variação da variação do consumo nas últimas três horas.

O apêndice B apresenta as matrizes de modelos escolhidos tipo Box e Jenkins para cada bairro. Verifica-se que para as três regiões de consumo todos os modelos aparecem na matriz de modelos escolhidos (ver tabelas 19, 20 e 21).

4.3 PREVISÃO DAS PRÓXIMAS 24 HORAS

4.3.1 Modelos tipo RNA e RLM

Para a previsão das próximas 24 horas utilizou-se a técnica *multi-step*, ou previsão com realimentação, de forma que o valor previsto de consumo para a próxima hora será entrada de consumo passado para previsão de consumo duas horas adiante, e assim sucessivamente. Este método de previsão aplica-se para qualquer valor de horas adiante, porém não indicado para previsões de muitos passos devido ao erro que se acumula a cada passo de previsão.

No caso do modelo híbrido, para cada previsão, verifica-se na matriz de modelos escolhidos segundo critério determinado de menor erro (RMSE ou MAPE), o modelo indicado para determinada hora.

Utilizaram-se os dados dos últimos dois meses, ou seja, registros pertencentes ao conjunto de teste utilizado para os modelos de redes neurais e regressão linear múltipla para realizar testes de previsão nas próximas 24 horas.

A figura 25 ilustra a previsão nas próximas 24 horas para o bairro de Uvaranas a partir de 23/12/2003 09h00min com os modelos RNh tangente (5 neurônios na camada oculta), RNd1 tangente (5 neurônios na camada oculta), RNd2 tangente (5 neurônios na camada oculta), RLMh, RLMd1 e híbrido. Os erros RMSE e MAPE para cada modelo encontram-se na tabela 7.

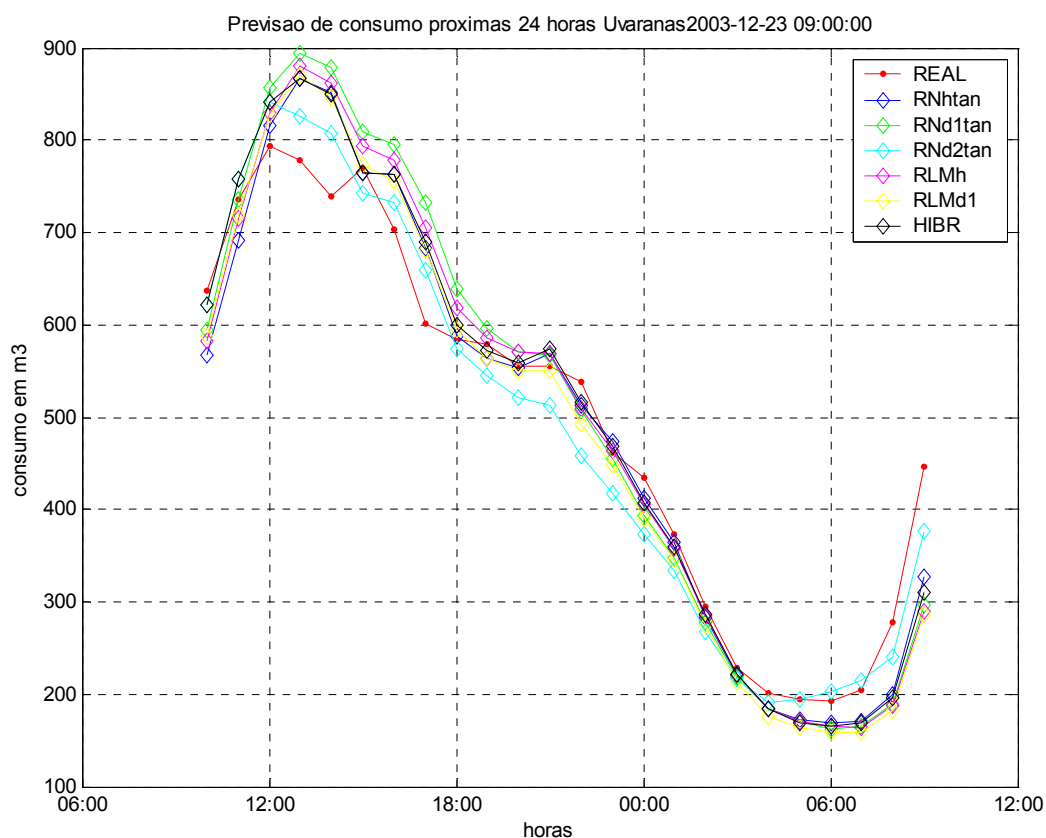


Figura 25 – Previsão de consumo das próximas 24 horas – Uvaranas.

Tabela 7 – Erros RMSE e MAPE para previsões das próximas 24 horas com modelos de redes neurais, regressão linear múltipla e híbrido para previsão a partir de 9:00 de 23/12/2003 – Uvaranas.

Modelo \ Erro	RMSE	MAPE
RNhl	60,191	9,758
RNht	51,142	8,377
RNd1l	57,206	10,369
RNd1t	66,547	10,786
RNd2l	56,604	10,850
RNd2t	41,150	7,279
RLMh	60,189	9,829
RLMd1	57,226	10,431
RLMd2	56,495	10,806
HIBRIDO	51,779	8,482

Neste caso, a RNd2t aparece como melhor opção para previsão das próximas 24 horas.

A figura 26 ilustra a previsão nas próximas 24 horas para o bairro de Uvaranas a partir de 26/12/2003 09h00min com os modelos RNh tangente, RNd1 tangente, RNd2 tangente, RLMh, RLMd1 e híbrido. Os erros RMSE e MAPE para cada modelo encontram-se na tabela 8.

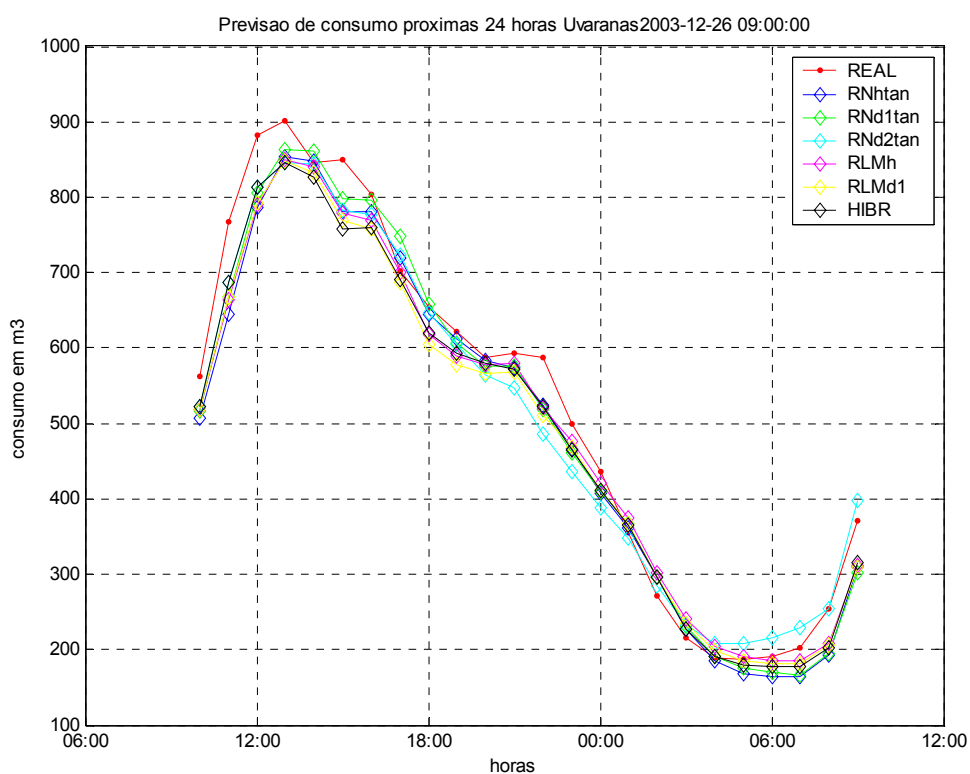


Figura 26 – Previsão de consumo das próximas 24 horas – Uvaranas.

Tabela 8 – Erros RMSE e MAPE para previsões das próximas 24 horas com modelos de redes neurais, regressão linear múltipla e híbrido previsão a partir de 9:00 de 23/12/2003 – Uvaranas.

Modelo \ Erro	RMSE	MAPE
RNhl	43,457	6,970
RNht	47,025	7,934
RNd1l	46,620	7,533
RNd1t	42,307	7,552
RNd2l	49,877	7,976
RNd2t	43,114	7,184
RLMh	43,539	6,980
RLMd1	46,724	7,545
RLMd2	49,889	7,971
HÍBRIDO	42,123	7,013

Dado que a previsão das próximas 24 horas pode ser realizada a partir de qualquer hora do dia, o desempenho dos modelos varia de acordo com a hora inicial da previsão. No apêndice A os gráficos das figuras 33, 35 e 37 mostram o RMSE médio para previsão das próximas 24 horas para todos os modelos tipo RNA, RLM e híbrido a partir de cada hora para as três regiões de consumo.

4.3.2 Modelos tipo Box e Jenkins

Para os modelos tipo Box e Jenkins, o procedimento foi similar ao utilizado para RNA e RLM, utilizando técnica *multi-step* de previsão e o modelo híbrido utilizando o modelo que reproduz o menor erro para a determinada hora e dia da semana. Para teste foram utilizados os últimos dois meses das séries temporais de cada bairro, e para cálculo do erro a ser inserido no modelo auto-regressivo utilizou-se o período de dados até uma hora antecedente ao período de teste.

A figura 27 ilustra a previsão nas próximas 24 horas para o bairro de Uvaranas a partir de 23/12/2003 09h00min com os modelos AR(2), AR(5) e híbrido. Os erros RMSE e MAPE para cada modelo encontram-se na tabela 9.

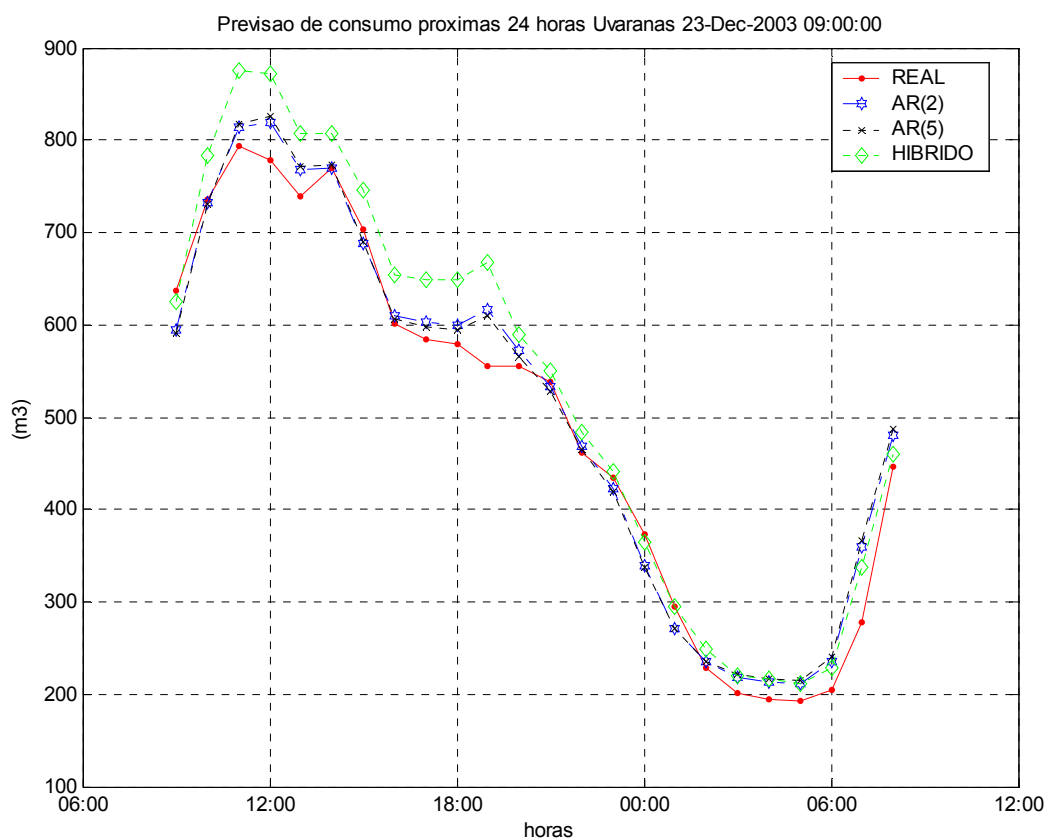


Figura 27 – Previsão de consumo das próximas 24 horas – Uvaranas.

Tabela 9 – Erros RMSE e MAPE para previsões das próximas 24 horas com modelos de Box e Jenkins a partir de 9:00 de 23/12/2003 – Uvaranas.

Modelo \ Erro	RMSE	MAPE
AR(1)	49,525	9,292
AR(2)	30,008	6,172
AR(5)	31,740	6,667
AR(25)	44,151	6,959
AR(27)	44,730	7,047
AR(29)	44,705	7,041
ARMA(2,1)	59,379	9,429
HÍBRIDO	49,312	8,170

No apêndice B os gráficos das figuras 39, 41 e 43 mostram o RMSE médio para previsão das próximas 24 horas de todos os modelos tipo BOX E JENKINS e híbrido a partir de cada hora para as três regiões de consumo.

Em se tratando dos modelos de redes neurais artificiais e regressão linear múltipla, para cada região de consumo foram geradas inicialmente nove matrizes 7x24 com os erros RMSE e nove matrizes de mesmo tamanho com os erros MAPE para previsão da próxima hora a partir de uma dada hora. A partir destas nove matrizes foi gerada uma matriz dos melhores modelos para cada hora e dia da semana, com a qual se utilizou para compor o modelo híbrido, totalizando 10 modelos diferentes para cada região de consumo. Estas matrizes para os bairros de Uvaranas, Distrito Industrial e Centro Histórico encontram-se no apêndice A.

Tomando-se os dados dos últimos dois meses, foi simulada a previsão do consumo para as próximas 24 horas para cada um dos 10 modelos a partir de cada hora contida neste período, independente do dia da semana. Desta simulação, obteve-se uma matriz 10 x 24 x 2 da média de erros RMSE e MAPE de previsão para as próximas 24 horas com cada modelo a partir de cada hora. A partir desta matriz observa-se que, dependendo da hora inicial de previsão, um diferente modelo pode ser mais adequado. As comparações entre modelos, bem como as tabelas dos modelos mais indicados para cada hora encontram-se no apêndice A.

Para cada região de consumo, modelos distintos são mais adequados para previsão das próximas 24 horas a partir de uma hora específica. No caso de Uvaranas, verifica-se uma predominância nos modelos de redes neurais com função de ativação tangente hiperbólica, especialmente para o RNd2, com uso da primeira e segunda diferença. O RNd2, porém, quando utilizado na hora indevida, como o caso das 04h00min às 07h00min e das 19h00min às 23h00min, pode ocasionar grandes erros. Este modelo parece ser melhor em desempenho quando utilizado em horas onde existe uma variação maior de consumo, como o caso das 08h00min às 17h00min, como mostra a tabela 15 (apêndice A).

Para o Distrito Industrial os modelos de redes neurais também predominaram sobre os demais, intercalando entre o RNh com função de ativação tangente hiperbólica e o RNd2 com função de ativação linear, tabela 16 (apêndice A). Os modelos de regressão linear múltipla também apresentam bons resultados praticamente acompanhando o RNd2, tal que sua escolha pode vir a ser preferível dado a facilidade de obtenção.

Para o Centro Histórico os modelos lineares prevaleceram, principalmente os de regressão linear múltipla, com especial atenção para o RLMh. O modelo híbrido aparece como possível solução em alguns casos, tabela 17 (apêndice A).

A partir da simulação da previsão das próximas 24 horas a partir de cada hora dos últimos dois meses de dados, obtiveram-se os histogramas do erro MAPE do melhor modelo para cada previsão de 24 horas, bem como a frequência em que cada modelo aparece com melhor desempenho. Cada modelo é representado por um índice, de acordo com a tabela 10.

Tabela 10 – Índices dos modelos do tipo RNA e RLM.

ÍNDICE	MODELO	ÍNDICE	MODELO
1	RNh	6	RNd2 tangente
2	RNh tangente	7	RLMh
3	RNd1	8	RLMd1
4	RNd1 tangente	9	RLMd2
5	RNd2	10	Híbrido

O primeiro histograma da figura 28 indica que o erro de previsão das próximas 24 horas para o bairro de Uvaranas concentra-se em torno de 10 por cento. As colunas no segundo histograma indicam a proporção com que cada modelo (modelo 1 a modelo 10) obteve os melhores desempenhos. Verifica-se que os modelos 2 (RNh tangente), 4 (RNd1 tangente) e 6 (RNd2 tangente) alcançam melhores resultados para a previsão nas próximas 24 horas.

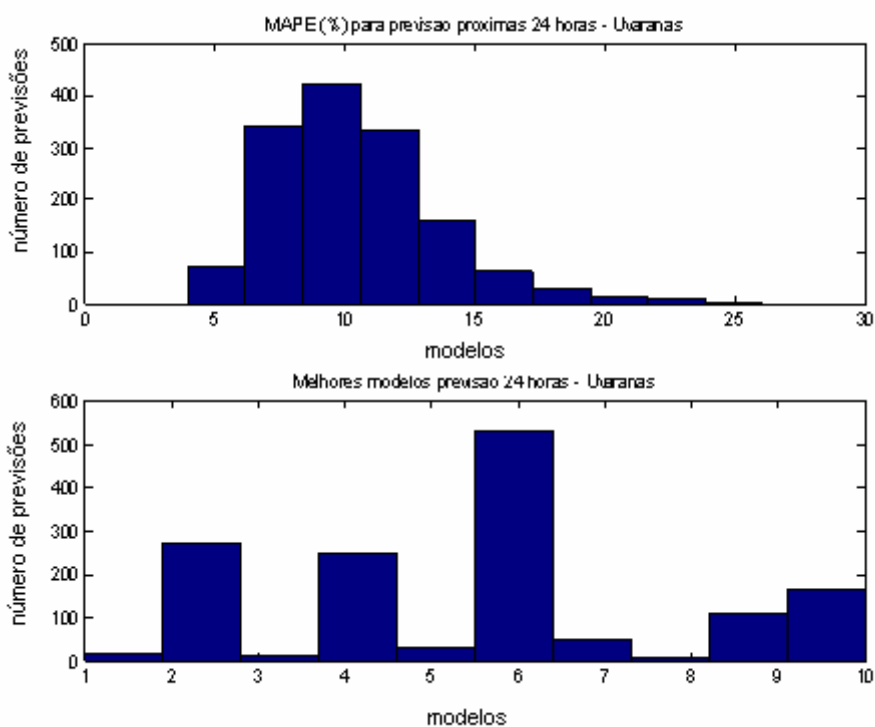


Figura 28 – Histogramas dos resultados da previsão de consumo das próximas 24 horas – Uvaranas .

O primeiro histograma da figura 29 indica que o erro de previsão das próximas 24 horas para o Distrito Industrial concentra-se em torno de 13 por cento. As colunas do segundo histograma indicam que o modelos 2 (RNh tangente) obteve melhores resultados para a previsão para as próximas 24 horas, seguido pelo modelo 10 (híbrido).

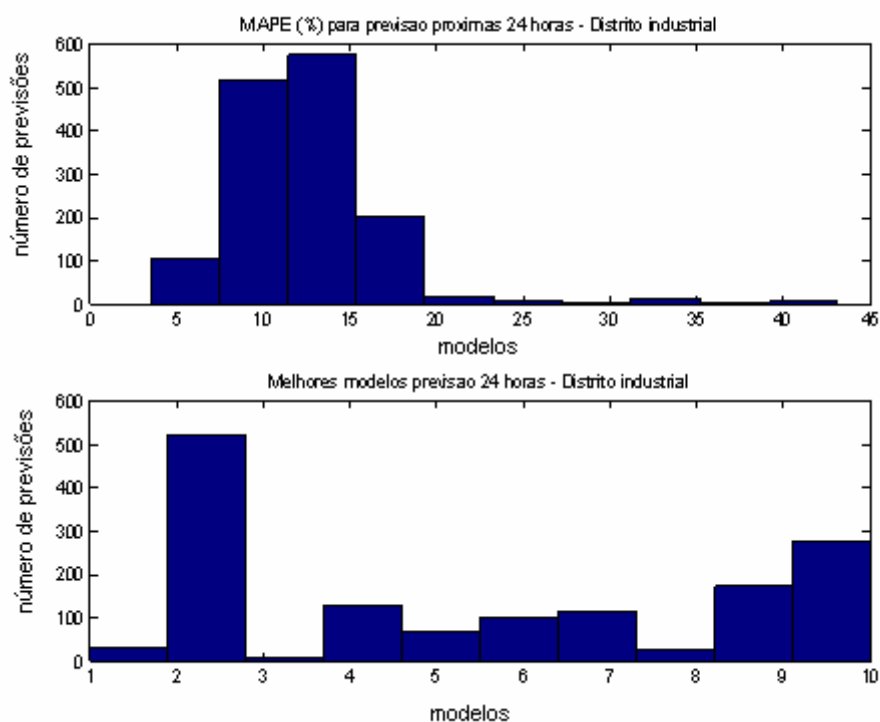


Figura 29 – Histogramas dos resultados da previsão de consumo das próximas 24 horas – Distrito Industrial.

O primeiro histograma da figura 30 indica que o erro de previsão das próximas 24 horas para o Centro Histórico concentra-se em torno de 18 por cento. As colunas do segundo histograma indicam que o modelos 7 (RLMh), 8 (RLMd1) e 9 (RLMd2) obtiveram os melhores resultados para a previsão para as próximas 24 horas, seguido pelo modelo 10 (híbrido).

Convém salientar que os resultados de desempenho dos modelos e conseqüentemente, da escolha do melhor modelo variam com o horizonte de previsão, ou número de passos para frente requeridos na previsão. Para ilustrar, foi simulada a previsão do consumo nas próximas 12 horas com os 10 modelos de redes neurais, regressão linear múltipla e híbrido para o bairro de Uvaranas. Os histogramas estão ilustrados na figura 31. Verifica-se que, para previsão das próximas 12 horas o erro decaiu sensivelmente, e o modelo híbrido se apresenta como melhor opção de escolha mais freqüentemente que na previsão das próximas 24 horas.

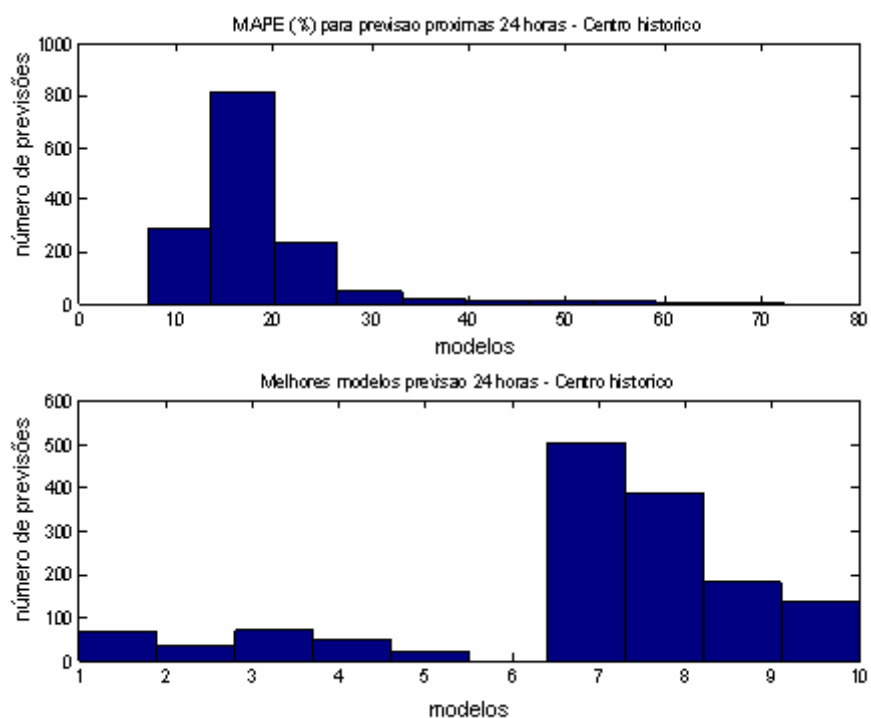


Figura 30 – Histogramas dos resultados da previsão de consumo das próximas 24 horas – Centro Histórico.

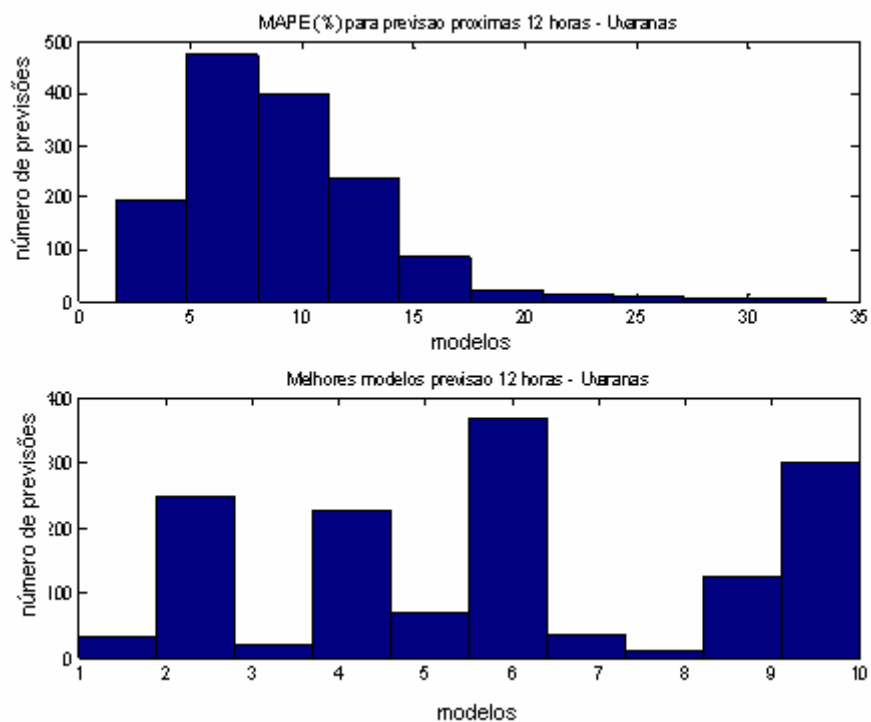


Figura 31 – Histogramas dos resultados da previsão de consumo das próximas 12 horas – Uvaranas.

Os modelos tipo Box e Jenkins , apesar de mostrarem resultados sensivelmente inferiores para previsão horária em relação aos modelos tipo RNA e RLM, para previsão nas próximas 24 horas, alcançaram melhores resultados em erro MAPE e RMSE em todas as regiões de consumo estudadas, o que demonstra serem mais estáveis, com erros mais delimitados.

Comprovando os gráficos de autocorrelação parcial para cada região de consumo (figuras 14,15 e 16), o melhor modelo para Uvaranas, de uma forma geral foi o AR(25), pra o Distrito Industrial o modelo híbrido obteve melhores resultados e para o Centro Histórico o melhor modelo foi o AR(25).

O uso nestes modelos da parcela de ajuste $aj_{(ds,h)}$ (equações 18 e 19) dependente da hora e dia da semana mostrou ser uma boa alternativa para séries temporais influenciadas pela data. Para determinação do tamanho ótimo da janela de dados passados para cálculo desta parcela de ajuste, foram simulados com o modelo AR(25) a previsão das próximas 24 horas a partir de cada hora dos dois últimos meses de dados de Uvaranas (conjunto de dados não utilizado para estimação dos parâmetros dos modelos), variando-se o número de semanas de dados históricos utilizados para calcular a matriz de ajuste segundo hora e dia da semana. Verificaram-se resultados com erro MAPE até 0.2 % menor para algumas horas, em particular para a previsão a partir das 18h00min, caso a janela de dados fosse obtida a partir das últimas dezessete semanas. A figura 32 mostra a variação do erro para 09h00min, 10h00min, 11h00min, 12h00min, 15h00min e 18h00min conforme número de semanas de dados.

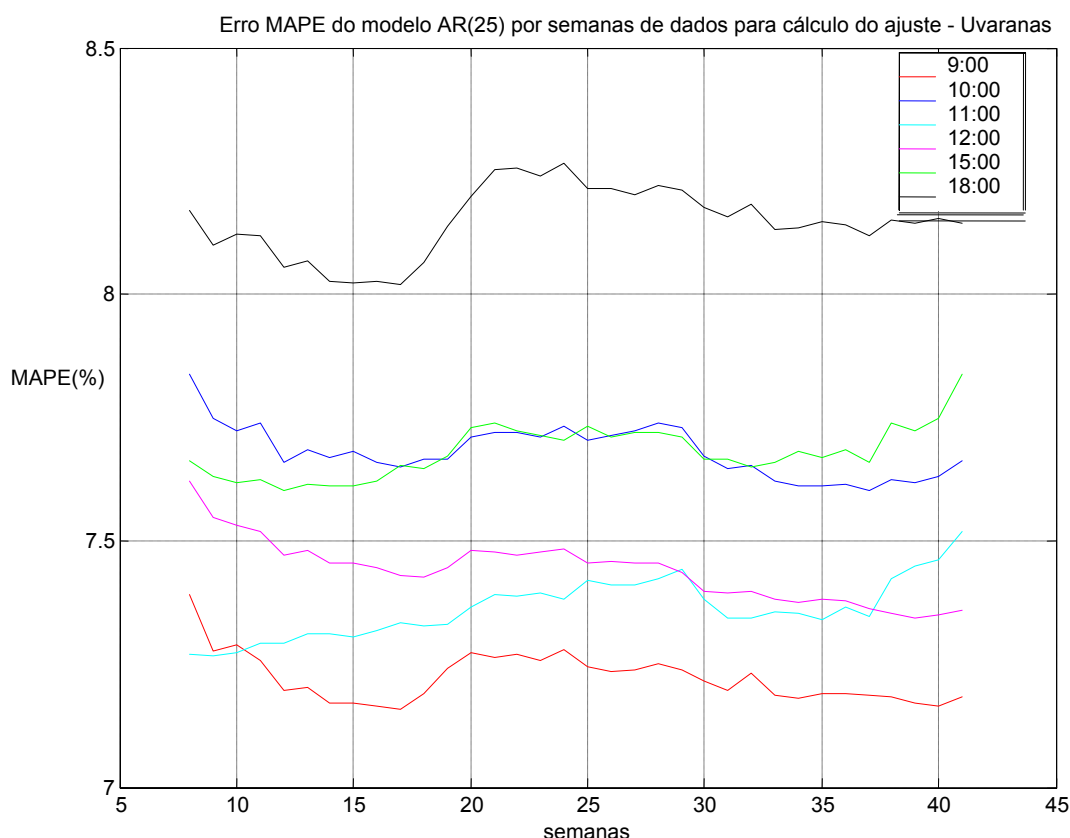


Figura 32 – Variação do erro de previsão de consumo das próximas 24 horas em relação ao número de semanas para cálculo da parcela de ajuste – Uvaranas

A determinação do tamanho ótimo da janela de dados passados para cálculo da parcela de ajuste tende a incluir certa sazonalidade ao modelo, corrigindo seus resultados a partir dos meses mais próximos. Para uma análise mais segura, esta simulação deveria ser feita para cada mês do ano, o que resultaria em um número de semanas ideal para cálculo da parcela de ajuste para cada mês e hora de previsão. Esta simulação ainda não foi possível pela falta de dados suficiente para tal. Outra análise a ser feita seria a verificação se há vantagem ou não na limitação da janela de dados para cálculo da parcela de ajuste em detrimento ao uso de uma quantidade maior de dados, o que confere ao modelo uma maior robustez. A confiança excessiva na sazonalidade também deve ser considerada com certa reserva, pelo fato que nos últimos anos o clima vem sofrendo alterações relevantes ocasionadas pelo efeito estufa (Braga *et. al*, 2002).

5 CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

Foram aplicadas as técnicas de redes neurais artificiais, regressão linear múltipla e modelos do tipo Box e Jenkins para previsão de consumo de água para três áreas com perfis de consumo distintos da cidade de Ponta Grossa, Paraná. Através da análise do erro para previsão horária, buscou-se obter um modelo híbrido utilizando os diferentes modelos de redes neurais e regressão linear múltipla, e um modelo híbrido utilizando os diferentes modelos do tipo Box e Jenkins .

Através da aplicação de diversas técnicas de previsão para regiões com características de consumo distintas verifica-se que nem sempre uma técnica ou um modelo único é uma boa solução para todos os casos. Cada região deve ser estudada separadamente, sem discriminação de modelos, buscando agregar os pontos fortes de cada modelo em uma combinação de modelos ou na escolha de um modelo mais adequado para uma dada ocasião.

Para previsões a curto prazo, o horizonte de previsão, bem como o momento a partir do qual se inicia a previsão exercem influência sobre o resultado final. Modelos distintos são mais adequados dependendo da hora inicial da previsão devido à variação característica de consumo ao longo do dia. Por apresentarem boas respostas na interpolação e respostas não tão boas na extrapolação, os modelos híbridos de redes neurais artificiais e regressão linear múltipla alcançaram resultados mais satisfatórios a medida que diminui-se o número de passos futuros de previsão, além de que, mesmo que se utilize sempre o melhor modelo para previsão da próxima hora de acordo com o dia da semana e hora atual, a combinação de modelos utilizados em série pode inserir erros que prejudicam o resultado final.

Os modelos do tipo Box e Jenkins com ajuste, por possibilitar uma correção de erro para cada previsão de consumo de acordo com a hora e dia da semana, alcançaram resultados mais satisfatórios que os de regressão linear múltipla e os projetos de redes neurais adotados (redes tipo *MLP* com até 5 neurônios na camada oculta) nas três regiões de consumo estudadas para previsão das próximas 24 horas. Para Uvaranas, obteve-se um erro MAPE para a previsão das próximas 24 horas abaixo dos dez por cento. Para o Distrito Industrial o erro permaneceu na faixa

de treze por cento, e para o Centro Histórico o erro obtido foi por volta de dez por cento, caracterizando uma boa estimativa de previsão para os fins propostos.

Um banco de dados bem estruturado é de extrema importância a fim de que seja viável a filtragem dos registros inválidos e, conseqüentemente, favorecer a obtenção de modelos mais confiáveis. Um período maior de dados é necessário para verificar o desempenho dos modelos ao longo do tempo e definir uma forma de readaptá-los automaticamente às variações de consumo que ocorrem a longo prazo. Um banco de dados mais extenso também é capaz de tornar vantajoso o uso de informações adicionais, tais como a inserção do mês ou estação do ano.

Uma maior atenção em relação à disponibilidade dos dados meteorológicos seria necessária a fim de tornar o conjunto de entradas dos modelos mais completo propiciando a investigação de novas metodologias de previsão com melhor desempenho e conseqüentemente possibilitando uma melhor explicação das parcelas de consumo envolvidas.

Dada a vasta variedade de modelos aplicáveis, os estudos futuros sobre uma análise mais detalhada do comportamento das séries temporais e identificação de características não lineares destas poderiam servir de meio para direcionar a escolha dos modelos de previsão a fim de evitar tempo despendido em pesquisa a ser repetido para cada nova série.

O presente trabalho verifica que mesmo utilizando um banco de dados elementar, é possível se obter uma previsão a curto prazo com resultados satisfatórios a ponto de serem aproveitados na otimização dos sistemas de abastecimento de água.

De posse de um modelo de previsão de demanda, diversas possibilidades de trabalhos futuros serão viáveis, envolvendo uma melhoria da operação da planta de sistemas de abastecimento de água, tais como a programação ideal de operação de bombas e válvulas, alarmes de vazamentos, comparações com *softwares* de simulação de rede hidráulica a fim de detectar anormalidades na rede, refino dos modelos de previsão, entre outros.

Tratando-se de cunho ambiental, torna-se mister o investimento em otimização de uso dos recursos impactantes ao meio ambiente, aqui envolvidos a água e a

energia, por parte das empresas de saneamento, cujo custo com este projeto é irrisório comparando-se aos benefícios envolvidos.

REFERÊNCIAS

Amaral, A. M. P., 2000, Consumo total e residencial de água tratada: aplicação de um modelo de séries temporais em Piracicaba. SP. Piracicaba, 92p. Tese – ESALQ/USP.

An A., Shan, N., Chan C., Cercone, N., Ziarko, W., 1996. Discovering rules for water demand prediction: an enhanced rough-set approach. Engineering Applications of Artificial Intelligence, vol. 9, pp. 645-653.

Baxter, C. W., Zhang Q., Stanley S. J., Shariff R., Tupas R-R. T., Stark H.L., 2001. Drinking water quality and treatment: the use of artificial neural networks. Can. J. Civ. Eng. 28 (Suppl. 1): 26-35

Box, G. E. P., Jenkins G. M., Reinsel G. C., 1994. Time series analysis: forecasting and control. 3^a edição, Prentice Hall.

Braga, B., Hespanhol I., Conejo, J. G. L., de Barros, M. T. L., Veras Jr., M. S., Porto, M. F. ^a, Nucci, N. L. R., Juliano, N. M. de A., Eiger, S., 2002. Introdução à engenharia ambiental. Prentice Hall, São Paulo.

Chen, J., Wang, W, Huang, C., 1995. Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting. Electric Power Systems Research vol. 34, 187-196.

Coelho, L. S., Trautwein Jr. B., Vieira G. E., 2003. Avaliação de redes neurais e ajustamento exponencial aplicados à previsão de consumo de água a curtíssimo prazo. VI Simpósio Brasileiro de Automação Inteligente, Bauru, SP, pp. 54-59.

Dash, P. K., Liew, A. C., Rahman, S. Dash, S., 1995. Fuzzy and neuro-fuzzy computing models for electric load forecasting. Engineering Application of Artificial Intelligence Vol. 8 no. 4, pp. 423-433.

Florent, G., Bernardt, B., Denis, G., 1998. Drinking water demand using an hourly timestep. National School for Water and Environmental Engineering – Strasbourg, França. The Fifth International Conference on Neural Information Processing.

Galvão, C. O., Valença, M. J. S., 1999. Sistemas Inteligentes: aplicações a recursos hídricos e sistemas ambientais. Ed. Universidade/UFRGS/ABRH, Porto Alegre, RS.

Harold, L. Stark, Stephen, J. Stanley, Ian D. Buchanan, 2001. Water demand forecasting using artificial neural networks. Department of Civil and Environmental

Engineering – University of Alberta, Alberta, Canada.

Haykin, S., 2001. Redes neurais: princípios e prática. 2^a edição, Bookman, Porto Alegre – RS.

Johnson, R. A., Wichern, D. W., 1998. Applied multivariate statistical analysis. Prentice hall, 4^a edição.

Klir, G. J., Yuan, B., 1995. Fuzzy sets and fuzzy logic: theory and application. Prentice Hall

Lertpalangsuntia, N., Chan, C. W., Mason, R., Tontiwachwuthikul, P., 1999. A toolset for construction of hybrid intelligent forecasting systems: application for water demand prediction. Artificial Intelligence in Engineering, vol.13, pp. 21-42.

Liu, K., Subbarayan, S., Shoults, R. R., Manry, M. T., Kwan, C., Lewis, F. L., Naccarino, J., 1996. Comparison of very short-term load forecasting techniques. IEEE Transactions on Power Systems; Vol. 11, no. 2, pp. 877-822 .

Liu, J., Savenije, H. H. G., Xu, J., 2003. Forecast of water demand in Weinan city in China using WDF-ANN model. Physics and Chemistry of the Earth, vol. 28, pp. 219-224.

Morettin, P. A., Toloí, C. M. C., 2004. Análise de séries temporais. São Paulo, Ed. Edgard Blücher.

Ozawa, K., Niimura, T., 1999. Fuzzy time-series model of electric power consumption. Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering, Edmonton, Alberta, Canada.

Petrella, R., 2002. O manifesto da água: argumentos para um contrato mundial. Ed. Vozes, Petrópolis – RJ.

Reis, A. J. R., Silva, A. P. A., 2003. Previsor neural de carga elétrica baseado na transformada wavelet discreta. Proceedings of the Brazilian Conference on Neural Networks, VI Congresso Brasileiro de Redes Neurais, São Paulo, pp. 343-348.

Rodrigues, D. D., 2002. Previsão de fornecimento de energia elétrica: um estudo de caso do ponto de vista do barramento. Dissertação de Mestrado – Universidade Federal de Minas Gerais, Belo Horizonte, MG.

Silva, R. T., Rocha, W. S., 1999. Caracterização da demanda urbana de água. Brasília,DF.

Silva, R., Conejo, J. G. L., 1998. Indicadores de perdas nos sistemas públicos de abastecimento. Brasília – DF.

Tucci, C. E. M., 1998. Modelos hidrológicos. Porto Alegre: Ed. Universidade/UFRGS/Associação Brasileira de Recursos Hídricos.

Valença M., Ludermir T., 2001. Multivariate modelling of water resources time series using constructive neural networks. Proceedings of the Brazilian Conference on Neural Networks – V Congresso Brasileiro de Redes Neurais, Rio de Janeiro, pp. 163-168.

Yassuda E. R., Oliveira W. E., Gaglianone S., Nogami P. S., Pereira B. E. B., Martins J. A., 1976. Técnica de abastecimento e tratamento de água – vol. 1. 2ª edição, CETESB, São Paulo – SP.

Yoshimoto, P. M., Tardel Filho, J., Sarzedas, G.L., 1999. Controle da pressão na rede. Brasília, DF.

Zhang, B., Dong, Z., 2001. An adaptive neural-wavelet model for short term load forecasting. Electric Power Systems Research, vol. 59, pp. 121-129.

Zhou, S. L., McMahon, T. A., Walton, A., Lewis J., 2000. Forecasting daily urban water demand: a case study of Melbourne. Journal of Hydrology, vol. 236, pp. 153-164.

Zhou, S. L., McMahon. T. A., Wang Q. J., 2001. Frequency analysis of water consumption for metropolitan area of Melbourne. Journal of Hydrology, vol. 247, pp. 72-84.

Zhou, S. L., McMahon. T. A., Walton A., Lewis J., 2002. Forecasting operational demand for an urban water supply zone. Journal of Hydrology, vol. 259, pp. 189-202.

APÊNDICE A – TESTES COM MODELOS DE RNA E RLM

Tabela 11 – Índices dos modelos tipo RNA e RLM.

ÍNDICE	MODELO	ÍNDICE	MODELO
1	RNh	6	RNd2 tangente
2	RNh tangente	7	RLMh
3	RNd1	8	RLMd1
4	RNd1 tangente	9	RLMd2
5	RNd2	--	--

Tabela 12 – Matriz de modelos tipo RNA e RLM escolhidos segundo critério de erro RMSE para a previsão horária do bairro de Uvaranas.

Dia da semana /hora	Domingo	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado
0:00	2	4	8	5	4	4	4
1:00	5	4	2	4	6	2	2
2:00	2	8	9	2	6	9	4
3:00	1	4	1	4	7	7	2
4:00	2	7	6	6	3	7	6
5:00	6	7	2	2	8	1	2
6:00	2	2	6	6	9	3	6
7:00	2	4	6	6	6	6	2
8:00	2	8	6	2	2	6	2
9:00	2	6	6	6	2	6	6
10:00	6	6	6	6	6	6	6
11:00	6	6	4	4	4	6	5
12:00	5	8	4	1	2	5	5
13:00	6	9	2	1	1	5	6
14:00	2	3	2	6	2	2	6
15:00	2	2	8	4	2	5	4
16:00	3	2	6	6	3	2	4
17:00	2	4	2	2	2	2	2
18:00	7	8	3	3	6	7	2
19:00	2	9	2	4	4	2	7
20:00	7	2	2	4	2	2	1
21:00	3	3	3	6	6	6	4
22:00	2	1	9	4	8	4	4
23:00	2	7	1	1	4	6	4

Tabela 13 – Matriz de modelos tipo RNA e RLM escolhidos segundo critério de erro RMSE para a previsão horária do Centro Histórico.

Dia da semana /hora	Domingo	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado
0:00	4	4	6	6	6	4	2
1:00	6	6	2	2	2	5	2
2:00	6	7	6	6	6	6	2
3:00	2	6	2	4	1	2	6
4:00	6	4	6	2	4	2	6
5:00	4	6	6	2	7	6	2
6:00	4	2	7	9	4	4	4
7:00	6	4	4	4	2	6	2
8:00	6	6	6	6	6	6	6
9:00	4	2	2	6	6	2	9
10:00	4	4	8	7	4	4	6
11:00	6	8	4	4	2	7	1
12:00	1	2	4	4	3	4	8
13:00	7	4	6	3	8	2	6
14:00	2	2	9	6	6	4	2
15:00	6	8	6	2	6	4	6
16:00	4	6	8	8	8	2	2
17:00	6	2	6	6	6	4	3
18:00	9	6	1	3	9	6	3
19:00	6	6	4	8	7	2	6
20:00	2	7	2	7	8	2	5
21:00	2	1	8	3	1	1	5
22:00	6	4	2	2	6	7	6
23:00	6	7	6	1	7	6	2

Tabela 14 – Matriz de modelos tipo RNA e RLM escolhidos segundo critério de erro RMSE para a previsão horária do Distrito Industrial.

Dia da semana /hora	Domingo	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado
0:00	8	4	2	6	6	4	6
1:00	9	4	6	2	2	4	6
2:00	7	2	6	2	2	6	2
3:00	7	2	2	2	4	4	4
4:00	7	4	6	2	4	4	4
5:00	1	2	4	2	4	2	6
6:00	6	2	2	9	4	2	7
7:00	4	2	4	4	6	4	4
8:00	6	6	6	6	4	2	9
9:00	2	4	6	4	2	6	4
10:00	6	2	4	6	6	7	4
11:00	2	7	2	6	2	6	4
12:00	4	4	6	6	2	1	4
13:00	2	4	2	5	4	8	4
14:00	2	2	2	5	9	2	4
15:00	4	4	6	2	4	1	4
16:00	2	4	5	6	7	4	6
17:00	4	2	2	5	5	4	2
18:00	8	6	2	9	1	4	4
19:00	6	6	2	6	1	2	3
20:00	2	2	2	7	6	5	7
21:00	2	4	4	4	6	9	6
22:00	9	5	5	4	6	4	6
23:00	2	4	1	2	4	4	1

Tabela 15 – Modelos tipo RNA e RLM escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Uvaranas segundo critério de erro RMSE.

Seleção/hora	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00
RMSE	76,352	76,113	67,553	70,958	68,536	59,31	59,465	59,797
Modelo	4	4	6	6	6	9	9	9
Seleção/hora	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00
RMSE	64,169	67,326	70,005	69,391	72,415	72,803	75,858	77,071
Modelo	10	6	6	6	6	6	6	6
Seleção/hora	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
RMSE	78,658	72,221	72,115	72,657	79,742	75,411	74,851	75,968
Modelo	6	2	2	2	4	2	2	4

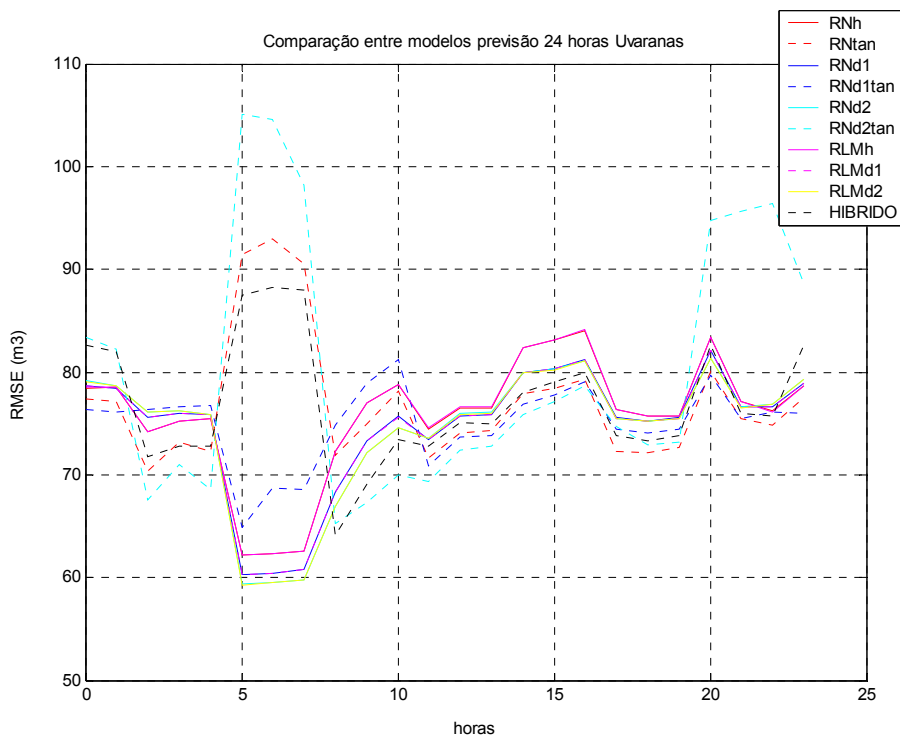


Figura 33 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Uvaranas.

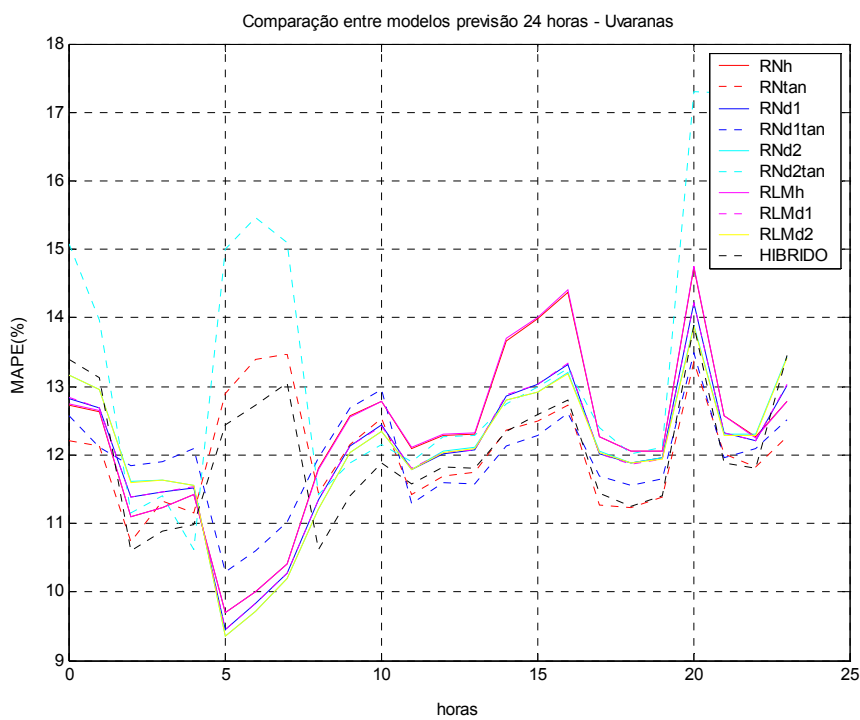


Figura 34 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Uvaranas.

Tabela 16 – Modelos tipo RNA e RLM escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Distrito Industrial segundo critério de erro RMSE.

Seleção/hora	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00
RMSE	22,179	22,495	23,808	24,788	25,612	24,234	24,688	24,197
Modelo	5	5	5	5	5	2	2	2
Seleção/hora	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00
RMSE	20,297	20,638	21,547	23,470	23,642	23,450	22,257	22,453
Modelo	2	2	2	5	5	5	9	9
Seleção/hora	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
RMSE	22,692	24,047	24,124	23,955	22,831	23,344	23,476	22,011
Modelo	9	5	5	5	5	5	5	5

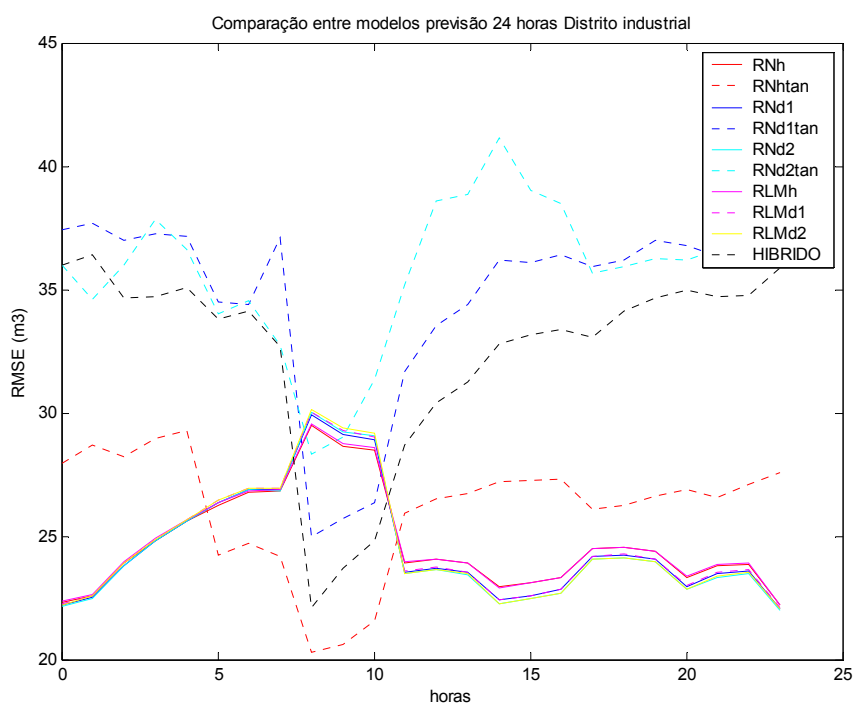


Figura 35 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Distrito Industrial.

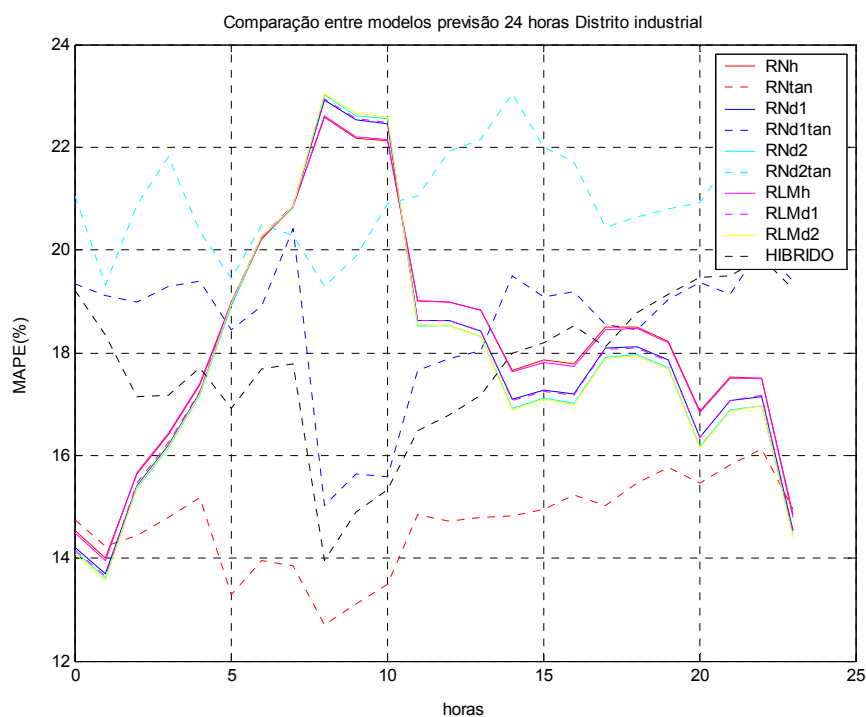


Figura 36 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Distrito Industrial.

Tabela 17 – Modelos tipo RNA e RLM escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Centro Histórico segundo critério de erro RMSE.

Seleção/hora	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00
RMSE	14,465	14,533	14,596	15,044	15,755	13,466	13,59	14,5
Modelo	1	1	10	10	1	7	7	7
Seleção/hora	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00
RMSE	14,628	14,697	14,944	13,650	13,994	14,053	13,963	13,876
Modelo	7	7	7	8	10	8	8	8
Seleção/hora	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
RMSE	14,297	14,301	14,332	14,525	14,802	14,805	14,835	14,461
Modelo	9	7	7	7	7	3	1	9

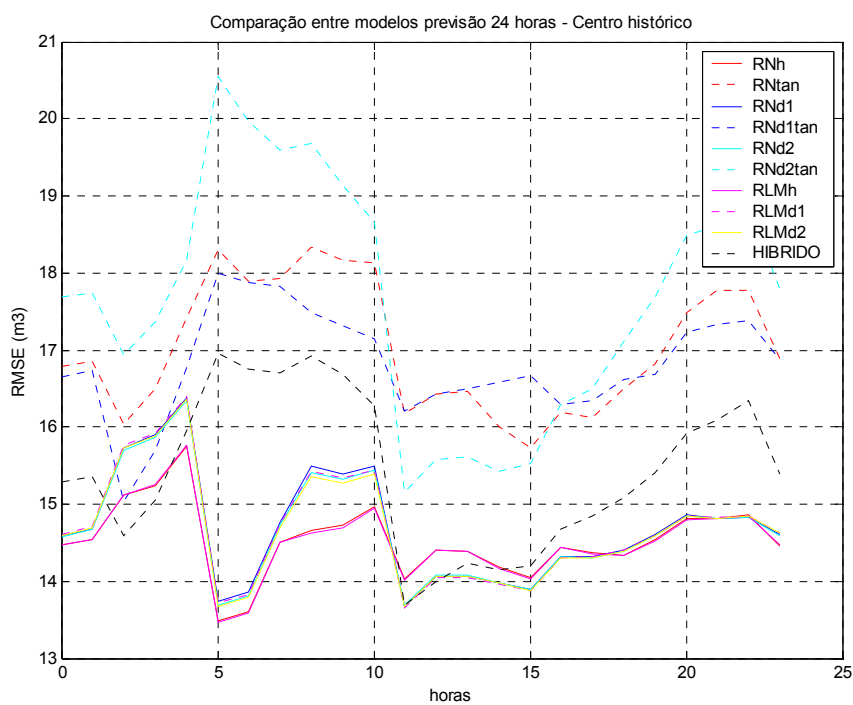


Figura 37 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Centro Histórico.

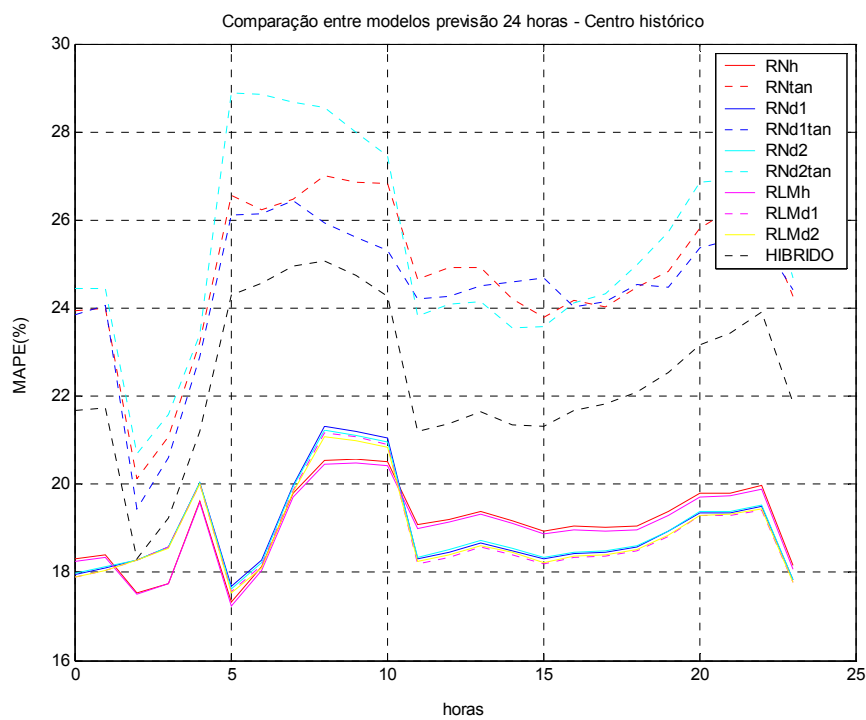


Figura 38 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Centro Histórico.

APÊNDICE B – TESTES COM MODELOS TIPO BOX E JENKINS

Tabela 18 – Índices dos modelos tipo Box e Jenkins .

ÍNDICE	MODELO	ÍNDICE	MODELO
1	AR(1)	5	AR(27)
2	AR(2)	6	AR(29)
3	AR(5)	7	ARMA(2,1)
4	AR(25)	--	--

Tabela 19 – Matriz de modelos tipo Box e Jenkins escolhidos segundo critério de erro RMSE para a previsão horária do bairro de Uvaranas.

Dia da semana /hora	Domingo	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado
0:00	7	6	4	4	4	2	6
1:00	6	4	4	4	4	5	6
2:00	3	1	1	5	3	5	7
3:00	3	3	1	1	1	1	2
4:00	1	1	4	1	7	1	1
5:00	1	1	1	1	1	1	1
6:00	6	1	5	4	6	2	4
7:00	6	4	5	4	4	4	3
8:00	6	1	1	2	1	1	2
9:00	7	3	2	1	4	6	2
10:00	2	7	3	4	4	3	4
11:00	4	3	4	6	2	3	1
12:00	7	7	1	2	1	1	7
13:00	4	1	7	1	2	7	3
14:00	2	1	4	2	1	1	1
15:00	4	2	1	5	4	5	1
16:00	5	4	3	1	4	1	5
17:00	3	1	1	5	4	4	5
18:00	7	4	1	1	2	6	3
19:00	1	3	1	5	7	7	7
20:00	6	6	4	4	5	6	6
21:00	7	6	1	1	4	4	4
22:00	1	1	5	7	5	5	2
23:00	1	1	5	4	1	5	5

Tabela 20 – Matriz de modelos tipo Box e Jenkins escolhidos segundo critério de erro RMSE para a previsão horária do Centro Histórico.

Dia da semana /hora	Domingo	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado
0:00	4	3	4	5	6	1	2
1:00	1	1	7	6	4	1	6
2:00	6	6	4	1	6	4	6
3:00	7	4	5	7	7	3	2
4:00	5	6	6	5	5	6	6
5:00	1	6	6	1	1	6	1
6:00	4	1	2	2	6	7	3
7:00	6	1	1	4	1	6	1
8:00	1	4	6	3	5	3	2
9:00	1	2	3	1	2	5	4
10:00	1	1	4	5	4	1	4
11:00	1	6	1	4	3	1	6
12:00	1	1	6	7	6	6	5
13:00	4	5	6	1	5	1	1
14:00	2	1	5	7	5	5	3
15:00	4	1	5	6	4	4	1
16:00	6	5	5	2	1	1	1
17:00	3	1	2	6	7	1	4
18:00	3	2	4	6	5	4	4
19:00	1	1	1	4	6	7	7
20:00	4	6	1	1	6	4	1
21:00	6	1	6	6	4	6	4
22:00	2	5	6	1	4	4	1
23:00	1	1	1	7	1	4	5

Tabela 21 – Matriz de modelos tipo Box e Jenkins escolhidos segundo critério de erro RMSE para a previsão horária do Distrito Industrial.

Dia da semana /hora	Domingo	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado
0:00	4	6	4	6	6	5	6
1:00	5	6	5	1	5	5	4
2:00	6	5	4	5	1	1	1
3:00	1	1	3	7	2	4	1
4:00	2	3	1	1	1	2	1
5:00	2	3	7	6	4	1	7
6:00	3	1	3	1	7	7	1
7:00	6	4	4	4	6	4	5
8:00	5	1	4	6	1	3	3
9:00	5	3	4	1	1	7	3
10:00	1	3	4	2	7	6	3
11:00	1	1	1	1	2	3	5
12:00	4	5	4	6	1	4	2
13:00	2	2	2	7	7	5	2
14:00	1	1	4	1	5	6	1
15:00	1	3	6	2	5	1	1
16:00	6	4	7	6	4	5	1
17:00	1	4	4	3	7	4	4
18:00	3	7	7	1	1	2	4
19:00	1	2	6	3	4	5	6
20:00	4	5	5	6	6	5	4
21:00	4	5	6	5	2	4	4
22:00	1	5	1	1	4	4	6
23:00	7	1	4	5	7	3	6

Tabela 22 – Modelos tipo Box e Jenkins escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Uvaranas segundo critério de erro RMSE.

Seleção/hora	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00
RMSE	55,189	54,963	53,64	55,671	55,511	53,831	54,801	54,589
Modelo	4	4	1	4	4	3	3	4
Seleção/hora	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00
RMSE	52,244	50,273	50,329	49,678	51,226	51,871	54,335	53,133
Modelo	4	4	4	4	4	4	4	4
Seleção/hora	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
RMSE	58,207	55,257	55,338	56,303	57,596	57,393	57,207	56,399
Modelo	4	4	4	4	4	8	4	4

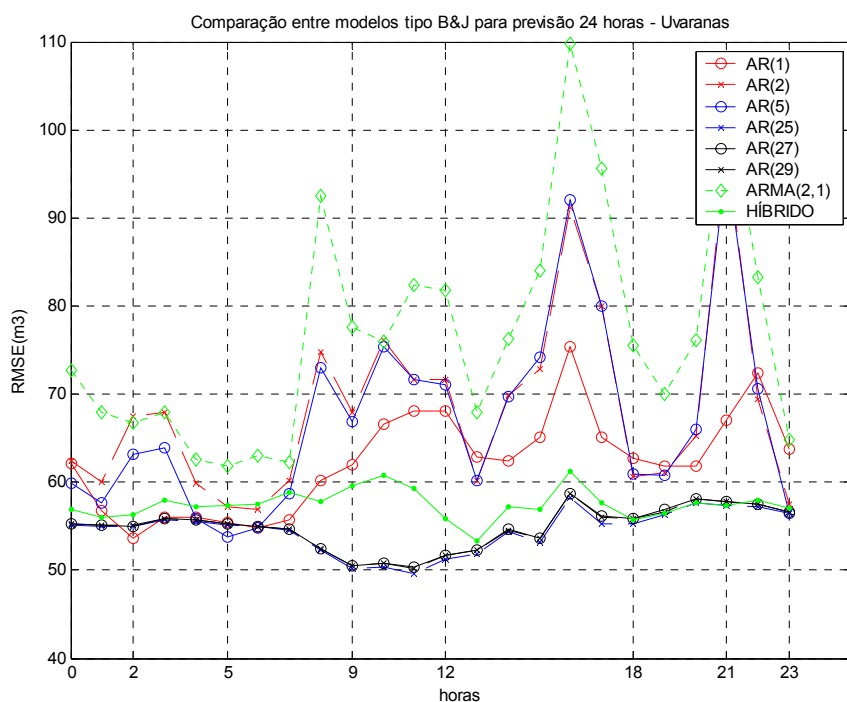


Figura 39 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Uvaranas.

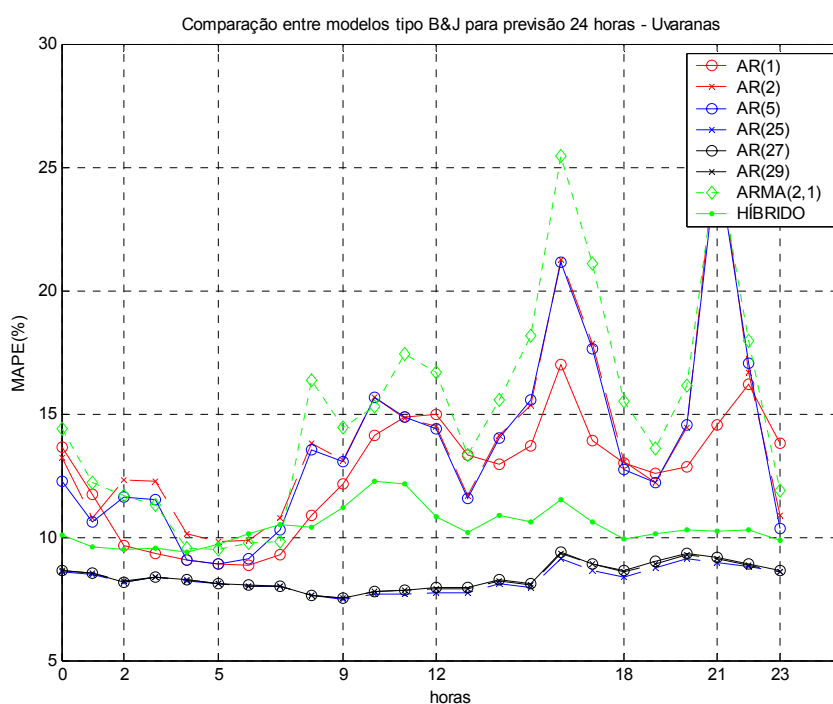


Figura 40 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Uvaranas.

Tabela 23 – Modelos tipo Box e Jenkins escolhidos para previsão das próximas 24 horas a partir de uma dada hora inicial de previsão para Distrito Industrial segundo critério de erro RMSE.

Seleção/hora	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00
RMSE	18,571	18,175	18,093	18,696	18,827	18,796	19,099	19,784
Modelo	8	8	8	8	8	8	8	8
Seleção/hora	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00
RMSE	20,355	19,713	19,770	19,571	19,766	19,818	19,094	18,917
Modelo	4	4	4	8	4	4	8	4
Seleção/hora	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
RMSE	19,760	21,232	19,246	19,115	18,893	19,172	19,273	18,857
Modelo	5	4	4	8	1	4	8	8

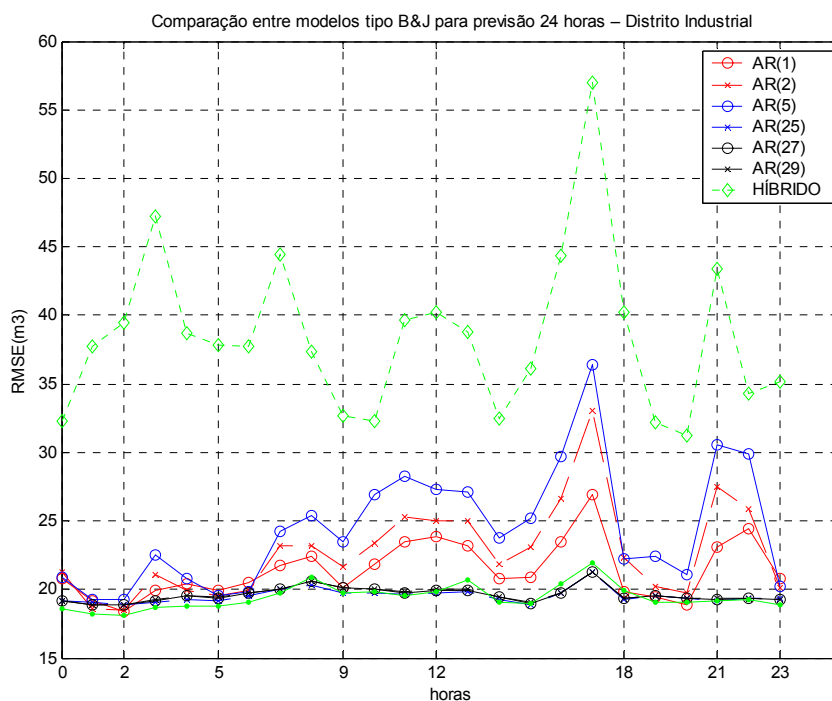


Figura 41 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Distrito Industrial.

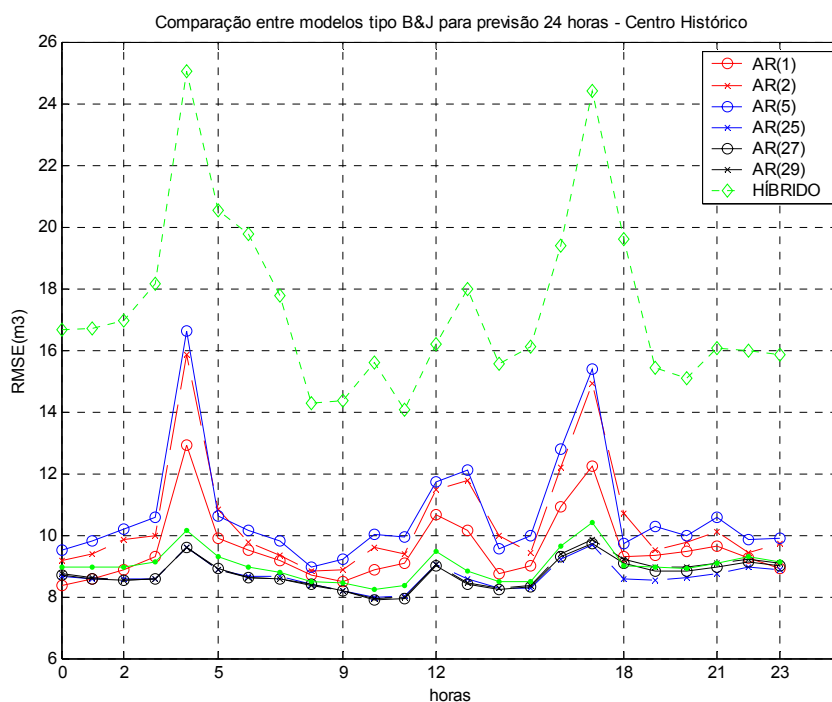


Figura 43 – RMSE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Centro Histórico.

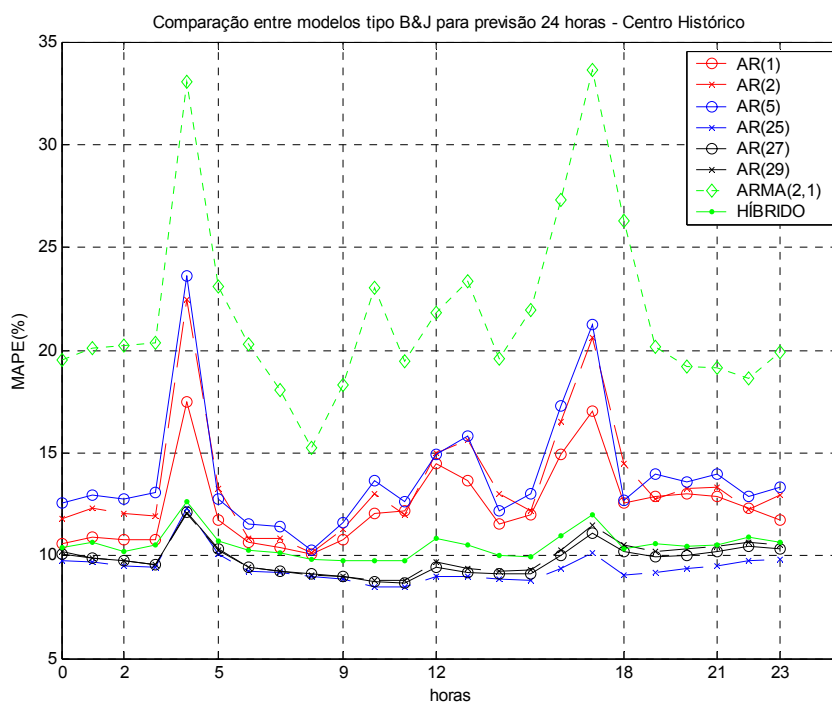


Figura 44 – MAPE para previsão de consumo das próximas 24 horas a partir de uma hora inicial para Centro Histórico.

