CLÁUDIA CRISTINA DOS SANTOS

PREVISÃO DE DEMANDA DE ÁGUA NA REGIÃO METROPOLITANA DE SÃO PAULO COM REDES NEURAIS ARTIFICIAIS E CONDIÇÕES SOCIOAMBIENTAIS E METEOROLÓGICAS

Tese apresentada à Escola Politécnica da Universidade de São Paulo, para obtenção do título de Doutor em Engenharia Civil

São Paulo

2011

CLÁUDIA CRISTINA DOS SANTOS

PREVISÃO DE DEMANDA DE ÁGUA NA REGIÃO METROPOLITANA DE SÃO PAULO COM REDES NEURAIS ARTIFICIAIS E CONDIÇÕES SOCIOAMBIENTAIS E METEOROLÓGICAS

Tese apresentada à Escola Politécinca da Universidade de São Paulo, para obtenção do título de Doutor em Engenharia Civil

Área de Concentração:
Engenharia Hidráulica e Sanitária

Orientador:
Prof. Dr. Augusto José Pereira Filho

São Paulo

2011

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 14 de julho de 2011.

Assinatura do autor _____

Assinatura do orientador _____

Santos, Claudia Cristina dos
Previsão de demanda de água na Região Metropolitana de São Paulo com redes neurais artificiais e condições sócio-ambientais e meteorológicas / C.C. dos Santos. -- ed.rev. -- São Paulo, 2011.
118 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Hidráulica e Ambiental.

1.Abastecimento de água (Sistema) 2.Redes neurais 3.Demandas (Previsão) I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Hidráulica e Ambiental II.t.

"O que faço, é uma gota no meio de um oceano. Mas sem ela, o oceano será menor."

Madre Tereza de Calcutá

Dedico este trabalho a minha querida mãe, uma grande mulher, corajosa e de um amor inconfundível que ensinou aos filhos a nunca desistir e sempre seguir em frente.

Ao meu querido esposo Wagner, por estar sempre presente, me apoiando e me encorajando nos momentos mais difíceis.

À memória do meu querido pai Agenor.

AGRADECIMENTOS

É uma benção de Deus, depois de ter vivido tantos momentos de tristeza, alegrias, incapacidade, incerteza, euforia, e cansaço, chegar ao fim desta batalha é poder agradecer e dizer muito ***obrigada*** a todos que contribuíram para isto.

Ao grande criador, por simplesmente estar viva, e poder ter chegado até aqui com muita saúde. A Nossa Senhora, que sempre intercede por mim.

Aos meus pais, Agenor e Maria, pelo amor com que fui criada e por me ensinarem que sonhar e concretizar os sonhos só depende de nossa vontade.

Ao meu querido esposo Wagner, por sempre acreditar em mim, apoiar meus sonhos, minhas idéias e representar minha segurança em todos os aspectos, pelo amor, incentivo, apoio incondicional, companheirismo e suporte emocional, além dos sacrifícios e concessões.

Aos meus irmãos, Claudio, Valéria e Viviane, pela união, carinho, companheirismo e incentivo constante.

Aos meus sobrinhos, Camila, Vinícius, Isabela, Letícia, Enzo, Giovanna e Carolina, fontes de alegrias, peço desculpas pelas vezes que não fui presente.

A minha sogra e cunhados, pela compreensão e apoio.

Ao Prof. Dr. Augusto José Pereira Filho, orientador, amigo, pelos anos de convivência, paciência, valiosas orientações e apoio nos momentos difíceis, obrigada por contribuir por meu crescimento pessoal e intelectual.

Ao Instituto Nacional de Pesquisas Espaciais - INPE pela permissão e oportunidade de continuar meus estudos e aperfeiçoamento tanto pessoal quanto profissional, assim como pela utilização de suas instalações.

Ao meu querido amigo Nelson, o grande incentivador de minha carreira cientifica.

Ao grande amigo Paulo Gurgel, pela ajuda, idéias, incentivo, paciência e convivência.

Aos meus colegas e amigos, João Roberto, Graziela, Sidney, Cristina, Tamara, Alcione e Eloísa pelo apoio, ajuda, incentivo e amizade nos bons e maus momentos.

Aos colegas Marcos Adami pela ajuda com as macros em excel e José Demísio com as RNA´s e todos apoio dado a esse trabalho.

Aos professores, José Rodolfo Scarati Martins e Mario Thadeu Leme de Barros, pelas críticas e sugestões dadas ao trabalho na qualificação.

Aos funcionários da pós-graduação do PHD pela disponibilidade e atendimento.

Aos professores do departamento PHD pelo meu desenvolvimento profissional.

A Companhia de Saneamento Básico do Estado de São Paulo (SABESP), Companhia de Tecnologia de Saneamento Ambiental (CETESB) e Fundação Centro Tecnológico de Hidráulica (FCTH), pelo fornecimento dos dados utilizados neste trabalho.

A Wagner Ginez, Viviana Borges e Kamel Zahed Filho pela ajuda e colaboração para entendimento dos dados de consumo de água da SABESP.

As pessoas que passaram pela minha vida e deixaram alguma marca, peço desculpa pelos momentos que ausente, não dei a devida atenção.

Enfim quero, agradecer a todas as pessoas que, de algum modo colaboraram de forma direta ou indireta para a realização deste trabalho.

A todos, muito obrigada!

RESUMO

Santos, C. C. **Previsão de Demanda de Água na Região Metropolitana de São Paulo com Redes Neurais Artificiais e Condições Socioambientais e Meteorológicas.** (2011). 121p. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo, São Paulo.

O presente trabalho apresenta a previsão de demanda de água em sistemas urbanos de abastecimento através de Rede Neural Artificial (RNA) utilizando dados de consumo de água e variáveis meteorológicas e socioambientais. A RNA utilizada foi uma de três camadas chamada de rede de múltiplas camadas alimentadas adiante com o algoritmo de treinamento LLSSIM (Hsu et al., 1996). Neste estudo, foram utilizados os dados de consumo de água (SABESP) e meteorológicos (IAG/USP) para o período de 2001 a 2005 para Região Metropolitana de São Paulo (RMSP). As variáveis socioambientais e meteorológicas que podem afetar o consumo de água foram analisadas. A ETA Cantareira e o setor Itaim Paulista foram utilizados para avaliar a relação entre o consumo e as variáveis antrópicas e meteorológicas para o ano de 2005. Esses conjuntos de dados foram utilizados para o treinamento, o teste e a previsão da RNA. Para a ETA Cantareira, foram criados 8 modelos e para o setor Itaim Paulista 57, sendo que os modelos 9 a 57 correspondem à previsão ideal. O desempenho dos modelos foi avaliado pelo o erro médio, erro médio absoluto, erro médio quadrático, o coeficiente de correlação, exatidão, viés, POD, FAR, CSI e POFD. Para a ETA Cantareira o melhor desempenho ocorreu para a média de 12 horas e para o Itaim Paulista a média de 6 horas. Na previsão ideal observou-se que a memória do sistema é um fator importante, principalmente quando se tem dois intervalos de tempo anterior. Os resultados mostraram a importância da memória, pois ela ajuda a melhorar o desempenho da previsão A previsão horária foi obtida com níveis de erros aceitáveis. Comparando os resultados de todas as configurações dos modelos, observou-se que há uma tendência para pequenos erros. Finalmente, conclui-se que o método proposto pode ser utilizado para previsão de consumo obtendo uma boa previsão.

Palavras-chaves: Previsão de demanda. Rede Neural artificial. Sistemas de Abastecimento de Água.

ABSTRACT

Santos, C. C. **Water Demand Forecasting in the Metropolitan Area São Paulo with Artificial Neural Network and Socioenvironmental and Meteorological Conditions.** (2011). 121p. Thesis (Doctoral) – Escola Politécnica, Universidade de São Paulo, São Paulo.

This work is concerned with the prediction of water demand in urban water supply systems using water consumption, meteorological and socioenvironmental variables in an Artificial Neural Network (ANN) system. The ANN is a three layer feed-forward network with the LLSSIM training algorithm (Hsu et. al., 1996). In this study, water consumption (SABESP) and meteorological (IAG USP) data sets between 2001 and 2005 were used for studying the Metropolitan Area São Paulo (MASP). Possible socio-environmental and meteorological conditions affecting water consumption in the MASP were analyzed. Two water treatment stations (ETA), namely, Cantareira and the Itaim Paulista were used to evaluate the relationship between water consumption against anthropic and meteorological conditions for the year 2005. These data sets were also used for training, testing and forecasting of the water consumption model with the ANN. For the Cantareira ETA, 8 model configurations were tested and 57 for the Itaim Paulista ETA. In this late case, configurations 9 to 57 were for ideal forecasts. The various model configurations were evaluated by the mean error, mean absolute error and mean square root error, correlation coefficient, bias, POD, FAR, CSI e POFD. The best performance for the Cantareira ETA was obtained for a 12-hour average of the input variables, and for the Itaim Paulista ETA, for the 6-hour average. The ANN model configurations fed with variables of previous three times steps (memory) performed best, followed by two previous time steps. The results indicate the importance of these memory to improving the performance of the forecasting. The hourly forecasting was obtained with acceptable error levels. Comparing the results of all model configurations, there is an overall tendency for minor errors. The proposed method can be used to demand forecast a good prediction.

Key-words: Water consumption Forecasting. Artificial Neural Network. Water Supply System.

LISTA DE FIGURAS

Figura 3.1: Métodos quantitativos e qualitativos e alguns de seus modelos (adaptada de Trautwein Júnior, 2004).....9

Figura 4.1: Sistema adutor da RMSP e áreas de abrangência por sistemas produtores. As cores indicam o sistema de produção (tons claros) e rede de adução (tons escuros). Os contornos (branco) indicam limites municipais. Contribuição (%) por sistema adutor para o total de 64,2 m^3s^{-1} indicado na legenda. (Fonte: Ginez e Pereira Filho, 2003).....34

Figura 4.2: a) Distribuição espacial das estações de qualidade do ar e estações meteorológicas da CETESB na RMSP. (Fonte: Pereira, Filho e Ginez, 2004). b) Localização dos postos pluviométricos da rede telemétrica do Alto Tietê (Fonte: http://www.saisp.br/site/sobreAT.htm). Os pontos vermelhos correspondem a localização das estações e dos postos pluviométricos respectivamente. 36

Figura 4.3: Rede Neural Artificial *feedforward* típica de três camadas.....38

Figura 4.4: Diagrama de bloco do fluxo do trabalho.....45

Figura 51: Falhas existentes no banco de dados CETESB e FCTH para o ano de 2004, as células em branco é onde existem dados e as coloridas são onde não têm dados. (a) exemplo de uma planilha do banco de dados da CETESB, Para todas as estações com dados meteorológicos, o F significa falha, a cor laranja indica que não existem dados para a temperatura, a azul para a umidade e rosa para a velocidade. (b) exemplo de planilha para o banco de dados da RTAT para o posto da Barragem Ponte Nova.....54

Figura 5.2: Porcentagem de falhas (%): para o período 2001 a 2005: a) Sistema de Abastecimento RMSP, cada nome identifica uma ETA; b) Estações Meteorológicas da CETESB, as siglas correspondem a: EST: estação, e o número depois de EST corresponde ao número da estação, DIR: direção do vento, VEL: velocidade do vento, TEMP: temperatura, HUM: umidade, RAD: radiação, PRESS: pressão e UVA: raio ultravioleta e c) Estações pluviométricas da Rede Telemétrica do Alto Tietê onde cada nome corresponde a uma estação diferente. As cores indicam o ano. 57

Figura 5.3: Produção de água do Sistema Produtor de Água da RMSP para o período de 2001 a 2005. 61

Figura 5.4. Anomalias da produção de água média mensal e da produção total. A linha azul corresponde a ETA ABV-Guarapiranga, a linha abóbora a ETA Garau – Cantareira, a verde a ETA Alto Tietê e a rosa pontilhada ao consumo total das três ETA´s. 62

Figura 5.5: Evolução temporal da produção de água média mensal total do Sistema Produtor de Água da RMSP. A linha azul é a produção médio para cada mês e a linha rosa é consumo médio total.....63

Figura 5.6: Evolução temporal da produção de água médio diário. a) Consumo médio diário total do Sistema Produtor de Água da RMSP. A linha azul é o consumo médio para cada dia da semana e a linha rosa é consumo médio total. b) Consumo dos três principais ETA´s da RMSP, linha vermelha ETA Alto Tietê, linha rosa ETA Cantareira e linha azul ETA Guarapiranga. 64

Figura 5.7: Evolução temporal da produção de água médio horário: a) Produção total do sistema Produtor. b) ETA Cantareira, c) ETA Guarapiranga e d) ETA Alto Tiete. Dias da semana estão indicados pela cor das linhas na legenda. 66

Figura 5.8: Evolução temporal da produção de água médio horário em dias de férias (linha vermelha) e feriado (linha azul) para a das três ETA, ETA Cantareira (topo), ETA Guarapiranga (meio) e ETA Alto Tiete (baixo). 67

Figura 5.9. Anomalias do consumo total (CT), temperatura (T), umidade relativa (UR), precipitação (R) e pressão (P). 68

Figura 5.10: Evolução temporal do consumo de água dos três sistemas adutores e do total de precipitação (topo), da temperatura (meio) e da umidade relativa (baixo). A linha azul se refere ao consumo total em todos os gráficos. Eixo das ordenadas da direita corresponde a variável meteorológica e, o da esquerda, ao consumo total. .. 70

Figura 5.11: Evolução temporal do consumo de água médio horário para 5 dias quentes (linha vermelha) e 5 dias frios (linha azul) na região das ETA Cantareira (topo), ETA Guarapiranga (meio) e ETA Alto Tiete (baixo). 71

Figura 5.12: Variações das correlações, da ETA Cantareira e do setor Itaim Paulista, em relação às variáveis meteorológicas e antrópicas. 74

Figura 5.13: Evolução temporal do consumo de água para o modelo RNA-C-8, onde a linha azul é o consumo observado, e a linha rosa é o consumo previsto pelo modelo. a) treinamento, b) previsão 77

Figura 5.13: Digrama de espalhamento da série temporal de consumo de água estimado pelo treinamento (a) e previsão (b). Estão indicados o ajuste linear (linha vermelha) e o coeficiente de correlação (R^2). 78

Figura 5.14: Índices de desempenho Exatidão dos modelos RNA para os limites de consumo de 30,53, 31,24, 31,97, 32,69 e 33,42 m^3s^{-1} indicados pelas curvas azuis, vermelhas, rosas, verdes e roxas respectivamente. (a) treinamento, (b) teste e (c) previsão. 79

Figura 5.15: Valores da frequência de bias (viés) dos modelos RNA para os limites de consumo de 30,53, 31,24, 31,97, 32,69 e 33,42m^3s^{-1} indicados pelas curvas azuis, vermelhas, rosas, verdes e roxas respectivamente. (a) treinamento, (b) teste e (c) previsão 81

Figura 5.16: Valores da probabilidade de detecção (POD), linha azul, a taxa de falso alarme (FAR), linha vermelha, índice crítico de sucesso (CSI), linha rosa, e

probabilidade de falsa detecção (POFD), linha azul claro, para os modelos de RNA do sistema da Cantareira. (a) treinamento, (b) teste e (c) previsão. 83

Figura 5.17: Evolução temporal do consumo de água para os modelos RNA-IP-1 e RNA-IP-2, onde a linha azul é o consumo observado, linha vermelha é o consumo previsto pelo RNA-IP-1 e linha verde modelo RNA-IP-2. a) treinamento, b) teste e c) previsão.....85

Figura 5.18: Digrama de espalhamento da série temporal de consumo de água observado e estimado pelo: a) RNA-IP-3 - treinamento, b) RNA-IP-4 - treinamento, c) RNA-IP-3 - teste, d) RNA-IP-4 - teste, e) RNA-IP-3 - previsão e b) RNA-IP-4 - previsão. Estão indicados o ajuste linear (linha vermelha) e o coeficiente de correlação (R^2). 87

Figura 5.19: Evolução temporal do consumo de água para os modelos RNA-IP-5 a 8, onde a linha azul é o consumo observado, linha vermelha é o consumo previsto. a) treinamento – RNA-IP-5, b) previsão – RNA-IP-5, c) treinamento – RNA-IP-6, d) previsão – RNA-IP-6, e) treinamento – RNA-IP-7, f) previsão – RNA-IP-7, g) treinamento – RNA-IP-5 e h) previsão – RNA-IP-5. 89

Figura 5.20: Valores da probabilidade de detecção (POD), linha azul, taxa de falso alarme (FAR), linha vermelha, índice crítico de sucesso (CSI), linha rosa, e probabilidade de falsa detecção (POFD), linha azul clara, para os modelos de RNA do setor Itaim Paulista. (a) treinamento, (b) teste e (c) previsão. 95

Figura 5.21: Evolução temporal do consumo de água para os modelos RNA-IP-9 a 32 e RNA-IP-33 a 55, onde a linha azul é o consumo observado, linha vermelha é o consumo previsto por RNA-IP-9 a 32 e linha verde modelo RNA-IP-33 a 55. a) treinamento e b) previsão.....98

Figura 5.22: Evolução do consumo de água para os modelos RNA-IP-32, 55, 56 e 57 e consumo observado. a) treinamento e b) previsão 99

Figura 5.23: Evolução do erro médio (EM), erro médio absoluto (MAE), erro médio quadrático (RMSE), coeficiente de correlação (r) para os modelos RNA-IP-9 a 57. (a) treinamento (erros); (b) previsão (erros); (c) treinamento (coeficientes) e (d) previsão (coeficientes). 101

Figura 5.24: Evolução exatidão (E) (azul), da freqüência de bias (Viés) (vermelha), da probabilidade de detecção (POD) (rosa), da taxa de falso alarme (FAR) (verde clara), do índice crítico de sucesso (CSI) (linha verde) e da probabilidade de falsa detecção (POFD) (roxa) para os modelos RNA-IP-9 a 57. (a) treinamento e (b) previsão.....102

LISTA DE TABELAS

Tabela 4.1: Características dos sistemas produtores da SABESP na RMSP. Para cada Sistema Produtor está indicada sua capacidades de tratamento (m^3 s^{-1}), produção (m^3 s^{-1}) e população. (Fonte: Ginez e Pereira Filho, 2003).....34

Tabela 4.2: Estações de qualidade do ar e estações meteorológicas da CETESB (Figura 3). A letra "x" indica a disponibilidade de uma variável meteorológica: temperatura do ar (T), umidade relativa (U), pressão (P), precipitação (C), vento (V) e radiação solar (R). Na última coluna estão indicados os respectivos postos de medição do SCOA (Fonte Pereira Filho e Ginez, 2004).....35

Tabela 4.3: Modelos de RNA MLP para o sistema Cantareira. Va são as variáveis antropicas (hora, dia da semana, tipo do dia, estação do ano), Vm são as variáveis meteorológicas (temperatura, umidade, chuva, pressão, direção e velocidade do vento), Q o consumo de água e t o tempo. 46

Tabela 4.4: Modelos de RNA MLP para o setor Itaim Paulista. Va são as variáveis antropicas (hora, dia da semana, tipo do dia, estação do ano), Vm são as variáveis meteorológicas (temperatura, umidade, chuva, pressão, direção e velocidade do vento), Q o consumo de água e t o tempo. 46

Tabela 4.5: Modelos de RNA MLP para previsão ótima do setor Itaim Paulista. Va são as variáveis antropicas (hora, dia da semana, tipo do dia, estação do ano), Vm são as variáveis meteorológicas (temperatura, umidade, chuva, pressão, direção e velocidade do vento), Q o consumo de água e t o tempo. Previsão com memória de 2 tempos anteriores.....47

Tabela 4.6: Modelos de RNA MLP para previsão ótima do setor Itaim Paulista. Va são as variáveis antrópicas (hora, dia da semana, tipo do dia, estação do ano), Vm são as variáveis meteorológicas (temperatura, umidade, chuva, pressão, direção e velocidade do vento), Q o consumo de água e t o tempo. Previsão com memória de 3 tempos anteriores.....48

Tabela 4.7: Tabela de contingência para contagem dos dados previstos e observados do consumo de água. 50

Tabela 5.1: Características dos sistemas produtores da SABESP na RMSP entre 2001 e 2005. Estão indicados, o número de dados, as médias, desvios padrão, mínimos e máximos de consumo de água (m^3 s^{-1}).....56

Tabela 5.2: Correlação de variáveis meteorológicas, temperatura (T), umidade (UR), chuva (R), pressão (P), direção do vento (Dir) e velocidade do vento (Vel), consumo de água (Q), hora (H), dia da semana (DS), estação do ano (EA) e tipo do dia (TD) do setor Itaim Paulista para identificar as melhores variáveis para uma previsão de consumo de água. Os valores em azul correspondem a valores maiores 0,10 e a coluna cinza indica as correlações com a variável de interesse o consumo (Q).....73

Tabela 5.3: Correlação de variáveis meteorológicas, temperatura (T), umidade (UR), chuva (R), pressão (P), direção do vento (Dir) e velocidade do vento (Vel), consumo

de água (Q), hora (H), dia da semana (DS), estação do ano (EA) e tipo do dia (TD) para ETA Cantareira para identificar as melhores variáveis para uma previsão de consumo de água. Os valores em azul correspondem a valores maiores 0,10 e a coluna cinza indica as correlações com a variável de interesse o consumo (Q).....73

Tabela 5.4: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para o treinamento dos modelos de RNA´s do sistema produtor da Cantareira. A coluna vermelha são os piores valores e a coluna azul os melhores valores.....76

Tabela 5.5: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para o teste dos modelos de RNA´s do sistema produtor da Cantareira. A coluna vermelha apresenta os piores valores e a coluna azul os melhores resultados.....76

Tabela 5.6: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para a previsão dos modelos de RNA´s do sistema produtor da Cantareira. A coluna vermelha são os piores valores e a coluna azul os melhores valores. 76

Tabela 5.7: Modelos de RNA´s que tiveram os menores e maiores valores de Exatidão (E), e a faixa de porcentagem de valores previsto corretamente , para o treinamento, teste e previsão, do sistema produtor da Cantareira. 80

Tabela 5.8: Modelos de RNA´s que tiveram valores subestimados ou superestimados devido a avaliação da freqüência de bias (viés), para o treinamento, teste e previsão, do sistema produtor da Cantareira.....82

Tabela 5.9: Média do POD, FAR, CSI e POFD, para o treinamento, teste e previsão. As células em azul indicam os melhores resultados, e as vermelhas os piores.....84

Tabela 5.10: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para o treinamento, teste e previsão dos modelos RNA-IP 1, 2, 3 e 4 do setor Itaim Paulista. A coluna vermelha são os piores valores e a coluna azul os melhores valores.....90

Tabela 5.11: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para o treinamento e previsão para os modelos RNA-IP 5, 6, 7 e 8 do setor Itaim Paulista. A coluna vermelha são os piores valores. 91

Tabela 5.12: Valores de Exatidão (E) e o limite utilizado para o treinamento, teste e previsão dos modelos RNA-IP-1 a 8. Os números em vermelhos são os menores valores e os azuis os melhores. A célula cinza apresenta valores abaixo de 90%. ..92

Tabela 5.13: Valores de frequência de bias (Viés) e o limite utilizado para o treinamento, teste e previsão dos modelos RNA-IP-1 a 8. Os números em vermelhos são os menores valores que ocorreram, e em azul os maiores. A célula cinza escura mostra os valores <0,5, a amarela =1 e cinza claro valores >1. 93

Tabela 5.14: Média do POD, FAR, CSI e POFD, para o treinamento, teste e previsão. As células em azul indicam os melhores resultados, e as vermelhas os piores. 96

Tabela 5.15: Média do E, Viés, POD, FAR, CSI e POFD, para o treinamento, teste e previsão. As células em azul indicam os melhores resultados, e as vermelhas os piores. 104

LISTA DE ABREVIATURAS E SIGLAS

AR	—	Auto Regressivos
ARIMA	—	Auto-Regressivo Integrado de Média Móvel
BIAS	—	Freqüência de Bias
BJ	—	Box & Jenkins
C	—	Cantareira
CETESB	—	Companhia de Tecnologia de Saneamento Ambiental
CORR	—	Coeficiente de correlação
CSI	—	Índice Crítico de Sucesso
CT	—	Consumo de Água total
DAEE	—	Departamento de Águas e Energia Elétrica
DAN2	—	Rede Neural Dinâmica
Dir	—	Direção do Vento
DS	—	Dia da Semana
E	—	Exatidão
$e(t)$	—	Mapeamento do erro
EA	—	Estação do Ano
EM	—	Erro Médio
EQM	—	Erro Quadrático Médio
ETA	—	Estação de Tratamento de Água
f	—	Função
f(.)	—	Função de Transferência
FAR	—	Taxa de Falso Alarme
FCTH	—	Fundação Centro Tecnológico de Hidráulica
FFNN	—	Rede Neural de Retro Alimentação (feedforward)
g_{non}	—	Função de mapeamento não linear;
GRNN	—	Rede Neural de Regressão Generalizada
H	—	Hora
hab.	—	Habitantes
IA	—	Inteligência Artificial
IAG	—	Instituto de Astronomia, Geofísica e Ciências Atmosférica
IBGE	—	Instituto Brasileiro de Geografia e Estatística

IFCS	—	Intelligent Forecasters Construction Set
IP	—	Itaim Paulista
l/d/pessoa	—	Litro/dia/pessoa
LF	—	Lógica Fuzzy
LLSSIM	—	Linear Least Squares Simplex
MAE	—	Erro Médio Absoluto
MLP	—	Percepton Multilayer
MLR	—	Regressão Linear Múltipla
MSE	—	Mean Square Error
n_a	—	Números de entradas
n_b	—	Números de saídas
NRMSE	—	Erro Médio Quadrático Normalizado
ONU	—	Organização das Nações Unidas
P	—	Pressão
PIB	—	Produto Interno Bruto
POD	—	Probabilidade de Detecção
POFD	—	Probabilidade de Falsa Detecção
Q	—	Consumo de Água
r	—	Coeficiente de correlação
Rad	—	Radiação Solar
R	—	Chuva
R^2	—	Coeficiente de determinação
RBNN	—	Rede Neural de Base Radial
RLM	—	Regressão Linear Múltipla
RMSE	—	Erro Médio Quadrático
rmse	—	Raiz Quadrada do Erro Médio
RMSP	—	Região Metropolitana de São Paulo
RNA	—	Rede Neural Artificial
RNB	—	Redes Neurais Biológicas
RTAT	—	Rede Telemétrica do Alto Tietê
SABESP	—	Companhia de Saneamento Básico do Estado de São Paulo
SAM	—	Sistema Adutor Metropolitano ou Sistema Integrado de Abastecimento de Água da Região Metropolitana de São

		Paulo
SCADA	—	Supervisory Control and Data Acquisition System
SCOA	—	Sistema de Controle do Abastecimento
S_h	—	Sinal de entrada efetivo
T	—	Temperatura do ar.
t	—	Tempo
TD	—	Tipo do Dia
$t_k(p)$	—	Saída da rede
UR	—	Umidade Relativa
V	—	Vento
Va	—	Variáveis Antrópicas
Vel	—	Velocidade do Vento
Vm	—	Variáveis Meteorológicas
$W_1, \ldots, W_n$	—	Pesos sinápticos
WDF-ANN	—	Water Demand Forescast using Artificial Neural Network
x	—	Variável independente
$x_1, \ldots, x_n$	—	Sinal de entrada
y	—	Variável dependente

SUMÁRIO

1. INTRODUÇÃO **1**

2. OBJETIVOS **5**

2.1. JUSTIFICATIVA **5**

3. REVISÃO BIBLIOGRÁFICA **8**

3.1. Previsão de Demanda 8
3.1.1. Previsão de Demanda de Água 10
3.2. Redes Neurais Artificiais 17
3.2.1. Modelos de Rede Neural Artificial para demanda de água 22

4. DADOS E METODOLOGIA **31**

4.1. Área de Estudo 31
4.1.2. Consumo da RMSP 32
4.2. Dados Utilizados 35
4.3. Redes Neurais Artificiais 37
4.3.1. Formulações 37
4.3.2. Algoritmo de treinamento 40
4.4. Previsão de Demanda de Água 44
4.5. Critérios de Avaliação 48

5. RESULTADOS **52**

5.1. Análise dos dados disponíveis 52
5.2. Análise do consumo de água 60
5.3. Análise do consumo de água comparadas as variáveis meteorológicas 68
5.5. Análise utilizando Rede Neural Artificial 75
5.5.1. ETA Cantareira 75
5.5.2. Itaim Paulista 84

6. CONCLUSÕES E RECOMENDAÇÕES **105**

6.1. Conclusões 105
6.2. Recomendações para trabalhos futuros 110

7. REFERÊNCIAS BIBLIOGRÁFICAS **111**

ANEXO **118**

1. INTRODUÇÃO

A água é essencial à vida, sendo um elemento necessário em quase todas as atividades humanas. Neste contexto, destacam-se a geração de energia elétrica, uso doméstico, uso industrial, irrigação, navegação, recreação, aquicultura, piscicultura, pesca e também, para assimilação e afastamento de esgotos (Setti et al., 2001). A disponibilidade de água em quantidade e qualidade representa um dos principais desafios em áreas urbanas brasileiras, principalmente nas grandes metrópoles. A crescente demanda de água e o risco de desabastecimento estão associados ao aumento da população urbana. Muitas metrópoles brasileiras têm crescido desordenadamente; coberturas vegetais são removidas e solos são impermeabilizados; em conseqüência, surgem alterações hidrológicas (Setti et al., 2001) e meteorológicas tais como o aumento, da temperatura do ar e de riscos de enchentes (Pereira Filho, 1999). Esses fatores também influenciam o consumo de água (Pereira Filho et al., 2004a).

A ocupação de áreas de proteção de mananciais deteriora a qualidade da água e impõe riscos adicionais ao abastecimento urbano (Teixeira, 2000). O espaço geográfico influencia o uso da água. Diferenças climáticas, geológicas, topográficas e outras existem entre regiões e podem alterar a sua quantidade e qualidade (Giansanti, 1993). O planejamento do uso da água faz-se necessário por causa da demanda populacional crescente e disponibilidade de recurso na natureza. A produção de água é variável no tempo e no espaço. As reduções de perdas nos sistemas de distribuição de água e do desperdício resultam em uma economia considerável de água (Zahed Filho, 1990).

Os usos, domésticos, comerciais, públicos e industriais da água nas cidades, vilas e pequenos núcleos urbanos são considerados urbanos. A demanda urbana de água é constituída pelo uso que se refere às atividades que dão origem ao núcleo urbano: indústria, comércio, prestação de serviços públicos e privados. Os sistemas de abastecimento urbanos distribuem água para diversos usos que variam do consumo

humano ao controle do fogo, e dos processos industriais à irrigação de jardins (Setti et al.,2001).

A operação dos sistemas de abastecimento e distribuição de água é um dos componentes principais deste consumo de água. A operação desses sistemas requer ajustes frequentes em resposta a variação da demanda, a fim de minimizar os custos da distribuição (Zahed Filho,1990). A demanda varia ao longo do tempo. Por exemplo, ela depende da hora do dia, do dia da semana, do mês e da estação do ano. Essas variações podem ser afetadas pelas condições meteorológicas, feriados e atividades domésticas e industriais (Mukhopadyay et al., 2001).

A demanda de água depende ainda de fatores como o tamanho da cidade, características da população, natureza e tamanho dos estabelecimentos comerciais e industriais e do custo do abastecimento. Nas últimas décadas, houve um aumento no desenvolvimento de modelos estocásticos, tipicamente de regressão múltipla e de série temporal, para previsão do consumo urbano de água. Séries temporais de consumo urbano de água anual têm sido relacionadas à população, à renda familiar, ao preço da água, à chuva, à temperatura do ar e à evaporação (Wong, 1972; Young,1973; Willsie e Pratt, 1974).

A escassez de recursos hídricos para abastecimento urbano atinge várias cidades particularmente as regiões metropolitanas, e essa situação não é diferente na Região Metropolitana de São Paulo (RMSP). As cidades que cresceram na bacia do Alto Tietê são atualmente atendidas por sistema de abastecimento de forma limitada (Teixeira, 2000). O crescimento populacional e a expansão urbana caótica restringem as projeções de demandas futuras (Secco, 2002). Nos anos 80 e 90, investimentos limitados em projetos de abastecimento resultaram no rodízio de água na RMSP. Nessa época, um plano diretor metropolitano foi elaborado para o abastecimento de água com base nas projeções dos censos do Instituto Brasileiro de Geografia e Estatística (IBGE). Esse plano permitiu o fim do rodízio em meados de 1988, mas ainda remanesceram algumas dificuldades relacionadas a acidentes e à manutenção da malha adutora (Secco, 2002).

Em 2000, houve uma redução dos índices pluviométricos na RMSP. A estiagem prolongada de meados de 2000 provocou o racionamento no sistema Alto Cotia, e depois, no sistema Guarapiranga. Este fato se agravou no segundo semestre de 2001 e afetou também o sistema Cantareira, que atende 55% da demanda de água na RMSP (Secco, 2002). Os níveis operacionais das represas desse sistema atingiram os menores valores na história do abastecimento da RMSP, devido a redução dos índices pluviométricos.

No período de 2003 a janeiro de 2004, o volume de água nas represas do Sistema Cantareira foi o menor de quase 70 anos. Por outro lado, nos últimos dias de janeiro e nos primeiros dias de fevereiro de 2004, houve várias enchentes na RMSP, mas mesmo assim, os níveis dos reservatórios não subiram acentuadamente. Esse fato se deve aos eventos serem locais e provocados por ilhas de calor e circulação de brisa marítima (Pereira Filho et al., 2004b). A ilha de calor gera circulações que convergem sobre o centro urbano e, assim, com a umidade suprida pela brisa do mar, há uma maior freqüência de chuvas sobre a RMSP (Pereira Filho, 1999; Borges e Pereira Filho, 2000) e menos sobre as bacias que suprem os mananciais do sistema de Abastecimento da Companhia de Saneamento Básico do Estado de São Paulo (SABESP).

Desta forma, o gerenciamento da água na RMSP requer um sistema de informação e diagnóstico de prognóstico para orientar as tomadas de decisão. Nesse sentido, questões tais como: quais os impactos das condições meteorológicas no consumo, quais as características do consumo regional e em que dias da semana a demanda é maior ainda requerem maior análise.

A previsão de demanda também é um fator importante para o gerenciamento do consumo de água, principalmente quando se pretende criar ou expandir um sistema de abastecimento urbano. Existem vários métodos de previsão de demanda, porém, métodos que utilizam Redes Neurais Artificiais – RNA ainda existem poucos na literatura (Adamowski,2008, Mingqi, 2009). A principal motivação para o uso de RNA´s é a necessidade de desenvolver modelos de previsão de demanda de água para sistemas urbanos de abastecimento como metodologias específicas para problema de previsão no Brasil especificamente para RMSP. Por esta razão, que

estudar os principais efeitos das variáveis socioambientais e meteorológicas e o desempenho do uso de RNA's são as principais motivações nesta tese. O trabalho visa fazer uma previsão da demanda de água em sistemas urbanos de abastecimento, usando a técnica de redes neurais artificiais para obter resultados mais confiáveis e precisos em previsões de curto prazo.

2. OBJETIVOS

O principal objetivo do trabalho é desenvolver um modelo de previsão de demanda de água a curto prazo para sistemas de abastecimentos urbanos, por meio de redes neurais artificiais (RNA), considerando a escala de discretização horária. O trabalho de pesquisa envolve estudos na área de RNA e a análise das variáveis que influenciam na demanda.

Pretende-se, com este estudo, verificar o desempenho das redes neurais artificiais em prover prognósticos confiáveis de valores futuros de demanda de água. Para o estudo de caso do modelo de RNA proposto, foi utilizado como entrada um conjunto de dados observados de consumo de água obtidos do Sistema Adutor Metropolitano de São Paulo.

2.1. JUSTIFICATIVA

Face a crise de abastecimento de água nas grandes metrópoles, há necessidade de se prever o consumo de água, para melhorar e otimizar a operação de sistemas de abastecimento. O clima influencia os seres vivos, sabendo-se que, num dia de altas temperaturas o consumo de água tende a aumentar (Pereira Filho et al. 2004a). Por outro lado, depois de uma tempestade associada a enchentes, o consumo também aumenta. Outro fato relacionado à chuva é o aumento do consumo após um período contínuo de chuva e a chegada do sol, quando as pessoas aproveitam para lavarem roupas, carros e calçados. Dessa forma, uma avaliação de quais variáveis hidrometeorológicas podem afetar mais o consumo de água da RMSP seria muito útil, podendo inclusive ser variáveis de entrada para ativar uma Rede Neural Artificial (RNA). As RNA´s têm sido utilizadas com sucesso para modelar relações envolvendo séries temporais em várias áreas (VEMURI, 1994). O sucesso das RNA para modelar sistemas dinâmicos em vários campos da ciência e da engenharia sugere que ela seja uma boa solução para o problema de previsão do consumo de

água. A RNA aprende e, portanto generaliza o que se torna uma vantagem em relação aos modelos tradicionais de modelagem de consumo de água. Além disto, depois de treinada para operar em um sistema específico, a RNA pode facilmente ser novamente treinada para lidar com pequenas modificações nas condições operativas do ambiente (Santos, 2001).

O principal objetivo de uma empresa de abastecimento urbano de água é combinar o abastecimento e a demanda em um serviço aceitável aos consumidores. Para que isso ocorra são necessários ajustes muito freqüentes em resposta às variações da demanda a fim de minimizar os custos. A variação da demanda está relacionada com o período do dia, dia da semana, mês e a estação do ano. Além disso, o consumo varia nos fins de semana e dias de feriados, de acordo com as atividades domésticas e industriais regulares dos consumidores. O prognóstico considera os fatores anteriores, que são necessários para otimização do sistema em períodos sucessivos de controle. Freqüentemente, os operadores do sistema fazem isto fundamentado na experiência prática. Usam informação como o dia da semana, hora do dia, eventos especiais (feriado, um evento esportivo), do dia anterior e de consumos da hora anterior. Somente recentemente a análise entre consumo de água e dos dados climáticos monitorados pelas empresas começaram a ser utilizada para previsão (Mukhopadhyay et al., 2001; Zhou et al., 2000).

Os principais fatores associados à definição do consumo são: posição econômica e social dos consumidores, tipos de residências, uso do solo, as características das instalações hidráulicas individuais, método de medida de consumo, preço da água e as condições climáticas. Desta forma, o uso das variáveis climáticas para auxiliar na previsão inclui um trabalho adicional que é o de desenvolver modelos de previsão utilizando variáveis climáticas que possam melhorar a previsão do consumo. Assim, a vantagem de utilizar as variáveis climáticas como entrada da RNA é que elas podem auxiliar e melhorar a previsão. Segundo Zahed Filho (1990), a não linearidade das equações físicas envolvidas na otimização de redes de adução, a dificuldade na obtenção de curvas reais de custo e a dimensão e complexidade dos sistemas têm sido empecilhos para a obtenção de um modelo eficaz que faça prescindir o julgamento e a experiência do operador. Neste caso, a RNA pode tornar-se uma boa ferramenta para auxiliar nas tomadas de decisões, pois tem a

vantagem de ser uma estrutura matemática não-linear capaz de arbitrariamente representar processos não-lineares que relacionam entradas e saídas de um sistema (HAYKIN, 1994). As RNA permitem modelar processos que envolvem séries temporais que envolvem sinais de entrada e saída que apresentem alguma complexidade (VEMURI, 1994). Desta forma, o sucesso das RNA em muitos campos da ciência e da engenharia sugere sua aplicabilidade, também, na previsão do consumo de água, especialmente nos casos onde o conhecimento dos processos seja muito limitado. A elaboração de uma RNA com o objetivo de ajustar a tendência do crescimento ou da redução da curva do consumo em que admita a linearidade da variação do consumo com o tempo, para um período futuro, em que a informação do instante atual e de um instante anterior seja utilizada é que as curvas de consumos são claramente não lineares e a RNA utiliza funções que conseguem ajustar isto. Outra vantagem é que depois de estudos realizados para uma determinada região, esta rede poderá ser adaptada rapidamente para outras regiões.

Outras vantagens das RNA´s , de grande valor, para modelagem da previsão são:

- as redes neurais são mais apropriadas para problemas de previsão dinâmica, porque os pesos envolvidos podem ser atualizados quando observações recentes estão disponíveis;
- pequenos erros na entrada não produzem mudanças significativas na saída por causa do processamento distribuído;
- elas armazenam em registros os dados criados, não sendo necessário guardar na memória todos os dados processados anteriormente e
- elas não requerem qualquer outro dado de entrada a não ser aqueles vetores de entrada e saída necessários ao treinamento.

3. REVISÃO BIBLIOGRÁFICA

Este capítulo tem o objetivo fornecer subsídios que fortaleçam a importância da previsão de demanda através de Redes Neurais Artificiais (RNA), apresentando alguns trabalhos relevantes a esta pesquisa.

3.1. Previsão de Demanda

A previsão é a estimativa de um valor ou variável num tempo futuro utilizando dados anteriores (Martins e Laugeni, 1999), sendo que as previsões de demanda são fundamentais para auxiliar na determinação dos recursos necessários para uma empresa. A demanda é a quantidade de insumo que um consumidor adquire por determinado preço em determinado período de tempo. A escolha do método a ser adotado para a previsão de demanda depende da natureza do produto e de vários fatores, tais como disponibilidade de dados históricos, horizonte de previsão a longo, médio ou curto prazo, precisão necessária, orçamento disponível e padrão dos dados existentes (sazonal, cíclico ou tendência) (DAVIS et al., 1997). As etapas para um modelo de previsão de demanda são:

- objetivo do modelo;
- coleta e análise dos dados;
- seleção da técnica de previsão;
- obtenção das previsões e
- monitoração do modelo.

As técnicas dos modelos de previsão podem ser de duas categorias:

- qualitativos e
- quantitativos.

As técnicas qualitativas estão baseadas em estimativas e opiniões de especialistas, sendo utilizadas quando não existem dados, e as quantitativas são usadas para

definir explicitamente como a previsão é determinada. A lógica é claramente determinada, e as operações são matemáticas. Dois tipos básicos de modelos são usados: os de séries temporais e causais. As técnicas de séries temporais utilizam dados históricos de demandas como base na determinação de padrões que podem ser repetidos no futuro. Já os modelos causais buscam relacionar as demandas (variável dependente) com outros fatores tais como, PIB, inflação, clima, e perfil de população, denominados variáveis independentes. A figura 3.1 mostra as técnicas dos métodos quantitativos e qualitativos e alguns de seus modelos.

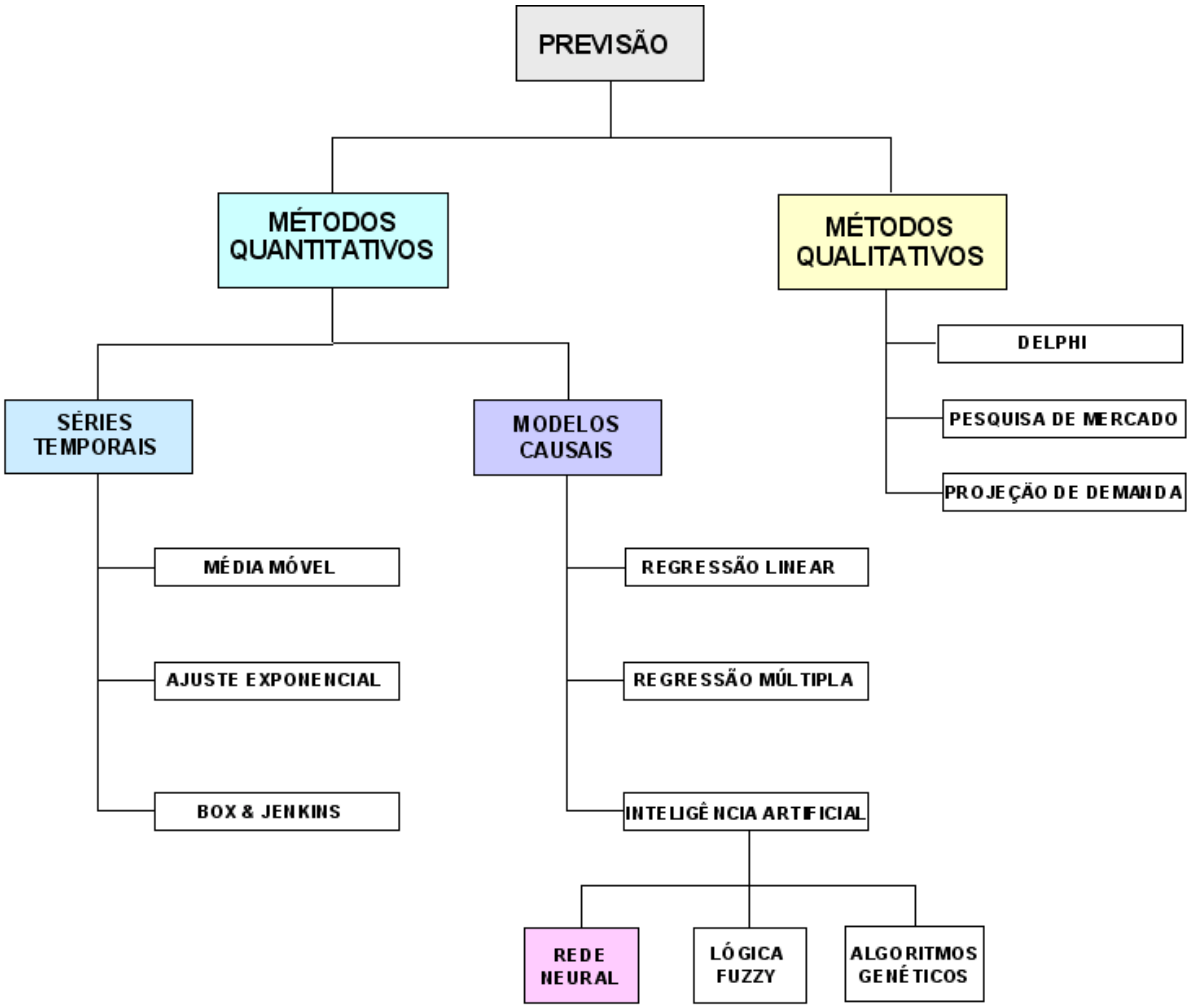


Figura 3.1: Métodos quantitativos e qualitativos e alguns de seus modelos (adaptada de Trautwein Júnior, 2004).

Os modelos quantitativos utilizam os históricos de informações disponíveis e são os mais usados como ferramenta de apoio e suporte à decisão. Os dados históricos da demanda têm quatro componentes (Shafer and Meredith, 1998):

1. tendência é a direção a longa distância da série, incluindo qualquer quantidade constante de demanda nos dados;
2. variação sazonal são flutuações regulares que se repetem em períodos quase sempre coincidindo com o calendário anual, mensal ou semanal;
3. variação cíclica: óbvia somente em séries que transpõem vários anos; um ciclo pode ser definido como uma oscilação de longo-prazo, ou um impulso de dados sobre a linha de tendência durante um período de pelo menos três períodos completos; e
4. variação aleatória, as variações desse tipo são sem uma causa específica e sem um padrão, portanto são tratados como erros aleatórios.

3.1.1. Previsão de Demanda de Água

As variações de demanda dependem do consumo de água, ou seja, a quantidade de água consumida, mais as perdas ocorridas no sistema. Desta maneira, é importante estudar o histórico do consumo, porque, através do histórico, é possível construir curvas típicas de consumo e obter parâmetros inerentes ao consumo dos setores de abastecimento que subsidiam os modelos de previsão de demanda.

A quantidade de água consumida em um sistema de abastecimento varia constantemente com o tempo. A previsão de demanda pode ser de longo ou curto prazo. A previsão de longo prazo está associada aos parâmetros que colaboram para planejar projetos de novos sistemas. A operação em tempo real está associada a previsão de curto prazo.

Segundo Tsutiya (2006) existem vários fatores que afetam a quantidade de água consumida e os mais importantes são:

- localidade;
- condições climáticas;
- hábitos e padrão de vida da população;
- natureza da cidade;

- sistema de fornecimento e cobrança (serviço medido ou não);
- qualidade da água fornecida;
- preço da água (tarifa);
- pressão na rede distribuidora;
- consumo residencial, comercial, industrial e público;
- perdas no sistema;
- existência de rede de esgotos e
- outros fatores.

O consumo de água tem variações que são:

- anual: o consumo tende a crescer com o decorrer do tempo, devido ao aumento populacional;
- mensal: as variações que ocorrem ao longo dos meses. Esse fator está intimamente ligado ao clima, ou seja, a sazonalidade;
- diária: são as variações de consumo que ocorrem no dia a dia
- horária: são as variações que ocorrem nas diversas horas do dia e,
- instantânea: são as variações que ocorrem devido a acidentes não previstos, tais como, incêndios, vazamentos, manutenção da rede, entre outras.

Balling Jr. e Gober (2007) estudaram o consumo de água anual no período entre 1980 e 2004 na cidade de Phoenix, Arizona, e verificaram que o consumo foi influenciado por variáveis climáticas. A correlação entre o uso de água, a temperatura média anual, precipitação total anual e valores do índice médio anual de seca hidrológica de Palmer foram +0,55, -0,69, -0,52, respectivamente, durante o período de estudo. O consumo anual de água aumenta com temperatura alta, baixa precipitação e tempo seco. A análise multivariada, utilizando dados climáticos mensais, indica que o consumo de água anual é controlado na maior parte do tempo pelo tempo seco, as temperaturas de outono, e precipitação do verão. Coeficientes do modelo indicam que as condições de temperatura, precipitação e / ou seca certamente impactam no consumo de água, embora o valor da resposta do consumo de água anual às mudanças no clima foi relativamente baixa para um ambiente urbano em que a maioria de consumo de água residencial foi considerável para fins externos, ou seja, são utilizadas nos sistemas de irrigação mecânica ou rega, devido

ao fato de Phoenix ser uma cidade árida. Então o clima e o consumo da água estão ligados por um conjunto complexo de processos comportamentais, sobre os quais sabemos muito pouco, mas que são cruciais para a concepção de programas para o uso mais eficiente da água na área urbana.

A demanda de água é controlada pelas variações diária, semanal e ciclos sazonais. Existe um grande número de pesquisas realizadas que tratam da modelagem anual ou mensal do uso urbano da água, sendo que estudos para o uso diário existem poucos. A principal aproximação feita na modelagem do uso diário da água é dividir o uso total da água dentro do tempo para analisar o uso em relação a sazonalidade e a não susceptibilidade ao tempo.

Zhou et al. (2001) estudaram a estimativa do intervalo médio da recorrência do consumo de água máximo diário para um, dois, três e cinco dias consecutivos, para a cidade de Melbourne (Austrália). A série do consumo diário foi obtida supondo o consumo médio de 356 litros por pessoa por dia. O estudo envolveu três etapas: calibração de um modelo de simulação de demanda de água diário para os meses de maior consumo, estimativa do consumo de água para uma série temporal e, por fim, o cálculo do intervalo médio de recorrência dos eventos extremos. A conclusão dessa pesquisa foi que a aproximação determinística/probabilística da variação das curvas de freqüência para a área de Melbourne, podem ser usadas para melhorar as estratégias do abastecimento de água que dependem das estimativas da demanda. Dessa forma, as variáveis do tempo são essenciais para explicar as diferenças no consumo da água de ano para ano.

Para prever o consumo diário de água, Zhou et al. (2000) criaram um modelo com base num conjunto de equações que representam os efeitos de quatro fatores: a tendência, a sazonalidade, a correlação e auto-correlação climática. O uso básico da água foi estimado pelos meses de menor consumo. A tendência do consumo básico a longo prazo, de ano para ano, foi representada por uma função polinomial dependente do tempo. O uso sazonal foi modelado pela sazonalidade climática e por componentes de permanência considerando os seis meses de verão e de inverno separadamente. O modelo desenvolvido foi testado usando um procedimento de

validação cruzada, e uma série de dados independentes durante o período do verão. A eficiência do modelo foi R^2 =89,6% e erro padrão foi de $\pm$8 considerado aceitável.

Considerando ainda a previsão do consumo diário, Maidment et al. (1985) desenvolveram um modelo de previsão de curto prazo baseado na análise de uma série temporal utilizando o modelo de Box/Jenkins. O modelo baseia-se em três proposições: (1) o consumo total pode ser dividido em consumo básico, que é consumo não suscetível ao tempo e observado como o consumo médio nos meses de inverno, e o consumo sazonal, que é suscetível ao tempo e observado como a diferença entre o consumo básico e o consumo total durante os outros meses do ano; (2) na ausência de chuva, o consumo sazonal segue uma característica padrão durante o ano que é dependente das condições de temperatura; e (3) a ocorrência de chuva causa uma queda imediata no consumo sazonal que diminui gradualmente no decorrer do tempo. Os dados utilizados no modelo foram os dados diários de Austin, Texas, do período de 1975-1981 que correspondem a 97% da variação do consumo de água diário municipal durante o período e as previsões do consumo diário foram realizadas para um prazo de duas semanas.

Maidment e Miaou (1986) aplicaram a metodologia de Maidment et al. (1985) para o consumo diário em nove cidades, três na Florida, três na Pensilvânia, e três no Texas. O coeficiente de determinação R^2 para as nove cidades foram em média 0,96 no Texas, 0,73 na Florida, e 0,61 na Pensilvânia. Eles concluíram que, como proporção do consumo médio anual, o consumo sazonal médio para as três cidades em cada estado foi de 23% no Texas, 15% na Florida, e 5% na Pensilvânia. A resposta do consumo de água em relação à chuva e à temperatura do ar foi similar para as cidades no interior de cada estado. Nas funções de resposta existiu um pequeno impacto em relação ao tamanho da cidade. A resposta do consumo água em relação à chuva primeiramente dependeu da ocorrência da chuva e em segundo de sua quantidade. Observaram também que existe uma resposta não linear do consumo de água em relação às mudanças de temperatura.

A previsão da demanda de água, com antecedência de 24 h, pode ser executado por um modelo matemático que combine dados de demanda anterior e outras informações, tal como previsões de tempo. Zhou et al (2002) propuseram uma

metodologia para a previsão de demanda horária da água num intervalo de 24h usando registros do consumo de água e de informação climática. O modelo envolve dois módulos: diário e horário. O módulo diário é formado por um conjunto de equações que representam os efeitos da sazonalidade, correlação climática, e autocorrelação. O módulo horário foi desenvolvido para separar a estimativa do consumo diário do consumo horário. Os modelos foram calibrados utilizando dados horários e diários para um período de seis anos, e validados com dados independentes para um período de sete meses. Sobre este último período, o modelo de previsão horário explicou 66% da variação do pico do consumo horário com um desvio padrão de 162 litro/pessoa/dia.

Estudos mais específicos sobre o efeito das variáveis climáticas foram realizados separando o ano em duas estações: inverno e verão (Howe e Linaweaver, 1967; Carver e Boland, 1980). Outros estudos modelaram a sazonalidade utilizando uma função de regressão para cada mês (Morgan e Smolen, 1976; Yamauchi e Huang, 1977; Cassuto e Ryan, 1979). Muitas variáveis foram utilizadas nos modelos para avaliar o impacto do clima no uso da água. A temperatura e a precipitação são as duas variáveis meteorológicas mais comuns, e possivelmente as mais eficazes (Morgan e Smolen, 1976; Weeks e McMahon, 1973).

Pereira Filho et al. (2004) avaliaram o consumo de água em relação às variáveis meteorológicas observadas na RMSP nos anos de 1999 e 2000. Os resultados obtidos sugerem que o consumo máximo médio mensal tem correlação com a média mensal da temperatura máxima. Além disso, o consumo médio diário tende a aumentar com os totais mensais mais baixos da precipitação e diminuir com os mais altos. A média mensal dos mínimos do consumo apresentou uma relação cíclica com a temperatura média mínima mensal. Santos et. al. (2006[a], 2007[b]) e Santos e Pereira Filho (2009) também utilizaram dados de chuva, temperatura, umidade e dias da semana para avaliar o consumo de água da Região Metropolitana de São Paulo.

A maioria das pesquisas a respeito do impacto das mudanças climáticas sobre os sistemas de abastecimento de água foram realizados levando-se em conta sistemas

individuais de modo que suas conclusões se aplicassem apenas a um sistema particular. Os avanços recentes da informática, hidrologia regional e do entendimento do funcionamento do sistema de abastecimento, permitiram o estudo da sensibilidade do funcionamento desse sistema em relação as mudanças climáticas de uma forma geral.

Vogel et al. (1997) desenvolveram um modelo hidroclimatológico regional para bacia hidrográfica no nordeste dos Estados Unidos, que relaciona escoamento anual às características climáticas e geomorfológicas de 166 estações. O modelo hidroclimatológico regional para bacia é, então, combinado com a relação analítica entre o armazenamento, a confiabilidade, a elasticidade e o rendimento do sistema de abastecimento de água. A sensibilidade de vários índices de desempenho do sistema de abastecimento tais como: o rendimento, a confiabilidade e a elasticidade do sistema, é derivada em função das condições climáticas, hidrológicas e do armazenamento. Os resultados do modelo permitiram determinar a sensibilidade do comportamento do sistema de abastecimento para o regime das mudanças climáticas. Os estudos de caso em quatro bacias de Nova Iorque e um sistema de abastecimento de Massachusetts indicam que a aproximação anual do modelo regional simples pode reproduzir os resultados aproximados de um modelo hidroclimatológico muito mais detalhado mensalmente.

Gato et al. (2007) utilizaram dados de consumo básico e correlacionaram esses dados com o dia da semana e os fatores climáticos como temperatura e precipitação. Os resultados revelaram que o consumo básico independe de fatores climáticos, sendo afetados pelo dia da semana e pelo final de semana.

Zahed Filho (1990) obteve um modelo de previsão de demandas que permite a utilização de modelos de otimização da operação associados à modelos de simulação do escoamento no sistema. O modelo foi aplicado a alguns reservatórios do Sistema Adutor Metropolitano de São Paulo, onde verificou a precisão e colaboração no auxílio ao modelo de otimização, quanto à melhora no índice de atendimento à demanda e na redução do número de manobras necessárias.

Narchi (1987) avaliou a demanda doméstica como principal parcela da demanda urbana de água. Ele discutiu a importância de sua previsão no planejamento e projeto de sistemas de abastecimento de água, apresentando métodos comuns para essas previsões. Além disso, apresentou um estudo realizado levando em conta os consumidores da cidade de São Paulo, no qual ajustou um modelo de regressão entre o consumo doméstico de água e variáveis como o tamanho e o nível de renda da família e o valor venal e a área construída do imóvel. Ele também avaliou o efeito do preço da água sobre a demanda doméstica, fazendo estimativas de elasticidade-preço da demanda, concluindo que a área construída dos imóveis e o tamanho da família foram as variáveis que melhor explicaram a demanda doméstica na cidade de São Paulo.

Barreto (2008) realizou um levantamento do perfil do consumo residencial e usos finais da água, a partir de uma amostra definida por um conjunto de residências, localizadas na zona oeste da cidade de São Paulo. Foram identificados os consumos diários desagregados por ponto de utilização, que permitiram definir o perfil de consumo e usos finais.

Secco (2002) apresentou a operação do abastecimento de água da RMSP por meio de estudos de casos em subsistemas de adução do sistemas produtores de água tratada e do conjunto de fatores que podem ser responsáveis por condições críticas de abastecimento em todas as fases do processo de captação, produção e distribuição. Ela avaliou e propôs alternativas de melhoria do sistema por meio de vários estudos de caso que abordaram o déficit de abastecimento público da região, tais como racionamento de água, deficiências estruturais das instalações, riscos de colapso, entre outros. A operação do sistema adutor da RMSP é uma atividade complexa e demanda um alto grau de segurança, face à responsabilidade da manutenção do abastecimento de milhões de habitantes.

Um dos grandes problemas dos sistemas de abastecimento é o custo da energia elétrica. Atender as condições de distribuição de água levando em consideração as restrições hidráulicas com o mínimo custo de energia, é uma tarefa difícil. As condições de operação, em tempo real, geram a necessidade de alimentação de informações operacionais automáticas em um curto espaço de tempo. Vicente

(2005) e Borges (2003) propuseram modelos para melhor o desempenho operacional do Sistema Adutor Metropolitano de São Paulo. Vicente (2005) utilizou um modelo de otimização proposto com solução linear e obteve como resultado uma redução no custo de energia elétrica. Borges (2003) usou um modelo de previsão de demanda de água para desenvolver uma interface entre um modelo de rede hidráulica e um modelo de previsão de demanda de água utilizando dados operacionais em tempo real; concluiu que é possível estabelecer regras operacionais mais eficientes e com isto obter redução do custo de energia elétrica.

3.2. Redes Neurais Artificiais

As RNA´s simulam os sistemas naturais utilizando-se de estruturas análogas às Redes Neurais Biológicas (RNB). A complexidade da conectividade entre os neurônios é responsável pelas características atribuídas à inteligência. A agilidade e a eficiência com que o cérebro realiza suas funções constituem-se no objetivo a ser alcançado pelos sistemas de Inteligência Artificial (IA), que se baseiam na simulação computacional de aspectos da inteligência humana, levando em consideração a capacidade que o cérebro humano tem de aprender e tomar decisões baseado em sua aprendizagem. Portanto, uma RNA pode ser definida como uma estrutura computacional que tem como objetivo permitir a implementação de modelos matemáticos que representem, de forma simplificada, a forma como o cérebro humano processa as informações que adquire. Assim, a RNA pode ser interpretada como um esquema de processamento que é capaz de armazenar conhecimento baseado em aprendizagem (experiência) e disponibilizar este conhecimento para as aplicações desejadas.

Segundo Zhang et al. (1998), apesar da RNA mostrar características satisfatórias, a construção de um modelo de RNA para um problema particular de previsão não é tarefa trivial. Os problemas que afetam a modelagem e o desempenho da RNA devem ser considerados com cuidado. A determinação de uma estrutura apropriada para rede é crucial, ou seja, o número de camadas e o número de nós em cada camada. Outras questões importantes para um bom desempenho da RNA são: a

seleção das funções de ativação dos neurônios da camada intermediária e de saída, o algoritmo de treinamento, os métodos de normalização dos dados, a definição dos conjuntos de treinamento e teste e análise do desempenho da rede.

A RNA *feedforward* é a rede mais popular e amplamente utilizada para previsão (Wasserman ,1989, Rumelhart et al. 1986a, 1986b, 1994 e 1995). Ela consiste de uma ou mais camadas de unidades de processamento não-lineares. As conexões entre unidades de camadas adjacentes são definidas por um conjunto de pesos sinápticos. No entanto, para ser considerada *feedforward*, uma rede neural precisa que suas saídas se conectem somente as unidades da próxima camada. O potencial das redes *feedforward* só foi revelado com o surgimento das redes *percepton* multicamadas (MLP) em 1980. As redes MLP são usados em uma variedade de problemas, especialmente na previsão, devido à sua capacidade de mapeamento de entrada-saída arbitrária.

Uma rede MLP é geralmente composta de várias camadas. A primeira camada é a camada de entrada onde a informação externa é recebida. A última camada é a de saída, onde a solução do problema é obtida. A camada de entrada e a camada de saída são separadas por uma ou mais camadas intermediárias chamadas camadas escondidas, onde ocorre o processamento da RNA. Ao projetar uma MLP, é necessário determinar as seguintes variáveis:

- número de nós de entrada;
- número de camadas escondidas e nós escondidos e
- número de nós de saída.

Para um problema de previsão, as entradas de uma RNA geralmente são variáveis independentes ou preditoras (explicativas). A relação estimada pela RNA é observada na equação 3.1:

$$y = f(x_1, x_2, \ldots, x_p) \tag{3.1}$$

onde: x_1, x_2, …, x_p são as variáveis independentes e y é a variável dependente.

Então a RNA funciona como um modelo de regressão não-linear. Por outro lado, para um problema de previsão de séries temporais, as entradas são tipicamente as observações passadas da série de dados e a saída é o valor futuro. A função que RNA executa para esse mapeamento é a descrita na equação 3.2:

$$y_{t+1} = f(y_t, y_{t-1}, \ldots, y_{t-p}) \tag{3.2}$$

onde y_t é a observação no tempo t. Assim, a RNA é equivalente a um modelo não linear autoregressivo para problemas de previsão de séries temporais.

O número de nós de entrada corresponde ao número de variáveis de entrada utilizadas para prever valores futuros. Para previsão simples o número de entradas é normalmente transparente e relativamente fácil de escolher, mas em um problema de previsão de séries temporais, o número de nós de entrada correspondem ao número de observações defasadas no tempo usada para descobrir os padrões subjacentes de uma série para prever valores futuros. No entanto, atualmente não existe maneira sistemática de determinar esse número. A seleção deste parâmetro deve ser incluída no processo de construção do modelo. O ideal seria ter um pequeno número de nós essenciais para revelar as características embutidas nos dados. Os nós de entrada podem afetar a capacidade de aprendizagem e previsão da rede , assim, o número de nós de entrada é provavelmente a variável de decisão mais crítica para uma previsão de séries temporais, uma vez que contém importantes informações sobre a estrutura de autocorrelação linear e / ou não-linear dos dados.

Os nós da camada escondida desempenham um papel muito importante nas aplicações das RNA. É nestes nós que a rede detecta o recurso, para capturar o padrão dos dados, e executar os complicados mapeamentos não lineares entre as variáveis de entrada e saída. É claro que, sem os nós da camada escondida, perceptrons simples, os nós de saída linear são equivalentes aos modelos lineares de previsão estocástica. Os trabalhos existentes mostram que uma única camada oculta é suficiente para RNA´s aproximarem qualquer função complexa não-linear com precisão (Cybenko, 1989;. Hornik et al, 1989), a maioria dos autores utilizam

somente uma camada escondida para efeitos de previsões. No entanto, uma camada escondida pode exigir um número muito grande de nós escondidos, o que não é desejável, porque o tempo de treinamento e a capacidade de generalização da rede vão piorar. A questão da determinação do melhor número de nós na camada escondida ainda é complexo e exige estudos. De modo geral, as redes com poucas camadas escondidas são preferíveis porque elas geralmente têm melhor capacidade de generalização e menos problema de superestimar valores. Mas em alguns casos, as redes com poucos nós escondidos podem não ter energia suficiente para modelar e aprender os com dados.

O número de nós de saída é relativamente fácil de especificar como está diretamente relacionada ao problema estudado. Para um problema de previsão de séries temporais, o número de nós de saída, muitas vezes corresponde ao horizonte de previsão.

O treinamento da rede neural é um problema de minimização não linear, na qual os pesos de uma rede são iterativamente modificados para minimizar o erro médio quadrado ou total entre os valores desejados e os valores de saída real para todos os nós de saída de todos os padrões de entrada. Existem muitos métodos diferentes para otimização (Fletcher, 1987) que oferece várias opções para o treinamento da RNA. Mas não existe nenhum algoritmo que garanta uma solução ótima global para um problema de otimização não-linear. Todos os algoritmos de otimização, na prática, inevitavelmente, sofrem de problemas ótimos locais e o máximo que pode-se fazer é usar o melhor método de otimização que obtem-se o melhor ótimo local, já que a verdadeira solução global não está disponível.

Então antes que a RNA seja usada para executar qualquer tarefa, ela deve ser treinada. Basicamente, o treinamento consiste no processo de determinação dos pesos, ou seja, o conhecimento adquirido pela rede é armazenado nos nós na forma de pesos. É através das sinapses que a RNA pode levar os mapeamentos não-lineares complexos da camada de entrada para o de saída. O treinamento de uma rede MLP é supervisionado (Haykin, 2001), onde a resposta desejada da rede (valor alvo) para cada padrão de entrada está sempre disponível, ou seja, como se a rede já tivesse conhecimento sobre o ambiente. Os dados de entrada estão na forma de

vetores ou de padrões de treinamento, sendo que cada elemento correspondente a um vetor de entrada em um nó de entrada na camada de entrada da rede. Portanto, o número de nós de entrada é igual à dimensão dos vetores da entrada. Para um problema de previsão qualquer, o número de nós de entrada está bem definido e corresponde ao número de variáveis independentes associadas ao problema. Já para uma previsão de séries temporais o número adequado de nós de entrada não é fácil de determinar, porque o vetor de entrada quase sempre tem ao longo da série uma janela móvel que prejudica definir o número de nós de entrada.

O total de dados disponíveis é normalmente dividido em um conjunto de treinamento (dados dentro da amostra) e um conjunto de teste (fora da amostra). O conjunto de treinamento é usado para estimar os pesos enquanto o conjunto de teste é utilizado para medir a capacidade de generalização da rede. O processo de treinamento geralmente segue o seguinte padrão, primeiro o conjunto de treinamento é inserido na entrada e os valores de ativação da entrada são ponderados e acumuladas nos nós da primeira camada escondida. Então o total é transformado por uma função de ativação, que por sua vez torna-se um atributo para os nós da próxima camada, segue assim até que os valores de ativação da saída são encontrados. O algoritmo de treinamento é usado para encontrar os pesos que minimizam o erro global, sendo assim um problema de minimização dos erros.

Desta forma, as RNA´s podem ser entendidas como conjuntos bem estruturados de unidades de processamentos, interligadas por canais de comunicação, cada qual tendo um determinado peso correspondente a um valor numérico. Elas consistem de várias unidades de processamento ou neurônios artificiais interconectados entre si, formando uma determinada disposição estrutural de camadas (entrada, intermediárias e saída) e conexões entre as camadas. As RNB usam milhões de neurônios em suas tarefas. Já as RNA estão restritas a algumas centenas. A RNA tem sido usada para modelagem de séries temporais em várias áreas de aplicações incluindo sistemas dinâmicos (Kirby, Miranda, 1994), processamento de sinais não linear (Lapedes e Farber, 1987), reconhecimento, identificação e classificação de padrões (Pao, 1989) entre outras aplicações. A complexidade observada e encontrada nas séries temporais sugere o uso de redes neurais, mostrando ser

capaz de modelar relação não lineares complexas sem *a priori* conhecer hipótese da natureza da relação.

3.2.1. Modelos de Rede Neural Artificial para demanda de água

As informações sobre a quantidade de água consumida são estruturadas como séries temporais, que podem ser associadas a outras séries que influenciam o consumo de água (por exemplo: variações climáticas, precipitação, tipo de população). Desta forma, esses conjuntos de dados podem ser utilizados para treinamento de uma RNA, para auxiliar na previsão do crescimento de seu consumo e também, identificar quais fatores influenciam nesse crescimento.

O sistema de suporte a decisão é uma ferramenta usada geralmente nos setores de negócio, indústria e governo para melhorar a qualidade e a consistência da tomada de decisão. Para tarefas de planejamento e programação, é importante que os fatores de decisão sejam previstos exatamente de modo que os responsáveis pelas decisões possam selecionar a melhor das alternativas. A demanda é um fator de decisão cuja avaliação é muito importante para a otimização dos custos. Assim, um grande erro na previsão pode acarretar um aumento nos custos operacionais e degradar a segurança dos sistemas de distribuição. Em geral, os modelos de previsão de demanda são projetados baseando-se na análise das relações entre as variáveis do tempo, eventos especiais e das demandas anteriores para estimar as tendências futuras. Na prática, os modelos adotam uma técnica simples que tendem a produzir erros elevados na previsão, especialmente quando o problema tem diferentes padrões. É aconselhável dividir o problema em sub-problemas e aplicar então técnicas inteligentes de módulos múltiplos para trabalhar cooperativamente de modo a melhorar a previsão.

Lertpalangsunti et al. (1999) criaram uma ferramenta denominada de “Intelligent Forecasters Construction Set (IFCS)” que foi elaborada para várias aplicações. Ela pode ser utilizada com as técnicas de lógica fuzzy e rede neural artificial. O usuário pode construir sua previsão utilizando regras, procedimentos e diagramas de fluxo,

que são organizados em uma hierarquia. A flexibilidade do IFCS permite a adição subseqüente de outros módulos. O IFCS foi usado para desenvolver um sistema de previsão de demanda de água baseado nos dados reais obtidos do sistema de distribuição de água da cidade de Regina no Canadá. A previsão da demanda desenvolvida com o sistema de IFCS é útil para otimização dos custos de operação das plantas de distribuição.

Os modelos estócasticos de demanda de água geralmente são desenvolvidos com os coeficientes de séries temporais usando historicamente dados de demanda de água disponível, juntamente com outras variáveis relevantes. Mas a identificação da estrutura encontra dificuldade para maioria das aplicações. Chang e Makkeasom (2006) desenvolveram um modelo de previsão de consumo que utilizou RNA para previsão de padrões de demanda de água. O modelo de RNA apresentou características não-lineares aprendidas a partir de dados históricos, da mesma forma como os humanos aprendem com a experiência. A não-linearidade, alta complexidade associados a incerteza com as demandas de água favoreceram ao uso potencial das RNAs e superam os métodos convencionais de série de temporais para a previsão de consumo. O modelo de RNA aprendeu corretamente, e verificou o desempenho das previsões utilizando os dados de entrada não utilizados durante o treinamento, por ser robusta a RNA obteve um tempo baixo para o processamento. Existiram erros ou ruído nos dados de entrada. As cidades, de Oviedo (crescimento rápido) e Winter Springs (crescimento lento) na Área Metropolitana de Orlando, na Flórida, foram investigados com respeito aos dados mensais. Tais práticas de reconhecimento padrão ajudaram na identificação de estratégias para expansão e operação dos sistemas de distribuição de água em longo prazo para empresas de água daquela região.

Um dos objetivos de uma gestão eficiente de abastecimento de água é o abastecimento regular de água na pressão exigida pelos consumidores. Sendo assim, a previsão de consumo de água em áreas urbanas é de fundamental importância para a gestão de abastecimento de água. Essa previsão também é relevante nos processos de revisão de preços, bem como para a gestão operacional de uma rede de água. Herrera et al. (2010) descrevem e comparam uma série de modelos para a previsão de demanda de água. Os modelos foram obtidos utilizando

dados de séries temporais de consumo de água em uma área urbana de uma cidade no sudeste da Espanha. Além dos modelos utilizados, eles propuseram um modelo simples baseado no perfil de demanda ponderada resultante de análise exploratória de dados. Para avaliação de todos os modelos, eles utilizaram uma metodologia experimental para dados de série temporal horária que continha a demanda de água detalhada num setor de uma rede de abastecimento de água. A precisão dos resultados obtidos, juntamente com o tamanho da área de demanda, sugeriu que esse era um ambiente adequado para a tomada de decisões e gestão.

Os métodos estocásticos, como o modelo auto-regressivo integrado de média móvel (ARIMA), têm sido aplicados para previsão de demanda. A maioria das aproximações estatísticas usam informações das condições do tempo e de feriados para a análise de intervenção, sendo que se for usado somente os dados da série temporal pode causar grandes erros na previsão. Entretanto, esta aproximação freqüentemente gera grandes erros de previsão. Liu et al. (1996) compararam os modelos de lógica fuzzy (LF), rede neural (RN) e auto-regressivos (AR) para previsão de curto prazo e concluíram que as aproximações encontradas com os modelos LF e RN são bons para previsão em curto prazo. Os desempenhos dos modelos baseados em LF e RN são superiores àqueles baseados no AR. Os modelos LF e RN podem prever pontos múltiplos dos dados contanto que existam bastantes dados históricos para o treinamento. Ao contrário, o modelo AR não necessita de um pré-treinamento, mas pode somente prever alguns pontos de dados.

Liu et al. (2003) propuseram um novo método para modelar e prever o consumo de água chamado de “Water Demand Forescast using Artificial Neural Network” (previsão de demanda de água usando rede neural artificial) - WDF-ANN, para estimar o consumo de água. Este modelo combina rede neural artificial e técnicas econométricas. O modelo não requer muitos dados de entrada. Os resultados indicam que o modelo estimou a demanda doméstica de água da cidade de Weinan na China. O coeficiente de correlação foi superior a 90% tanto no treinamento como na verificação.

BuHamra et al. (2003) combinaram dois modelos, Box/Jenkins (BJ) e redes neurais artificiais (RNA), para modelar séries temporais do consumo de água no Kuwait. O modelo BJ foi usado para estimar os dados não registrados do consumo de água de maio 1990 a dezembro 1991 devido à invasão do Iraque no Kuwait em agosto 1990. Uma rede neural supervisionada foi projetada, treinada e verificada para modelar e prever o consumo da água de janeiro 1980 a dezembro 1999. Eles observaram que as variáveis estimadas pelo BJ e usadas na RNA forneceram um modelo melhor de RNA do que com métodos tradicionais.

Os dados semanais do consumo de água de 48 casas do Kwait foram coletados por um período de um ano. Baseado nestes dados, Mukhopadhyay et al. (2001) estimaram que a média do consumo de água per capita numa residência do Kwait estava na faixa de 182-2018 litro/dia/pessoa, com uma média de 814 litro/dia/pessoa. Com um modelo de regressão linear e outro de RNA, com os dados observados, observaram que o consumo de água depende do número de banheiros e de quartos da residência, do tamanho do jardim, da renda familiar, da temperatura atmosférica, da umidade relativa, e do número de pessoas que vivem na residência. O modelo de regressão linear múltipla do consumo semanal de água nas residências junto com os parâmetros de tempo (temperatura, umidade, e precipitação) e as variáveis como o número de pessoas que vivem na residência, o tamanho do jardim, o número de quartos e banheiros na residência, e a renda familiar foram utilizados para estudar qual o grau de dependência do consumo de água com esses parâmetros. A aplicabilidade da RNA para prever o consumo semanal de água nas vilas residências também foi estudada. Os fatores que contribuíram significativamente para o consumo de água em uma residência após o estudo realizado com o modelo de regressão linear múltipla foram usados como neurônios de entrada na RNA e o consumo semanal em uma residência foi considerado como neurônio de saída.

An et al. (1996) utilizaram um método alternativo para a aquisição de conhecimento do consumo de água através da concepção de um algoritmo que aprende e aperfeiçoa as regras de decisão a partir de um conjunto de amostras de treinamento, ou de dados observados, este algoritmo também cria regras de classificação através de informação incompleta. Como o consumo instantâneo de água no meio urbano é

determinado por vários tipos de consumidores (indústrias, comércios e domésticos), além também de ser influenciado por fatores como condições climáticas, variações sazonais, dia da semana e feriado. Nesse trabalhos, utilizaram dezoito fatores que podem afetar a consumo de água numa cidade. O primeiro fator utilizado foi o dia da semana, que foi escolhido baseado na observação de que no fim de semana os consumos totais são normalmente inferiores aos de durante a semana. Além disso, as segundas-feiras são conhecidas por serem dias de maior consumo porque muitas pessoas lavam suas roupas, além de que, nas segundas-feiras de verão as pessoas regam mais seus gramados porque elas não fazem isto no final de semana. Os outros 17 fatores utilizados foram as de condições climáticas tais como: temperatura, umidade, precipitação, vento e horas de sol, agrupados em três dias consecutivos.

Firat et al. (2009[a], 2010[b]) utilizaram várias técnicas de Rede Neural Artificial (RNA), tais como Redes Neurais de Regressão Generalizada (GRNN), Rede Neurais Retro Alimentação (feedforward) (FFNN) e Rede Neural de Base Radial (RBNN) para avaliar seus desempenhos para previsão do consumo de água mensal utilizando fatores socioeconômicos e climáticos que afetam o uso da água. O conjunto de dados foi dividido em dois subgrupos, treinamento e teste. Para identificar o melhor ajuste para o modelo de previsão o desempenho dos modelos, tanto para o treinamento quanto para o teste, foram comparados com os valores de consumo observados. Para verificar o desempenho dos modelos foram calculados para todos os modelos o Erro Médio Quadrático Normalizado (NRMSE), a eficiência (E) e o coeficiente de correlação (CORR). Os modelos com os melhores ajustes também foram treinados e testados pelo modelo de Regressão Linear Múltipla (MLR). Os resultados indicaram que GRNN supera todos os outros métodos na modelagem do consumo mensal de água.

As companhias de saneamento têm feito muitos investimentos na automação de seus sistemas de abastecimento de água nos últimos anos. Esses investimentos visam proporcionar dados em tempo atual de vazão e pressão. Isto é feito para obter uma completa otimização dos sistemas, sendo necessário modelar o comportamento do consumo ao longo do dia nas áreas de distribuição. Falkenberg et al. (2003) e Falkenberg (2005) utilizaram uma RNA para prever o consumo de determinados bairros para 24 horas futuras. Utilizaram um período de três meses de

dados para modelar a série temporal. Estes dados representam os consumos horários de um bairro da cidade de Ponta Grossa no estado do Paraná. O primeiro passo foi prever o consumo da hora seguinte, desta forma foram testadas quatro combinações para definir qual era a melhor combinação de entradas para a rede, para isto variou as funções de ativação e o número de neurônios na camada escondida, que ficou entre um e cinco neurônios. Os primeiros 85% dos dados foram utilizados para o treinamento e os 15% restantes para verificação. Para avaliação dos erros foi utilizado o erro quadrático médio (EQM).

Em outro trabalho, Falkenberg et al. (2003) também tentaram prever o consumo de água numa determinada área de distribuição com 24 horas de tendências. Neste trabalho, a análise de consumo baseou-se num banco de dados de cinco meses contendo as leituras das vazões medidas a cada 5 minutos de um sistema "Supervisory Control and Data Acquisition system (SCADA)". A série temporal foi modelada por meio de uma RNA e de um modelo de regressão linear múltipla (RLM). Os resultados dos dois modelos foram comparados. Os dados foram filtrados para obter uma consistência da relação do perfil de demanda. Baseado em estudos sobre os principais fatores determinantes das características de consumo, foram estipuladas as entradas dos modelos de RNA e RLM. Foi utilizada uma RNA do tipo "feedforward" com uma camada escondida, sendo que, a função de ativação utilizada para camada escondida foi a tangente hiperbólica e para camada de saída a função linear. Para o treinamento foi utilizado o algoritmo Levemberg-Marquardt, minimizando o MSE. Os dados foram divididos de forma que 70% foram utilizados para o treinamento e 30% para a verificação. Para a análise dos dados, utilizou a raiz quadrada do erro quadrático médio (rmse) para os dados de verificação e o erro absoluto percentual médio.

A previsão de consumo para curtíssimo prazo foi objetivo do estudo de Trautwein (2004). Ele analisou dois métodos de previsão de consumo de água, o método de ajuste exponencial e método de redes neurais. A previsão foi realizada com séries históricas de consumo de duas zonas de pressão diferentes de um Sistema de Abastecimento de Água com características distintas. Os resultados comprovaram a eficiência das previsões de consumo em intervalo de horas, mas mostraram que não existe um modelo ideal. Odan (2010) também utilizou redes neurais para fazer a

previsão de consumo na cidade de Araraquara, utilizando redes percepton múltipla camadas (MLP), rede neural dinânica (DAN2) e duas RNAs híbridas. O melhor modelo de previsão foi os que utilizaram a DAN2, mostrando melhor desempenho.

As previsões de demanda de água são necessárias para a concepção, exploração e gestão dos sistemas de abastecimento urbano de água. Bougadis et al. (2005) estudaram o desempenho relativo de modelos de regressão, análise de séries temporais e redes neurais artificiais (RNA) estudando a previsão do pico de demanda de água de curto prazo. A importância das variáveis climáticas (chuva e temperatura do ar máxima, além da demanda de água passados) sobre a gestão da demanda de água também foi analisada. A análise numérica foi realizada com dados da cidade de Ottawa e Ontário no Canadá. A infraestrutura de abastecimento de água existentes não era capaz de atender à demanda do crescimento populacional, assim, foi necessário um realizar um estudo para determinar o efeito do manejo de pico da demanda de água no dimensionamento e realização de instalações para o desenvolvimento de um estratégia de expansão. Três RNAs diferentes, três modelos de regressão e sete modelos de séries temporais foram desenvolvidos e comparados. Os modelos de RNAs sempre superaram os de regressão e de modelos de séries temporais desenvolvidas neste estudo. Foi constatado que a demanda de água baseado na semana é mais significativamente correlacionada com a quantidade de chuva do que com a sua ocorrência.

Silva et al. (2001) utilizaram uma RNA para prever o consumo dos reservatórios de água da Companhia de Saneamento Básico do Estado de São Paulo (SABESP) utilizando o software chamado Neugents da Computer Associates. Neste estudo, utilizaram como variáveis de influência, a hora do dia e os valores de consumos de horas anteriores. Supondo-se um erro acumulado de 10% do volume previsto, o que equivaleria a cerca de 2% do volume do reservatório, o horizonte máximo de previsão seria de 2 horas com risco de falha em torno de 10% do tempo ou de 4 horas, com risco de 20%, supondo-se um erro de 15%, riscos de falhas de 10%, para um horizonte é 4 horas.

As previsões do pico de demanda de água diário são necessários para o custo de operação, gerenciamento sustentável e expansão de infra-estruturas de

abastecimento urbana de água. Adamowski (2008) compara uma regressão linear múltipla, análise de séries temporais e redes neurais artificiais (RNAs) como técnicas de modelagem para previsão do pico diário de verão de demanda de água. A análise foi realizada para um período de 10 anos de dados de pico de demanda de água diária e variáveis meteorológicas (temperatura máxima diária e precipitação diária) para os meses de verão de maio a agosto de cada ano para uma área de alto consumo de água existente na cidade de Ottawa, Canadá. Trinta e nove modelos de regressão linear múltipla, nove modelos de séries temporais, e 39 modelos de redes neurais artificiais foram desenvolvidos e seus desempenhos foram comparados. A abordagem de rede neural artificial é mostrada por produzir uma melhor previsão de demanda de água para o pico de dia de verão do que a regressão linear múltipla e análise de séries temporais. Os melhores resultados foram obtidos quando há pico de demanda de água do dia anterior e a temperatura máxima a partir do dia atual e anterior, e as ocorrência / não ocorrência de precipitação a partir de cinco dias anteriores, foram utilizados como dados de entrada. Constatou-se também que o pico de demanda diário de água no verão é mais correlacionada com a ocorrência de chuvas e não a quantidade de chuva em si, e que a atribuição de um sistema de ponderação para os dias sem precipitação anteriores não resultam em modelos mais precisos.

Adamowski and Karapataki (2010) compararam várias modelos de regressão linear e três tipos de redes neurais artificiais de múltiplas camadas (cada um dos quais utiliza um tipo diferente de algoritmo de aprendizagem) como métodos para a modelagem de previsão de pico de demanda de água semanais para Chipre. A análise utilizou 6 anos de dados de pico semanal de demanda de água e variáveis meteorológicas (temperatura máxima semanal e precipitação total semanal) de duas regiões diferentes (Athalassa e Jardim Público) na cidade de Nicósia, Chipre. 20 modelos de regressão linear múltipla, 20 de redes neurais artificiais (RNA) Levenberg-Marquardt, foram desenvolvidos 20 modelos de RNA resiliente back-propagation, e 20 modelos de gradiente conjugado Powell-Beale e seus desempenhos foram comparados. Para ambas as regiões Athalassa e Jardim Público, em Nicósia, o método Levenberg- Marquardt foi o que obteve uma previsão mais precisa de demanda de água máxima semanal que os outros dois tipos de RNAs e regressão linear múltipla. Constatou-se também que o pico de demanda de

água semanal em Nicósia é melhor correlacionada com a ocorrência de precipitação e não propriamente a quantidade de chuva.

Zhang et al. (2006) descreve o uso de RNA para a previsão de demanda de água em curto prazo. As técnicas estatísticas são utilizadas para analisar e identificar as variáveis de entrada relevantes para os modelos RNA's. Um estudo de caso é apresentado em que RNA é utilizado para previsão de demanda de água de dois dias a frente para o sistema de abastecimento de água Companhia Louisville, Louisville Kentucky. Estudos anteriores separaram o consumo de água diário em não-sazonais e sazonal. A diferença desse trabalho é que o método separa todo o ano para a temporada de verão e inverno. A partição é efetuada de acordo com os diferentes efeitos significativo do clima sobre a demanda de água em curto prazo. No inverno, fatores climáticos têm poucos efeitos sobre a demanda de água, portanto, o modelo de previsão RNA foi simplificado, utilizando apenas a demanda histórica como entrada. A precisão da previsão alcançou 97,21%, considerado adequado para o gerenciamento do sistema de abastecimento de água no inverno. Na temporada de verão, foram selecionados fatores climáticos, tais como temperatura, umidade relativa, ponto de orvalho, velocidade do vento e precipitação que, foram incorporados aos modelos de RNA. A diferença de demanda durante a semana e fim de semana também foi considerada. O melhor modelo de RNA para previsão de verão usou a temperatura máxima, umidade relativa, precipitação e demanda históricos como fatores de produção para a previsão de demanda para próximos dois dias. A precisão da previsão foi de 95,89%. Um modelo de persistência foi empregado para a temporada de inverno e verão para comparação dos efeitos na previsão de demanda. A comparação dos resultados mostrou que os modelos RNA demonstraram uma forte capacidade de extrair as relações não-lineares entre o consumo de água e parâmetros climáticos. A abordagem de RNA simplificou o processo da modelagem de previsão de demanda de curto prazo significativamente em relação aos modelos de regressão convencional, assim os resultados para previsão com RNA mostraram serem bem promissores para a previsão de demanda.

4. DADOS E METODOLOGIA

O manejo de água para otimização dos sistemas urbanos de abastecimento de água é fundamental para a sociedade. O crescimento elevado da demanda está extrapolando as previsões e a capacidade de oferta de água. Justifica-se, deste modo, o desenvolvimento e aplicação de modelagens matemáticas para a obtenção de políticas de planejamento otimizadas para uma gestão adequada deste recurso. Apresenta-se, neste capítulo, a área de estudo, os dados e um modelo de otimização para operação de sistemas de abastecimento de água, utilizando uma Rede Neural Artificial (RNA).

4.1. Área de Estudo

A RMSP é formada por 39 cidades, sendo composta pela capital do estado de São Paulo e mais 38 cidades. A população da região em 2008 era de aproximadamente 19,7 milhões de habitantes (Seade, 2009), ou seja, 47,9% do total do estado, possuindo a densidade demográfica mais elevada, 2.479,6 habitantes por km^2, diante da estadual que é de 165,8 hab/km^2.

A disponibilidade hídrica por habitante é de 201m^3/hab./ano, considerada extremamente baixa segundo a recomendação da ONU que é de 2500 m^3/hab.ano (Porto, 2003), Essa situação de escassez é considerada dramática, ocorrendo devido ao fato da RMSP estar localizada numa região de cabeceiras de rios e por ser o maior aglomerado urbano do País.

A RMSP situa-se num planalto a 750m em relação ao nível do mar. O clima paulista é caracterizado por fatores fisiográficos e atmosféricos. Os fatores fisiográficos têm origem física, como o relevo, forma e altitude e natureza do terreno e os climáticos, como as massas de ar, frentes frias e convecção. O clima na RMSP é considerado úmido e subtropical (tipo Cfa segundo Köppen) com diminuição de chuvas no

inverno e temperatura média anual de 19ºC, tendo invernos brandos e verões com temperaturas moderadamente altas, aumentadas pelo efeito da poluição e da altíssima concentração de edifícios. O mês mais quente, fevereiro, tem temperatura média de 22,5℃ e o mês mais frio, julho, de 16℃.

Devido à proximidade do mar, a maritimidade é uma constante do clima local sendo responsável pela moderação das temperaturas. No inverno, porém, o penetração de fortes massas de ar polar, acompanhadas de excessiva nebulosidade, às vezes fazem com que as temperaturas permaneçam muito baixas. Temperaturas máximas entre 14ºC e 16ºC são comuns até mesmo durante o outono e no início da primavera. Outro fator que influência o clima é o efeito das ilhas de calor, causado por excessos de edificações, asfalto, concreto e poucas áreas verdes também influenciam no clima da região.

A umidade tem índices considerados aceitáveis durante todo o ano, sendo a média anual de 78%. O índice de umidade abaixo de 30% são considerados prejudiciais para a saúde da população segundo a Organização Mundial da Saúde (OMS). A umidade relativa na RMSP chega a atingir níveis críticos no inverno devido a poluição, combinada com a falta de ventos e a menor ocorrência de chuvas de maio à setembro a baixa umidade prejudica a dispersão de poluentes, ocasionando o fenômeno de inversão térmica.

A precipitação anual média é de 1486mm, concentrados principalmente no verão. As estações do ano são relativamente bem definidas: o inverno é ameno e seco, e o verão, moderadamente quente e chuvoso. Outono e primavera são estações de transição.

4.1.2. Consumo da RMSP

A população da RMSP é abastecida de água por oito estações de tratamento de água (ETA) da SABESP. que produzem uma vazão de cerca de 63 m^3/s^1. Para uma população que está distribuída em 39 municípios, sendo que 28 são atendidos pelas

oito ETA´s da Sabesp (Sabesp, 2001) e os demais por meio de seus serviços autônomos de distribuição. O conjunto de sistemas produtores, chamando de Sistema Adutor Metropolitano – SAM, ou Sistema Integrado de Abastecimento de Água da Região Metropolitana de São Paulo, composto por reservatórios de regularização, captações, adutoras de água bruta e tratada, estações de tratamento, reservatórios, elevatórias e rede de distribuição (Sabesp, 2002). Cada sistema produtor possui uma área de atendimento proporcional à sua produção e geograficamente próxima à respectiva estação de tratamento de água. O sistema adutor da Sabesp distribui 99% da água tratada na RMSP (Ginez e Pereira Filho, 2003).

O sistema Cantareira é responsável por 50% do abastecimento de todo Sistema Produtor Metropolitano de São Paulo. A capacidade nominal deste sistema é 33 m^3/s, e a ETA Guaraú está localizada na parte norte da RMSP. O segundo principal sistema é o Guarapiranga com capacidade de 15 m^3/s na ETA ABV. O sistema Alto Tietê é o terceiro sistema produtor, com capacidade de 10 m^3/s (Porto, 2003). A figura 4.1 mostra o sistema integrado que é composto das seguintes ETA´s: Cantareira, Guarapiranga, Alto Tietê, Rio Grande, Rio Claro, Alto Cotia, Baixo Cotia e Ribeirão da Estiva. A tabela 4.1 mostra a capacidade de cada um dos sistemas de abastecimento.

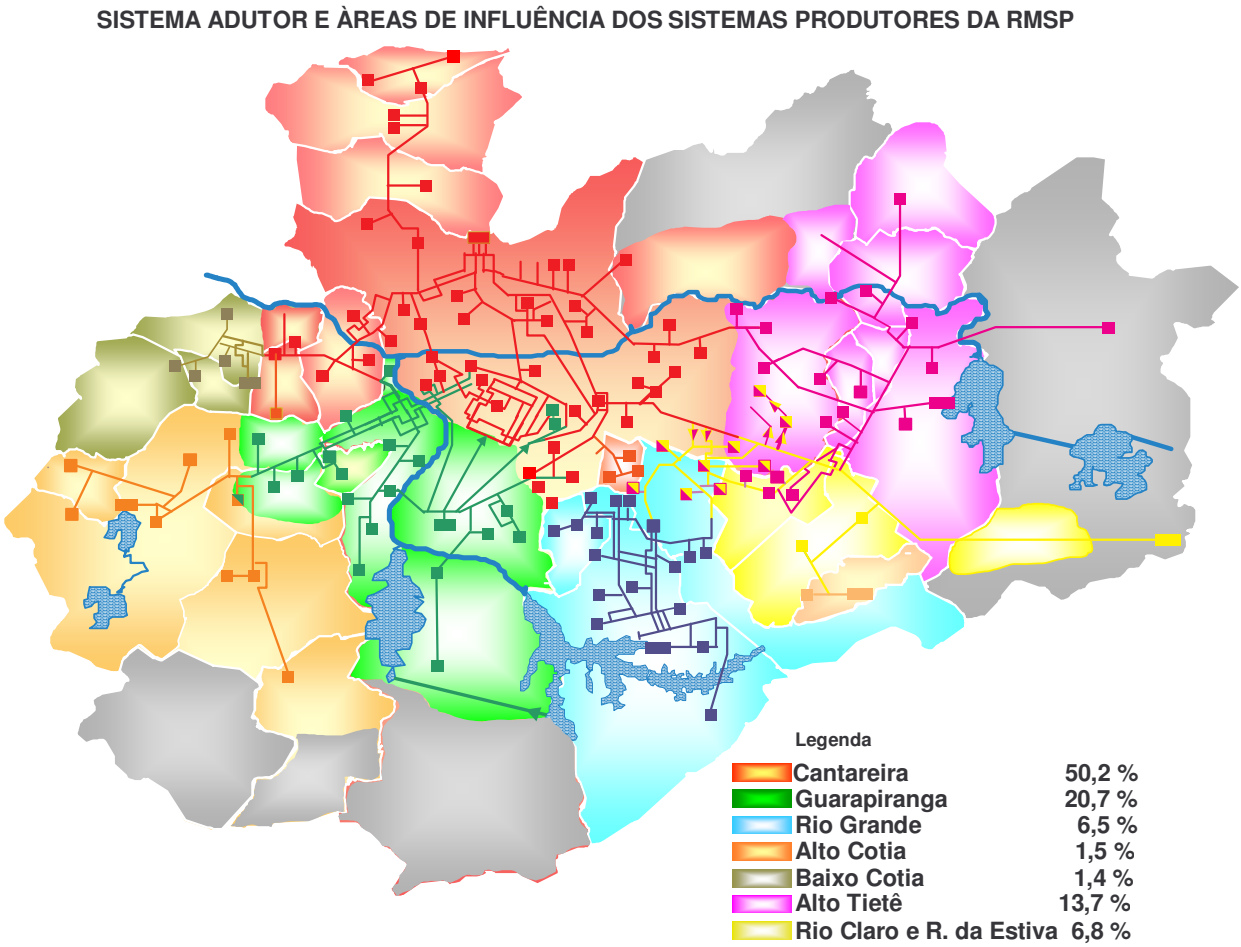


Figura 4.1: Sistema adutor da RMSP e áreas de abrangência por sistemas produtores. As cores indicam o sistema de produção (tons claros) e rede de adução (tons escuros). Os contornos (branco) indicam limites municipais. Contribuição (%) por sistema adutor para o total de 64,2 m^3s^{-1} indicado na legenda. (Fonte: Ginez e Pereira Filho, 2003).

Tabela 4.1: Características dos sistemas produtores da SABESP na RMSP. Para cada Sistema Produtor está indicada sua capacidades de tratamento (m^3 s^{-1}), produção (m^3 s^{-1}) e população. (Fonte: Ginez e Pereira Filho, 2003)

Sistema Produtor	*Disponibilidade do Manancial*	*Capacidade ETA (2001)*	*Produção Março 20002*	*População (milhões)*
Cantareira	**31,3**	**33,0**	**30,1**	**8,8**
Guarapiranga	**14,3**	**14,0**	**13,6**	**4,0**
Alto Tietê	**8,1**	**10,0**	**8,7**	**2,5**
Rio Grande	**4,2**	**4,2**	**4,3**	**1,3**
Rio Claro	**3,6**	**4,0**	**3,1**	**0,9**
Alto Cotia	**0,9**	**1,3**	**0,8**	**0,2**
Baixo Cotia	**0,6**	**1,1**	**0,9**	**0,3**
Ribeirão da Estiva	**0,1**	**0,1**	**0,1**	**0,1**
Total	**63,1**	**67,7**	**61,5**	**18,0**

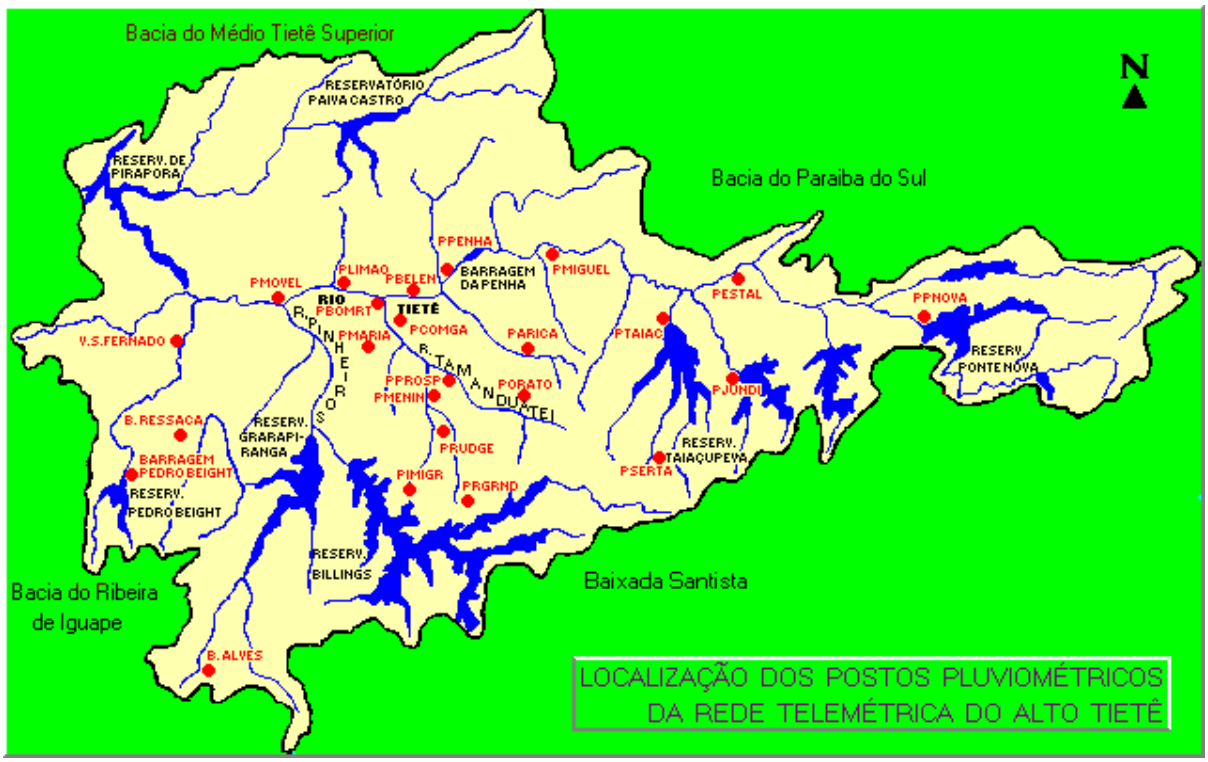
4.2. Dados Utilizados

As análises do consumo de água, analisados nesse trabalho, contemplam o período de 2001 a 2005 e envolvem dados das estações telemétricas da SABESP, da Companhia de Tecnologia de Saneamento Ambiental – CETESB e do Departamento de Águas e Energia Elétrica (DAEE), e os dados da estação meteorológica do IAG. A figura 4.2a mostra a localização e a distribuição espacial das estações de medição da qualidade do ar, estações meteorológicas da CETESB e a localização dos pontos da rede telemétrica do Alto Tietê (RTAT) que se encontram na figura 4.2b. A localização das estações da CETESB e os respectivos pontos de medição de vazão do Sistema de Controle do Abastecimento - SCOA estão relacionados na tabela 4.2 (Pereira Filho e Ginez, 2004). Observa-se que as variáveis medidas pela CETESB não são as mesmas para todas as estações.

Tabela 4.2: Estações de qualidade do ar e estações meteorológicas da CETESB (Figura 3). A letra "x" indica a disponibilidade de uma variável meteorológica: temperatura do ar (T), umidade relativa (U), pressão (P), precipitação (C), vento (V) e radiação solar (R). Na última coluna estão indicados os respectivos postos de medição do SCOA (Fonte Pereira Filho e Ginez, 2004).

Bairro	T	U	P	C	V	R	SCOA
Parque Dom Pedro II	x	x			x		51
Santana					x		32
Centro							29
Lapa					x		40
Pinheiros							41
Taboão da Serra							111
Santo Amaro							1
Ibirapuera	x	x			x	x	1
Moóca					x		33
Penha							14
São Miguel Paulista	x	x	x		x		23
São Caetano do Sul	x	x			x		71,72,73

(a)



(b)

Figura 4.2: a) Distribuição espacial das estações de qualidade do ar e estações meteorológicas da CETESB na RMSP. (Fonte: Pereira, Filho e Ginez, 2004). b) Localização dos postos pluviométricos da rede telemétrica do Alto Tietê (Fonte: http://www.saisp.br/site/sobreAT.htm). Os pontos vermelhos correspondem a localização das estações e dos postos pluviométricos respectivamente.

4.3. Redes Neurais Artificiais

As redes neurais artificiais vêm sendo utilizadas para resolver problemas de previsão de séries temporais, devido à sua capacidade de generalização e habilidade de aprendizado, tornando-se capazes de identificar e assimilar as características das séries. Outro fator que leva a utilização das RNA's é o processo de calibração automática que supre uma deficiência de calibração dos parâmetros de outros modelos.

4.3.1. Formulações

Existem diferentes tipos de redes neurais, e cada uma delas tem uma arquitetura, método de aprendizagem e desempenho diferentes. Mas a dinâmica e a estrutura de uma rede corresponde prontamente a sua função. Neste trabalho, a rede utilizada é do tipo *feedforward multi-camadas* (Santos, 2001), que são utilizadas para reconhecimento de padrões, filtragem de sinal, compressão de dados e comparações de padrões heteroassociativos. Uma rede neural *feedforward* típica de três camadas é apresentada na figura 4.3. A camada de entrada é onde os padrões são apresentados a rede; a camada escondida é a extratora de característica, ou seja, onde ocorre a maior parte do processamento; e a camada de saída onde o resultado final é concluído e apresentado. Os elementos de processamento em cada camada são chamados de nós ou unidades. Cada um desses nós é conectado aos nós das camadas vizinhas. Os parâmetros associados com cada uma dessas conexões são chamados pesos. O peso $W_{h,\ i}$ representa a intensidade da conexão entre as camadas, de entrada e escondida, enquanto $W_{o,\ h}$ representa a intensidade da conexão das camadas, escondida e saída. Todas as conexões nodais são feitas apenas em uma direção (Hsu et al., 1995).

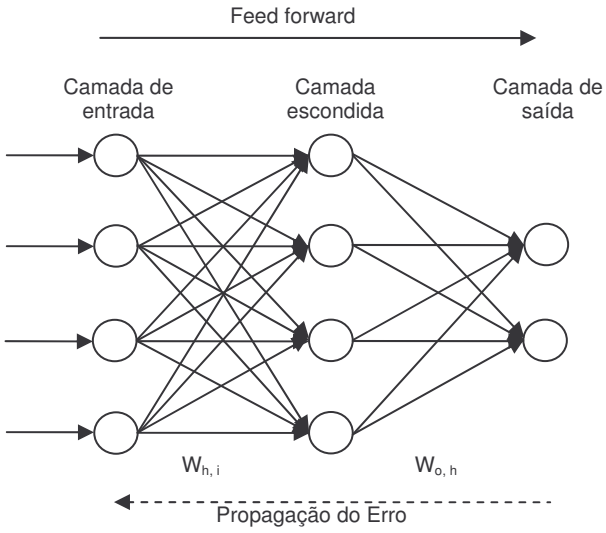


Figura 4.3: Rede Neural Artificial *feedforward* típica de três camadas

Cada nó *h* recebe um sinal de entrada de cada um dos nós *i* da camada anterior. Cada sinal de entrada (x_h) tem associado a ele um peso ($w_{h,i}$). Então, é feita uma média ponderada entre os sinais de entrada e os pesos das conexões. O sinal de entrada efetivo (S_h) para o nó *h* é a somatória dos pesos de todos os sinais de entrada. O resultado é aplicado a uma função de ativação ou função de transferência. O sinal de saída do nó é produzido através da passagem do sinal de entrada efetivo, S_h, por meio de uma função de ativação não linear, algumas vezes chamada de função de transferência ou função limiar. Representada por:

$$S_h = \sum_{i=0}^{n_0} w_{hi} x_i \tag{4.1}$$

onde: S_h é o sinal de entrada efetivo;

$w_{h,i}$ é peso e

x_i é o sinal de entrada.

A função de ativação mais usada em RNA's é a função sigmóide. A função sigmoidal é caracterizada pelo formato S, alongado, contínuo e monotonicamente crescente, limitada por duas assíntotas horizontais. A forma particular de definição para um neurônio *j* é dada pela função logística:

$$y_j = f(S_h) = \frac{1}{[1+\exp(S_h)]} \tag{4.2}$$

onde o S_h pode variar na faixa de $\pm\infty$, enquanto y_h é limitada entre 1 e 0.

Um método utilizado para treinamento desse tipo de RNA é o algoritmo de erro *backpropagation*. Durante o treinamento, a rede opera em uma seqüência de dois passos. Primeiro, um padrão é apresentado à camada de entrada da rede. A atividade resultante flui através dela, camada por camada, até que a resposta seja produzida pela camada de saída. No segundo passo, a saída obtida é comparada á saída desejada para esse padrão particular. Se esta não estiver correta, o erro é calculado. O erro é propagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro que é retropropagado. As redes que utilizam *backpropagation* trabalham com uma variação da regra delta, apropriada para redes multicamadas, denominada de Regra Delta Generalizada. A regra delta essencialmente implementa um gradiente decrescente no quadrado da soma do erro para funções de ativação linear. Entretanto, o erro da superfície pode não ser tão simples, e suas derivadas mais difíceis de serem calculadas. Nesses casos devem ser utilizadas redes com camadas intermediárias. Ainda assim, as redes ficam sujeitas aos problemas de mínimos locais.

O treinamento das redes multicamadas com *backpropagation* pode demandar muitos passos no conjunto de treinamento, resultando em um tempo de treinamento consideravelmente longo. Se for encontrado um valor mínimo local, o erro para o conjunto de treinamento pára de diminuir e estaciona em um valor maior que o aceitável. Uma maneira de aumentar a taxa de aprendizado sem levar à oscilação é modificar a regra delta generalizada para incluir uma constante que determine o efeito das mudanças passadas dos pesos na direção atual do movimento no espaço de pesos.

4.3.2. Algoritmo de treinamento

O algoritmo *Linear Least Squares SIMplex* – LLSSIM (Hsu et al., 1985) é um algoritmo híbrido para treinamento de RNA´s do tipo *feedforward* de três camadas. Esse algoritmo é uma combinação do algoritmo dos Mínimos Quadrados Linear e da otimização do algoritmo *Simplex* com múltiplas inicializações. A vantagem deste algoritmo é que, com poucas iterações, obtêm-se os mínimos globais ou próximos do global. Para implementar essas duas estratégias de treinamento, o conjunto de pesos é dividido. Os pesos, na entrada da camada escondida, são calculados a partir do algoritmo de otimização Simplex e, na saída desta camada, através da estimativa dos Mínimos Quadrados. A divisão dos pesos reduz o espaço de busca, diminuindo o tempo do processo de treinamento. O Simplex melhora as características da busca global através de múltiplas entradas aleatórias, evitando-se os mínimos locais (Hsu et al., 1996).

Havendo *m* conjuntos de padrões de entrada e saída, com n_0 entradas, n_2 saídas e n_1 neurônios escondidos, num treinamento cíclico (batch) (GUPTA et al.; 1997), o padrão de entrada é representado por $[x_1(p), x_2(p), \cdots, x_{n0}(p)]^T$, a saída da rede por $[z_1(p), z_2(p), \cdots, z_{n2}(p)]^T$, e a saída alvo por $[t_1(p), t_2(p), \cdots, t_{n2}(p)]^T$. Então, a função custo a ser minimizada, pode ser expressa por:

$$F[w] = \frac{1}{2} \sum_{p=1}^{m} \sum_{k=1}^{n_2} \left(t_k(p) - z_k(p)\right)^2 \qquad (4.3)$$

onde,

$$z_k(p) = f\left(\sum_{j=0}^{n_1} w_{kj}^0 \, y_j(p)\right) \qquad (4.4)$$

$$y_j(p) = f\left(\sum_{i=0}^{n_0} w_{ji}^h \, x_i(p)\right) \qquad (4.5)$$

onde,

$x_i(p)$ representa o i-ésimo sinal de entrada do p-ésimo padrão;

x_0 é a tendência de entrada;

w_{ji}^{h} é o peso da entrada da camada escondida conectada a i-ésima entrada do j-ésimo neurônio escondido;

$y_j(p)$ representa a saída do j-ésimo neurônio da camada escondida;

w_{kj}^{0} é o peso da saída da camada escondida conectado à resposta do j-ésimo neurônio escondido do k-ésimo neurônio de saída;

$z_k(p)$ representa a saída do k-ésimo neurônio de saída;

$t_k(p)$ é a saída alvo do p-ésimo padrão, e

f(.) é a função de transferência não linear, monótona e limitada entre 0 e 1.

A função custo pode ser reescrita como,

$$\begin{aligned} F[w] &= \frac{1}{2}\sum_{p=1}^{m}\sum_{k=1}^{n_2}\left(t_k(p)-z_k(p)\right)^2 \\ &= \frac{1}{2}\sum_{p=1}^{m}\sum_{k=1}^{n_2}\left[t_k(p)-f\left(\sum_{j=0}^{n_1} w_{kj}^{0}\, f\left(\sum_{i=0}^{n_0} w_{ji}^{h} x_i(p)\right)\right)\right]^2 \end{aligned} \tag{4.6}$$

e a função de transferência *f(.)* é a função sigmóide na forma da função logística:

$$f(u) = \frac{1}{1+\exp(-u)} \tag{4.7}$$

O objetivo do treinamento da rede é obter os pesos w_{ji}^{h} e w_{kj}^{0} que minimizem o valor da função custo (4). A estratégia de treinamento do LLSSIM utiliza a divisão dos pesos em dois grupos, os pesos na entrada da camada escondida w_{ji}^{h} e os pesos na saída da camada escondida w_{kj}^{0}. Define-se TS_k (p) como sendo o valor do alvo *p* transformado pelo inverso da função logística do nó de saída k:

$$TS_k(p) = \ln\left\{t_k(p)/[1-t_k(p)]\right\} \tag{4.8}$$

A partir da equação (4.8) acima, defini-se uma nova função de erro.

$$F[w] = \frac{1}{2}\sum_{p=1}^{m}\sum_{k=1}^{n_2}\left(TS_k(p) - s_k^o(p)\right)^2$$
$$= \frac{1}{2}\sum_{p=1}^{m}\sum_{k=1}^{n_2}\left[TS_k(p) - \sum_{j=0}^{n_1} w_{kj}^0 f\left(\sum_{i=0}^{n_0} w_{ji}^h x_i(p)\right)\right]^2 \quad (4.9)$$

Nesta representação os alvos transformados $TS_k(p)$ são lineares em relação aos pesos da saída da camada escondida w_{kj}^0, e os pesos da entrada da camada escondida são não lineares w_{ji}^h. Assumindo que os valores dos pesos da entrada da camada escondida são conhecidos, os pesos ótimos da saída da camada escondida, condicionados a partir dos valores dos pesos de entrada desta camada, podem ser calculados explicitamente usando o método dos mínimos quadrados linear (LLS), fazendo a derivada da função F_1 relativa aos pesos da camada escondida e igualando a zero, obtendo-se:

$$\frac{\partial F_1}{\partial w_{kj}^0} = -\sum_{p=1}^{m}\left(TS_k(p) - s_k^0(p)\right)y_j(p) = 0 \quad (4.10)$$

A equação (4.10) pode ser reescrita como:

$$\sum_{p=1}^{m} TS_k(p)\, y_j(p) = \sum_{p=1}^{m}\sum_{l=0}^{n_1} w_{kl}^0 y_l(p) y_j(p)$$
$$= \sum_{p=1}^{m} y_j(p) \sum_{l=o}^{n_1} y_l(p) w_{kl}^0 \quad (4.11)$$

Definindo-se:

$$R_{jl} = \sum_{p=1}^{m}\sum_{l=0}^{n_1} y_j(p) y_l(p) \quad (4.12)$$

$$Q_j = \sum_{p=1}^{m} TS_k(p) y_j(p) \quad (4.13)$$

Os pesos na saída da camada escondida, w_k^0, são obtidos de um sistema de equações lineares:

$$w_k^0 = R^{-1}Q \qquad (4.14)$$

onde,

$w_k^0 = [w_{k0}^0, w_{k1}^0, \cdots, w_{kn1}^0]^T$ são os pesos condicionais ótimos para saída da camada escondida, pois seus valores dependem dos valores selecionados pelos pesos de entrada desta camada.

A estrutura da RNA é utilizada para modelar processos não lineares que relacionam entradas e saídas, como por exemplo, o consumo de água urbano. O conjunto de dados utilizados foi normalizado de acordo com uma função sigmóide

As séries utilizadas foram normalizadas entre [0, 1]. A normalização nesta faixa evita a saturação do sinal de saída da RNA (SMITH, 1993). O consumo de saída no instante t, $z(t)$, relaciona as entrada $x(t-j)$ e saídas $z(t-j)$ em instantes anteriores numa estrutura geral de modelo não linear:

$$z(t) = g_{non}(z(t-1) - \cdots, z(t-n_a), x(t-1), \cdots, x(t-n_b)) + e(t) \qquad (4.15)$$

onde,

g_{non} é a função de mapeamento não linear;

$e(t)$ é o mapeamento do erro a ser minimizado;

n_a e n_b são os números de entradas e saídas em instantes de tempo anteriores que contribuem na saída atual e

t, o tempo.

Esta estrutura recebe a notação de RNA (n_a, n_b, n_h, n_0) onde,

$n_a + n_b$ é o número de nós na camada de entrada,

n_h é o número de nós na camada escondida e

n_0 é o número de nós na camada de saída.

Na identificação de um modelo de RNA, os valores de n_a, n_b e n_h devem ser

selecionados, e os pesos da rede w_{ji}^{h} e w_{kj}^{0} estimados de maneira que a previsão do erro seja minimizada.

4.4. Previsão de Demanda de Água

As análises das informações coletadas foram realizadas em três etapas: análise preliminar, análise de correlação e análise do efeito das variáveis climatológicas e outras variáveis sobre a demanda de água. Após o estudo das variáveis climatológicas que influenciam no consumo de água, serão avaliadas aquelas que mais contribuem para variação do consumo.

Uma rede de três camadas foi desenvolvida tendo como vetores de entrada, o consumo de água nas últimas 24h, a precipitação, a umidade, a pressão, a temperatura, além de outras variáveis que possam influenciar no consumo, tendo como vetor de saída o consumo de água, horário ou diário. As variáveis foram normalizadas (Lachtermacher e Fuller, 1995) devido a diferença de unidades entre elas. A normalização ficou na faixa entre 0 e 1. Do total dos dados, 70% foram utilizados para o treinamento da RNA e os 30% restantes para teste e verificação.

Após a análise dos dados, o primeiro processamento realizado foi a normalização (pré-processamento) dos dados de entrada para o treinamento, teste e previsão da RNA; posteriormente, os resultados obtidos foram desnormalizados (pós-processamento). Os valores de N, número total de padrões ou exemplos, contidos no conjunto de treinamento são variáveis de acordo com o período de dados disponíveis. As séries, de entrada e saída identificadas, foram classificadas como atributos com valores reais por se tratarem de variáveis do consumo, meteorológicas e temporais. Nesse caso, os dados foram normalizados no intervalo de [0,1]. A figura 4.4 apresenta o diagrama do processo utilizado para realizar o trabalho.

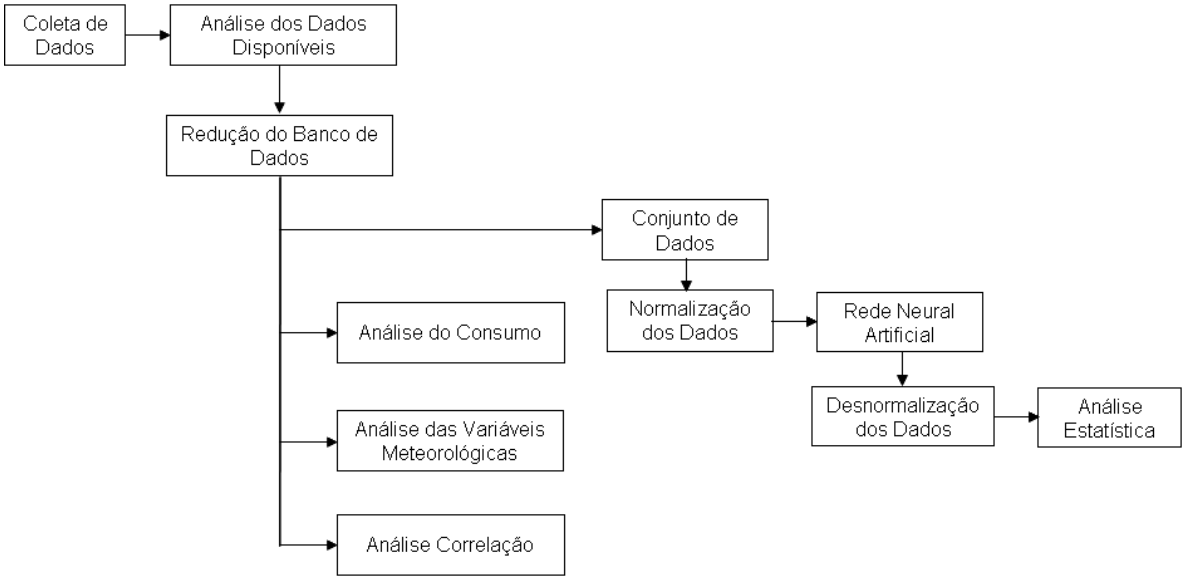


Figura 4.4: Diagrama de bloco do fluxo do trabalho.

Os dados dos oito sistemas produtores de abastecimento da RMSP foram utilizados para avaliar a relação do consumo com as variáveis meteorológicas e antrópicas no período de 2001 a 2005. Depois de realizada essa avaliação, utilizou-se o consumo do sistema produtor da Cantareira e do setor do Itaim Paulista do ano de 2005 para testar a RNA. O sistema Cantareira foi escolhido por ser o principal sistema produtor da RMSP e o Itaim por ser um setor que é predominantemente residencial. Utilizou-se somente os dados do ano de 2005 para testar a RNA devido ao grande volume de dados e também a erros existentes nos dados dos outros anos.

As denominações para a entrada dos modelos de previsão correspondem as seguintes variáveis: H-hora; DS-dia da semana; EA-estação do ano e TD- tipo do dia, denominadas variáveis antrópicas (Va). T- temperatura, UR- umidade relativa, R-chuva, P-pressão, Dir-direção do vento e Vel velocidade do vento são chamadas de variáveis meteorológicas (Vm) e consumo de água (Q). Entre parênteses estão definidos os tempos para avaliação da rede, sendo (t) o tempo, atual e anterior a previsão e (t-1), (t-2), (t-3), ..., (t-24) referem-se respectivamente a: 1h; 2h; 3h; e 24 horas antes do instante atual, já (t+1), (t+2), ..., (t+24) são referentes a: 1h; 2h, ...,e 24h respectivamente, após o instante atual.

A tabela 4.3 apresenta os modelos da RNA MLP para o Sistema Cantareira. Os modelos RNA-C-1, 2 e 3 utilizam dados horários. No primeiro modelo, utiliza-se

somente as variáveis antrópicas e meteorológicas no instante t; o segundo utiliza as variáveis anteriores, mas as Vm no instante (t-1), e o terceiro modelo todas as anteriores mais o consumo no instante (t-1). A média de 6 horas foi utilizada para os modelos RNA-C-4, 5, 6 e 7 e a média de 12 horas para RNA-C-8 e 9.

Tabela 4.3: Modelos de RNA MLP para o sistema Cantareira. Va são as variáveis antropicas (hora, dia da semana, tipo do dia, estação do ano), Vm são as variáveis meteorológicas (temperatura, umidade, chuva, pressão, direção e velocidade do vento), Q o consumo de água e t o tempo.

MODELO	ENTRADA	SAÍDA
RNA-C-1	Va(t) e Vm(t)	Q(t)
RNA-C-2	Va(t), Vm(t) e Vm(t-1)	Q(t)
RNA-C-3	Va(t), Vm(t), Vm(t-1) e Q(t-1)	Q(t)
RNA-C-4	Va(t+6), Vm(t), Vm(t-6), Q(t) e Q(t-6)	Q(t+6)
RNA-C-5	Va(t+12),Vm(t), Vm(t-6),Vm(t-12), Q(t), Q(t-6) e Q(t-12)	Q(t+12)
RNA-C-6	Va(t+18),Vm(t), Vm(t-6),Vm(t-12), Vm(t-18), Q(t), Q(t-6), Q(t-12) e Q(t-18)	Q(t+18)
RNA-C-7	Va(t+18),Vm(t), Vm(t-6),Vm(t-12), Vm(t-18), Vm(t-24), Q(t), Q(t-6), Q(t-12),Q(t-18) e Q(t-24)	Q(t+24)
RNA-C-8	Va(t+12), Vm(t), Vm(t-12), Q(t) e Q(t-12)	Q(t+12)
RNA-C-9	Va(t+24), Vm(t), Vm(t-12), Vm(t-24),Q(t), Q(t-12) e Q(t-24)	Q(t+24)

Os modelos da RNA MLP para o setor de Itaim Paulista estão nas tabelas 4.4 a 4.6 Na tabela 4.4 mostra que os modelos RNA-IP-1 e 2 são idênticos, porém têm o número de camadas escondidas diferentes, enquanto o RNA-IP-3 não usa as Va. O modelo RNA-IP-4 também não utiliza as Va, somente a temperatura, umidade e consumo como entrada. Os modelos anteriores utilizam a média horária e os modelos RNA-IP-5, 6, 7 e 8 a média de 6 horas.

Tabela 4.4: Modelos de RNA MLP para o setor Itaim Paulista. Va são as variáveis antropicas (hora, dia da semana, tipo do dia, estação do ano), Vm são as variáveis meteorológicas (temperatura, umidade, chuva, pressão, direção e velocidade do vento), Q o consumo de água e t o tempo.

MODELO	ENTRADA	SAÍDA
RNA-IP-1	Va(t+1), Vm(t), Vm(t-1), Q(t) e Q(t-1)	Q(t+1)
RNA-IP-2	Va(t+1), Vm(t), Vm(t-1), Q(t) e Q(t-1)	Q(t+1)
RNA-IP-3	Vm(t), Vm(t-1), Q(t) e Q(t-1)	Q(t+1)
RNA-IP-4	T(t), T(t-1), UR(t), UR(t-1), Q(t) e Q(t-1)	Q(t+1)
RNA-IP-5	Va(t+6), Vm(t), Vm(t-6), Q(t) e Q(t-6)	Q(t+6)
RNA-IP-6	Va(t+12),Vm(t), Vm(t-6),Vm(t-12), Q(t), Q(t-6) e Q(t-12)	Q(t12)
RNA-IP-7	Va(t+18),Vm(t), Vm(t-6),Vm(t-12), Vm(t-18), Q(t), Q(t-6), Q(t-12) e Q(t-18)	Q(t+18)
RNA-IP-8	Va(t+18),Vm(t), Vm(t-6),Vm(t-12), Vm(t-18), Vm(t-24), Q(t), Q(t-6), Q(t-12),Q(t-18) e Q(t-24)	Q(t+24)-

Os modelos para previsão ótima estão nas tabelas 4.5 e 4.6. A previsão ótima é um teste em que utiliza dados observados como se fossem previstos. O objetivo é saber

como os modelos reagiriam se os dados fossem de modelos de previsão. A tabela 4.5 apresenta os modelos para previsão ótima utilizando duas memórias anteriores para o tempo atual. A previsão foi realizada com dados diários, de hora em hora, até completar 24 horas.

Tabela 4.5: Modelos de RNA MLP para previsão ótima do setor Itaim Paulista. Va são as variáveis antropicas (hora, dia da semana, tipo do dia, estação do ano), Vm são as variáveis meteorológicas (temperatura, umidade, chuva, pressão, direção e velocidade do vento), Q o consumo de água e t o tempo. Previsão com memória de 2 tempos anteriores

MODELO	ENTRADA	SAÍDA
RNA-IP-9	Va(t+1), Vm(t), Vm(t-1), Q(t) e Q(t-1)	Q(t+1)
RNA-IP-10	Va(t+2), Vm(t), Vm(t+1), Q(t) e Q(t+1)	Q(t+2)
RNA-IP-11	Va(t+3), Vm(t+1), Vm(t+2), Q(t+1) e Q(t+2)	Q(t+3)
RNA-IP-12	Va(t+4), Vm(t+2), Vm(t+3), Q(t+2) e Q(t+3)	Q(t+4)
RNA-IP-13	Va(t+5), Vm(t+3), Vm(t+4), Q(t+3) e Q(t+4)	Q(t+5)
RNA-IP-14	Va(t+6), Vm(t+4), Vm(t+5), Q(t+4) e Q(t+5)	Q(t+6)
RNA-IP-15	Va(t+7), Vm(t+5), Vm(t+6), Q(t+5) e Q(t+6)	Q(t+7)
RNA-IP-16	Va(t+8), Vm(t+6), Vm(t+7), Q(t+6) e Q(t+7)	Q(t+8)
RNA-IP-17	Va(t+9), Vm(t+7), Vm(t+8), Q(t+7) e Q(t+8)	Q(t+9)
RNA-IP-18	Va(t+10), Vm(t+8), Vm(t+9), Q(t+8) e Q(t+9)	Q(t+10)
RNA-IP-19	Va(t+11), Vm(t+9), Vm(t+10), Q(t+9) e Q(t+10)	Q(t+11)
RNA-IP-20	Va(t+12), Vm(t+10), Vm(t+11), Q(t+10) e Q(t+11)	Q(t+12)
RNA-IP-21	Va(t+13), Vm(t+11), Vm(t+12), Q(t+11) e Q(t+12)	Q(t+13)
RNA-IP-22	Va(t+14), Vm(t+12), Vm(t+13), Q(t+12) e Q(t+13)	Q(t+14)
RNA-IP-23	Va(t+15), Vm(t+13), Vm(t+14), Q(t+13) e Q(t+14)	Q(t+15)
RNA-IP-24	Va(t+16), Vm(t+14), Vm(t+15), Q(t+14) e Q(t+15)	Q(t+16)
RNA-IP-25	Va(t+17), Vm(t+15), Vm(t+16), Q(t+15) e Q(t+16)	Q(t+17)
RNA-IP-26	Va(t+18), Vm(t+16), Vm(t+17), Q(t+16) e Q(t+17)	Q(t+18)
RNA-IP-27	Va(t+19), Vm(t+17), Vm(t+18), Q(t+17) e Q(t+18)	Q(t+19)
RNA-IP-28	Va(t+20), Vm(t+18), Vm(t+19), Q(t+18) e Q(t+19)	Q(t+20)
RNA-IP-29	Va(t+21), Vm(t+19), Vm(t+20), Q(t+19) e Q(t+20)	Q(t+21)
RNA-IP-30	Va(t+22), Vm(t+20), Vm(t+21), Q(t+20) e Q(t+21)	Q(t+22)
RNA-IP-31	Va(t+23), Vm(t+21), Vm(t+22), Q(t+21) e Q(t+22)	Q(t+23)
RNA-IP-32	Va(t+24), Vm(t+22), Vm(t+23), Q(t+22) e Q(t+23)	Q(t+24)

Na tabela 4.6, os modelos para previsão ótima são diferentes dos anteriores, porque utilizam três memórias anteriores para o momento que se quer prever. Observa-se também que os modelos RNA-IP-56 e 57 foram realizados com características diferentes dos outros modelos. Eles fizeram a previsão para 24 horas, sendo que, no modelo RNA-IP-56 utilizou-se todas as memórias de tempos anteriores a previsão e o RNA-IP-57 não utilizou os valores do consumo. Após definidas as configurações de RNA, os conjuntos de dados foram divididos, alguns em três partes, sendo 70% para treinamento, 15% para teste e os outros 15% para validação. Para os modelos de previsão os conjuntos foram divididos em duas partes, 70% para treinamento e 30% para a previsão.

Tabela 4.6: Modelos de RNA MLP para previsão ótima do setor Itaim Paulista. Va são as variáveis antrópicas (hora, dia da semana, tipo do dia, estação do ano), Vm são as variáveis meteorológicas (temperatura, umidade, chuva, pressão, direção e velocidade do vento), Q o consumo de água e t o tempo. Previsão com memória de 3 tempos anteriores

MODELO	ENTRADA	SAÍDA
RNA-IP-33	Va(t+2), Vm(t), Vm(t-1), Vm(t+1), Q(t), Q(t-1)e Q(t+1)	Q(t+2)
RNA-IP-34	Va(t+3), Vm(t), Vm(t+1), Vm(t+2), Q(t), Q(t+1) e Q(t+2)	Q(t+3)
RNA-IP-35	Va(t+4), Vm(t+1), Vm(t+2), Vm(t+3), Q(t+1), Q(t+2) e Q(t+3)	Q(t+4)
RNA-IP-36	Va(t+5), Vm(t+2), Vm(t+3), Vm(t+4), Q(t+2), Q(t+3) e Q(t+4)	Q(t+5)
RNA-IP-37	Va(t+6), Vm(t+3), Vm(t+4), Vm(t+5), Q(t+3), Q(t+4) e Q(t+5)	Q(t+6)
RNA-IP-38	Va(t+7), Vm(t+4), Vm(t+5), Vm(t+6), Q(t+4), Q(t+5) e Q(t+6)	Q(t+7)
RNA-IP-39	Va(t+8), Vm(t+5), Vm(t+6), Vm(t+7), Q(t+5), Q(t+6) e Q(t+7)	Q(t+8)
RNA-IP-40	Va(t+9), Vm(t+6), Vm(t+7), Vm(t+8), Q(t+6), Q(t+7) e Q(t+8)	Q(t+9)
RNA-IP-41	Va(t+10), Vm(t+7), Vm(t+8), Vm(t+9), Q(t+7), Q(t+8) e Q(t+9)	Q(t+10)
RNA-IP-42	Va(t+11), Vm(t+8), Vm(t+9), Vm(t+10), Q(t+8), Q(t+9) e Q(t+10)	Q(t+11)
RNA-IP-43	Va(t+12), Vm(t+9), Vm(t+10), Vm(t+11), Q(t+9), Q(t+10) e Q(t+11)	Q(t+12)
RNA-IP-44	Va(t+13), Vm(t+10), Vm(t+11), Vm(t+12), Q(t+10), Q(t+11) e Q(t+12)	Q(t+13)
RNA-IP-45	Va(t+14), Vm(t+11), Vm(t+12), Vm(t+13), Q(t+11), Q(t+12) e Q(t+13)	Q(t+14)
RNA-IP-46	Va(t+15), Vm(t+12), Vm(t+13), Vm(t+14), Q(t+12), Q(t+13) e Q(t+14)	Q(t+15)
RNA-IP-47	Va(t+16), Vm(t+13), Vm(t+14), Vm(t+15), Q(t+13), Q(t+14) e Q(t+15)	Q(t+16)
RNA-IP-48	Va(t+17), Vm(t+14), Vm(t+15), Vm(t+16), Q(t+14), Q(t+15) e Q(t+16)	Q(t+17)
RNA-IP-49	Va(t+18), Vm(t+15), Vm(t+16), Vm(t+17), Q(t+15), Q(t+16) e Q(t+17)	Q(t+18)
RNA-IP-50	Va(t+19), Vm(t+16), Vm(t+17), Vm(t+18), Q(t+16), Q(t+17) e Q(t+18)	Q(t+19)
RNA-IP-51	Va(t+20), Vm(t+17), Vm(t+18), Vm(t+19), Q(t+17), Q(t+18) e Q(t+19)	Q(t+20)
RNA-IP-52	Va(t+21), Vm(t+18), Vm(t+19), Vm(t+20), Q(t+18), Q(t+19) e Q(t+20)	Q(t+21)
RNA-IP-53	Va(t+22), Vm(t+19), Vm(t+20), Vm(t+21), Q(t+19), Q(t+20) e Q(t+21)	Q(t+22)
RNA-IP-54	Va(t+23), Vm(t+20), Vm(t+21), Vm(t+22), Q(t+20), Q(t+21) e Q(t+22)	Q(t+23)
RNA-IP-55	Va(t+24), Vm(t+21), Vm(t+22), Vm(t+23), Q(t+21), Q(t+22) e Q(t+23)	Q(t+24)
RNA-IP-56	Va(t+24), Vm (t-1), Vm(t+1), Vm(t+2), ...,Vm(t+23), Q(t-1), Q(t+1), Q(t+2),..., e Q(t+23)	Q(t+24)
RNA-IP-57	Va(t+24), Vm (t-1), Vm(t+1), Vm(t+2), ... e Vm(t+23)	Q(t+24)

4.5. Critérios de Avaliação

A verificação dos resultados é mais confiável quando se tem boa quantidade e qualidade de dados. Quanto maior o número de amostras mais confiável são os resultados da verificação. A não homogeneidade dos dados é um grande problema no desempenho da previsão. Desta forma, para melhor quantificar as previsões realizadas foi utilizado, para avaliação, o erro médio (EM), erro médio absoluto (MAE), erro médio quadrático (RMSE), o coeficiente de correlação (r) e o coeficiente de determinação, que estão definidos nas equações 4.16, 4.17, 4.18, 4.19 e 4.20 respectivamente:

$$EM = \frac{1}{n}\sum_{i=1}^{n}(F_i - O_i) \tag{4.16}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|F_1 - O_1\right| \tag{4.17}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_1 - O_1)^2} \tag{4.18}$$

$$r = \frac{\sum_{i=1}^{n}(F_1 - \overline{F})(O_1 - \overline{O})}{\sqrt{\sum_{i=1}^{n}(F_1 - \overline{F})^2}\ \sqrt{\sum_{i=1}^{n}(O_1 - \overline{O})^2}} \tag{4.19}$$

$$R^2 = r^2 \tag{4.20}$$

sendo:

F: valor previsto do consumo;

O: valor observado do consumo;

$\overline{F}$ e $\overline{O}$: média aritmética de ambas variáveis e

n: número de intervalos de tempo.

As tabelas de contingência são usadas para estudar a relação entre duas ou mais variáveis, descrevendo as frequências das categorias de uma das variáveis em relação as categorias de outras, ou seja, é uma tabela de frequências classificadas de acordo com os valores das variáveis em questão. É uma tabela de dupla entrada em que as linhas correspondem a uma variável e as colunas á outra variável. A tabela 4.7 mostra uma tabela de contingência onde, as linhas correspondem aos dados previstos, e as colunas aos dados observados.

Tabela 4.7: Tabela de contingência para contagem dos dados previstos e observados do consumo de água.

		OBSERVADO		
		evento	não evento	total
PREVISTO	evento	A	B	A+B
	não evento	C	D	C+D
	total	A+C	B+D	n= A+B+C+D

onde:

- A é o número de acertos do evento;
- B é o número de eventos previstos, mas que não foram observados;
- C é o número de eventos que foram observados e não foram previstos;
- D é o número de acertos do não-evento;
- (A+C) é o total de eventos observados;
- (B+D) é o total de não-eventos observados;
- (A+B) é o total de eventos previstos;
- (C+D) é o total de não-eventos previstos e
- n é o tamanho do conjunto ou amostra.

A tabela de contingência é útil para verificar que tipos de erro estão ocorrendo. Uma previsão perfeita deveria produzir apenas acertos e não falsos alarmes e erros. Para verificar o desempenho das previsões (Wilks, 2006) será calculado também, a exatidão (E), a freqüência de bias (viés), a probabilidade de detecção (POD), a taxa de falso alarme (FAR), o índice crítico de sucesso (CSI) e a probabilidade de falsa detecção (POFD), descritas pelas equações 4.21, 4.22, 4.23, 4.24, 4.25 e 4.26

$$E = \frac{A + D}{n} \tag{4.21}$$

$$\text{VIÉS} = \frac{A + B}{A + C} \tag{4.22}$$

$$\text{POD} = \frac{A}{A + C} \tag{4.23}$$

$$FAR = \frac{B}{A + B} \tag{4.24}$$

$$CSI = \frac{A}{A + C + B} \tag{4.25}$$

$$POFD = \frac{B}{D + B} \tag{4.26}$$

O FAR indica a porcentagem de valores que não foram previstos corretamente; o POD a porcentagem de valores que foram previsto corretamente; o CSI indica as porcentagens dos valores previstos ou observados corretamente e o POFD as porcentagens dos valores que não foram previstos. Os valores perfeitos para essas são:

- ✓ FAR = 0
- ✓ POD = 1
- ✓ CSI= 1
- ✓ POFD = 0

5. RESULTADOS

Os resultados obtidos foram gerados através dos dados de consumo de água obtidos do Sistema de Controle e Operação de Abastecimento (SCOA) da Sabesp e dados meteorológicos da Estação Meteorológica do IAG. A RNA utilizada foi uma rede *feedforward* de 3 camadas e o algoritmo o LLSSIM (Hsu et al., 1995), que é uma combinação do algoritmo dos Mínimos Quadrados Linear e da otimização do algoritmo Simplex. O objetivo é obter uma solução para a previsão de consumo, utilizando como entrada o consumo e dados meteorológicos, esperando-se que a previsão possa ser utilizada para a otimização da operação do sistema de abastecimento.

Para demonstrar a aplicabilidade e funcionalidade do modelo proposto, apresentam-se os resultados do desempenho da RNA levando em conta a situação climatológica e socioeconômica da região. Serão apresentados neste capítulo, os resultados obtidos para o treinamento, teste e previsão da RNA, assim como, a estatística gerada para verificar o desempenho da rede. Também será mostrada uma análise da relação do consumo com as variáveis meteorológicas.

5.1. Análise dos dados disponíveis

A princípio, os dados disponíveis, para análise dos efeitos das variáveis meteorológicos no consumo e os dados para entrada da RNA foram de:

- consumo de água da Região Metropolitana de São Paulo fornecido pela SABESP;
- variáveis meteorológicas da rede de monitoramento da CETESB e da estação meteorológica do IAG e
- precipitação da rede telemétrica do Alto Tietê (RTAT) fornecidos pela Fundação Centro Tecnológico de Hidráulica (FCTH).

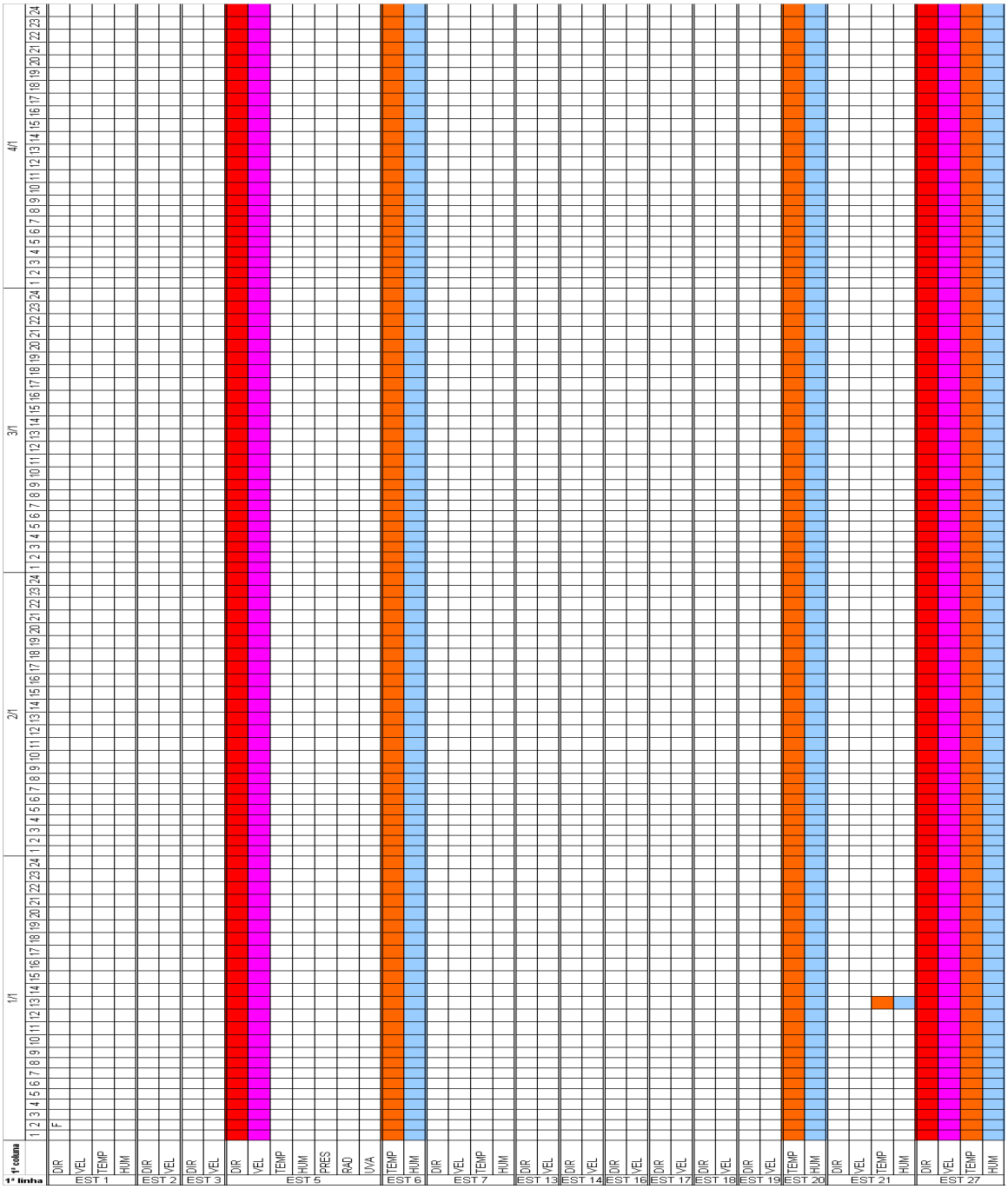
Todos estes dados estão em formato de planilhas, fato que levou a verificação se todos os dados estavam concatenados para o período de 2001 a 2005. O problema da não concatenação é que, algumas vezes, existem dados para um determinado período de uma variável, mas não para a outra variável. A opção para verificar a concatenação entre as variáveis foi verificar as falhas existentes para o período dos dados existentes. A figura 5.1 mostra exemplos das planilhas criadas para verificar as falhas dos bancos de dados da CETESB e do FCTH.

A figura 5.1a apresenta um trecho da planilha criada para os dados da CETESB para o ano de 2004. Outras planilhas semelhantes foram criadas para o período de 2001 a 2005. A função dessas planilhas é indicar onde não existem dados. A primeira linha da figura 5.1a corresponde as estações da CETESB que contêm dados meteorológicos. As variáveis meteorológicas disponíveis para cada uma dessas estações podem ser observadas na segunda linha. O dia e o mês encontram-se na primeira coluna da planilha, essas datas correspondem ao período de 01/01 a 31/12 de 2004. A hora pode ser vista na segunda coluna. Como essa planilha é muito grande observa-se na figura 5.1a somente os dados do período de 01/01 a 04/01. Observa-se que existem células coloridas e células em branco. A célula em branco é onde existem dados e a colorida é onde não têm dados, observa-se que existem longos períodos sem dados.

Um exemplo de planilha para os dados pluviométricos da RTAT pode ser visto na figura 5.1b. Observa-se os dados do posto da Barragem de Ponte Nova para o ano de 2004. Para outros postos e anos foram criadas planilhas semelhantes ao da Barragem de Ponte Nova. Na linha 1 observa-se a data, correspondente ao período de 01/01/2004 a 31/12/2004, assim como na figura 5.1a e que, devido ao tamanho da planilha é mostrado apenas o período de 01/01/2004 a 15/02/2004.

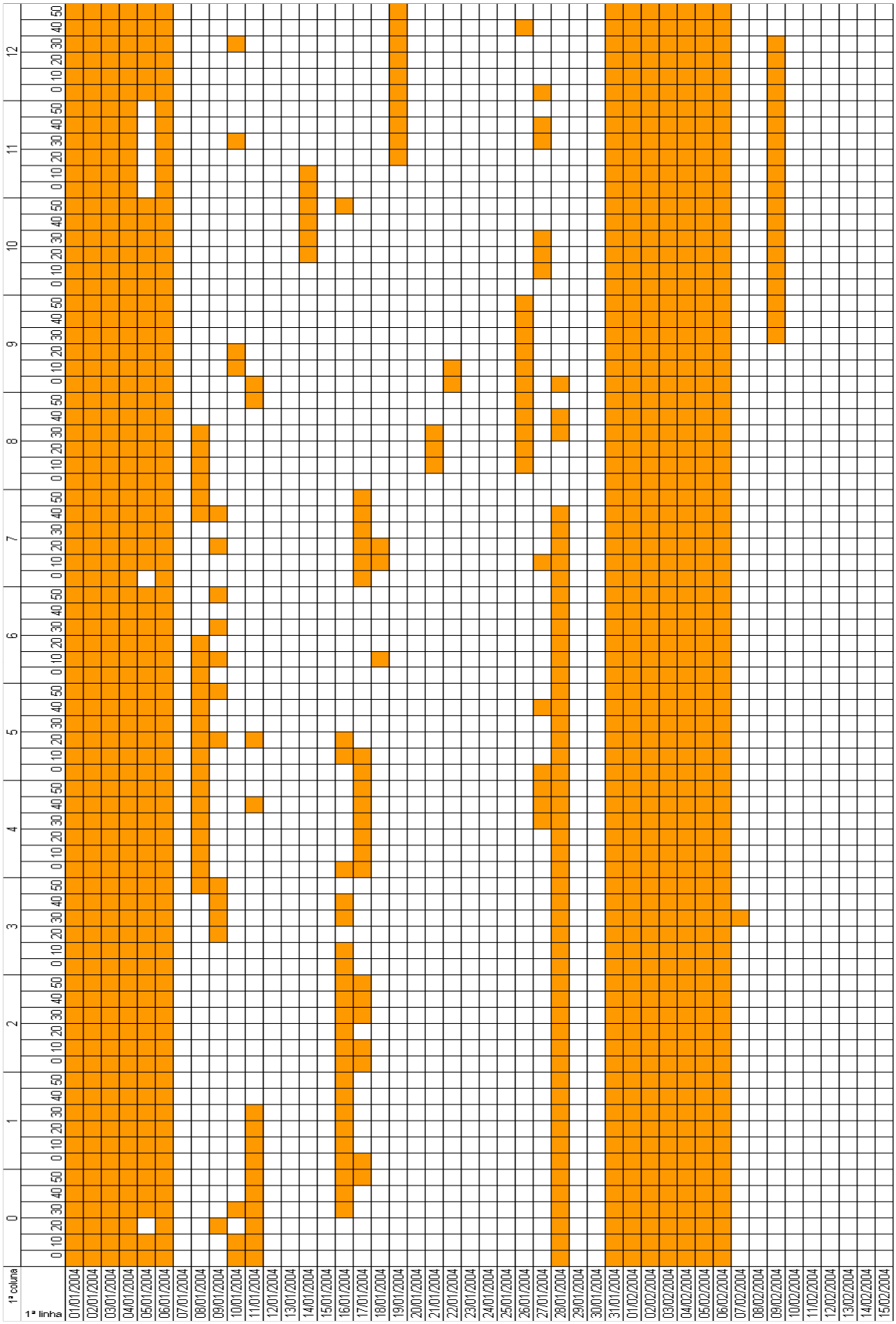
Na coluna 1 encontra-se a hora e na linha 2 os minutos. Novamente, quando se tem células em branco existem dados e quando estão pintadas faltam dados. Desta forma, nota-se que para alguns períodos existem dados para Barragem Móvel e não existem para os dados da CETESB, ou vice-versa. O grande problema nas falhas existentes nesses dados é seu preenchimento. Muitas vezes, existem períodos longos, onde há ausência de dados, tornando-se difícil estabelecer um critério para o

preenchimento dessas falhas. As planilhas completas, exemplificando o que foi dito, podem ser vistas no Anexo 1.



(a)

Figura 51: Falhas existentes no banco de dados CETESB e FCTH para o ano de 2004, as células em branco é onde existem dados e as coloridas são onde não têm dados. (a) exemplo de uma planilha do banco de dados da CETESB, Para todas as estações com dados meteorológicos, o F significa falha, a cor laranja indica que não existem dados para a temperatura, a azul para a umidade e rosa para a velocidade. (b) exemplo de planilha para o banco de dados da RTAT para o posto da Barragem Ponte Nova.



(b)

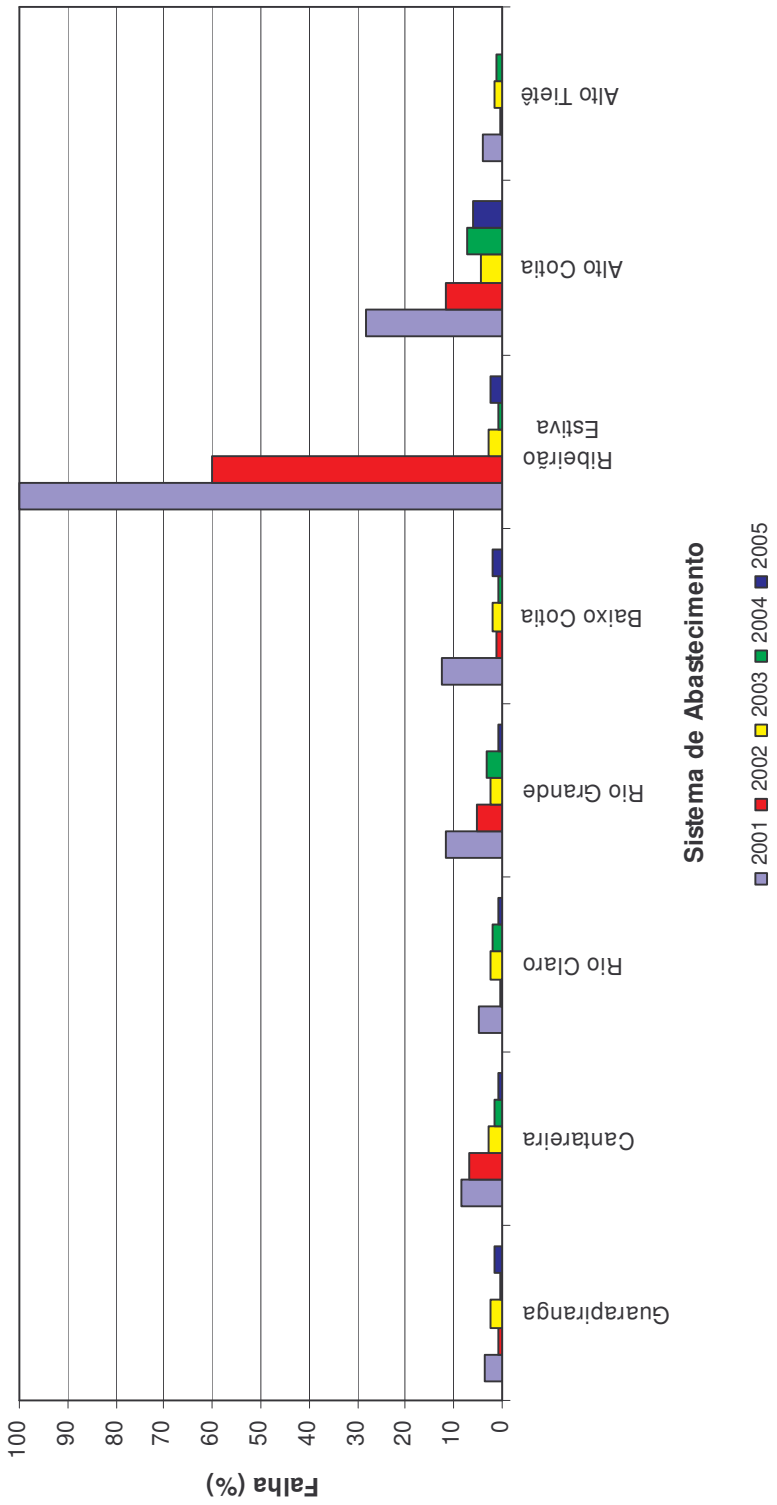
Continuação da figura 5.1

A porcentagem de falhas para as estações meteorológicas da CETESB, estações pluviométricas da RTAT e consumo de água são apresentadas na figura 5.2. Nota-se que o conjunto de dados da CETESB e RTAT apresentam uma quantidade significativa de falhas nos dados. Os dados faltantes compreendem períodos longos e de difícil preenchimento, principalmente os de chuva. Por outro lado, as falhas existentes no banco de dados do Sistema de Abastecimento são menores, sendo que as mais significativas encontram-se nas ETA´s Alto Cotia e Ribeirão da Estiva. Nessa última, as medições começaram em 2002. e as falhas da ETA Alto Cotia são devidas ao próprio sistema. Os dados da estação meteorológica do IAG estão consistidos e por esta razão não contêm falhas. Devido ao grande número de falhas das estações da CETESB e RTAT, optou-se em utilizar somente os dados da estação do IAG.

O número de dados, as médias, valores, mínimos e máximos de consumo e os desvios padrão para as ETAs estão na tabela 5.1. O menor banco de dados refere-se a ETA Ribeirão da Estiva devido aos erros de dados, falhas e ter iniciado as medições somente em 2002. O maior banco de dados obtido foi da ETA Rio Claro.

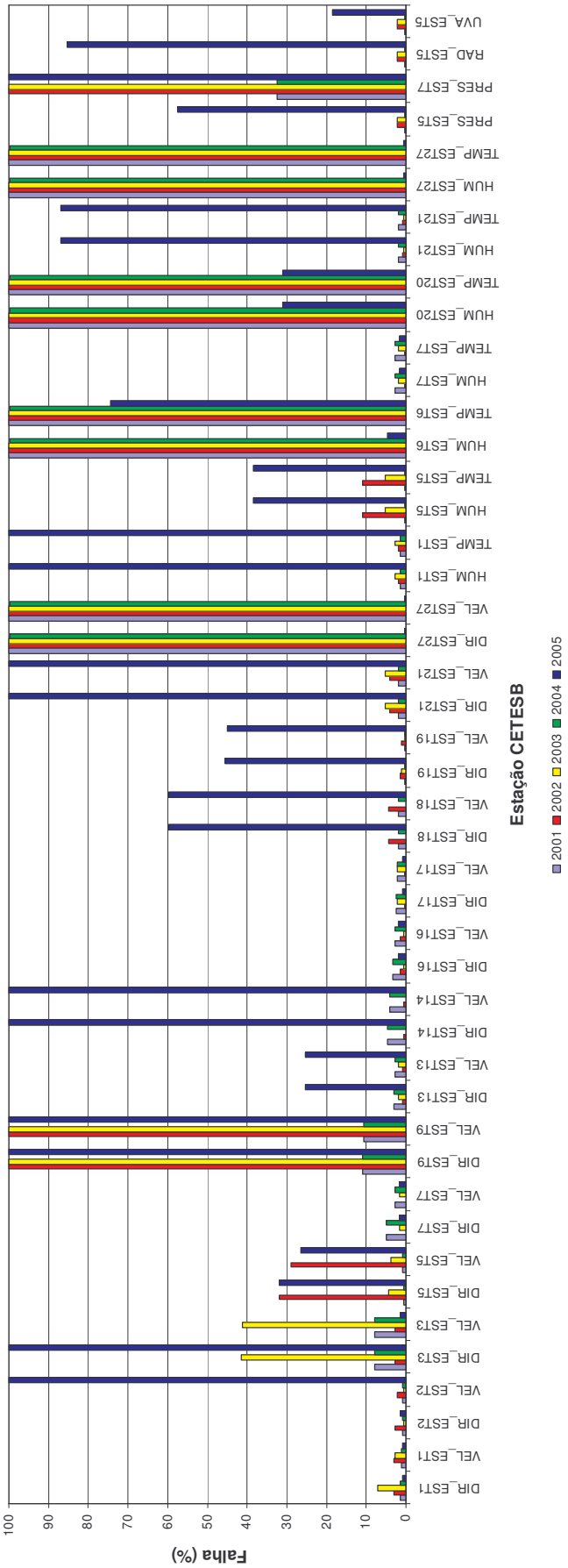
Tabela 5.1: Características dos sistemas produtores da SABESP na RMSP entre 2001 e 2005. Estão indicados, o número de dados, as médias, desvios padrão, mínimos e máximos de consumo de água ($m^3 s^{-1}$).

ETA	No de dados	Média ($m^3 s^{-1}$)	Desvio Padrão ($m^3 s^{-1}$)	Mínimo ($m^3 s^{-1}$)	Máximo ($m^3 s^{-1}$)
ABV - Guarapiranga	42811	13,05	1,06	9,01	19,41
Cantareira	41832	31,11	3,40	20,00	39,80
Alto Tietê	42732	9,30	1,55	6,00	17,38
Rio Grande	41775	4,55	0,35	2,18	5,55
Rio Claro	42937	1,19	0,27	0,10	2,00
Alto Cotia	38515	0,98	0,20	0,22	1,59
Baixo Cotia	38738	0,38	0,11	0,10	0,73
Ribeirão da Estiva	16287	0,09	0,01	0,01	0,12

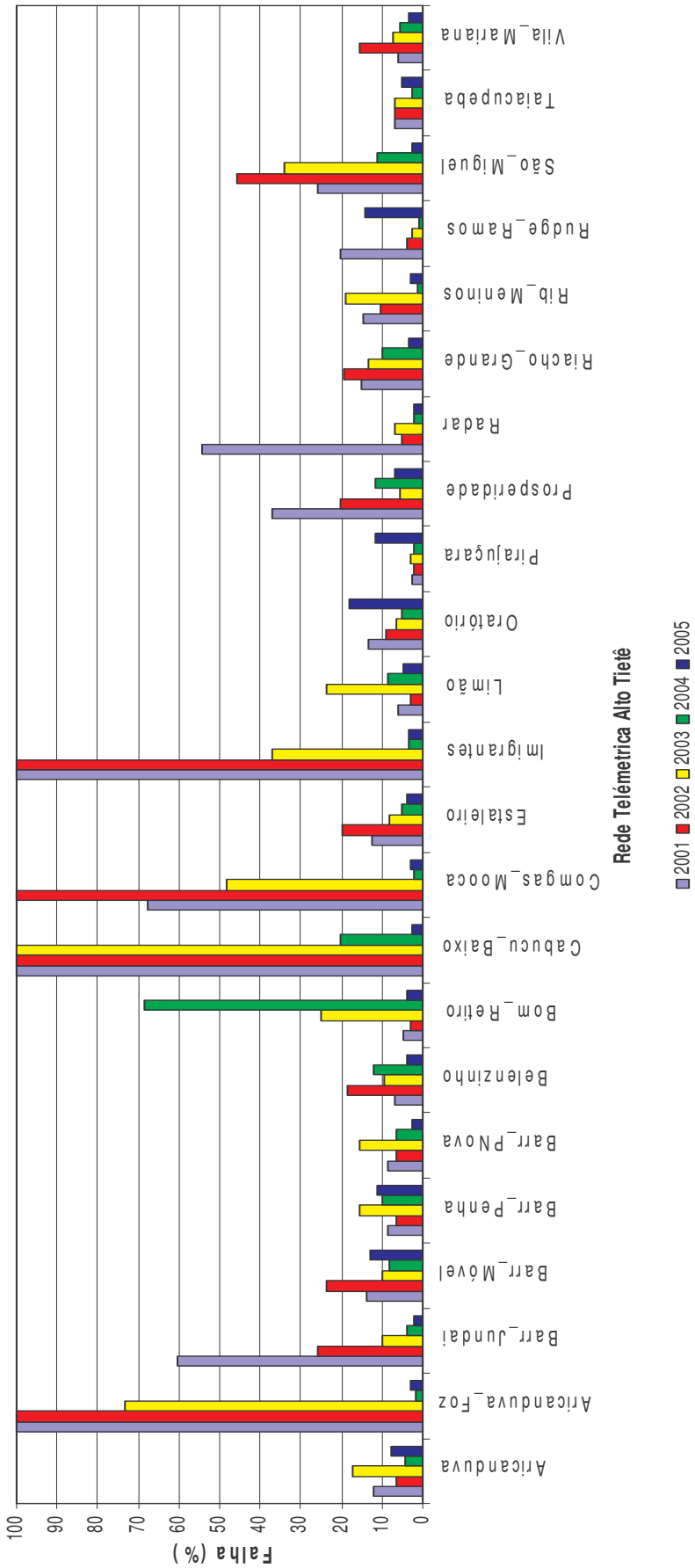


(a)

Figura 5.2: Porcentagem de falhas (%): para o período 2001 a 2005: a) Sistema de Abastecimento RMSP, cada nome identifica uma ETA; b) Estações Meteorológicas da CETESB, as siglas correspondem a: EST: estação, e o número depois de EST corresponde ao número da estação, DIR: direção do vento, VEL: velocidade do vento, TEMP: temperatura, HUM: umidade, RAD: radiação, PRESS: pressão e UVA: raio ultravioleta e c) Estações pluviométricas da Rede Telemétrica do Alto Tietê onde cada nome corresponde a uma estação diferente. As cores indicam o ano.



(b) continuação da figura 5.2



(c) continuação figura 5.2

5.2. Análise do consumo de água

A produção e distribuição de água em sistemas de abastecimento tendem a seguir um padrão com variações influenciadas pelas condições climáticas, dia da semana, demanda dos usuários, sazonalidade, consumo excepcional e outros. A água consumida varia continuamente em função do tempo, variações que podem ser anuais, mensais, diárias, horárias e instantâneas. As variações anuais tendem a ocorrer com o crescimento populacional, melhoria dos hábitos da população e do desenvolvimento populacional.

A variação anual do consumo total do Sistema Produtor de Água da RMSP encontra-se na figura 5.3. Observa-se o crescimento do consumo durante os anos devido ao aumento populacional e crescimento socioeconômico, mas, em 2004, ocorreu uma queda no consumo, provocada pelo fato que, no decorrer dos anos de 2003 e 2004, os índices pluviométricos ficaram abaixo da média, resultando em um reabastecimento insuficiente dos reservatórios, em particular no Sistema Cantareira, o maior sistema da Região Metropolitana de São Paulo. Em razão dessa estiagem, a SABESP implementou um programa para incentivar a população da RMSP a reduzir o consumo de água.

O ano de 2001 teve o menor valor de consumo devido ao racionamento ocorrido nesse ano. Os anos de 2000 e 2001 foram muito secos, com baixos índices pluviométricos ocorrendo uma redução considerável do nível da represa do Sistema Cantareira a exemplo de 2003 e 2004.

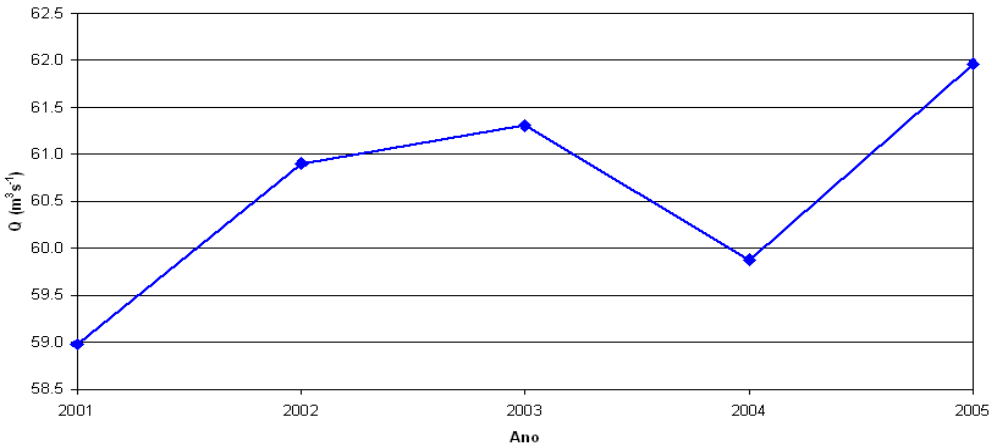


Figura 5.3: Produção de água do Sistema Produtor de Água da RMSP para o período de 2001 a 2005.

O consumo médio mensal tende a mostrar a sazonalidade existente. Os maiores consumos ocorrem no verão, e os menores, no inverno. Em geral, a tendência do consumo é diminuir a partir do mês de março e aumentar a partir do mês de novembro, meses que correspondem às estações outono/inverno e primavera/verão, respectivamente. No inverno, as temperaturas são mais baixas e com isto tem-se uma redução do consumo, ao contrário do verão em que as temperaturas são mais elevadas. Outro fator que gera um aumento do consumo no verão são as chuvas que frequentemente ocasionam enchentes. Já no inverno a redução de chuva provoca uma redução da umidade relativa deixando o clima mais seco, e o clima seco, também, eleva o consumo.

A Figura 5.4 mostra as anomalias do consumo médio mensal dos principais sistemas adutores (ABV – Guarapiranga, Garaú-Cantareira e Alto Tietê) e do consumo médio total. Nota-se que a evolução mensal de cada sistema é distinta e relacionada ao perfil populacional, outras peculiaridades de cada região e de sua localização. Observa-se que, devido ao perfil socioambiental, socioeconômico e às atividades desenvolvidas nessas regiões, os picos mínimos do sistema Guaraú ocorrem em maio; já no sistema ABV, ocorrem dois picos, um em julho e outro em outubro, e no Alto Tietê, em setembro.

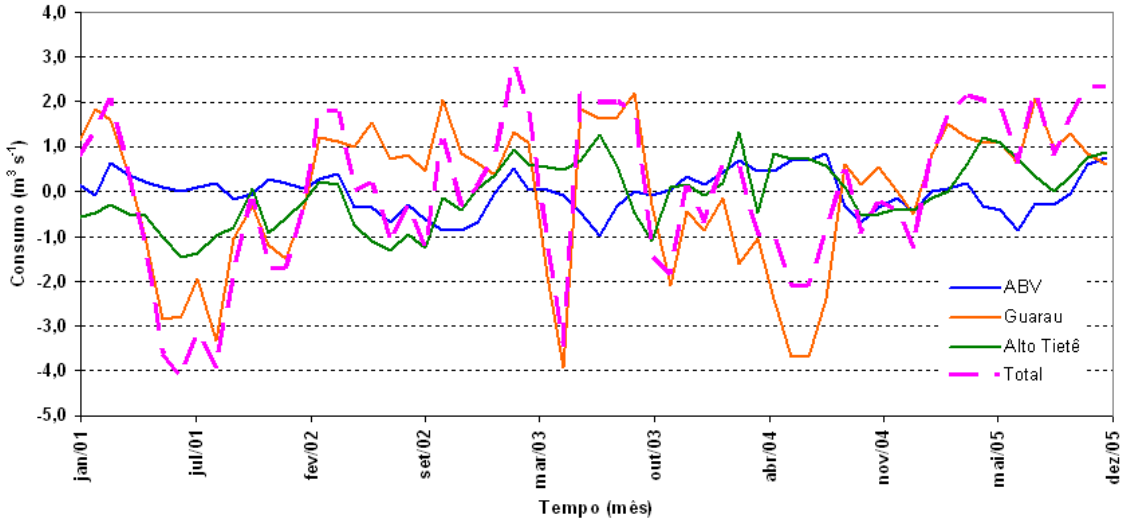


Figura 5.4. Anomalias da produção de água média mensal e da produção total. A linha azul corresponde a ETA ABV-Guarapiranga, a linha abóbora a ETA Garau – Cantareira, a verde a ETA Alto Tietê e a rosa pontilhada ao consumo total das três ETA´s.

Observa-se que o pico máximo ocorre em março devido ao retorno das atividades normais na cidade, considerando que janeiro e fevereiro é período de férias. A produção dos sistemas Guarapiranga e Alto Tietê têm a mesma evolução temporal, com algumas diferenças, embora a característica da densidade populacional das duas regiões sejam parecidas. Os sistemas ABV e Alto Tietê apresentam anomalias de produção muito menores do que o sistema Guaraú, que apresenta uma variação maior para os valores abaixo da média de longo prazo, quando comparado aos valores acima da média.

A evolução da produção média mensal do consumo total do Sistema Produtor de Água e a média anual do consumo para o período de 2001 a 2005 são mostradas na figura 5.5. Em geral, a tendência do consumo é diminuir a partir do mês de março e de aumentar a partir do mês de novembro, meses que correspondem às estações outono/inverno e primavera/verão, respectivamente. No inverno, as temperaturas são mais baixas e com isto teremos uma redução do consumo, ao contrário do verão em que as temperaturas são mais elevadas.

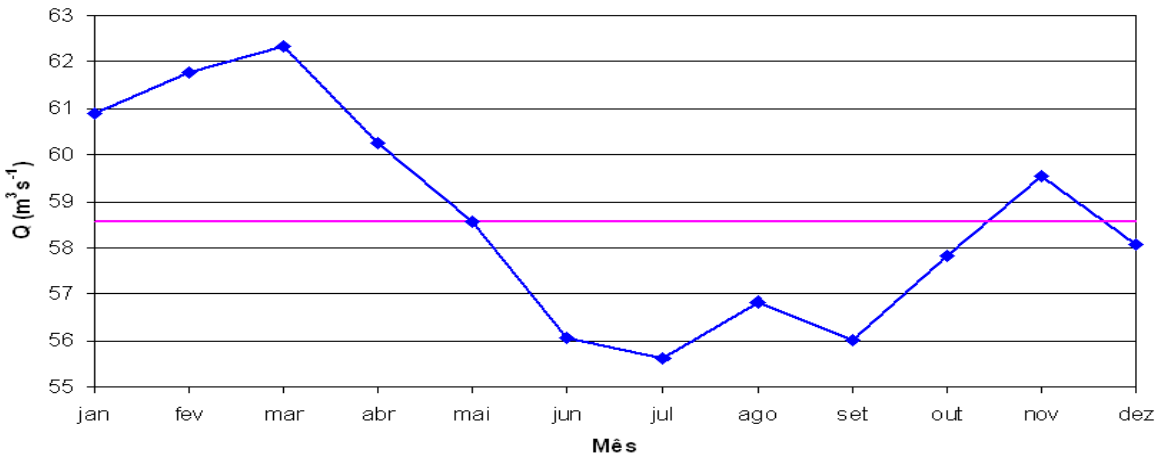
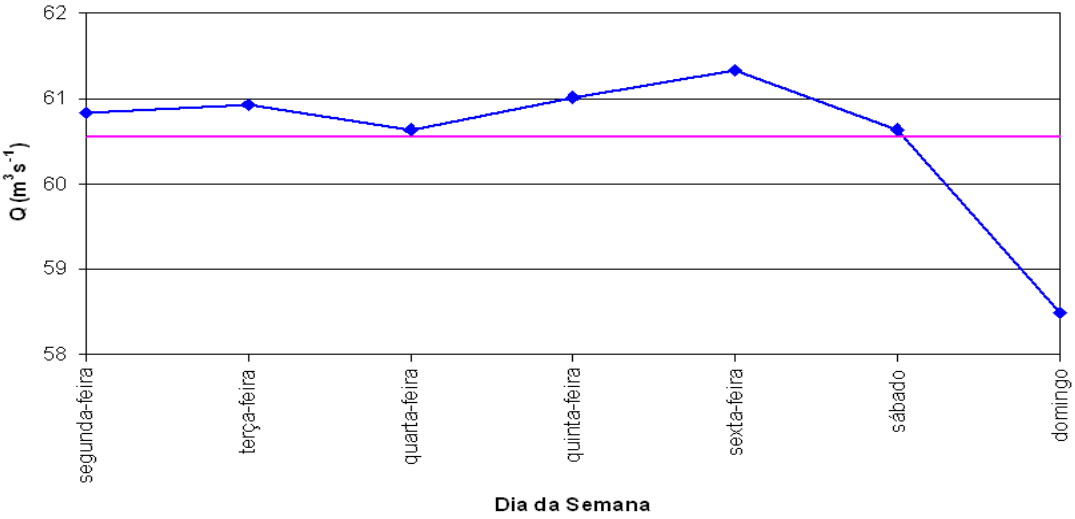


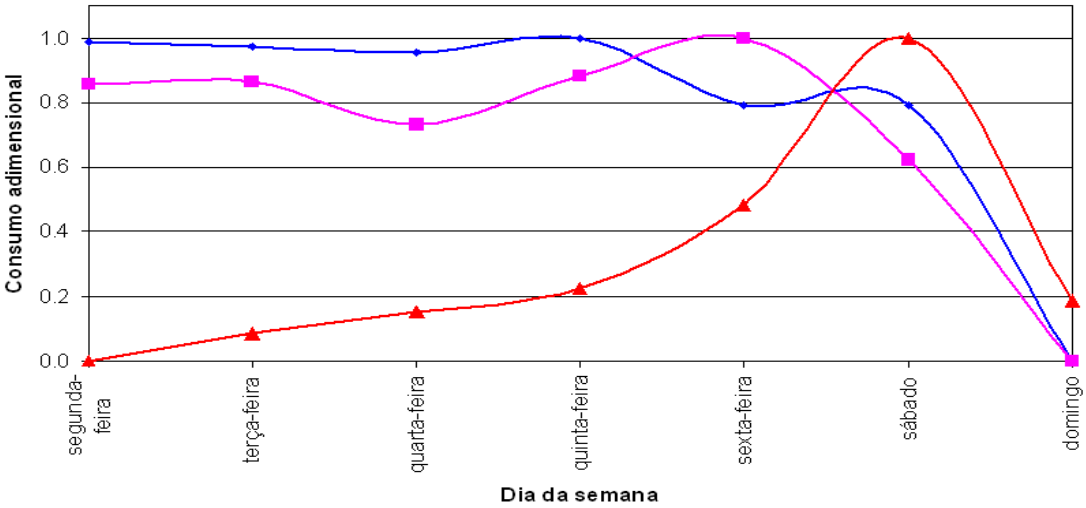
Figura 5.5: Evolução temporal da produção de água média mensal total do Sistema Produtor de Água da RMSP. A linha azul é a produção médio para cada mês e a linha rosa é consumo médio total

Observa-se que os consumos dos meses mais quentes sempre estão acima da média e os dos meses mais frios abaixo. O mês de maio tem o mesmo valor da média anual do consumo, o mês de março o maior e julho o de menor consumo. O mês de agosto tem um pico em relação aos meses de inverno, isto é consequência do tempo seco que ocorre nesse período provocando um aumento no consumo.

O comportamento do consumo nos dias da semana é variável. Então, em um dia qualquer da semana, o consumo pode ser diferente da média da região em função de parâmetros que interferem no sistema. Isso evidencia a limitação do uso de uma curva média em aplicações de operações em tempo real. Na figura 5.6a, observa-se a produção total (linha azul) e média (linha rosa) para os dias da semana. Nota-se que o único dia da semana que está abaixo da média é o domingo; já a sexta feira é o dia de maior consumo, sendo que as quartas-feiras e os sábados são dias de consumo próximos da média. Isto não é regra, porque essa variação depende de fatores inerentes na região de cada sistema. O domingo é o dia em que as atividades, tanto residencial, industrial e comercial estão reduzidas, fato este que colabora para diminuir o consumo.



(a)



(b)

Figura 5.6: Evolução temporal da produção de água médio diário. a) Consumo médio diário total do Sistema Produtor de Água da RMSP. A linha azul é o consumo médio para cada dia da semana e a linha rosa é consumo médio total. b) Consumo dos três principais ETA´s da RMSP, linha vermelha ETA Alto Tietê, linha rosa ETA Cantareira e linha azul ETA Guarapiranga.

O efeito da localidade do sistema é observado na figura 5.6b. Nota-se que a produção da ETA Guarapiranga (linha azul) não oscila muito até quinta-feira e o consumo reduz-se a partir de sexta-feira até domingo, quando atinge o menor valor. A ETA Cantareira (linha rosa) tem um comportamento diferente do Guarapiranga; seu consumo é mais ou menos constante no início da semana, ocorrendo uma queda na quarta-feira e voltando a subir na quinta-feira, quando, na sexta-feira atinge seu valor máximo e no domingo o mínimo.

Entretanto, em linhas gerais, pode-se dizer que o consumo dos dois sistemas tem comportamentos mais ou menos parecidos, o que não se pode dizer da ETA Alto Tietê (linha vermelha). Como esta ETA abastece uma região dormitório, onde a população é mais presente no período noturno durante a semana, o consumo é totalmente diferente dos outros dois sistemas. Neste caso, o maior consumo ocorre no sábado que é quando a população permanece mais tempo em suas residências. Diferentes dos outros o menor consumo ocorre na segunda-feira e sobe com o passar da semana. Comparando a figura 6a e 6b, pode-se dizer que o consumo total segue o padrão da ETA Cantareira, e isto acontece por ela ser a maior do sistema e ter uma influencia maior no cálculo do consumo total.

O consumo horário varia com o decorrer do dia, o tipo de consumidor de cada região e ocorrências que possam acontecer. Em geral o pico do consumo horário ocorre a partir das 12 horas (figura 5.7a), onde passa a ser mais ou menos constante com pequenas variações de consumo até as 17 horas. O consumo começa a diminuir em torno das 18 horas, tornando-se quase constante no período entre 21 e 24 horas. O período de 1 às 6 horas da manhã apresenta uma redução do consumo, sendo que o mínimo ocorre às 6 horas da manhã e após este período passa a aumentar novamente.

As Figs. 5.7b, c e d mostram a evolução horária do consumo para diferentes dias da semana. A ETA Alto Tietê apresenta dois picos de consumo, durante o dia, por volta das 12h e 16h tendo o maior pico aos sábados. A região abastecida pela referida ETA é denominada de cidade dormitório onde a população tende a aumentar nos fins de semana. O consumo horário da ETA Alto Tietê não tem um ciclo, diurno e noturno bem definido como nas outras duas.

O consumo da ETA´s Cantareira e Guarapiranga, figura 5.7b e 5.7c respectivamente, têm o padrão de consumo horário bem semelhante a curva do consumo total do sistema produtor. Conforme exposto anteriormente o menor consumo ocorre aos domingos nas duas ETA´s e durante os outros dias da semana, elas tem variações em seus consumos horários, mas não obedecem o mesmo comportamento, em alguns dia da semana, o consumo pode ser maior ou menor em uma ou outra ETA.

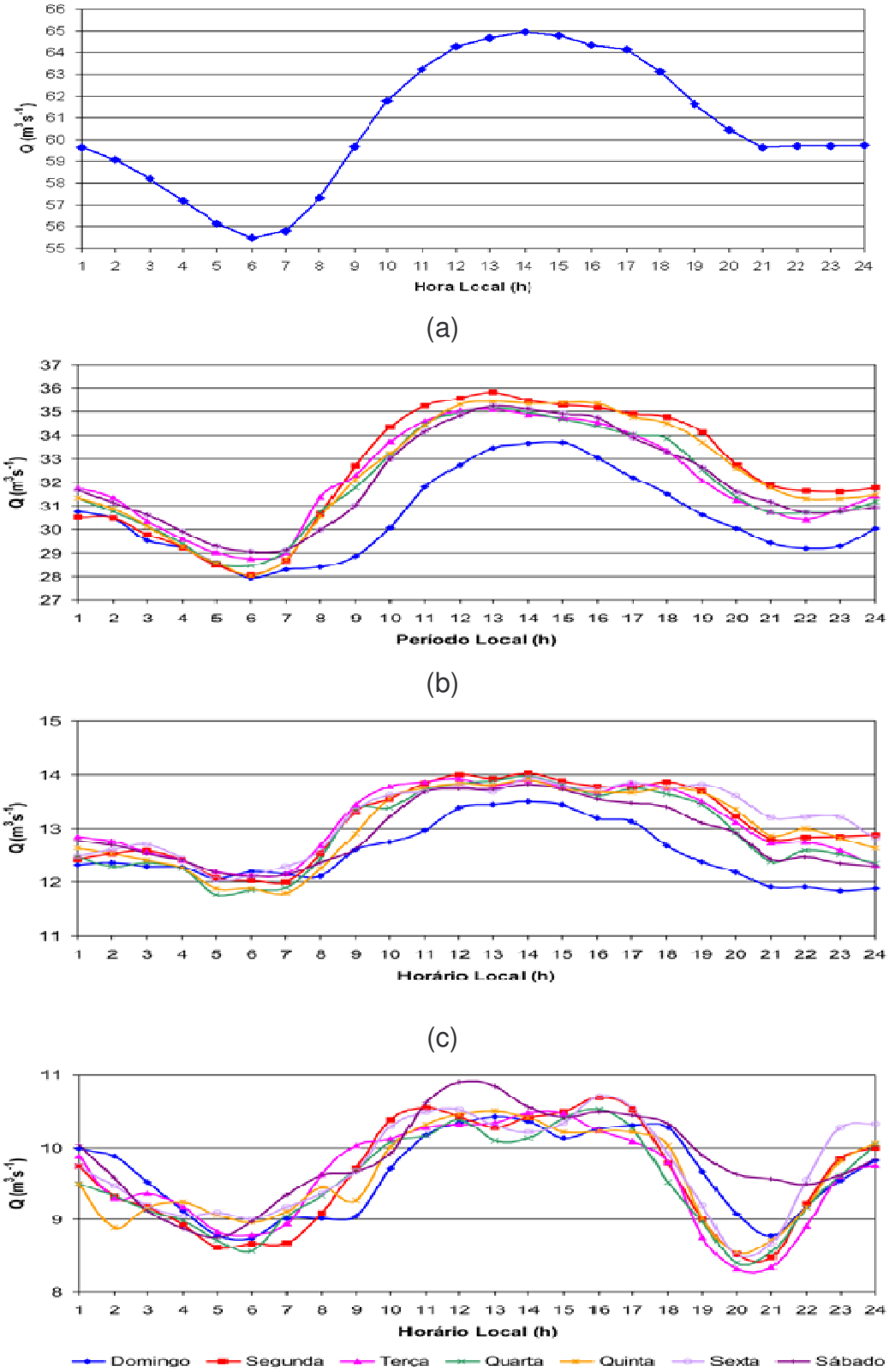


Figura 5.7: Evolução temporal da produção de água médio horário: a) Produção total do sistema Produtor. b) ETA Cantareira, c) ETA Guarapiranga e d) ETA Alto Tiete. Dias da semana estão indicados pela cor das linhas na legenda.

Os consumos nos dias de feriado e férias no ano de 2005 das ETAs Cantareira, Guarapiranga e Alto Tietê podem ser observados na Figura 5.8, indicando que o consumo de água, nos feriados, é menor do que no período de férias, na ETA Cantareira. Nas ETAs Guarapiranga e Alto Tietê, isto não ocorre, pois nota-se que, em determinados horários nos dias de feriado, o consumo é maior do que do período de férias. Sugere-se que as características populacionais na região da referida da ETA determinem este comportamento. Os horários de pico dos feriados e férias da ETA Cantareira e Guarapiranga coincidem, além de apresentarem um ciclo diuturno bem mais definido em relação a ETA Alto Tietê.

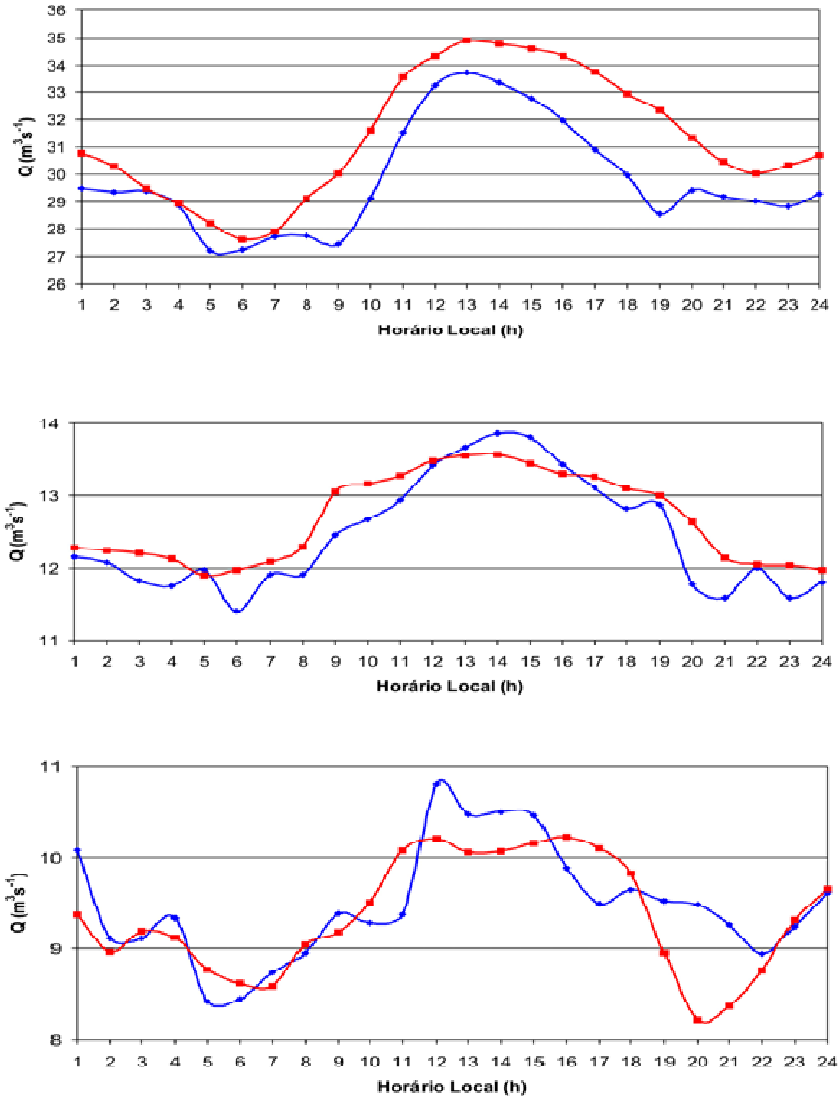


Figura 5.8: Evolução temporal da produção de água médio horário em dias de férias (linha vermelha) e feriado (linha azul) para a das três ETA, ETA Cantareira (topo), ETA Guarapiranga (meio) e ETA Alto Tiete (baixo).

5.3. Análise do consumo de água comparadas as variáveis meteorológicas

Eventos de precipitação, rajadas de vento e granizo sobre a RMSP no período de verão e os eventos intensos de baixa umidade no outono e inverno têm impacto significativo na população e com isto no consumo de água. A Figura 5.9 mostra o consumo de água total mensal (CT), a temperatura média do ar (T), umidade relativa (UR), precipitação (R) e pressão(P). O comportamento do consumo médio segue o ciclo anual da chuva/temperatura onde o máximo ocorre no início do ano (verão) e o mínimo por volta de junho/julho no inverno (período seco). Assim, o consumo máximo coincide com o período de temperaturas e umidade mais elevadas em São Paulo. Por outro lado o consumo mínimo ocorre quando a temperatura é baixa. Consumo e temperatura têm o mesmo comportamento, quando a temperatura diminui o consumo também diminui e vice versa com exceção dos meses de junho, agosto e dezembro de 2004 e abril de 2005.

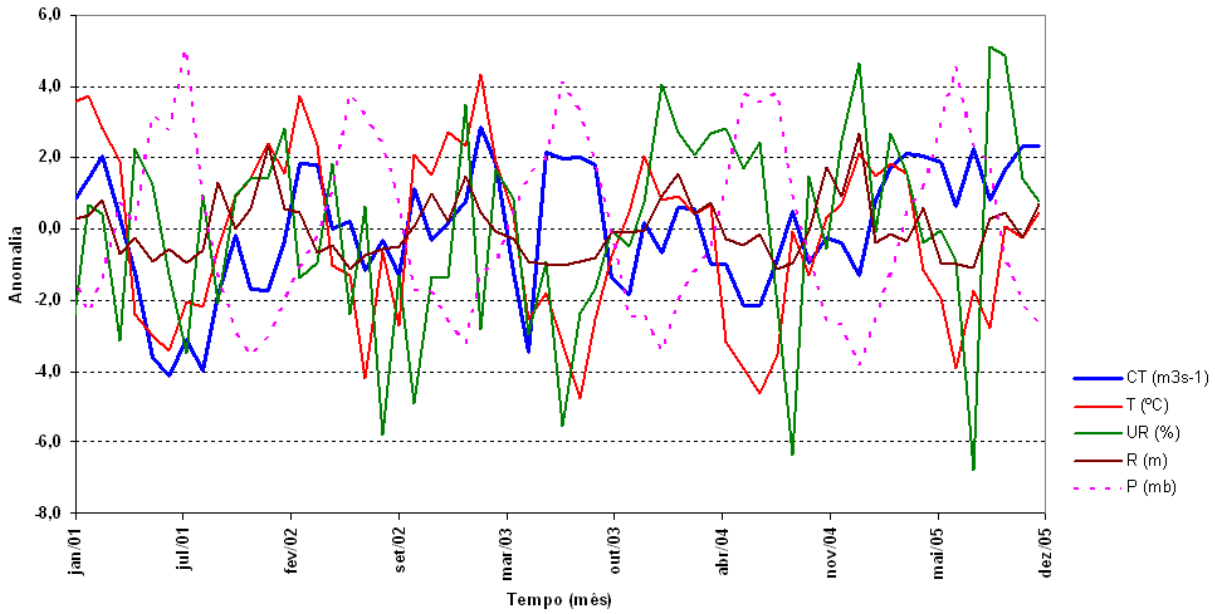


Figura 5.9. Anomalias do consumo total (CT), temperatura (T), umidade relativa (UR), precipitação (R) e pressão (P).

Quando a umidade diminui, o consumo aumenta; isto acontece porque o ar fica mais seco. Os menores consumos ocorreram geralmente em julho quando as

temperaturas são menores. O mês de maior consumo é março. Observa-se que, em 2003, ocorreu uma diminuição do consumo no mês de outubro por causa do racionamento de água. As temperaturas médias em 2004 foram anemas e o consumo também foi menor do que os anos anteriores. A umidade relativa menor está associada ao aumento de consumo de água, principalmente por razões fisiológicas. O ano de 2002 foi o mais seco da série. A umidade relativa tem as maiores variações da média, tanto para positivo como para negativo e como já foi citado anteriormente, esta variável tem grande influência no consumo de água.

Observa que a chuva e o consumo têm o mesmo sinal, ou seja, quando um aumenta o outro também aumenta. Este fato é ocasionado porque, após a chuva, a população tem que fazer limpezas, principalmente se ocorreu enchentes. Quando a pressão atmosférica diminui, o consumo aumenta, a exemplo do que ocorre com a temperatura e consumo.

Nota-se que o ano de 2002 foi menos chuvoso (seco), quando ocorreu racionamento de água. Observa-se que chuvas fracas e de longa duração tendem a diminuir o consumo; por outro lado, chuvas intensas elevam o consumo por causa da limpeza das áreas afetadas. Por outro lado, só após eventos de chuva prolongada, o consumo tende a aumentar em virtude da limpeza doméstica, particularmente de vestuário.

A figura 5.10 mostra a evolução mensal do consumo total médio e total da precipitação, temperatura e umidade relativa média. Nota-se um aumento de consumo no mês de agosto devido a baixa umidade relativa. Nota-se que o consumo aumenta ou diminui com o aumento ou diminuição da temperatura, exceto nos meses de janeiro e agosto em razão das férias e baixa umidade do ar, respectivamente. Por último, o consumo de água mensal tende a diminuir com o aumento da umidade relativa, precipitação e diminuição da temperatura.

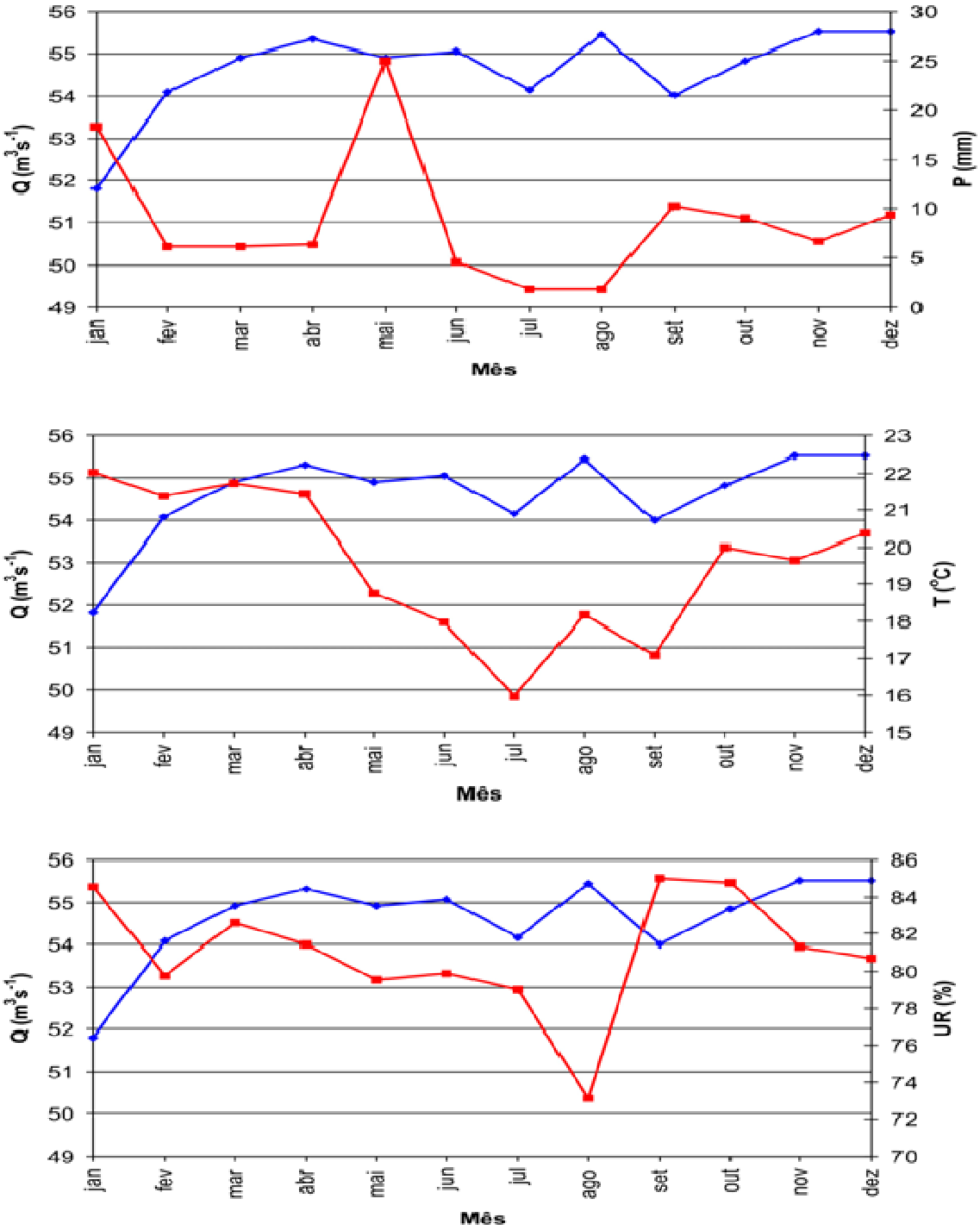


Figura 5.10: Evolução temporal do consumo de água dos três sistemas adutores e do total de precipitação (topo), da temperatura (meio) e da umidade relativa (baixo). A linha azul se refere ao consumo total em todos os gráficos. Eixo das ordenadas da direita corresponde a variável meteorológica e, o da esquerda, ao consumo total.

O efeito da temperatura média no consumo médio de água na região das três ETAs mais importantes (Cantareira, Guarapiranga e Alto Tietê), entre 06 e 10 de janeiro (período quente), e 06 e10 de julho (período frio) de 2005 é mostrado na figura 5.11.

O consumo diminui em dias frios e aumenta em dias quentes. Nota-se que o pico de consumo em dias quentes ocorreu por volta das 13 h e nos dias frios, por volta das 16h. A diferença é mais acentuada no período noturno. A ETA Guarapiranga também não apresenta um ciclo diuturno com grande amplitude. Notamos que a grande diferença entre estes dias encontra-se na ETA Alto Tietê.

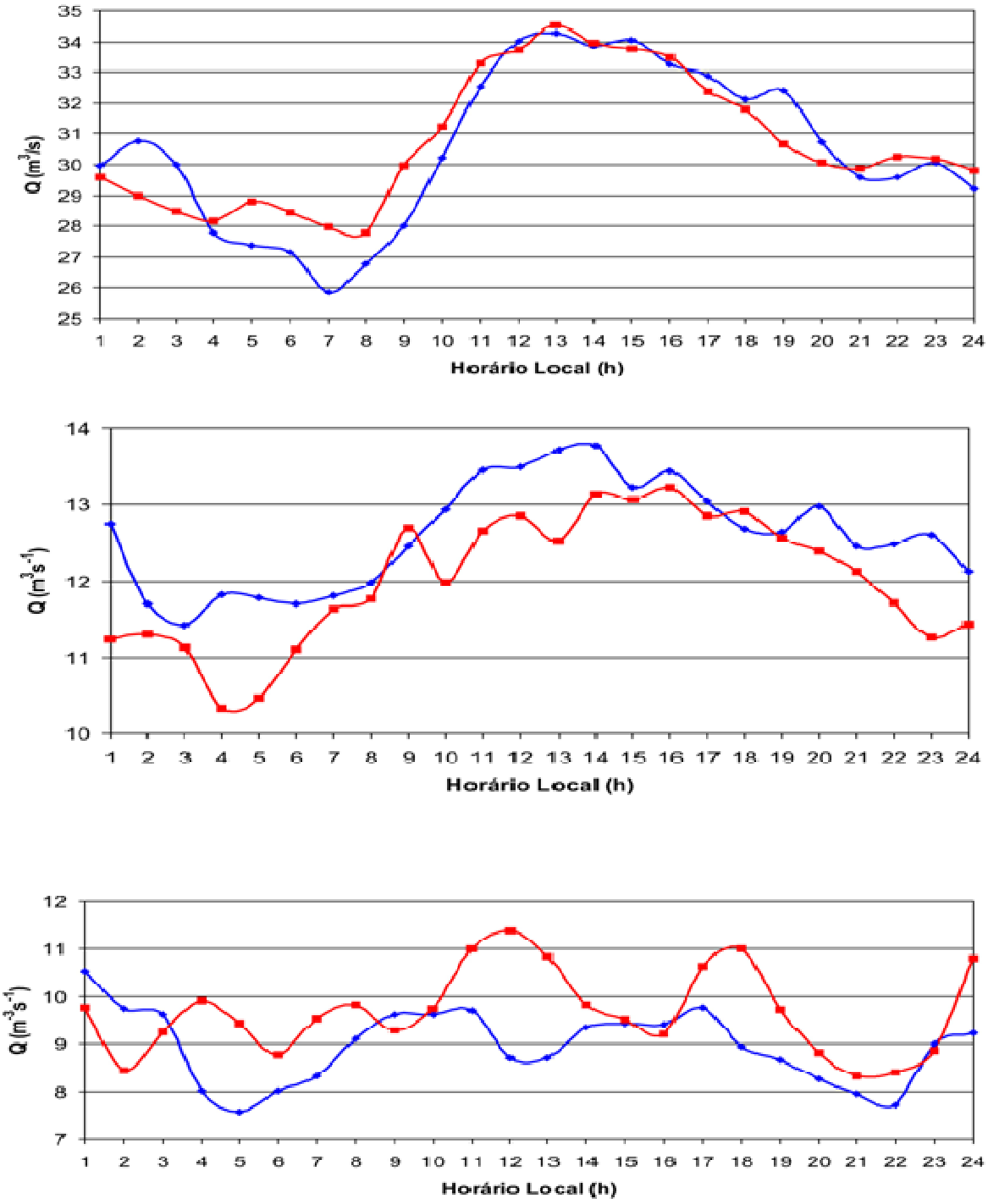


Figura 5.11: Evolução temporal do consumo de água médio horário para 5 dias quentes (linha vermelha) e 5 dias frios (linha azul) na região das ETA Cantareira (topo), ETA Guarapiranga (meio) e ETA Alto Tiete (baixo).

5.4. Análise de Correlação

A relacão linear entre duas variáveis aleatórias são muito importantes para uma boa previsão. A correlação refere-se à medida da relação entre duas variáveis, embora correlação não implique causalidade. O grau de correlação existente entre 11 variáveis do setor Itaim Paulista e ETA Cantareira estão nas tabelas 5.2 e 5.3 respectivamente. Os valores em azul são aqueles superiores ou iguais a $|0{,}10|$.

As correlações baixas indicam que nenhuma variável é dominante e que o conjunto das mesmas é que resulta no consumo. Dentro da coluna cinza, na tabela 5.2, são apresentadas as correlações existentes entre a variável principal, que é o consumo de água (Q), e as variáveis de interesse. Nota-se que a hora (H), estação do ano (EA), tipo do dia (TD), temperatura (T), umidade (UR), direção (Dir) e velocidade do vento (Vel) são variáveis que apresentam correlação com o consumo.

A temperatura e umidade em relação ao consumo (Q) têm a correlação maior e com sinal inverso, 0,60 e -0,67 respectivamente. Isto indica que a temperatura e o consumo crescem ou decrescem no mesmo sentido, e a umidade em sentido contrário. O fato das correlações serem baixas sugere que o uso da RNA pode ser uma ferramenta que consiga fazer uma previsão de consumo melhor do que com se fosse usada uma regressão simples, isto, porque a RNA tentará generalizar e aprender com os valores que serão fornecidos.

Na tabela 5.3, observa-se a correlação para ETA Cantareira, a exemplo da tabela 5.2, a coluna cinza corresponde a correlação existentes entre o consumo de água (Q) e as outras variáveis. Novamente a hora (H), estação do ano (EA), tipo do dia (TD), temperatura (T), umidade (UR), direção (Dir), velocidade do vento (Vel) e também o tipo do dia (TD) são as variáveis que melhor apresentam correlação com o consumo. A temperatura e umidade seguem tendo a melhor correlação com o consumo, 0,49 e -0,54 respectivamente.

Tabela 5.2: Correlação de variáveis meteorológicas, temperatura (T), umidade (UR), chuva (R), pressão (P), direção do vento (Dir) e velocidade do vento (Vel), consumo de água (Q), hora (H), dia da semana (DS), estação do ano (EA) e tipo do dia (TD) do setor Itaim Paulista para identificar as melhores variáveis para uma previsão de consumo de água. Os valores em azul correspondem a valores maiores 0,10 e a coluna cinza indica as correlações com a variável de interesse o consumo (Q).

VARIÁVEL	H	DS	EA	TD	T	UR	R	P	Dir	Vel	Q
Hora (H)	1	0.00	0.00	0.00	**0.16**	**-0.14**	0.00	0.01	**0.13**	0.06	**0.49**
Dia Semana (DS)	0.00	1	0.00	**0.65**	0.03	-0.01	0.00	-0.01	0.01	-0.01	0.04
Estação do Ano (EA)	0.00	0.00	1	0.00	**-0.21**	0.05	0.02	**0.14**	0.02	**0.10**	-0.04
Tipo do Dia (TD)	0.00	**0.65**	0.00	1	0.00	0.00	-0.01	0.00	0.04	0.03	0.04
Temperatura (T)	**0.16**	0.03	**-0.21**	0.00	1	**-0.73**	-0.01	**-0.48**	**0.36**	**0.31**	**0.60**
Umidade (UR)	**-0.14**	-0.01	0.05	0.00	**-0.73**	1	**0.10**	0.09	**-0.34**	**-0.39**	**-0.67**
Chuva (R)	0.00	0.00	0.02	-0.01	-0.01	**0.10**	1	**-0.11**	0.08	0.02	-0.05
Pressão (P)	0.01	-0.01	**0.14**	0.00	**-0.48**	0.09	**-0.11**	1	**-0.24**	-0.08	-0.01
Direção (Dir)	**0.13**	0.01	0.02	0.04	**0.36**	**-0.34**	0.08	**-0.24**	1	**0.23**	**0.27**
Velocidade (Vel)	0.06	-0.01	**0.10**	0.03	**0.31**	**-0.39**	0.02	-0.08	**0.23**	1	**0.40**
Consumo (Q)	**0.49**	0.04	-0.04	0.04	**0.60**	**-0.67**	-0.05	-0.01	**0.27**	**0.40**	1

Tabela 5.3: Correlação de variáveis meteorológicas, temperatura (T), umidade (UR), chuva (R), pressão (P), direção do vento (Dir) e velocidade do vento (Vel), consumo de água (Q), hora (H), dia da semana (DS), estação do ano (EA) e tipo do dia (TD) para ETA Cantareira para identificar as melhores variáveis para uma previsão de consumo de água. Os valores em azul correspondem a valores maiores 0,10 e a coluna cinza indica as correlações com a variável de interesse o consumo (Q).

VARIÁVEL	H	DS	EA	TD	T	UR	R	P	Dir	VEL	Q
HORA (H)	1	0.00	0.00	0.00	**0.15**	**-0.14**	0.00	0.01	**0.12**	0.06	**0.28**
Dia semana (DS)	0.00	1	-0.01	**0.64**	0.05	-0.03	-0.01	0.00	0.00	-0.01	**-0.16**
Estação do Ano (EA)	0.00	-0.01	1	-0.01	**-0.26**	0.02	-0.02	**0.22**	-0.01	0.08	0.06
Tipo do Dia (TD)	0.00	**0.64**	-0.01	1	0.02	-0.01	-0.01	-0.01	0.03	0.04	**-0.24**
Temperatura (T)	**0.15**	0.05	**-0.26**	0.02	1	**-0.71**	0.00	**-0.50**	**0.37**	**0.31**	**0.49**
Umidade (UR)	**-0.14**	-0.03	0.02	-0.01	**-0.71**	1	**0.10**	0.07	**-0.34**	**-0.39**	**-0.54**
Chuva (R)	0.00	-0.01	-0.02	-0.01	0.00	**0.10**	1	**-0.11**	0.07	0.02	-0.02
Pressão (P)	0.01	0.00	**0.22**	-0.01	**-0.50**	0.07	**-0.11**	1	**-0.25**	-0.08	-0.05
Direção (Dir)	**0.12**	0.00	-0.01	0.03	**0.37**	**-0.34**	0.07	**-0.25**	1	**0.25**	**0.21**
Velocidade (Vel)	0.06	-0.01	0.08	0.04	**0.31**	**-0.39**	0.02	-0.08	**0.25**	1	**0.25**
Consumo (Q)	**0.28**	**-0.16**	0.06	**-0.24**	**0.49**	**-0.54**	-0.02	-0.05	**0.21**	**0.25**	1

A figura 5.12 apresenta as variações das correlações, da ETA Cantareira e do setor Itaim Paulista, em relação às variáveis meteorológicas e antrópicas. Observa-se que os valores das correlações do setor Itaim Paulista em relação a ETA Cantareira são maiores com exceção as variáveis, estação do ano e chuva. Isto pode estar relacionado com os valores médios do consumo e ao tamanho do setor abastecido. A ETA Cantareira é um setor produtor. Ela gera e distribuir água para vários setores que tem características bem diferentes umas das outras, isto pode indicar o porquê de seus valores serem menor do que do setor Itaim Paulista que é um sistema

pequeno. O setor Itaim Paulista é considerado um setor residencial com características bem definidas.

A umidade e a temperatura, tanto para o Itaim Paulista como para ETA Cantareira, mostram os maiores valores de correlação se comparadas às outras variáveis. Essas variáveis têm comportamentos diferentes, sendo que a correlação da temperatura é sempre positiva e da umidade negativa.

Os valores observados na correlação identificam as seguintes classes de correlação:
c=±1: perfeita;
±1< c ≥±0,7: forte;
±0,7<c ≥±0,3: moderada;
±0,3<c ≥0: fraca e
c=0: nula.

Nota-se na figura 5.12 que as correlações ficaram entre moderada e fraca. Isto pode indicar que se pode ter uma relação não linear entre as variáveis, levando a acreditar novamente que a RNA seja uma boa solução para a previsão do consumo, devido a não linearidade das variáveis com o consumo.

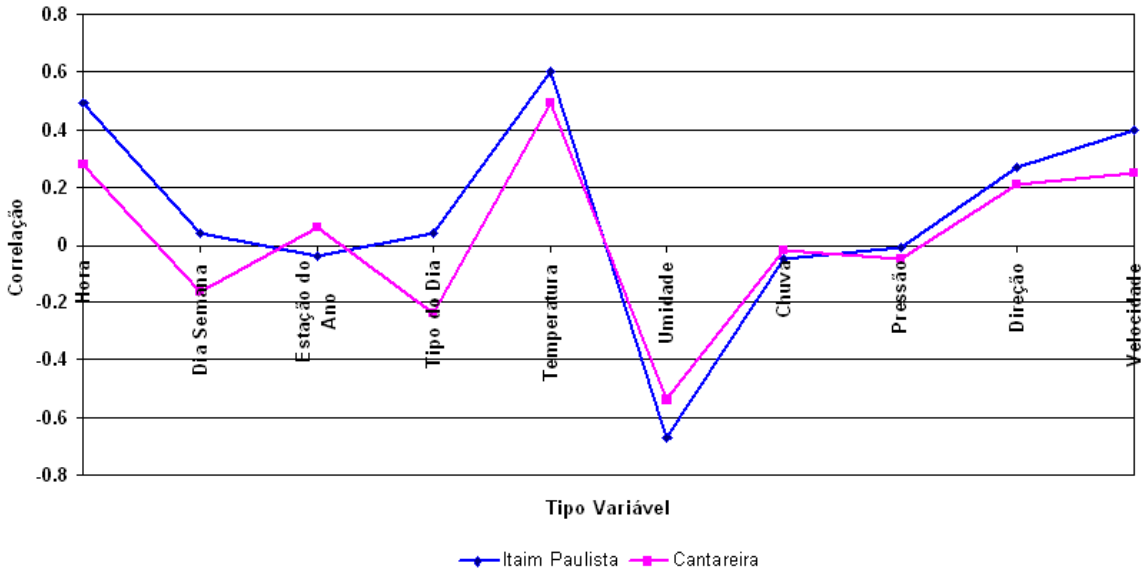


Figura 5.12: Variações das correlações, da ETA Cantareira e do setor Itaim Paulista, em relação às variáveis meteorológicas e antrópicas.

5.5. Análise utilizando Rede Neural Artificial

A RNA utilizada, neste trabalho, foi uma rede multicamadas percepton (MLP) e o software utilizado foi o LLSSIM. A RNA foi treinada com os dados de consumo da ETA Cantareira e do setor Itaim Paulista, além do dia da semana, tipo do dia, hora, estação do ano, temperatura, umidade, chuva, pressão, velocidade e direção do vento. As variáveis foram normalizadas entre 0 – 1 para que ficassem com a mesma escala de valor. O conjunto de dados foi divido em três partes: 70% para o treinamento, 15% para o teste e os outros 15% para validação. O treinamento, o teste e a previsão foram executados considerando diversas topologias.

5.5.1. ETA Cantareira

A ETA Cantareira foi escolhida por ser a mais importante do sistema produtor de água da Sabesp, abastecendo o maior número de população. Para gerar os dados para a RNA foram utilizados dados do ano de 2005. Para verificar o desempenho dos modelos de RNA, descritos na tabela 4.3, foram calculados o erro médio (EM), o erro médio absoluto (MAE), o erro médio quadrático (RMSE) e o coeficiente de correlação (r), que podem ser vistos nas tabelas 5.4, 5.5 e 5.6, respectivamente treinamento, teste e previsão.

Na tabela 5.4, observa-se que o melhor resultado para treinamento foi para o modelo RNA-C-8 (coluna cinza), onde RNA é Rede Neural Artificial, o C é Cantareira e 8 é o número do modelo. Esse modelo utilizou a média do consumo de 12 horas, e tem a configuração de 18-1-1, ou seja, tem 18 entradas, 1 camada escondida e 1 camada de saída. Os piores valores ocorreram para o modelo RNA-C-1 (coluna vermelha), o modelo utilizou dados horários e a configuração foi 10-1-1. Os modelos RNA-C-1 e 2 não utilizaram o consumo como entrada da RNA. Os demais modelos utilizar o consumo e as variáveis meteorológicas como entrada nos tempos (t) e (t-1).

Tabela 5.4: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para o treinamento dos modelos de RNA´s do sistema produtor da Cantareira. A coluna vermelha são os piores valores e a coluna azul os melhores valores.

	RNA-C-1	RNA-C-2	RNA-C-3	RNA-C-4	RNA-C-5	RNA-C-6	RNA-C-7	RNA-C-8	RNA-C-9
Erro Médio Absoluto (MAE)	1,565	1,308	1,308	1,344	1,127	1,105	1,147	0,767	1,117
Erro Médio (EM)	0,263	-0,177	-0,177	2,953	2,175	2,109	2,260	0,152	0,741
Erro Médio Quadrático (RMSE)	2,050	1,721	1,721	1,719	1,475	1,452	1,503	1,034	1,402
Coeficiente de Correlação (r)	0,694	0,763	0,763	0,698	0,787	0,793	0,775	0,824	0,790
Coeficiente de Determinação (R^2)	0,481	0,583	0,583	0,487	0,620	0,629	0,600	0,679	0,624

Para o teste foram utilizados somente os modelos RNA-C-1, 2 e 3. Observa-se (tabela 5.5) que os piores valores ocorreram para o modelo RNA-C-1 (vermelho), já os modelos RNA-C-2 e 3 (azul) tiveram valores praticamente iguais, obtendo os melhores resultados.

Tabela 5.5: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para o teste dos modelos de RNA´s do sistema produtor da Cantareira. A coluna vermelha apresenta os piores valores e a coluna azul os melhores resultados.

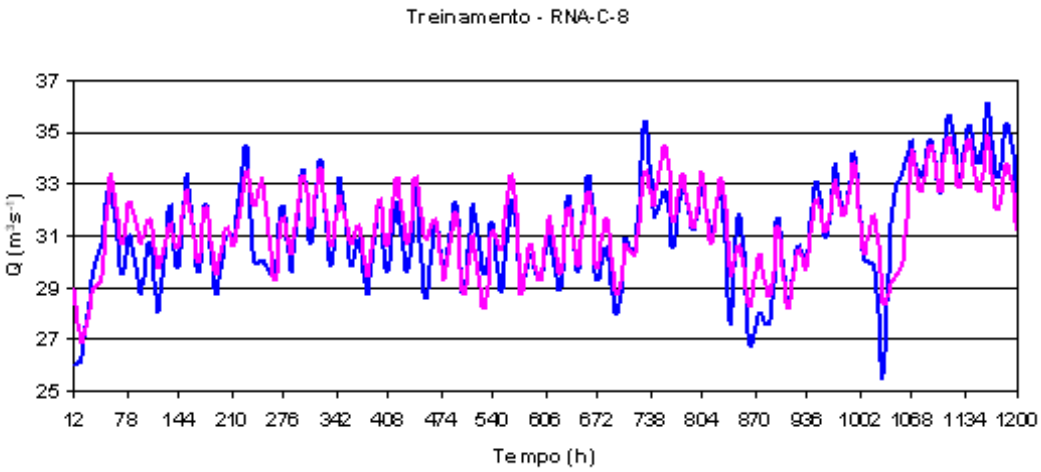
	RNA-C-1	RNA-C-2	RNA-C-3
Erro médio Absoluto (MAE)	1,824	1,490	1,489
Erro Médio (EM)	1,009	-0,728	-0,726
Erro médio quadrático (RMSE)	2,392	1,990	1,990
Coeficiente de Correlação (r)	0,674	0,707	0,707
Coeficiente de Determinação (R^2)	0,454	0,500	0,500

Assim como para o treinamento, na previsão (tabela 5.6) os piores e os melhores valores ocorreram respectivamente para RNA-C-1 (vermelho) e RNA-C-8 (azul). Os valores EM, MAE, RMSE e r são menores para o treinamento de todos os modelos, piorando para teste e previsão dos modelos. A previsão tem os maiores valores, e isto ocorre porque na previsão os valores do consumo não são fornecidos como saída da RNA, obrigando a RNA a calcular os valores sem ter referência para isto.

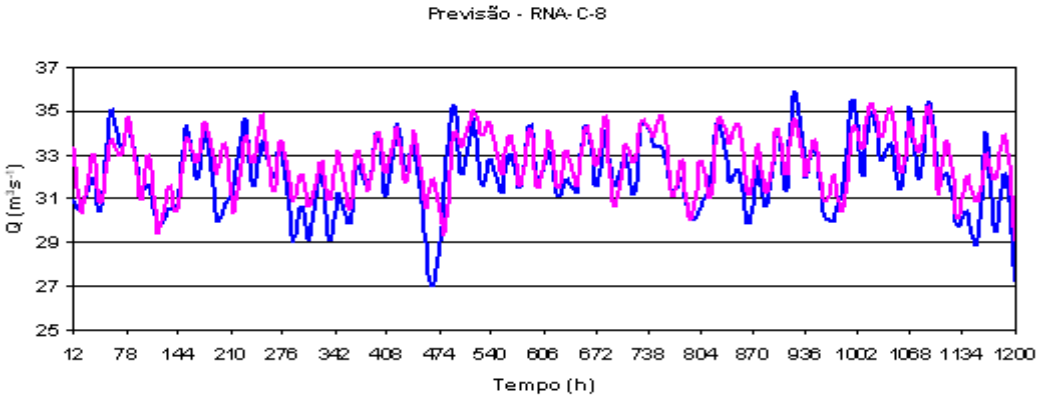
Tabela 5.6: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para a previsão dos modelos de RNA´s do sistema produtor da Cantareira. A coluna vermelha são os piores valores e a coluna azul os melhores valores.

	RNA-C-1	RNA-C-2	RNA-C-3	RNA-C-4	RNA-C-5	RNA-C-6	RNA-C-7	RNA-C-8	RNA-C-9
Erro médio Absoluto (MAE)	2,434	1,801	1,801	1,635	1,524	1,577	1,619	1,071	1,117
Erro Médio (EM)	1,725	-1,185	-1,187	0,906	0,663	0,936	1,070	0,683	0,741
Erro médio quadrático (RMSE)	3,171	2,368	2,367	0,226	0,161	1,994	2,084	1,360	1,402
Coeficiente de Correlação (r)	0,562	0,694	0,695	0,661	0,716	0,733	0,724	0,792	0,790
Coeficiente de Determinação (R^2)	0,316	0,482	0,483	0,437	0,513	0,538	0,524	0,628	0,624

A figura 5.13 apresenta a evolução temporal do consumo observado (linha azul) e consumo obtido pelo modelo RNA-C-8 (linha rosa). A evolução obtida para o treinamento é observada na figura 5.13a, e da previsão na figura 5.13b. Observa-se que alguns valores tendem a ser subestimados e outros superestimados. A previsão superestima mais do que o treinamento. Na média, os valores obtidos no treinamento se aproximam mais do valor observado do que os valores da previsão. Nota-se que os valores menores tendem a ser mais superestimados do que os valores maiores, principalmente na previsão. Uma justificativa, para esse fato pode ser que os valores observados estejam incorretos.



(a)

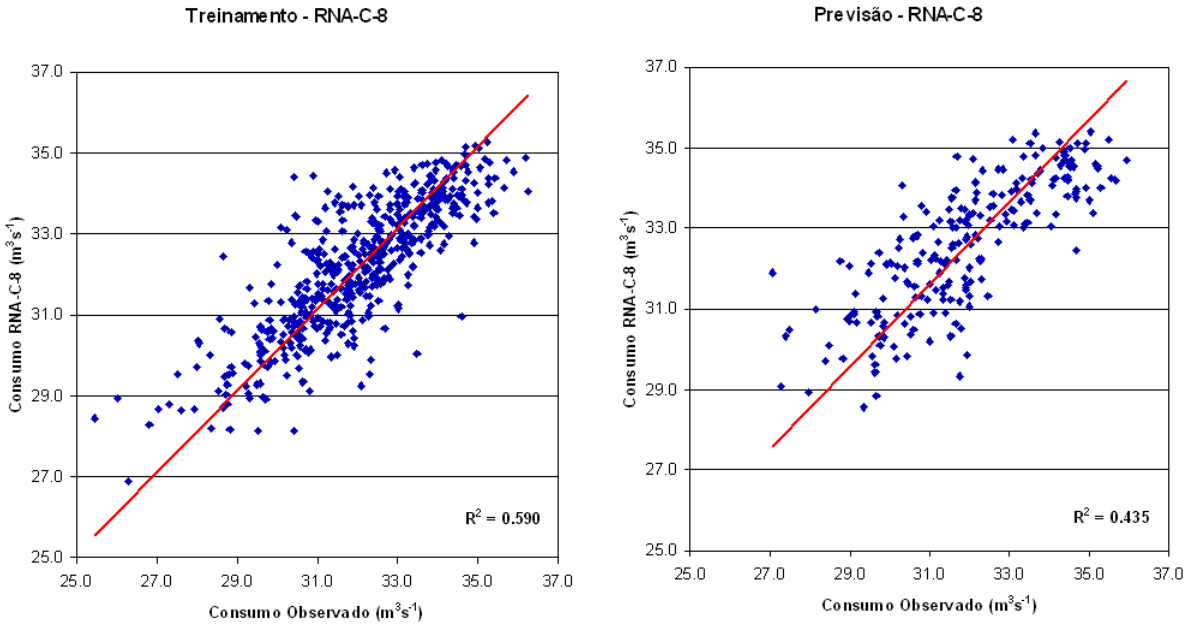


(b)

Figura 5.13: Evolução temporal do consumo de água para o modelo RNA-C-8, onde a linha azul é o consumo observado, e a linha rosa é o consumo previsto pelo modelo. a) treinamento, b) previsão

Os gráficos de espalhamento para o modelo RNA-C-8 estão representados na figura 5.13, a linha vermelha corresponde à linha de tendência. Observa-se que o

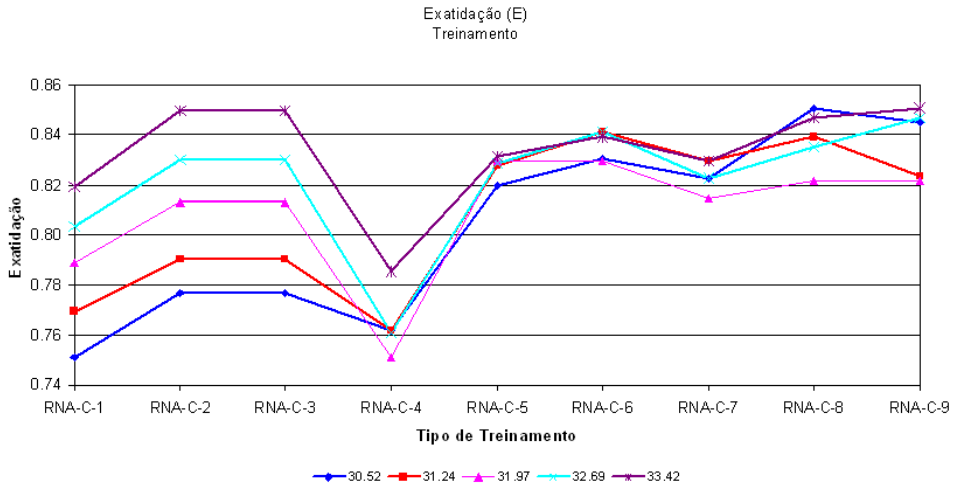
coeficiente de correlação R^2 do treinamento (0,590) é maior que o da previsão (0,435). Esses valores indicam que o treinamento tem uma correlação maior com os dados observados do que os obtidos para a previsão. A linha de tendência indica que os valores, tanto para o treinamento como para previsão, tem uma relação positiva, mas imperfeita entre os valores observados e obtidos pelo modelo. E indicam uma correlação moderada. O espalhamento do treinamento tem um comportamento parecido com o da previsão, porém a concentração de pontos está na faixa de 31 a 34 m^3s^{-1} o que não ocorre na previsão.



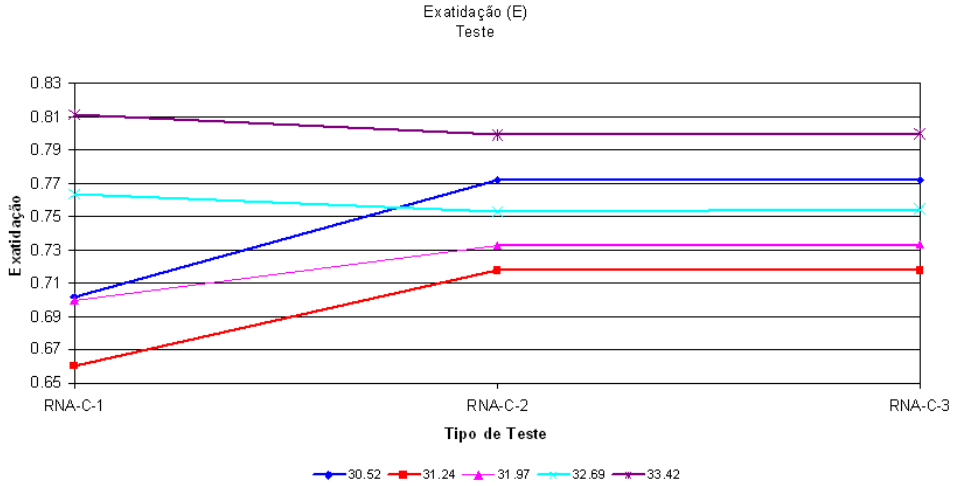
(a) (b)

Figura 5.13: Digrama de espalhamento da série temporal de consumo de água estimado pelo treinamento (a) e previsão (b). Estão indicados o ajuste linear (linha vermelha) e o coeficiente de correlação (R^2).

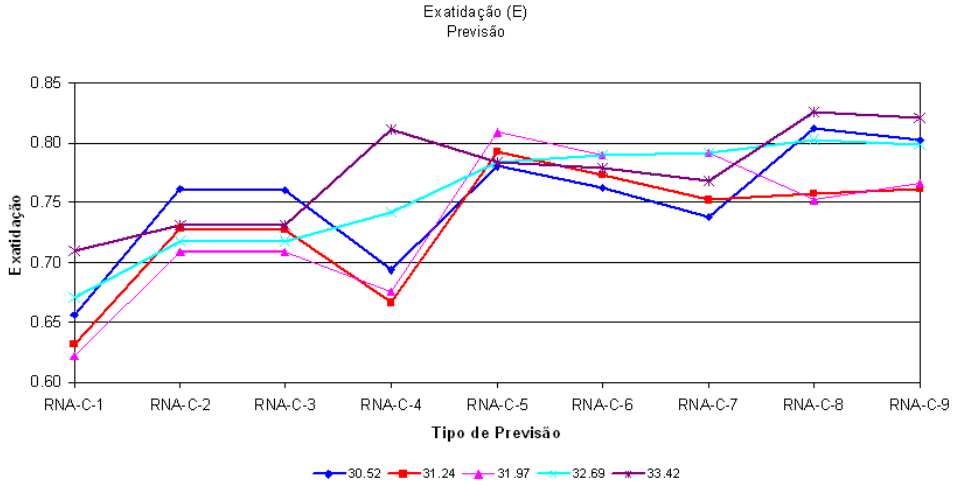
Para verificar a exatidão (E), a freqüência de bias (BIAS), probabilidade de detecção (POD), taxa de falso alarme (FAR), índice crítico de sucesso (CSI) e probabilidade de falsa detecção (POFD), foram criados limites de consumo. Esses limites foram estabelecidos utilizando a média, e a média ± 0,25 e ± 0,5 do desvio padrão. A exatidão é mostrada nas figuras 5.14 a, b e c que correspondem respectivamente ao treinamento, teste e previsão.



(a)



(b)



(c)

Figura 5.14: Índices de desempenho Exatidão dos modelos RNA para os limites de consumo de 30,53, 31,24, 31,97, 32,69 e 33,42 m^3s^{-1} indicados pelas curvas azuis, vermelhas, rosas, verdes e roxas respectivamente. (a) treinamento, (b) teste e (c) previsão.

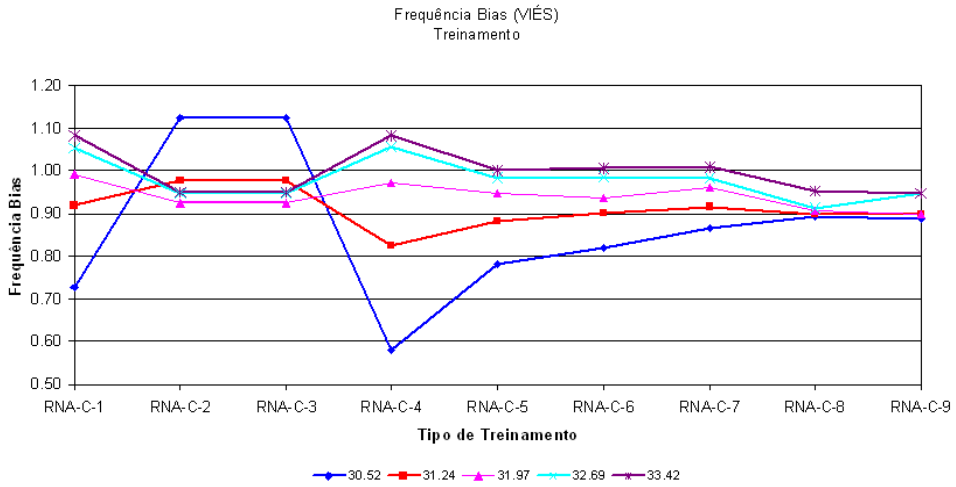
No treinamento (figura 5.14a), observa-se que os menores valores de exatidão ocorreram para o modelo RNA-C-1 e 4. O modelo RNA-C-4 obteve os menores valores de exatidão para todos os limites utilizados, quando comparados com os outros modelos. Sendo que, o menor valor para esse modelo ocorreu para o limite. O menor valor do modelo RNA-C-1 ocorreu para limite de 30,52 m^3s^{-1}. Para esses limites, pode-se dizer que 75% dos valores foram previstos corretamente. Os maiores valores de exatidão ocorreram para as faixas de 30,52 m^3s^{-1} e 33,42 m^3s^{-1}, respectivamente para os modelos RNA-C-8 e 9, onde 85% dos valores foram previstos corretamente.

Observa-se que o menor valor de exatidão para o teste (figura 5.14b) foi do modelo RNA-C-1 para o limite de 31,24 m^3s^{-1}, e o maior para o limite 33,42 m^3s^{-1}, também do modelo RNA-C-1. Onde, foram previstos corretamente 66% e 81% dos valores respectivamente. Na previsão (figura 5.14c), o menor valor ocorreu para o limite 31,97 m^3s^{-1} do modelo RNA-C-1 e 62% dos valores foram previstos corretamente. Os modelos RNA-C-8 e 9 para o limite de 33,42 m^3s^{-1} obteve o maior valor de exatidão, onde 83% dos valores foram previstos corretamente. A gama de valores previstos corretamente ficou entre 75% a 85% para o treinamento, 70% a 81% para o teste e 62% a 83% da previsão. A tabela 5.7 mostra os modelos da Cantareira que tiveram os maiores e menores valores de exatidão.

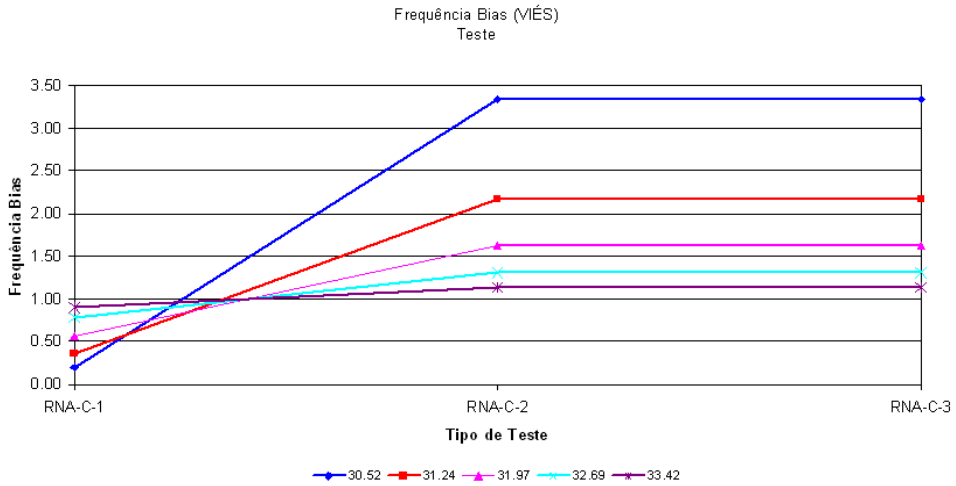
Tabela 5.7: Modelos de RNA´s que tiveram os menores e maiores valores de Exatidão (E), e a faixa de porcentagem de valores previsto corretamente , para o treinamento, teste e previsão, do sistema produtor da Cantareira.

	Menores	Maiores	Previsto Corretamente
Treinamento	RNA-C-1 e 4	RNA-C-8 e 9	75% a 85%
Teste	RNA-C-1	RNA-C-1	66% a 81%
Previsão	RNA-C-1	RNA-C-8	62% a 83%

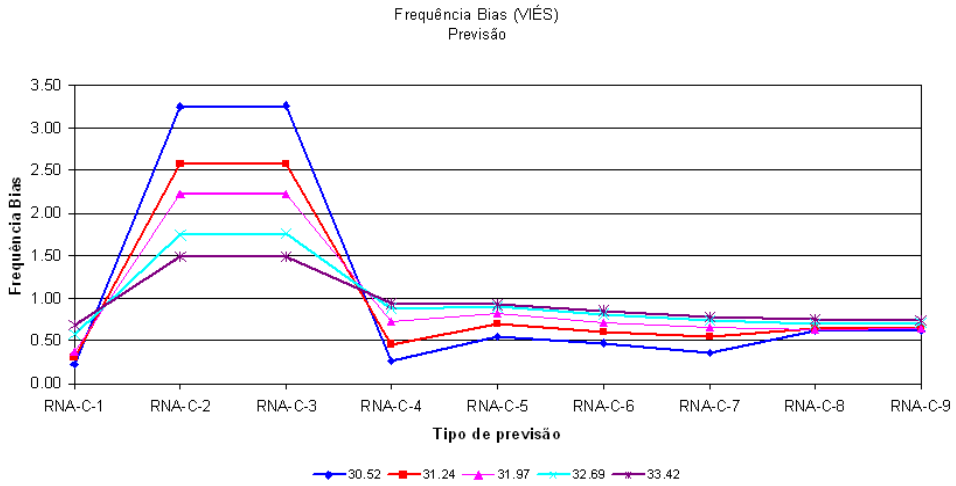
A frequência de bias (Viés) mede a tendência do sistema de previsão em superestimar e subestimar os valores previstos. Os valores de Viés>1 indicam que o sistema superestima os valores previstos, Viés<1 subestima e Viés=1 indica valores perfeitos. A figura 5.15 mostra os valores da frequência de bias (Viés) obtidos para os limites dos modelos de RNA do sistema Cantareira.



(a)



(b)



(c)

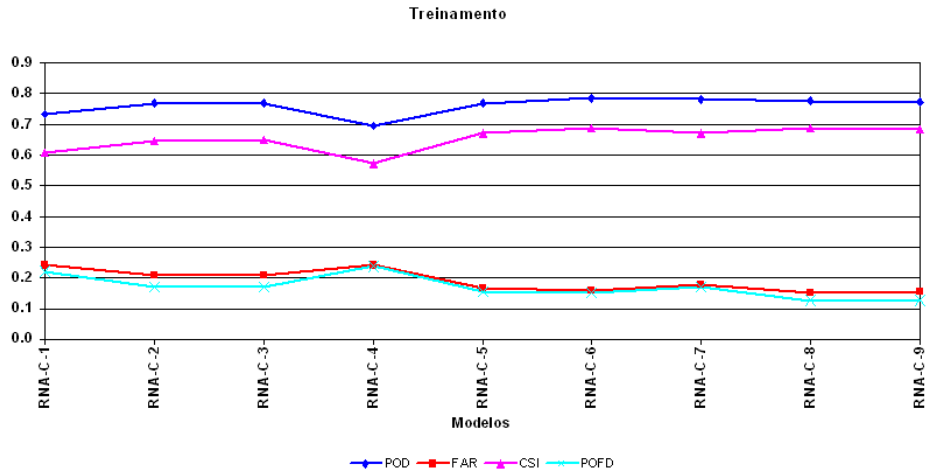
Figura 5.15: Valores da frequência de bias (viés) dos modelos RNA para os limites de consumo de 30,53, 31,24, 31,97, 32,69 e 33,42m^3s^{-1} indicados pelas curvas azuis, vermelhas, rosas, verdes e roxas respectivamente. (a) treinamento, (b) teste e (c) previsão.

O valor perfeito ocorreu para o modelo RNA-C-5 no valor limite de 33,42m^3s^{-1} para o treinamento (figura 5.15a). Para o teste e previsão não se obteve valor perfeito. No treinamento, os valores foram superestimados para os limites 30,52m^3s^{-1}, modelos RNA-C-2 e 3, 32,69m^3s^{-1}, modelos RNA-C-1 e 4, e 33,42 m^3s^{-1}, modelos RNA-C-1,4,6 e 7. Os valores abaixo de 1 estão entre a faixa de 0,58 a 0,99. Indicando que os valores de alguns limites e modelos podem ter sido subestimados ou ficado próximos dos valores observados. No teste (figura 5.15b), os modelos RNA-C- 2 e 3 superestimaram todos os limites e o modelo RNA-C-1 subestimou. A figura 5.15c mostra que novamente os modelos RNA-C-2 e 3, para a previsão, superestimaram os valores de todos os limites. E os modelos RNA-C-1, e RNA-C-4 a 9 subestimaram todos, sendo que, os menores valores ocorreram para os modelos RNA-C-1, 4 e 7. Os valores subestimados ficaram na faixa entre os valores 0,22 a 0,95, mostrando que alguns valores foram subestimados, mas ficaram bem próximos dos valores observados. O resumo dos modelos onde a frequência de bias (viés) subestimou e superestimou encontra-se na tabela 5.8.

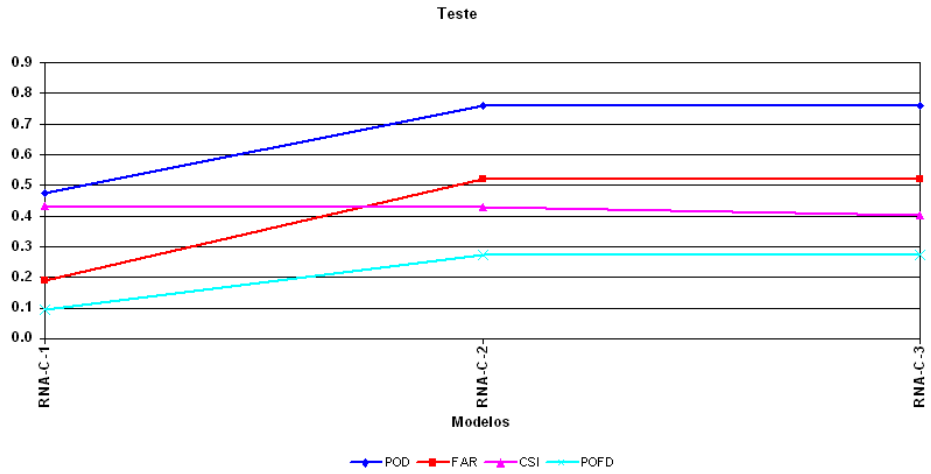
Tabela 5.8: Modelos de RNA´s que tiveram valores subestimados ou superestimados devido a avaliação da freqüência de bias (viés), para o treinamento, teste e previsão, do sistema produtor da Cantareira.

	Subestimou	Superestimou
Treinamento	RNA-C-4	RNA-C-1, 2, 3, 4, 6 e 7
Teste	RNA-C-1	RNA-C-2 e 3
Previsão	RNA-C-1, 4 e 7	RNA-C-2 e 3

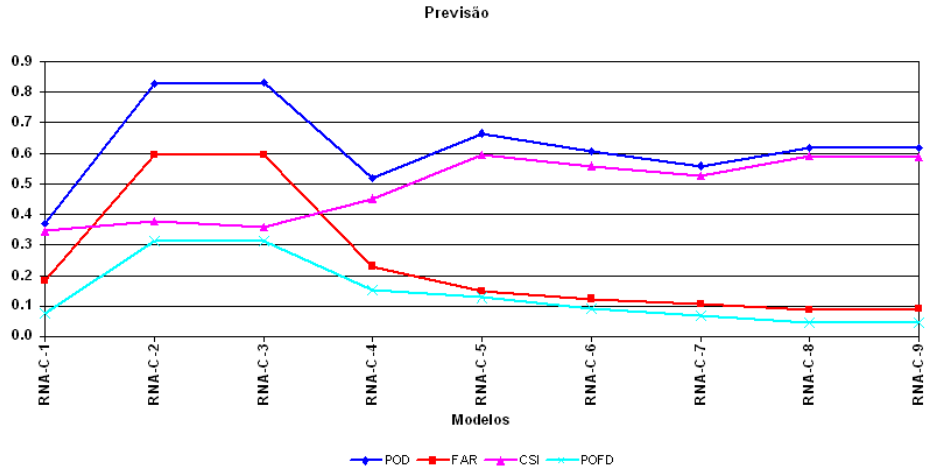
A figura 5.16 mostra a probabilidade de detecção (POD), a taxa de falso alarme (FAR), o índice crítico de sucesso (CSI) e probabilidade de falsa detecção (POFD) para os modelos de RNA do sistema Cantareira para o treinamento, teste e previsão. O POD indica os valores que foram previstos corretamente e o CSI, os valores que foram previstos ou observados corretamente. O FAR e o POFD mostram os valores que não foram previstos corretamente. Nota-se que esses índices tem os mesmos comportamentos quando comparados. Observa-se que o CSI e o POFD sempre são menores que POD e FAR em todos os modelos, indicando que na maior parte do tempo os valores previstos são maiores que os observados



(a)



(b)



(c)

Figura 5.16: Valores da probabilidade de detecção (POD), linha azul, a taxa de falso alarme (FAR), linha vermelha, índice crítico de sucesso (CSI), linha rosa, e probabilidade de falsa detecção (POFD), linha azul claro, para os modelos de RNA do sistema da Cantareira. (a) treinamento, (b) teste e (c) previsão.

No treinamento (figura 5.16a), o modelo RNA-C-4 tem os menores valores para POD e CSI e os maiores para FAR e POFD. No teste (figura 5.16b), o modelo RNA-C-1 obteve os menores valores para todos os índices. Os modelos RNA-C-2 e 3 tiveram os maiores valores para POD, FAR e POFD, e o menor para CSI em relação aos outros modelos. Os modelos RNA-C-2 e 3 tiveram o mesmo comportamento na previsão (figura 5.16c), tendo os maiores valores para POD, FAR e POFD, e menor para CSI. A tabela 5.9 apresenta a média do POD, FAR, CSI e POFD, para o treinamento, teste e previsão. As células em azul indicam os melhores resultados, e as vermelhas os piores. Nota-se que o teste tem os piores resultados quando comparados com o treinamento e a previsão. Isto pode ter ocorrido devido ao conjunto de dados utilizado para o teste.

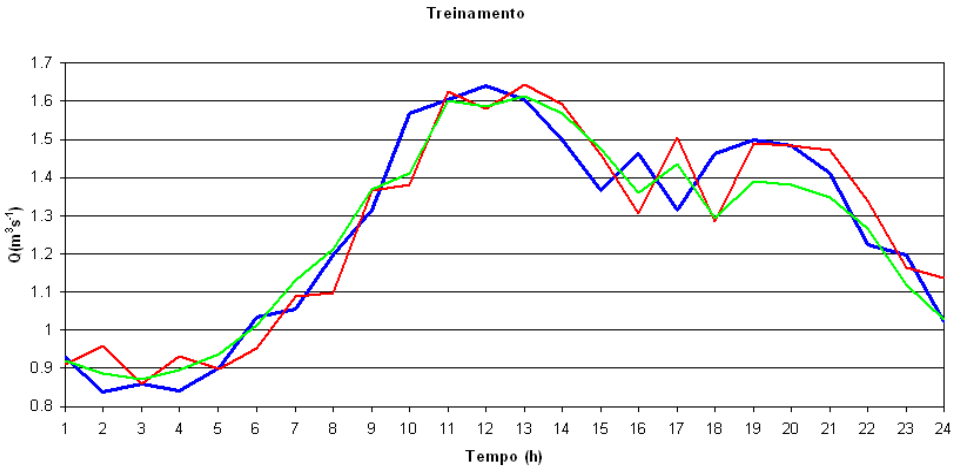
Tabela 5.9: Média do POD, FAR, CSI e POFD, para o treinamento, teste e previsão. As células em azul indicam os melhores resultados, e as vermelhas os piores.

	TREINAMENTO	TESTE	PREVISÃO
POD	76%	67%	62%
FAR	24%	41%	19%
CSI	65%	42%	53%
POFD	17%	22%	14%

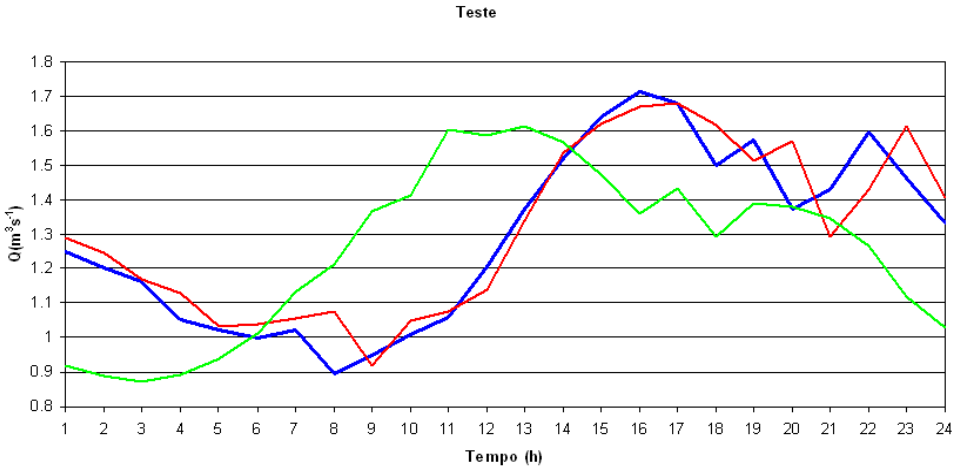
5.5.2. Itaim Paulista

O consumo do setor Itaim Paulista é predominantemente residencial (91,9%) (Tsutiya, 2006), e o setor faz parte do sistema Alto Tietê. A figura 5.17 mostra o comportamento do consumo obtido pelos modelos RNA-IP-1 e 2, para o treinamento, teste e previsão. Na denominação dos modelos, RNA significa Rede Neural Artificial, o IP a Itaim Paulista e 1 e 2 ao número do modelo.

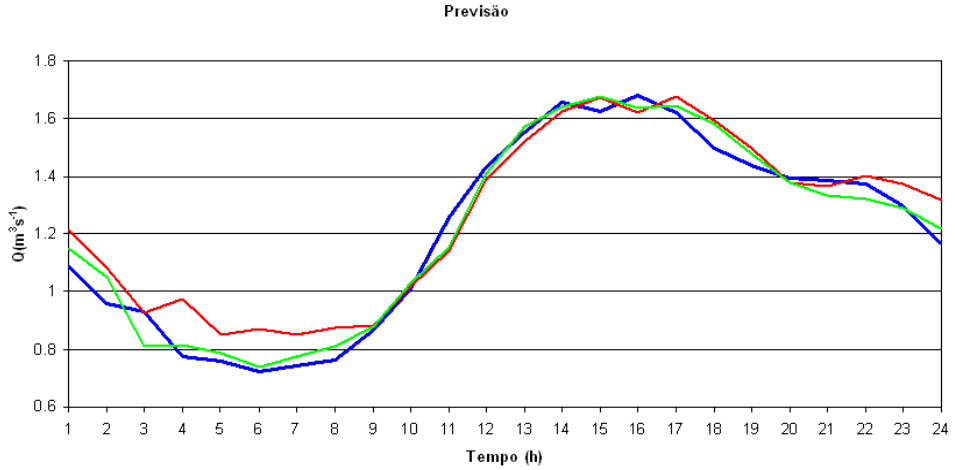
A configuração do modelo RNA-IP-1 é 18-1-1 e do RNA-IP-2 é 18-5-1, ou seja, existem 18 nós na camada de entrada, 1 nó na camada de saída e, respectivamente, 1 e 5 nós na camada escondida. Então, a diferença entre esses modelos é o número de nós na camada escondida. Esses testes foram realizados para verificar qual a influência do número de nós na camada escondida nos modelos.



(a)



(b)



(c)

Figura 5.17: Evolução temporal do consumo de água para os modelos RNA-IP-1 e RNA-IP-2, onde a linha azul é o consumo observado, linha vermelha é o consumo previsto pelo RNA-IP-1 e linha verde modelo RNA-IP-2. a) treinamento, b) teste e c) previsão.

Observa-se na figura 5.17 que os modelos tendem a acompanhar os valores observados. Na figura 5.17b nota-se que os valores obtidos pelo modelo RNA-IP-2 (linha verde), teve um deslocamento muito grande quando comparado aos valores observados (linha azul) e do modelo RNA-IP-1 (linha vermelha). A razão desde deslocamento pode ter ocorrido devido ao conjunto de dado utilizado. No treinamento (figura 5.17a), tanto o modelo RNA-IP-1 quanto o RNA-IP-2 tiveram uma pequena defasagem em relação aos valores observados, sendo que, o modelo RNA-IP-1 apresentou uma defasagem maior do que do modelo RNA-IP-2. Na previsão (figura 5.17c), os resultados dos modelos ficaram próximos dos valores observados. Os valores de consumo pequenos tendem a ser mais superestimados dos os valores maiores. Então, o aumento do número de nós na camada escondida melhora em alguns casos um pouco o resultado, mas em outros não. Observou-se que esse fato aumenta o tempo de processamento, mas não traz tanto benefícios para os resultados. Desta forma, optou-se que em utilizar somente um nó camada escondida para os demais testes realizados para o setor Itaim Paulista.

Os diagramas de espalhamento dos modelos RNA-IP-3 e 4 estão na figura 5.18. A diferença entre os modelos é que o modelo RNA-IP-3 utiliza somente as variáveis meteorológicas e o consumo como entrada e o modelo RNA-IP-4, a temperatura, a umidade relativa e o consumo como entrada. Os dois modelos não utilizaram as variáveis antrópicas na entrada da RNA. As configurações dos modelos são 14-1-1 e 6-1-1, respectivamente. Nota-se que no treinamento (figura 5.18a) e no teste (figura 5.18b) para os dois modelos têm-se um espalhamento maior que para a previsão (figura 5.18c), sendo que o espalhamento do teste é maior do que para o treinamento e previsão. O diagrama de espalhamento para previsão tende a ter uma correlação quase perfeita entre os dados previstos e os observados. O melhor R^2 ocorreu para previsão e o menor para o teste dos dois modelos. O modelo RNA-IP-3 tem um R^2 um pouco melhor que do modelo RNA-IP-4. Assim, pode-se dizer que quando utilizamos todas as variáveis meteorológicas disponíveis, elas influenciam mais do que quando utilizamos somente a temperatura e a umidade. Observa-se também em alguns valores os modelos tendem a subestimar e/ou superestimar alguns valores. Alguns valores estimados diferem significativamente dos valores observados.

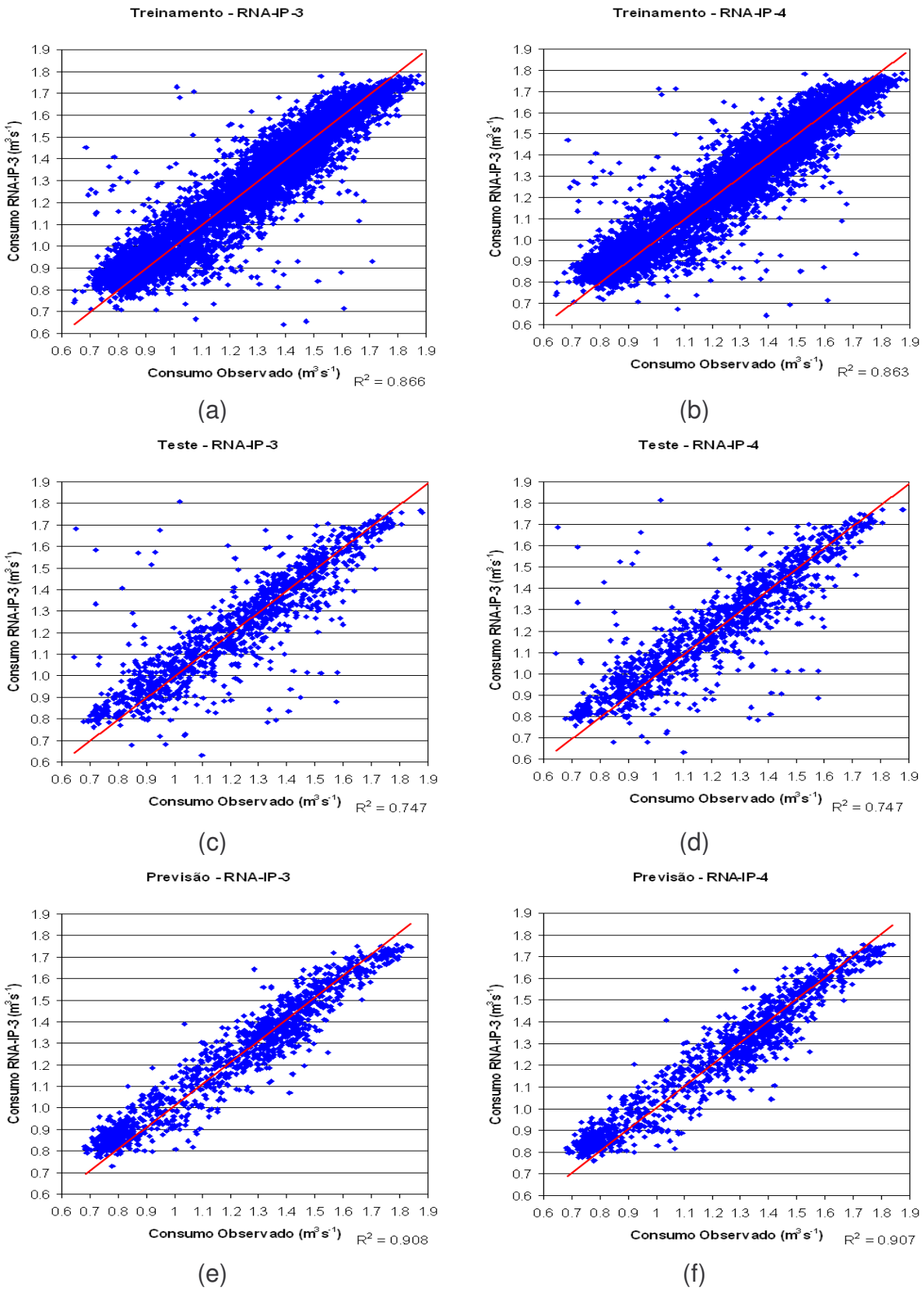


Figura 5.18: Digrama de espalhamento da série temporal de consumo de água observado e estimado pelo: a) RNA-IP-3 - treinamento, b) RNA-IP-4 - treinamento, c) RNA-IP-3 - teste, d) RNA-IP-4 - teste, e) RNA-IP-3 - previsão e b) RNA-IP-4 - previsão. Estão indicados o ajuste linear (linha vermelha) e o coeficiente de correlação (R^2).

Os modelos RNA-IP-5 a 8 utilizaram a média do consumo de 6 horas. O treinamento e a previsão foram realizados para 6, 12, 18 e 24 horas, e as configurações utilizadas foram 18-1-1, 25-1-1, 32-1-1 e 39-1-1, respectivamente. Os números da camada de entrada foram aumentando à medida que se acrescentou a memória das 6 horas anteriores, para cada passo a frente da previsão, e isto aconteceu para o treinamento e a previsão.

Na figura 5.19, nota-se que os valores estimados (linha vermelha) algumas vezes são subestimados e em outros superestimados. Em determinados momentos os valores observados e estimados coincidiram. Observa-se que os valores estimados e observados completam um ciclo a cada 24 horas. Nota-se que o treinamento e a previsão melhoram à medida que o tempo da previsão vai aumentando, isto ocorre devido ao acréscimo da memória do sistema.

Os valores estimados para o modelo RNA-IP-5 sempre é subestimado tanto para treinamento (figura 5.19a) quanto para previsão (figura 5.19b). Os valores estimados oscilam, para mais ou para menos ou coincidem com os valores observados para os outros modelos. Observa-se, também que em determinadas horas, ocorre uma pequena defasagem entre os valores estimados e observados para os modelos RNA-IP-6 a 8. Os piores valores ocorreram para o modelo RNA-IP-5, e os melhores para o modelo RNA-IP-8, que correspondem a previsão de 6 e 24 horas, respectivamente. Nota-se, que os valores passaram a ficar melhor a partir do modelo RNA-IP-6. Conclui-se, que a previsão para 6 horas será um bom instrumento para previsões futuras, mas o das 24 horas será mais precisa.

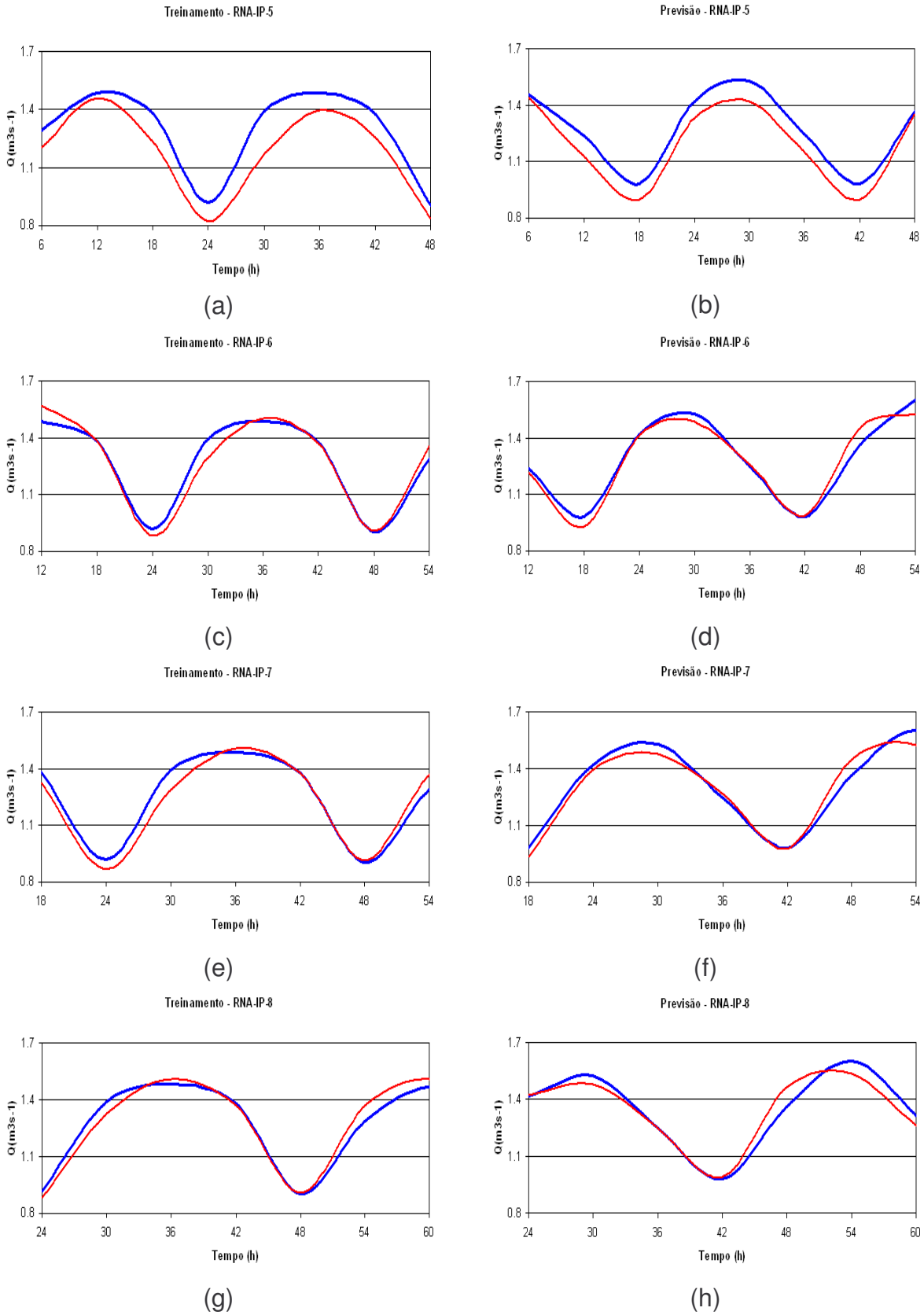


Figura 5.19: Evolução temporal do consumo de água para os modelos RNA-IP-5 a 8, onde a linha azul é o consumo observado, linha vermelha é o consumo previsto. a) treinamento – RNA-IP-5, b) previsão – RNA-IP-5, c) treinamento – RNA-IP-6, d) previsão – RNA-IP-6, e) treinamento – RNA-IP-7, f) previsão – RNA-IP-7, g) treinamento – RNA-IP-5 e h) previsão – RNA-IP-5.

Para os modelos RNA-IP-1 a 4, os resultados do erro médio (EM), do erro médio absoluto (MAE), do erro médio quadrático (RMSE) e do coeficiente de correlação (r) podem ser vistos na tabela 5.10, para o treinamento, o teste e a previsão. Nota-se que o modelo RNA-IP-2 tem os melhores resultados para o treinamento e previsão, e o pior resultado para o teste. No teste o melhor resultado ocorreu para o modelo RNA-IP-4. Na previsão o pior resultado ocorreu para o modelo RNA-IP-1. No treinamento não houve pior resultado, pois os valores obtidos estão bem próximos uns dos outros.

Tabela 5.10: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para o treinamento, teste e previsão dos modelos RNA-IP 1, 2, 3 e 4 do setor Itaim Paulista. A coluna vermelha são os piores valores e a coluna azul os melhores valores.

TREINAMENTO				
	RNA-IP-1	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**
Erro Médio Absoluto (MAE)	0.073	0.057	0.073	0.074
Erro Médio (EM)	0.005	0.003	0.004	0.004
Erro Médio Quadrático (RMSE)	0.101	0.081	0.101	0.102
Coeficiente de Correlação (r)	0.937	0.960	0.937	0.936
Coeficiente de Determinação (R^2)	0,878	0,922	0,878	0,876
TESTE				
	RNA-IP-1	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**
Erro Médio Absoluto (MAE)	0.084	0.297	0.082	0.082
Erro Médio (EM)	0.019	0.044	0.005	0.003
Erro Médio Quadrático (RMSE)	0.129	0.367	0.128	0.128
Coeficiente de Correlação (r)	0.882	0.132	0.882	0.882
Coeficiente de Determinação (R^2	0,778	0,017	0,778	0,777
PREVISÂO				
	RNA-IP-1	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**
Erro Médio Absoluto (MAE)	0.073	0.054	0.067	0.066
Erro Médio (EM)	0.037	0.016	0.021	0.013
Erro Médio Quadrático (RMSE)	0.090	0.070	0.084	0.083
Coeficiente de Correlação (r)	0.962	0.973	0.962	0.961
Coeficiente de Determinação (R^2)	0,925	0,847	0,926	0,924

A tabela 5.11 mostra o erro médio (EM), o erro médio absoluto (MAE), o erro médio quadrático (RMSE) e o coeficiente de correlação (r) para os modelos RNA-IP-5 a 8 para o treinamento e previsão. O pior resultado ocorreu para o modelo RNA-IP-5 tanto para o treinamento como para a previsão, como este modelo não utilizava as memórias passadas do sistema, os maiores valores de MAE, EM, RMSE e o menor r podem tem ocorrido por isto. Os valores para o treinamento, para os modelos RNA-IP-6, 7 e 8 ficaram parecidos indicando que a memória do sistema colabora para a

aproximação dos resultados da previsão em relação aos valores observados. O treinamento tem o MAE e RMSE maior, comparado ao da previsão, mas o erro médio é menor. O valor de r para a previsão foi melhor para treinamento.

Tabela 5.11: Valores de Erro Médio Absoluto (MAE), Erro Médio (EM), Erro Médio Quadrático (RMSE) e Coeficiente de correlação (r) para o treinamento e previsão para os modelos RNA-IP 5, 6, 7 e 8 do setor Itaim Paulista. A coluna vermelha são os piores valores.

TREINAMENTO				
	RNA-IP-5	**RNA-IP-6**	**RNA-IP-7**	**RNA-IP-8**
Erro médio Absoluto (MAE)	0.070	0.044	0.045	0.045
Erro Médio (EM)	0.011	0.006	0.006	0.006
Erro médio quadrático (RMSE)	0.089	0.060	0.060	0.060
Coeficiente de Correlação (r)	0.930	0.969	0.969	0.969
Coeficiente de Determinação (R^2)	0,865	0,939	0,938	0,938
PREVISÃO				
	RNA-IP-5	**RNA-IP-6**	**RNA-IP-7**	**RNA-IP-8**
Erro médio Absoluto (MAE)	0.044	0.028	0.028	0.030
Erro Médio (EM)	0.024	0.009	0.010	0.012
Erro médio quadrático (RMSE)	0.083	0.058	0.059	0.061
Coeficiente de Correlação (r)	0.884	0.935	0.931	0.929
Coeficiente de Determinação (R^2)	0,965	0,965	0,965	0,965

Para o cálculo da exatidão (E), da freqüência de bias (BIAS), da probabilidade de detecção (POD), da taxa de falso alarme (FAR), do índice crítico de sucesso (CSI) e da probabilidade de falsa detecção (POFD) foram utilizados os seguintes limites de consumo: a média, e a média ±1 e ±1,5 do desvio padrão. A tabela 5.12 mostra os resultados para E obtidos para os modelos RNA-IP-1 a 8, para o treinamento, teste e previsão. Para os modelos RNA-IP-5 a 8, o teste não foi realizado.

No treinamento, a exatidão indica que mais de 90% dos valores obtidos pelos modelos estão corretos. O modelo RNA-IP-5 é exceção, porque obteve três valores (célula cinza) abaixo de 90%. Os valores desse modelo estão acima de 80%, sendo que o pior valor foi de 83,5% (vermelho). O melhor valor ocorreu para os modelos RNA-IP-7 e 8 e foi de 99,6%. Para o teste, o modelo RNA-IP-4 mostrou o melhor resultado de exatidão, que foi de 97,6% (azul), e o pior ocorreu para o modelo RNA-IP-2, e foi de 60,5% (vermelho). Os modelos RNA-IP-1 a 4 obtiveram apenas três valores abaixo de 90% (célula cinza) para a previsão, e demais valores estão acima de 90%. Para esses modelos o maior valor foi 97,8% (azul) para o modelo RNA-IP-3 e o menor 87,2% (vermelho) para o modelo RNA-IP-3. O maior valor para

os modelos RNA-IP-5 a 8 foi de 98,5% (azul) e o menor 81,4% (vermelho). A exatidão para os modelos que utilizaram a média de 6 horas tende a ser menores do que dos modelos que utilizam a média horária. Os valores para o teste também, tendem a serem mais baixos que para o treinamento e previsão.

Tabela 5.12: Valores de Exatidão (E) e o limite utilizado para o treinamento, teste e previsão dos modelos RNA-IP-1 a 8. Os números em vermelhos são os menores valores e os azuis os melhores. A célula cinza apresenta valores abaixo de 90%.

TREINAMENTO									
LIMITE	**RNA-IP-1**	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**	**LIMITE**	**RNA-IP-5**	**RNA-IP-6**	**RNA-IP-7**	**RNA-IP-8**
0.80	0.949	0.951	0.950	0.951	**0.88**	**0.835**	0.955	0.935	0.947
0.95	0.917	0.943	0.911	0.915	**1.00**	0.995	0.995	**0.996**	**0.996**
1.23	0.917	0.949	0.921	0.919	**1.23**	0.867	0.905	0.901	0.912
1.52	0.944	0.951	0.942	0.943	**1.47**	0.889	0.921	0.922	0.920
1.66	0.961	0.967	0.962	0.961	**1.59**	0.973	0.973	0.972	0.971
TESTE									
LIMITE	**RNA-IP-1**	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**	**LIMITE**	**RNA-IP-5**	**RNA-IP-6**	**RNA-IP-7**	**RNA-IP-8**
0.80	0.943	0.913	0.943	0.944	**0.88**	–	–	–	–
0.95	0.881	0.650	0.887	0.887	**1.00**	–	–	–	–
1.23	0.894	**0.605**	0.901	0.899	**1.23**	–	–	–	–
1.52	0.951	0.778	0.956	0.954	**1.47**	–	–	–	–
1.66	0.974	0.882	0.975	**0.976**	**1.59**	–	–	–	–
PREVISÃO									
LIMITE	**RNA-IP-1**	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**	**LIMITE**	**RNA-IP-5**	**RNA-IP-6**	**RNA-IP-7**	**RNA-IP-8**
0.80	**0.872**	0.923	0.875	0.879	**0.88**	0.837	**0.985**	0.970	0.957
0.95	0.940	0.958	0.939	0.938	**1.00**	0.962	0.970	0.972	0.967
1.23	0.929	0.960	0.938	0.929	**1.23**	**0.814**	0.872	0.872	0.857
1.52	0.954	0.970	0.963	0.967	**1.47**	0.842	0.879	0.877	0.877
1.66	0.975	0.973	**0.978**	0.976	**1.59**	0.982	0.982	0.982	0.982

A frequência de bias (Viés) para o treinamento, teste e previsão para os modelos RNA-IP-1 a 8 estão na tabela 5.13. Os valores de Viés>1 estão nas células cinza claro, os Viés<0,5 nas células cinza escuro e os Viés=1 nas células amarelas. No treinamento, o valor perfeito ocorreu para os modelos RNA-IP-1, 3, 4, 5 e 6. O maior valor de Viés é observado no modelo RNA-IP-7 e é de 1,12 (azul) e o menor é 0,23 (vermelho) que ocorreu para o modelo RNA-IP-4.

No treinamento, os modelos RNA-IP-1 a 4 tende a superestimar os valores que estão próximos da média dos valores observados. Os valores baixos tendem a ser subestimados para os modelos RNA-IP-1, 2 e 4. Os modelos RNA-IP-6 a 8 a superestimar os valores altos e baixos, e o modelo RNA-IP-5 a subestimar. No teste, o valor perfeito ocorreu nos modelos RNA-IP-1,3 e 4. Os limites baixos foram

subestimados para todos os modelos, sendo que o menor valor foi 0,28 (vermelho). Os modelos RNA-IP-2, 3 e 4 tendem a superestimar os valores, sendo que o maior valor foi de 1,05 (azul), e ocorreu para modelo RNA-IP-2. Assim como no treinamento, os valores baixos tendem a ser subestimados.

O modelo RNA-IP-1, para previsão, tem o valor mais baixo de todos os Viés, e foi de 0,06 (vermelho), indicando que os valores foram subestimados. Os modelos RNA-IP-3, 4 e 5, também subestimaram os valores. Os valores perfeitos foram observados nos modelos RNA-IP-1, 3, 4 e 6. Os valores foram superestimados no para os modelos RNA-IP-5,6,7 e 8, sendo que o modelo RNA-IP-7 teve o maior valor, 1,07 (azul). Os valores dos limites mais altos tendem ser superestimados para os RNA-IP-2, 5, 6, 7 e 8. O limite 1,59 m^3s^{-1} de todos modelos superestimam os valores previstos.

Tabela 5.13: Valores de frequência de bias (Viés) e o limite utilizado para o treinamento, teste e previsão dos modelos RNA-IP-1 a 8. Os números em vermelhos são os menores valores que ocorreram, e em azul os maiores. A célula cinza escura mostra os valores <0,5, a amarela =1 e cinza claro valores >1.

TREINAMENTO									
LIMITE	**RNA-IP-1**	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**	**LIMITE**	**RNA-IP-5**	**RNA-IP-6**	**RNA-IP-7**	**RNA-IP-8**
0,80	0,29	0,82	0,25	**0,23**	**0,88**	0,66	1,07	**1,12**	1,01
0,95	0,95	0,93	0,93	0,93	**1,00**	1,00	1,00	1,01	1,01
1,23	1,01	1,01	1,02	1,02	**1,23**	0,93	0,95	0,92	0,94
1,52	0,99	0,99	0,99	0,99	**1,47**	0,96	0,95	0,95	0,96
1,66	1,00	1,01	1,00	1,00	**1,59**	1,03	1,03	1,03	1,02
TESTE									
LIMITE	**RNA-IP-1**	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**	**LIMITE**	**RNA-IP-5**	**RNA-IP-6**	**RNA-IP-7**	**RNA-IP-8**
0,80	**0,28**	0,51	0,38	0,38	**0,80**	–	–	–	–
0,95	0,84	**1,05**	0,88	0,89	**0,95**	–	–	–	–
1,23	0,95	0,85	1,01	1,02	**1,23**	–	–	–	–
1,52	1,00	0,94	1,00	1,01	**1,52**	–	–	–	–
1,66	0,99	0,97	1,00	1,00	**1,66**	–	–	–	–
PREVISÃO									
LIMITE	**RNA-IP-1**	**RNA-IP-2**	**RNA-IP-3**	**RNA-IP-4**	**LIMITE**	**RNA-IP-5**	**RNA-IP-6**	**RNA-IP-7**	**RNA-IP-8**
0,80	**0,06**	0,51	0,12	0,14	**0,88**	0,41	1,00	0,84	0,83
0,95	0,88	0,92	0,93	0,93	**1,00**	1,01	1,04	**1,07**	1,05
1,23	0,93	0,96	1,00	1,02	**1,23**	0,83	0,89	0,88	0,87
1,52	0,97	0,99	0,99	0,99	**1,47**	0,89	0,95	0,96	0,96
1,66	1,00	1,01	1,00	1,00	**1,59**	1,02	1,02	1,02	1,02

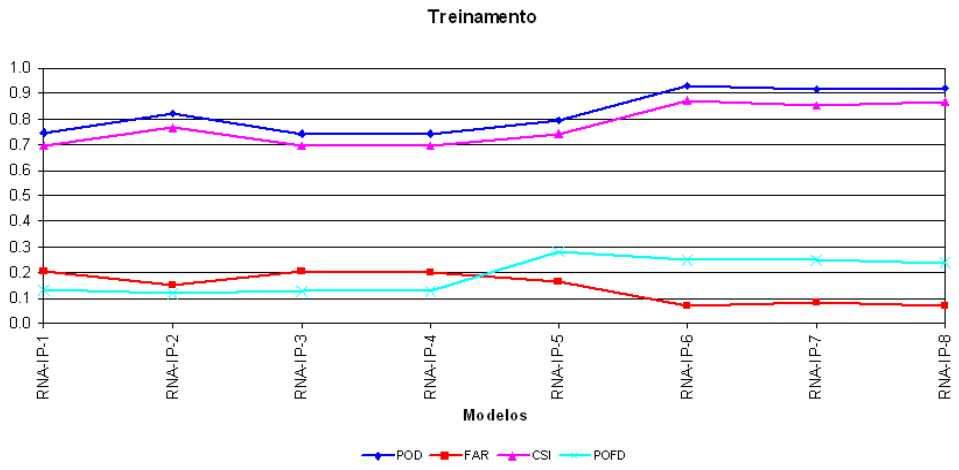
A probabilidade de detecção (POD), a taxa de falso alarme (FAR), o índice crítico de sucesso (CSI) e a probabilidade de falsa detecção (POFD), para os modelos

RNA-IP-1 a 8 do setor Itaim Paulista podem ser visto na figura 5.20. Observa-se que os valores de CSI (linha rosa) são sempre menores do que do POD (linha azul). O FAR (linha vermelha) e POFD (linha azul clara) têm comportamentos diferentes para o treinamento, teste e previsão.

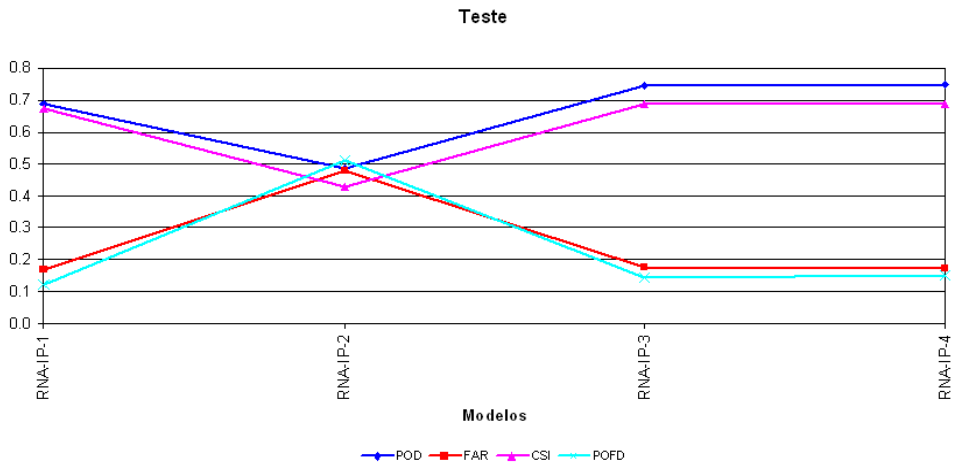
No treinamento (figura 5.20a), os modelos RNA-IP-1 a 4 têm valores de POD e CSI menores que dos modelos RNA-IP-5 a 8, indicando que esses modelos tendem a fazer uma previsão mais precisa que dos modelos RNA-IP-1 a 4. Observa-se que os valores de CSI dos modelos RNA-IP-6 a 8 estão próximos de 0,9 e os valores de POD acima de 90%. Os valores de FAR dos modelos RNA-IP-1 a 4 são maiores que dos modelos RNA-IP-5 a 8. Nota-se que os valores de POFD têm um comportamento inverso do FAR, onde os valores dos modelos RNA-IP-5 a 8 são maiores que dos modelos RNA-IP-1 a 4. Os valores de consumo baixos foram superestimados e os altos subestimados no treinamento.

Na figura 5.20b observa-se que para o teste os valores de FAR e CSI são maiores do que os valores de POD e CSI para o modelo RNA-IP-2. Esse resultado indica que esse modelo não fez boas previsões. Os valores de POD são maiores do que CSI, e são menores dos que os obtidos para o treinamento. Isto mostra que os valores obtidos para o treinamento são melhores que o do teste. Os valores de FAR são maiores que POFD para os modelos RNA-IP-1, 3 e 4.

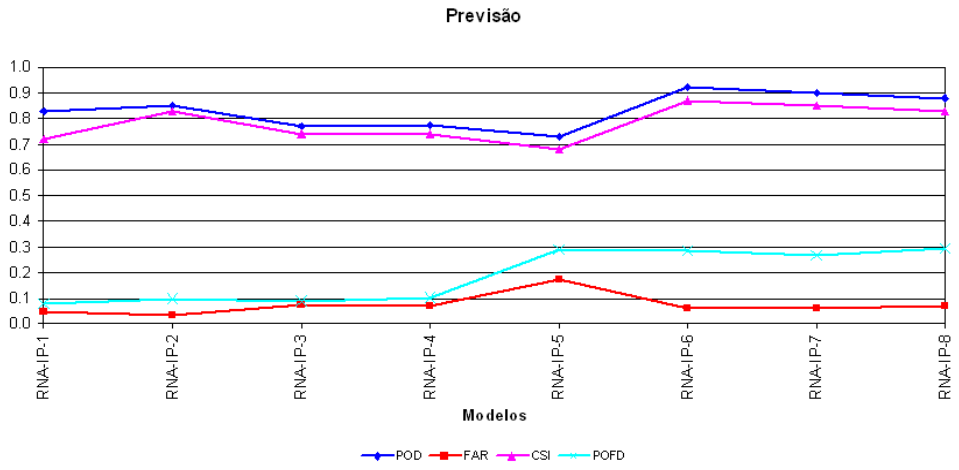
Novamente, na previsão (figura 5.20c), os valores de POD são maiores que CSI para todos os modelos. Assim como FAR é maior que POFD. Nota-se que os valores de POFD para os modelos RNA-IP-5 a 8 são bem maiores que os valores de FAR, isto não ocorre para os modelos RNA-IP-1 a 4, onde os valores estão próximos. Observa-se que isto ocorreu somente na previsão, no treinamento e no teste, os valores ficaram próximos. Os valores POD e CSI ficaram parecidos com os resultados obtidos pelo treinamento, e foram melhores que do teste. Desta forma, pode se dizer que os valores do treinamento e previsão são melhores do que de teste e isto pode ter corrido devido a divisão do banco de dados.



(a)



(b)



(c)

Figura 5.20: Valores da probabilidade de detecção (POD), linha azul, taxa de falso alarme (FAR), linha vermelha, índice crítico de sucesso (CSI), linha rosa, e probabilidade de falsa detecção (POFD), linha azul clara, para os modelos de RNA do setor Itaim Paulista. (a) treinamento, (b) teste e (c) previsão.

A tabela 5.14 mostra a média dos valores de CSI, FAR, CSI e POFD para os modelos RNA-IP-1 a 8. Observa-se que os piores valores (células vermelhas) ocorreram para o teste dos modelos RNA-IP-1 a 4, com exceção do POFD que ocorreu para a previsão dos modelos RNA-IP-5 a 8. O melhor valor médio (célula azul) do POD indica que 89% dos valores previstos pelos modelos RNA-IP-5 a 8 foram previstos corretamente, e a menor média foi para o teste dos modelos RNA-IP-1 a 4, e indicou que 66,6% dos valores foram previstos corretamente.

A melhor média do CSI ocorreu para o treinamento dos modelos RNA-IP-5 a 8 e indicam que foram previstos ou observados corretamente 83,5% dos valores e o pior valor ocorreu para os modelos RNA-IP-1 a 4 e foi de 61,9%. Os modelos RNA-IP-1 a 4 obtiveram a melhor média para a previsão e indica que 5,6% dos valores não foram previstos corretamente e o pior ocorreu para o teste desses mesmos modelos e indicaram que 25,1% não foram previstos corretamente. A média de POFD indica que 9,3% dos valores dos modelos RNA-IP-1 a 4 não foram previstos corretamente, esses valor indica a melhor média e ocorreu na previsão. O pior valor ocorreu para a previsão dos modelos RNA-IP-5 a 8 e foi de 28,2%.

Tabela 5.14: Média do POD, FAR, CSI e POFD, para o treinamento, teste e previsão. As células em azul indicam os melhores resultados, e as vermelhas os piores.

		TREINAMENTO	TESTE	PREVISÃO
POD	RNA-IP-1 a 4	76,4%	66,6%	80,4%
	RNA-IP-5 a 8	89%		85,7%
FAR	RNA-IP-1 a 4	19%	25,1%	5,6%
	RNA-IP-5 a 8	9,6%		9,1%
CSI	RNA-IP-1 a 4	71,4%	61,9%	75,7%
	RNA-IP-5 a 8	83,5%		80,8%
POFD	RNA-IP-1 a 4	12,5%	23,4%	9,3%
	RNA-IP-5 a 8	25,3%		28,2%

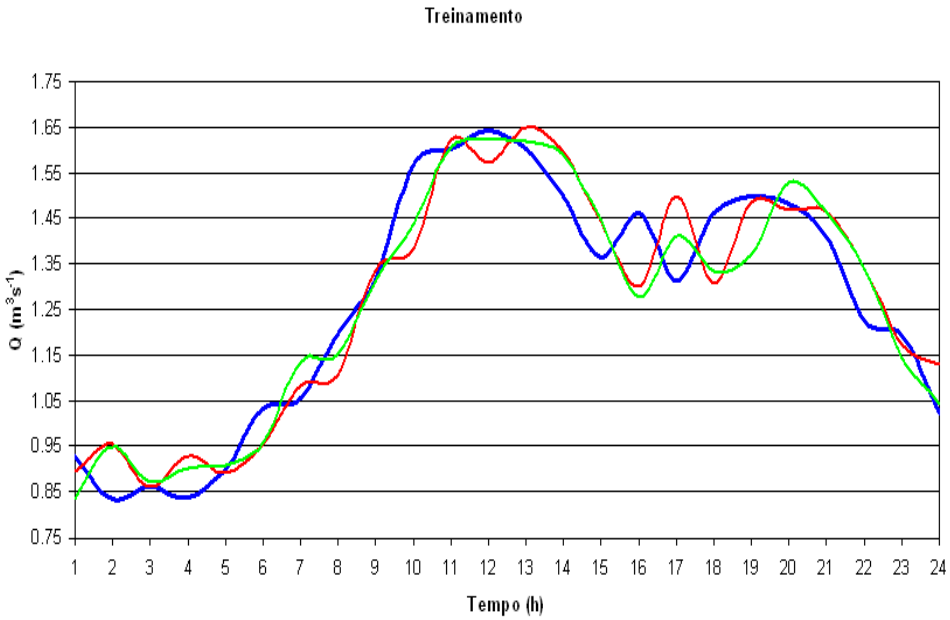
A previsão ótima foi utilizada para simular uma previsão onde os dados observados foram utilizados como se fossem valores previstos. Os dados são horários e foram feitas previsões de 1 a 24 horas. Os modelos RNA-IP-9 a 32 utilizam dados sem a memória e os modelos RNA-IP-33 a 55 utilizam a memória do sistema, RNA-IP-56 e 57 que tem duas particularidades, o primeiro utilizou todas as memórias do sistema

e o segundo utilizou todas as memórias do sistema com exceção das memórias do consumo anterior.

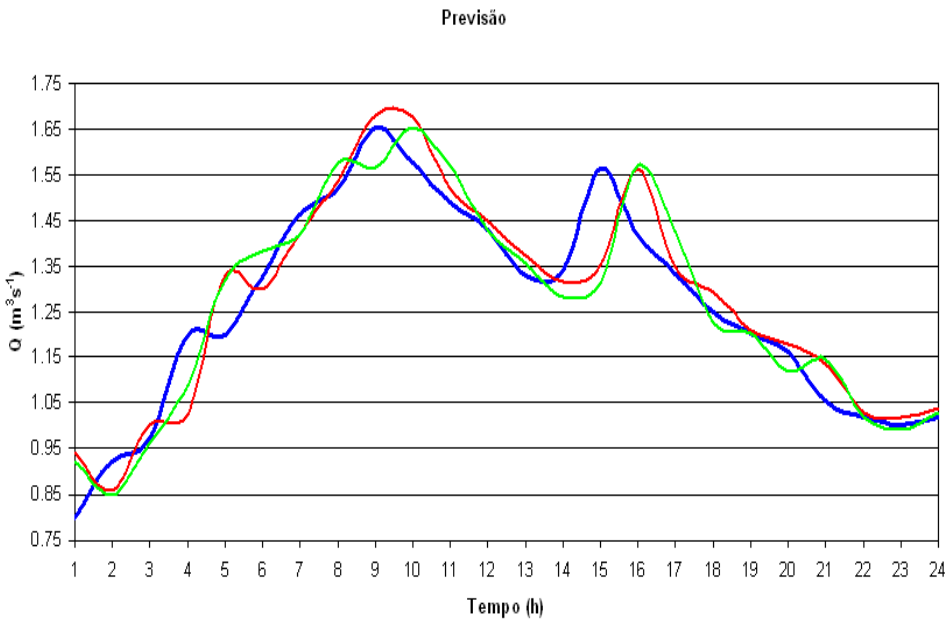
Os modelos apresentam uma previsão para 24 horas. A figura 5.21 mostra o treinamento e a previsão da RNA para os modelos RNA-IP-9 a 32 (linha vermelha) e RNA-33 a 55 (linha verde), cada hora corresponde ao valor previsto pelo respectivo modelo de sua hora. Nota-se que no treinamento e na previsão os valores previstos têm uma defasagem no tempo em relação aos valores observados. Os valores dos modelos RNA-IP-9 a 32 (linha vermelha) tendem a acompanhar mais a evolução do consumo observado (linha azul) do que os dos modelos RNA-IP-33 a 55.

Observa-se que, nos modelos RNA-IP-33 a 55, em determinadas horas ocorre um achatamento da curva tanto para o treinamento como para previsão, além de que eles tendem a oscilar mais do que dos modelos RNA-IP-9 a 32. Nota-se que os modelos tendem a seguir o padrão do consumo observado (linha azul), sendo que os modelos RNA-IP-33 a 55 têm valores mais próximos dos observado do que os modelos RNA-IP-9 a 32. Os picos observados tendem a oscilar tanto para mais como quanto para menos dos valores observados. Desta forma, observa-se que quando se aumenta a memória do sistema os valores tendem a melhor os resultados obtidos pelos modelos, por essa razão é que os valores obtidos pelos modelos RNA-IP-33 a 55 são um pouco melhor que dos modelos RNA-IP-9 a 32. Os valores obtidos pela previsão são melhores que os obtidos pelo treinamento dos modelos.

A figura 5.22 mostra os valores previstos para os modelos RNA-IP-32, 55, 55 a 57. Nota-se que em relação ao valor observado, o valor é maior para o modelo RNA-IP-32 e menor para RNA-IP-57 e o que mais se aproxima do valor observado é o modelo RNA-IP-55. O modelo RNA-IP-57 não utiliza valores de consumo anterior como entrada da RNA, levando o modelo a ter o maior erro, no treinamento e previsão. O modelo RNA-IP-56 utilizou todos os dados anteriores das variáveis meteorológicas e do consumo anterior e mostrou apenas que aumentou seu tempo de processamento e não melhorou seu resultado.

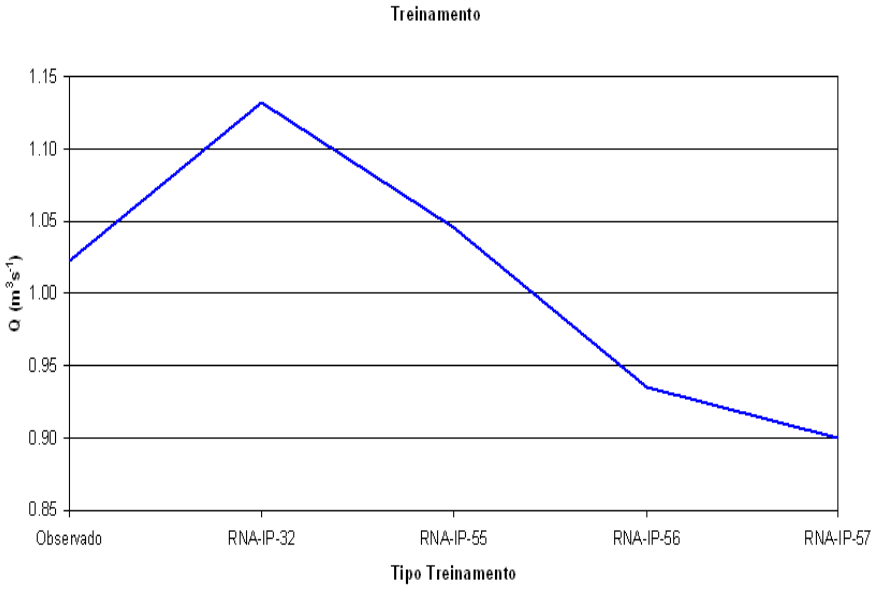


(a)

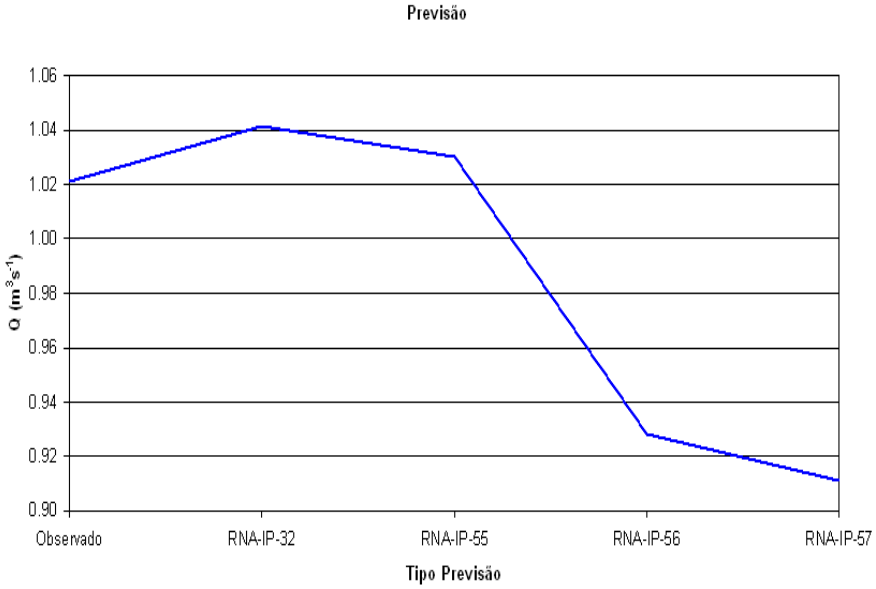


(b)

Figura 5.21: Evolução temporal do consumo de água para os modelos RNA-IP-9 a 32 e RNA-IP-33 a 55, onde a linha azul é o consumo observado, linha vermelha é o consumo previsto por RNA-IP-9 a 32 e linha verde modelo RNA-IP-33 a 55. a) treinamento e b) previsão.



(a)



(b)

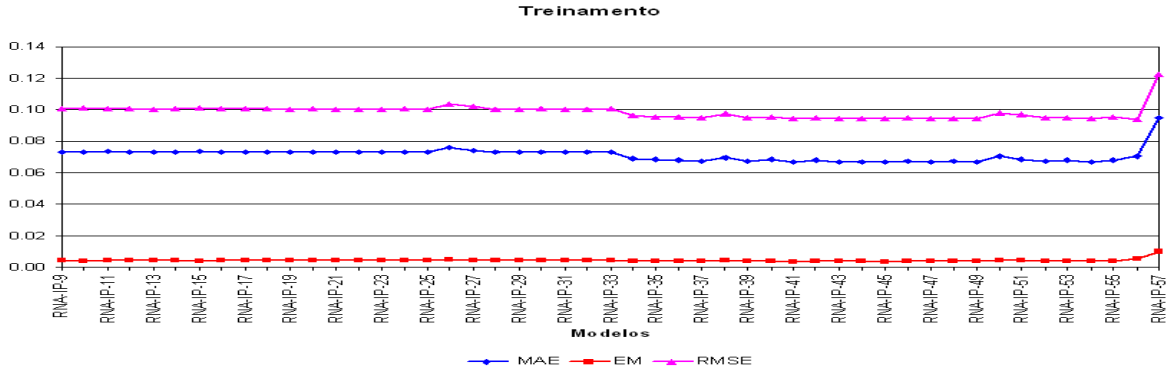
Figura 5.22: Evolução do consumo de água para os modelos RNA-IP-32, 55, 56 e 57 e consumo observado. a) treinamento e b) previsão

O erro médio (EM), erro médio absoluto (MAE), erro médio quadrático (RMSE), coeficiente de correlação (r) e coeficiente de determinação (R^2) para o treinamento e previsão são mostrados na figura 5.23. Nota-se que os valores oscilam mais para previsão (figura 5.23b) do que para o treinamento (figura 5.23a). No treinamento, os valores de EM, MAE, RMSE tendem a não ter muita variação com exceção do modelo RNA-IP-57 que é maior para EM, MAE e RMSE. As variações dos erros nos

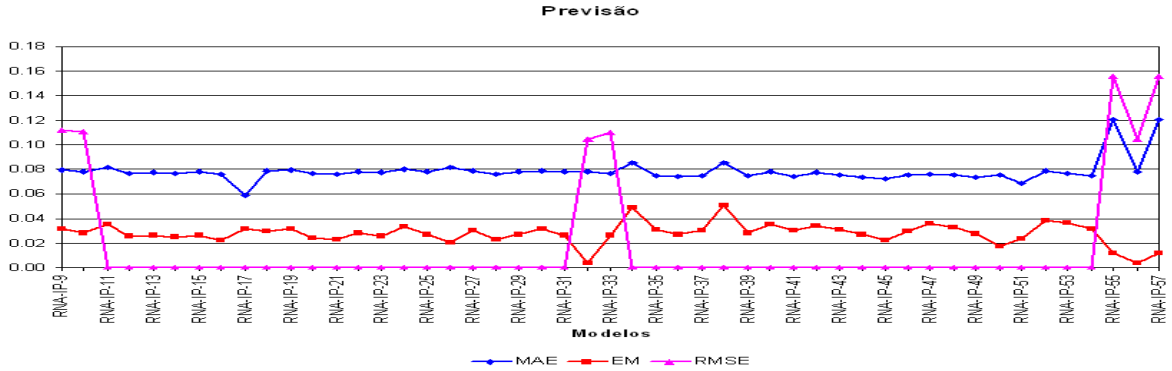
modelos RNA-IP-9 a 32 e RNA-IP-33 a 55 são pequenas. Os valores tendem a serem maiores para os modelos RNA-IP-5 a 32; do que para RNA-IP-33 a 55. Observa-se que na previsão (figura 5.23b), as oscilações dos erros obtidos pelos modelos são maiores do que do treinamento (figura 5.23a).

No treinamento, nota-se que ocorreu um pequeno aumento nos erros do modelo RNA-IP-26, com relação aos modelos RNA-IP-9 a 33. A partir do modelo RNA-IP-34 ocorre uma queda nos valores, com exceção dos modelos RNA-IP-38, 50, 56 e 57 que têm aumento nos seus valores. O maior valor para os erros ocorreu para o modelo RNA-IP-57. Na previsão, o EM tem mais flutuações nos valores quando comparados aos outros erros. Os maiores picos de EM ocorreram em RNA-IP-34 e 38 e os menores em RNA-IP-32 e 56. O MAE também tem muitas variações, mas oscilam menos que os valores de EM. O menor pico ocorreu para RNA-IP-17 e o maior para RNA-IP-55 e 57. Os valores de RMSE são constantes para maioria dos modelos, mas 7 modelos tiveram picos muitos altos e são eles: RNA-IP-9, 10, 32, 33, 55, 56 e 57. Os RMSE da previsão são zero para vários modelos, isto indica que as previsões são perfeitas. Assim, considerando os erros na maior parte do tempo eles são baixos, mesmo nos picos, indicando que as previsões podem ser consideradas boas. Os valores dos erros obtidos, tanto para o treinamento quanto para a previsão, indicam que os valores obtidos pelos os modelos são considerados bons, porque os erros foram pequenos.

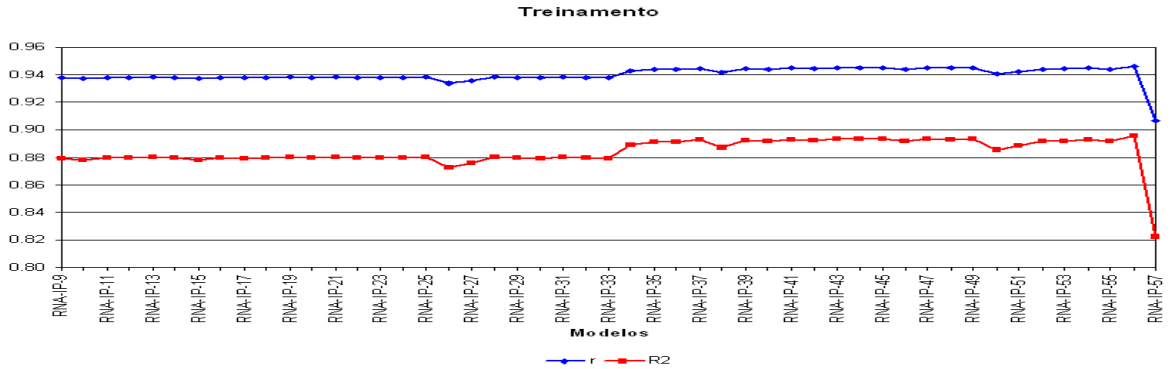
O coeficiente de correlação (r) e o coeficiente de determinação (R^2), para o treinamento (figura 5.23c), não tiveram muitas variações entre os modelos. A exceção foi o modelo RNA-IP-57 que teve um pico grande com relação aos outros modelos. Na previsão (figura 5.23 d), observa-se que alguns modelos tiveram valores diferentes da maioria dos modelos. Os modelos RNA-IP-17, 26, 32, 33, 34, 35, 38, 50, 51 e 55 a 57 tiveram oscilações em seus valores quando comparados com os outros modelos. O r também é um indicador, sendo que o valor ideal é 1 e como os valores estão próximos dele indica também uma boa previsão. Então, na figura 5.23 que os erros oscilaram mais na previsão do que no treinamento. O modelo RNA-IP-57 teve os maiores erros e os menores valores de r e R^2. Os valores flutuaram mais para os modelos RNA-IP-9 a 55 e os erros RNA-IP-5 a 32 são maiores quando comparados com os RNA-IP-33 a 55.



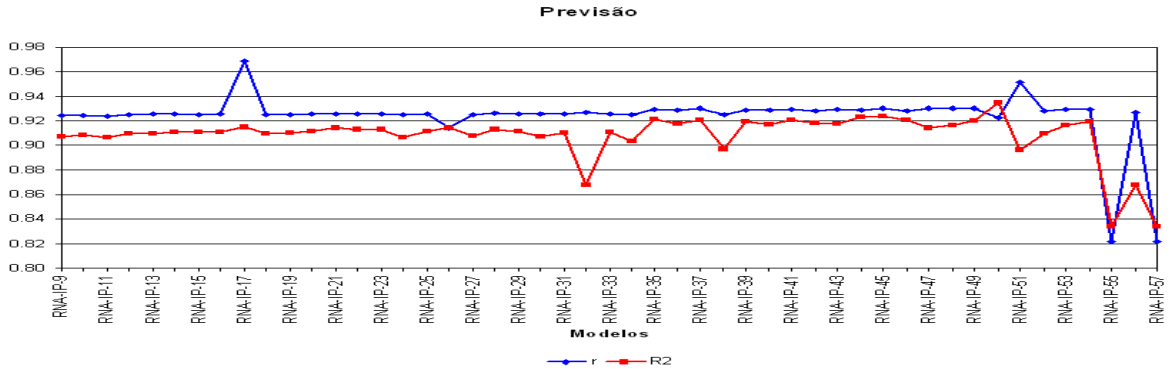
(a)



(b)



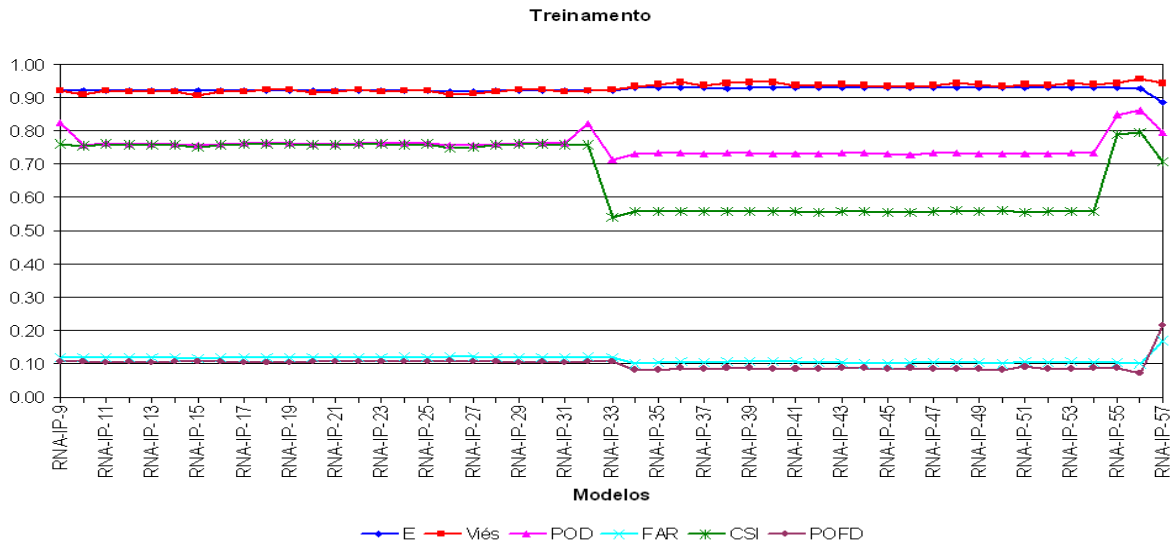
(c)



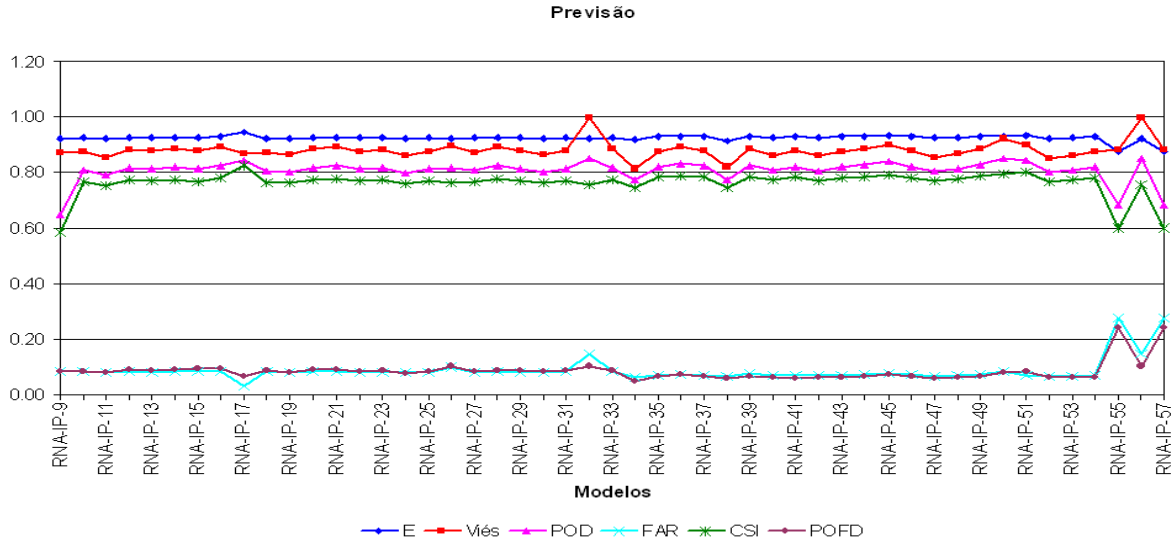
(d)

Figura 5.23: Evolução do erro médio (EM), erro médio absoluto (MAE), erro médio quadrático (RMSE), coeficiente de correlação (r) para os modelos RNA-IP-9 a 57. (a) treinamento (erros); (b) previsão (erros); (c) treinamento (coeficientes) e (d) previsão (coeficientes).

O cálculo da exatidão (E), da freqüência de bias (Viés), da probabilidade de detecção (POD), da taxa de falso alarme (FAR), do índice crítico de sucesso (CSI) e da probabilidade de falsa detecção (POFD) foi realizado utilizando os seguintes limites: a média e, a média ± 1 e ± 1,25 do desvio padrão. A figura 24 mostra os resultados para E, Viés, POD, FAR, CSI e PO, para o treinamento (figura 5.24a) e previsão (figura 5.24b), obtidos para os modelos RNA-IP-9 a 57.



(a)



(b)

Figura 5.24: Evolução exatidão (E) (azul), da freqüência de bias (Viés) (vermelha), da probabilidade de detecção (POD) (rosa), da taxa de falso alarme (FAR) (verde clara), do índice crítico de sucesso (CSI) (linha verde) e da probabilidade de falsa detecção (POFD) (roxa) para os modelos RNA-IP-9 a 57. (a) treinamento e (b) previsão.

Nota-se que, para o treinamento, os valores da exatidão (linha azul), para os modelos RNA-IP-33 a 55, têm um aumento em relação os modelos RNA-IP-9 a 32, sendo que isto não ocorre na previsão. Os modelos RNA-IP-56 a 57 sempre são menores que os outros modelos tanto para o treinamento como para a previsão. Os valores do treinamento tendem a ser mais constantes que o da previsão. No treinamento ocorre um aumento do valor na transição do modelo RNA-IP-32 para o RNA-IP-33 e novamente do RNA-IP-56 para o RNA-IP57.

Observa-se que para treinamento, os valores da frequência de bias (Viés) (linha vermelha), dos modelos RNA-IP-9 a 57 têm mais valores que são superestimados do para a previsão. Os valores baixos sempre são mais subestimados do que os valores altos, tanto para treinamento quanto para a previsão. Comparando as médias dos modelos nota-se que a previsão tende a subestimar mais os valores que o treinamento.

Observa-se que o POD (linha vermelha) e o CSI (linha verde), no treinamento (figura 5.24a) obteve resultados melhores que da previsão (figura 5.24b). Os valores de POD e CSI estão bem próximos uns dos outros, com exceção dos valores de CSI obtidos pelos modelos RNA-IP-33 a 54 que tiveram uma variação grande dos valores em relação ao POD, e em relação aos modelos RNA-IP-9 a 32. Na previsão, essa variação não ocorreu. Os valores de POFD (linha roxa) e FAR (linha azul clara) tiveram o mesmo comportamento e os valores estavam bem parecidos. Os valores oscilaram mais para a previsão do que para o treinamento.

Os valores da exatidão e viés foram maiores que para POD, CSI, FAR e POFD. Os menores valores foram para FAR e POFD. Os valores do treinamento são maiores que os obtidos para o treinamento. O modelo RNA-IP-9, quando comparado com os modelos RNA-IP-9 a 32, tem uma variação maior que os outros modelos. Na comparação dos modelos RNA-IP-33 a 55, o modelo RNA-IP-33 tem o menor valor em relação aos outros modelos e o maior acontece para o modelo RNA-IP-55. A maior variação ocorreu para os valores dos modelos RNA-IP-56 e 57.

A tabela 5.15 mostra a média dos valores obtidos pela E, Viés, POD, FAR, CSI e POFD, os piores resultados estão nas células vermelhas e as melhores nas células

azuis. O melhor resultado obtido pela exatidão (E) foi para o treinamento dos modelos RNA-IP-33 a 55 e o pior para previsão dos modelos RNA-IP-56 a 57. O viés teve se melhor resultado para a previsão dos modelos RNA-IP-56 a 57 e pior para a previsão dos modelos RNA-IP-33 a 55. Para os valores do POD, no treinamento, os modelos RNA-IP-56 a 57 tiveram a melhor capacidade de previsão que os outros modelos e os modelos RNA-IP-33 a 55, a pior capacidade de prever, também no treinamento. Os valores de FAR são aqueles em que verificam os valores que não foram observados. Os modelos RNA-IP-56 e 57, neste caso obtiveram as piores porcentagens de valores não observados para previsão e a melhor capacidade ocorreu também para previsão dos modelos RNA-IP-33 a 55. O CSI conseguiu prever melhor nos modelos RNA-IP-9 a 32, na previsão, e pior, no treinamento, dos modelos RNA-IP-33 a 55. O pior valor de POFD ocorre para o treinamento dos modelos RNA-IP-56 a 57 e o melhor para previsão dos modelos RNA-IP-33 a 55.

Tabela 5.15: Média do E, Viés, POD, FAR, CSI e POFD, para o treinamento, teste e previsão. As células em azul indicam os melhores resultados, e as vermelhas os piores.

		RNA-IP-9 a 32	RNA-IP-33 a 55	RNA-IP-56 a 57
E	Treinamento	92,1%	93%	90,6%
	Previsão	92,4%	92,4%	89,6%
VIÈS	Treinamento	91,9%	93,8%	94,8%
	Previsão	88,1%	87,2%	94%
POD	Treinamento	76,7%	73,6%	82,7%
	Previsão	80,7%	81%	76,6%
FAR	Treinamento	11,9%	10,4%	13,4%
	Previsão	8,2%	8%	21,1%
CSI	Treinamento	75,8%	56,5%	75%
	Previsão	76,2%	77%	67%
POFD	Treinamento	10,5%	8,7%	14,5%
	Previsão	8,7%	7,3%	17%

6. CONCLUSÕES E RECOMENDAÇÕES

Neste capítulo são apresentadas as conclusões deste trabalho, analisando a influência das variáveis, meteorológicas e antrópicas, e também, o desempenho das redes neurais em relação ao aprendizado e generalização observados nos testes efetuados. O capítulo também sugere os possíveis trabalhos futuros.

6.1. Conclusões

A proposta deste trabalho foi contribuir com um estudo para a previsão de demanda de água em sistemas urbanos de abastecimento, por meio de um modelo de redes neurais artificiais (RNA's), capaz de prever uma demanda de água univariada, com discretização horária. O modelo foi estruturado de forma a conter uma camada, de entrada, intermediária e saída.

Os estudos foram concentrados no Sistema de Abastecimento de Água da Região Metropolitana de São Paulo, especificamente na Estação de Tratamento de Água da Cantareira e no setor do Itaim Paulista, que faz parte do sistema Alto Tietê, considerado um setor predominantemente residencial.

A primeira etapa foi organizar os dados disponíveis. Os dados de consumo de água foram fornecidos pela SABESP, as variáveis meteorológicas pela CETESB e IAG e a precipitação pela FCTH. As variáveis meteorológicas fornecidas pela CETESB e de precipitação pela FCTH, continham uma quantidade muito grande de falhas, sendo pouco recomendável ou até certo ponto limitante fazer sua reconstrução através de técnicas de séries temporais. Então, optou-se em descartar esses dados e utilizar somente os dados das variáveis meteorológicas do IAG, que tinha um bom controle de qualidade. Os dados de consumo de água utilizados foram da SABESP, mesmo tendo algumas falhas foi possível fazer correção.

As condições climáticas, dia da semana, demanda dos usuários, sazonalidade, consumo excepcional e outros fatores influenciam os padrões do consumo. As variações anuais tendem a ocorrer com o crescimento populacional, observados no período em que foi estudado, com exceção dos anos de 2001 e 2004, devido ao racionamento de água ocorrido nesses períodos em decorrência da estiagem de anos anteriores.

A sazonalidade é observada por meio do consumo mensal, sendo que os maiores consumos são verificados no verão e os menores no inverno. As baixas temperaturas no inverno ajudam na redução do consumo, ao contrário do verão em que as temperaturas são elevadas e ocasionam esse aumento. O aumento e a redução das chuvas também influenciam no consumo de água. No verão, as chuvas são mais intensas, muitas vezes ocasionando enchentes, o que contribui para o aumento do consumo. A redução de chuva no inverno produz um clima mais seco, elevando também o consumo. Períodos muito longos sem chuva podem ocasionar um desabastecimento dos reservatórios, levando as companhias a criarem programas de racionamento de água. Os meses de maior e menor consumo são respectivamente, março e julho. No período de inverno, em agosto, ocorre um pico em relação aos outros meses devido a baixa umidade nesse período.

A localização do sistema e as peculiaridades desta região e o perfil populacional definem as características do consumo, influenciando em seus padrões. O comportamento do consumo nos dias da semana é variável. Então em um dia qualquer da semana, o consumo pode ser diferente da média da região, função de parâmetros que interferem no sistema. Esse fato evidencia a limitação do uso de uma curva média nas operações em tempo real. Em geral, o consumo mínimo ocorre no domingo devido a redução das atividades socioeconômicas. Dependendo da localização do sistema, o consumo máximo pode ocorrer em qualquer dia da semana. Para a ETA Alto Tietê, o valor máximo de consumo ocorre aos sábados, devido a região ser considerada dormitório.

O comportamento do consumo médio segue o ciclo anual da chuva e temperatura, onde o máximo ocorre no início do ano (verão) e o mínimo por volta dos meses de junho e julho, no inverno (período seco). Assim, o consumo máximo coincide com o

período de temperaturas e umidade mais elevadas em São Paulo. Por outro lado o consumo mínimo ocorre quando a temperatura é baixa. O consumo e a temperatura têm o mesmo comportamento; quando a temperatura aumenta ou diminui, o consumo tem mesma tendência.

O consumo horário varia com o decorrer do dia, perfil do consumidor de cada região e de ocorrências que possam ocorrer durante o período. O pico do consumo na média ocorre a partir das 12 horas. Durante a semana, o consumo horário pode sofrer variações. Os consumos em dias úteis, feriados, finais de semanas e férias são diferentes. Nos feriados, o consumo é sempre menor do que no período de férias e nos finais de semana; já nos dias úteis, a tendência de consumo é maior.

A correlação do consumo de água com as variáveis antrópicas e meteorológicas mostraram que o consumo do setor Itaim Paulista tem melhor correlação com a hora, temperatura, umidade, direção e velocidade do vento e para ETA Cantareira, com a hora, dia da semana, tipo do dia, temperatura, umidade, direção e velocidade do vento. Como a ETA Cantareira é a principal do sistema produtor da Cantareira, a que distribui água para todo o sistema, os valores da correlação foram menores do que do setor Itaim Paulista, embora estejam correlacionadas com mais variáveis do que o setor Itaim Paulista. Isto indica que os sistemas produtores tendem a ter mais variáveis correlacionadas a ele, mas valor de correlação menor. Isso ocorreu porque o setor Itaim Paulista tem influência e peculiaridades de uma área residencial e a ETA Cantareira, como abastece uma área mais complexa e engloba todos os tipos de consumo tem influências de vários tipos fazendo que ela tenha correlação com mais variáveis que o setor Itaim Paulista.

Os resultados obtidos pela RNA mostram que para grandes sistemas como a ETA Cantareira os resultados obtidos na previsão, utilizando a média do consumo de 1 hora, não resultam em bons resultados. Observou-se que quando foi utilizada a memória para os dados, médios e previsão de 1 hora os resultados melhoram um pouco. Os erros MAE, EM, RMSE e o “r”, obtidos para RNA-C-1 foram piores que para os outros modelos. Os resultados de E, BIAS, POD, FAR, CSI e POFD também não foram considerados bons. Os melhores resultados ocorreram para o modelo RNA-C-8 que utilizou a média de 12horas e a memória de variáveis. Observou-se

que, quando se utilizou a memória do sistema, os resultados melhoram. Pode-se considerar que os modelos fizeram boas previsões, podendo ser utilizados para fazer a previsão do ETA Cantareira, e obter um bom resultado para tomadas de decisões quanto o uso da água com antecipação de 12 horas. Então, recomenda-se que para sistema como a ETA Cantareira, seja utilizada uma RNA com a configuração do modelo RNA-C-8 para previsão de consumo, onde é feito a previsão para 12 horas à frente.

Foram realizados vários testes de previsão de consumo com a RNA para o setor Itaim Paulista. Os primeiros testes tiveram as mesmas características com exceção do número da camada escondida. O teste do modelo RNA-IP-2 observou que o aumento da camada escondida não melhorou o ajuste da curva em relação aos dados observados, mas no treinamento e previsão houve melhora em relação ao modelo RNA-IP-1. Desta forma, esta melhoria não compensa devido ao aumento verificado no tempo de processamento, sendo melhor manter uma camada escondida para os demais modelos.

O modelo RNA-IP-3 não utilizou as variáveis antrópicas (Va) como entrada, somente as Vm e os consumos, nos tempos t e t-1. O modelo RNA-IP-4 utilizou somente a temperatura, a umidade e consumo nos tempos t e t-1 como entrada. A saída para os dois modelos foi o consumo em t+1. Os resultados dos modelos mostraram que é viável utilizar essas configurações para previsão de consumo. Os piores foram verificados para RNA-IP-2 e RNA-IP-1, respectivamente para teste e previsão. No treinamento, os valores obtidos apresentaram pouca variação o que impediu de avaliar qual foi o melhor resultado. Os valores de RNA-IP-2 foram melhores que os outros modelos, mas não compensa utilizar esse modelo porque o tempo de processamento foi maior do que dos outros modelos e a melhora dos resultados não foram tão significantes.

Os modelos RNA-IP-5 a 8 utilizaram a média de 6 horas do consumo. A previsão foi realizada a cada 6 horas até completar 24 horas. A previsão para 6 horas tende a subestimar os valores, mas, com o passar do tempo da previsão, a curva tende a se ajustar melhor quando comparada aos dados observados. O modelo RNA-IP-6 obteve os piores resultados, isto porque os outros modelos utilizavam a memória do

sistema que colaborou para reduzir os erros e melhorar os resultados, tanto que os valores obtidos para os outros modelos foram praticamente iguais no treinamento, mas na previsão ocorreram pequenas variações nos valores. Assim, é melhor fazer a previsão para o setor Itaim Paulista para cada 6 horas, porque os resultados são melhores dos que obtidos para a média de 1 hora.

A previsão ótima foi feita para testar as previsões de 1 a 24 horas, utilizando os dados observados das variáveis como fossem dados de modelos de previsão. Os modelos RNA-IP-9 a 32 utilizaram dados de tempo t e t-1 e os modelos RNA-IP-33 a 55 utilizaram a memória do sistema, nos tempos t, t-1, t-2. Os modelos RNA-IP-56 e 57 têm duas particularidades, o primeiro utilizou todas as memórias do sistema e o segundo utilizou também as memórias com exceção das memórias do consumo anterior. Com exceção da defasagem apresentada nos valores previstos pelos modelos, em relação aos dados observados, os valores previstos pelos modelos RNA-IP-33 a 55 se ajustaram melhor do que os valores dos RNA-IP-9 a 32, indicando novamente que a memória é útil no modelo. Os resultados dos modelos RNA-IP-56 e 57 são menores do que os valores observados, indicando que o aumento da memória não melhora os valores previstos e sim aumenta o tempo de processamento. Então, para previsão ótima é melhor utilizar as configurações utilizadas nos modelos RNA-IP-33 a 55, pois assim teremos melhores resultados na previsão.

Observou-se que a memória do sistema é um fator importante para previsão. Para períodos de curto prazo é mais fácil obter resultados melhores para um sistema menor do que para um maior. Neste caso, recomenda-se previsões para períodos mais longos. Pode-se dizer que os objetivos foram alcançados. A previsão horária foi obtida com níveis de erros aceitáveis e comparados os resultados de todos os casos, observa-se a mesma tendência para todos, isto é, erros menores. Estes resultados podem ajudar na operação do sistema utilizado, conjuntamente com modelos de previsão meteorológica.

6.2. Recomendações para trabalhos futuros

Sugere-se, como continuidade deste trabalho, a implementação de variações do modelo de redes neurais artificiais, que produza uma previsão multivariada com discretização horária.

Novas topologias de redes neurais devem ser estudadas para a previsão de demanda de água, para que os resultados já obtidos neste estudo possam ser comparados a outros e permitir uma melhor avaliação dos resultados.

Utilizar diferentes algoritmos para treinamento da RNA, função base radial, algoritmos genéticos e auto-organização de rede.

Verificar a possibilidade de implementar políticas operacionais obtidas a partir de modelo hidráulico de otimização com base nas curvas de demanda previstas, por redes neurais artificiais.

Acrescentar na entrada da RNA aspectos operacionais do sistema de abastecimento para tentar, minimizar custo operacional e acoplar este modelo a um modelo de previsão meteorológica.

Investigar variáveis socioeconômicas, tais como características da habitação, valor da propriedade, uso da terra, situação econômica e renda.

Testar o modelo obtido para outras regiões tais como: comerciais e industriais.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ADAMOWSKI, J.F. Peak daily water demand forecast modeling using artificial neural networks. **Journal of Water Resources planning and management.** v. 134, nº. 2, 2008., ISSN 0733-9496/2008/2-119–128

ADAMOWSKI, J. ; KARAPATAKI, C. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ann learning algorithms. **Journal of Hydrologic Engineering**, v. 15, nº 10, p. 729–743, 2010. ISSN 1084-0699/2010/10

AN, A.; SHAN, N.; CHAN, C.; CERCONE, N.; ZIARKO, W. Discovering rules for water demand prediction: an enhanced rough-set approach. **Engineering Applications of Artificial Intelligence**, v.9, p. 645-653, 1996.

BALLING JR., R. C.; GOBER P. Climate variability and residential water use in the city of Phoenix, Arizona. **JOURNAL OF APPLIED METEOROLOGY AND CLIMATOLOGY**, v. 46, nº 7, 2007. ISSN 1558/8424

BARRETO, D. Perfil do consumo residencial e usos finais da água. **Ambiente Construído**, v. 8, n. 2, p. 23-40, 2008.

BORGES, A.S.; PEREIRA FILHO, A.J. Análise das condições meteorológicas de superfície antecedentes à precipitação de origem convectiva na cidade de São Paulo. **In: XI Congresso Brasileiro de Meteorologia**, SBMET, Rio de Janeiro, out./2000. II Seminário Brasileiro de Hidrometeorologia.

BORGES, V.M.N.A **Acoplamento de um modelo de previsão de demanda de água a um modelo simulador em tempo real – um estudo de caso:**sistema adutor metropolitano de São Paulo. São Paulo, 2003. 205p. Dissertação (Mestrado) – Escola Politécnica , Universidade de São Paulo.

BOUGADIS, J.; ADAMOWSKI, K.; DIDUCH, R. Short-term municipal water demand forecasting. **HYDROLOGICAL PROCESSES**, v.19, no. 1, p.137–148, 2005.

BuHAMRA, S.; SMAOUI, N.; GABR, M. The Box-Jenkins analysis and neural networks: prediction and time series modelling. **Applied Mathematical Modelling**, v. 27, p. 805-15, 2003.

CARVER, P.H., BOLAND, J.J. Short and long-run effects of price on municipal water use. **Water Resources Research**, v. 16(4), p. 609-16, 1980.

CASSUTO, A.E.; RYAN, S. Effect of price on the residential demand for water within an agency. **Water Resources Bulletin**, v. 15, nº 2, p. 345-53, 1979.

Chang, N. B., Makkeasorn, A. Water demand analysis in urban region by neural network models. **8th Annual Water Distribution Systems Analysis Symposium**, Cincinnati, Ohio, USA, August 27-30, 2006

CYBENKO,G. Approximation by superpositions of a sigmoidal function. **Mathematical Control Signals Systems**, v.2, p. 303–314, 1989.

DAVIS, M.M.; AQUILANO, N.J.; CHASE, R.B. **Fundamentos da administração da produção.** 3ª ed., Porto Alegre, Bookman, 2001.

FALKENBERG, A.V. **Previsão de consumo urbano de água em curto prazo**. Curitiba, 2005. 89p. Dissertação (Mestrado) –Universidade Federal do Paraná.

FALKENBERG, A.V.; DYMINSKI, A. S.; RIBEIRO, E.P. Redes neurais artificiais aplicadas à previsão de consumo de água. **In: Proceedings of the VI Brazilian Conference on Neural Networks** – VI Congresso Brasileiro de Redes Neurais, p. 319-24, São Paulo, SP, 2003.

FALKENBERG, A.V.; DYMINSKI, A. S.; RIBEIRO, E.P. Neural network applied on water demand forecastng. **In: XXIV Iberian Latin-American Congress on Computational Methods in Engineering**, Ouro Preto, MG, 2003.

FIRAT, M.; YURDUSEV, M. A.; ERKANA, M. Evaluation of artificial neural network techniques for municipal water consumption modeling. **Water Resource Management**, v.23, p. 617-632, 2009.

FIRAT, M.; YURDUSEV, M. A.; ERKANA, M. Comparative analysis of neural network techniques for predicting water consumption time series. **Journal Hydrology**, v.384, p.46-51, 2010.

FLETCHER, R. **Practical Methods of Optimization**, 2ª ed., Chichester, John Wiley & Sons ,1987.

GATO, S.; JAYASURIYA, N.; ROBERTS, P. Temperature and rainfall thresholds for base use urban water demand modeling. **Journal of Hydrology**, v.337, p. 364-376, 2007.

GIANSANTI, A.E. **Água como um bem econômico.** São Paulo, 1993. 431p. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo.

GINEZ, W.G.; PEREIRA FILHO, A.J. Sistema de controle da produção e adução d' água na região metropolitana de São Paulo: presente e futuro. **In: XV Simpósio Brasileiro de Recursos Hídricos**, Curitiba, PR, nov. 2003.

GUPTA, H.V.; HSU, K.; SOROOSHIAN, S. Superior training of artificial neural networks using weight-space partitioning. In: International Conference on Neural Networks (ICNN'97), Texas, **Proceedings of the IEEE 1997**, pp.1919-1923, 1997.

HAYKIN, S. **Neural networks: a comprehensive foundation.** Ontario, Prentice Hall International, Inc., 1994.

HERRERA, M.; TORGO, L.; IZQUIERDO, J.; PÉREZ-GARCÍA, R. Predictive models for forecasting hourly urban water demand. **Journal Hydrology**, v.387, p. 141-150, 2010.

HORNIK, K., STINCHCOMBE, M., WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Networks** , v. 2, p. 359–366, 1989.

HOWE, C.W.; LINAWEAVER JR., F.P. The impact of price on residential water demand and its relation to system design and price structure. **Water Resources Research**, v. 3 (1), p. 13-32, 1967.

HSU, K.; GUPTA, H. V; SOROOSHIAN, S. Artificial neural network modeling of the rainfall-runoff process. **Water Resources Research**, v. 31(10), p. 2517-2530, 1995.

HSU, K.L; GUPTA, H.V and SOROOSHIAN, S. **A superior training strategy for three-layer feedforward artificial neural networks.** Tucson, University of Arizona, 1996. (Technique report, HWR nº 96-030, Departament of Hydrology and Water Resources)

KIRBY, M.; MIRANDA, R. Nonlinear reduction of high-dimensional dynamical systems via neural network. **Physical Review Letters**, v. 72, nº 12, p.1822-25, 1994.

LAPEDES, A.; FARBER, R. Nonlinear signal processing using neural networks: prediction and signal modeling. Los Alamos, 1987. **Preprint**.

LACHTERMACHER, G., FULLER, J.D. Backpropagation in time- series forecasting. **Journal of Forecasting**, v.14, p. 381–393. 1995.

LERTPALANGSUNTI, N.; CHAN, C.W.; MASON, R.; TONTIWACHWUTHIKUL, P. A toolset for construction of hybrid intelligent forecasting systems: application for water demand prediction. **Artificial Intelligence in Engineering**, v. 13, p. 21-42, 1999.

LIU, J.; SAVENIJE, H.H.G.; XU, J. Forecast of water demand in Weinan City in China using WDF-ANN model. **Physics and Chemistry of the Earth**, v. 28, p. 219-24, 2003.

LIU, K.; SUBBARAYAN, S.; SHOULTS, R.R.; MANRY, M.T.; KWAN, C.; LEWIS, F.L.; NACCARINO, J. Comparison of very short-term load forecasting techniques. **IEEE Transactions on Power Systems**, v. 11 (2), p. 877-82, 1996.

MAIDMENT, D.R.; MIAOU, S.P. Daily water use in nine cities. **Water Resources Research**, vol. 22 (6), p. 845-51, 1986.

MAIDMENT, D.R.; MIAOU, S.P.; Crawford, M.M. Transfer function models of daily urban water use. **Water Resources Research**, vol. 21(4), p. 425-32, 1985.

MARTINI, P.G.; LAUGENI, F.P. **Administração da produção.** São Paulo, Saraiva, 1998.

McCULLOCH, W.S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics,** 5: pp.115-133, 1943.

MINGQI, C.; JUNPING, L. Water demand prediction model based on radial basis function neural network. In: The International Conference on Information Science and Engineering, 1., 2009, Nanjing, Jiangsu: **Proceedings** 2009. p. 5295 – 5298. ISBN: 978-0-7695-3887-7. Disponível em < http://www.computer.org/portal/web/csdl/doi/10.1109/ICISE.2009.1343>. Acesso em: 10 nov. 2010.

MORGAN, W.D.; SMOLEN, J.C. Climatic indicators in the estimation of municipal water demand. **Water Resources Bulletin**, v. 12 (3), p. 511-18, 1976

MUKHOPADHYAY, A.; AKBER, A.; AL-AWADI, E. Analysis of freshwater consumption patterns in the private residences of Kwait. **Urban Water,** n. 3, p. 53-62, 2001.

NARCHI, H. **Fatores intervenientes na demanda doméstica de água. estudo para uma amostra de consumidores da cidade de São Paulo.** São Paulo, 1987. 158p. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo.

ODAN, F.K. **Previsão de demanda para sistemas de abastecimento de água**. São Carlos, 2010. 77p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

PEREIRA FILHO, A.J. Radar measurements of tropical summer convection: urban feedback on flash floods. In: 29th Radar Conference, MAS, Montreal, Canadá, jul./1999. **Preprints** (MAS, Paper 17.4).

PEREIRA FILHO, A. J., L. C. T. RODRIGUES e W. GINEZ, 2004a: Impacto das condições meteorológicas no consumo de água na Região Metropolitana de São Paulo. **In:** 1º Seminário de Planejamento Urbano e Desastres Naturais, XIII Congresso Brasileiro de Meteorologia, Fortaleza, CE.

PEREIRA FILHO, A. J., M. T. L. BARROS, R. HALLAK E A. W. GANDU, 2004b: Enchentes na Região Metropolitana de São Paulo: aspectos de mesoescala e avaliação de impactos. **In:** 1º Seminário de Planejamento Urbano e Desastres Naturais, XIII Congresso Brasileiro de Meteorologia, Fortaleza, CE.

PAO, J.H. **Adaptive pattern recognition and neural networks.** Addison-Wesley, 1989.

PORTO, M. **Recursos hídricos e sanemento na região metropolitana de São Paulo:um desafio do tamanho da cidade.** Série Água do Brasil, 2003.1ª edição, Brasília.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, 323, p. 533-536, 1986.

RUMELHART, D.E., HINTON, G.E., WILLIAMS, R.J. Learning exinternal representation by back- propagating errors. **In:** Rumelhart, D.E., McCleland, J.L., the PDP Research Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, MA. 1986.

RUMELHART, D.E., DURBIN, R., GOLDEN, R., CHAUVIN, Y., 1995. Backpropagation: the basic theory. **In:** Chauvin, Y., Rumelhart, D.E. (Eds.), Backpropagation: Theory, Architectures, and Applications. Lawrence Erlbaum Associates, New Jersey, pp. 1–34.

SABESP. **Palestra proferida ao comitê de bacia hidrográfica do Alto Tietê.** Sabesp, São Paulo, ago. 2001.

SABESP **Desafios e oportunidades.** Planejamento Operacional, São Paulo, 2002.

SANTOS, C.C. **Modelagem de bacias urbanas com redes neurais artificiais.** São Paulo, 2001. 105p. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo.

SANTOS, C. C.; PERREIRA FILHO, A. J. Consumo de água na cidade de São Paulo. **In: Simpósio Brasileiro de Sensoriamento Remoto** (SBSR), 14, 2009, Natal. Anais http://marte.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2008/11.18.13.31/doc/4845-4850.pdf

SANTOS, C. C.; PEREIRA FILHO, A. J.; GINEZ, W. G. Características do consumo de água na região metropolitana de São Paulo e os impactos do tempo e do clima. **In: Congresso Brasileiro de Meteorologia** (CBMET), 16, 2006, Florianópolis. Anais... CD-ROM.

SANTOS, C. C.; PEREIRA FILHO, A. J.; GINEZ, W. G. Características sazonais do consumo de água na região metropolitana de São Paulo e os impactos de variáveis ambientais. **In: Encontro Técnico AESABESP**, 13., 2007, São Paulo. Anais...CD-ROM.

SECCO, C.K.Z. **Operação de sistemas de abastecimento com limitação da produção de água:** estudo de caso da região metropolitana de São Paulo. São Paulo, 2002. 204p. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo.

SETTI, A.A.; LIMA, J.E.F.W.; CHAVES, A.G.M.; PEREIRA, I.C. **Introdução ao gerenciamento de recursos hídricos.** Brasília, Agência Nacional de Energia Elétrica; Agência Nacional de Águas, 2001.

SHAFER, S.M; MEREDITH, J.R. **Operations management.** New York, John Wiley & Sons, 1998.

SMITH, M. **Neural networks for statistical modeling.** New York, Van Nostrand Reinhold, 1993.

SILVA, C.A.M.; ZAHED FILHO, K.; ARNAUT, W.L.B. Modelo previsional de consumos de água em tempo real baseado em redes neurais. In: ISA Show Brasil 2001. **Anais ISA Show Brasil 2001**. São Paulo: ISA South America, 2001.

TEIXEIRA, J.L. Futuro Ameaçado: vem aí a cobrança pelo uso da água. Será o fim do desperdício? **Problemas Brasileiros**, n. 338, mar./abr. 2000.

TRAUTWEIN JR., B. **Avaliação de métodos para previsão de consumo de água para curtíssimo prazo:** um estudo de caso em empresa de Saneamento. Curitiba, 2004. 107p. Dissertação (Mestrado) – Pontifícia Universidade Católica do Paraná.

TSUTIYA. M. T. **Abastecimento de Água.** Departamento de Engenharia Hidráulica e Sanitária da Escola Politécnica da Universidade de São Paulo. 3 ª Edição. São Paulo, 2006.

VEMURI, V.R. **Artificial neural networks:** forecsating times series. Los Alamitos, IEEE Computer Society Press, 1994.

VICENTE, R.V. **Modelo de operação para centros de controle de sistemas de abastecimento de água –** estudo de caso: sistema adutor metroploitano de São Paulo. 2005. 140p. Dissertação(Mestrado) – Escola Politécnica , Universidade de São Paulo.

VOGEL, R.M.; BELL, C.J.; FENNESSEY, N.M. Climate, streamflow and water supply in the northeastern United States. **Journal of Hydrology**, v. 198,p. 42-68, 1997.

WASSERMAN, P.D. **Neural computing: theory and practice**. New York, Van Nostrand, Reinhold, 1989.

WEEKS, C.R.; McMAHON, T.A. A comparison of water use, Australia and the US. **Journal of American Water Works Association**, v. 65 (4), p. 232-37, 1973.

WIDROW, B.; HOFF, M. E. Adaptative switching circuits. **Instituto of Radio Engineers, Western Eletronic Show and Convetion***,* p. 96-104, 1960.

WILLSIE, R.H.; PRATT, H.L. Water use relationships and projection corresponding with regional growth, Seattle region. **Water Resources Bulletin**, v. 2, n. 10, p. 360-71, 1974.

Wilks, D.S. **Statistical Methods in the Atmospheric Sciences.** Academic Press. Second Edition. 2006.

WONG, S.T. A model on municipal water demand: a case study of northeastern Illionois. **Land Economics**, v. 48 (1), p. 34-44, 1972.

YAMAUCHI, H.; HUANG, W. Alternative models for estimating the time series components of water consumption data. **Water Resources Bulletin**, v. 13 (3), p. 599-610, 1977.

YOUNG, R.A. Price elasticity of demand for municipal water: a case study Tucson, Arizona. **Water Resources Research**, v. 9 (4), p. 1068-72, 1973.

ZAHED FILHO, K. **Previsão de demanda de consumo em tempo real no desenvolvimento operacional de sistemas de distribuição de água.** São Paulo, 1990. 135p. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo.

ZHANG, G.; PATUWO, B. E..; HU, M.Y. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, v. 14, nº 1, p. 35-62, 1998

ZHANG, J.J.; SONG, R.; BHASKAR, N. R., FRENCH, M. N. Short-term water demand forecasting: a case study. **8th Annual Water Distribution Systems Analysis Symposium**, Cincinnati, Ohio, USA, August 27-30, 2006.

ZHOU, S.L.; McMAHON, T.A.; WANG, Q.J. Frequency analysis of water consumption for metropolitan area of Melbourne. **Journal of Hydrology**, vol. 247, p. 72-84, 2001.

ZHOU, S.L.; McMAHON, T.A.; WALTON, A.; LEWIS, J. Forecasting daily urban water demand: a case study of Melbourne. **Journal of Hydrology**, n. 236, p. 153-64, 2000.

ZHOU, S.L.; McMAHON, T.A.; WALTON, A.; LEWIS, J. Forecasting operational demand for an urban water supply zone. **Journal of Hydrology**, n. 259, p. 189-202, 2002.

ANEXO

O CD, anexo único, é parte integrante desta tese e contém as tabelas com registros de dados observados das estações telemétricas do Alto Tietê e CETESB.