UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS AMBIENTAIS

PAULO JOSÉ ALVES DE OLIVEIRA

PREVISÃO DA DEMANDA DE ÁGUA EM SISTEMAS URBANOS DE ABASTECIMENTO ATRAVÉS DE MODELOS ARIMA SAZONAIS

CAMPO GRANDE
2012

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS AMBIENTAIS

PAULO JOSÉ ALVES DE OLIVEIRA

PREVISÃO DA DEMANDA DE ÁGUA EM SISTEMAS DE ABASTECIMENTO ATRAVÉS DE MODELOS ARIMA SAZONAIS

Dissertação apresentada para obtenção do grau de Mestre no Programa de Pós-Graduação em Tecnologias Ambientais da Universidade Federal de Mato Grosso do Sul, na área de concentração em Saneamento Ambiental e Recursos Hídricos.

ORIENTADOR: Prof. Dr. Peter Batista Cheung

CO-ORIENTADOR: Prof. Dr. Johanes Jérson Janzen

Aprovada em:

Banca Examinadora:

Prof. Dr.
Orientador - Instituição

Prof. Dr.
Instituição

Prof. Dr.
Instituição

Campo Grande, MS
2012

AGRADECIMENTOS

Aos Professores Doutor Peter Batista Cheung e Johanes Jérson Janzen, pela orientação fornecida durante a elaboração deste trabalho e paciência durante o percurso.

À Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pela bolsa de estudo concedida.

A todos os colegas, professores e funcionários do Departamento de Hidráulica e Transportes da UFMS, pela colaboração.

A minha mãe pelo explícito incentivo durante toda a minha vida acadêmica.

Ao grande mestre Narumi Abe pelo apoio em meus primeiros passos na vida acadêmica e em programação.

Ao meu grande mestre Lauro Maymone Coelho Netto pela paciência e sabedoria influenciados na logística de minha vida.

Ao meu grande amigo Gustavo Kevilli Busanelli pelas inúmeras horas de aconselhamento e discussões filosóficas.

RESUMO

OLIVEIRA, P. J. A. de (2012). *Previsão de demanda de água em sistemas de abastecimento através de modelos arima sazonais. Campo Grande, 2012. 57 p. Dissertação (Mestrado) - Universidade Federal de Mato Grosso do Sul, Brasil.*

Previsões precisas da demanda de água em curto prazo são informações importantes na tomada de decisão do gerenciamento de recursos hídricos. Elas podem ser aplicadas em projetos, estudos futuros, e posteriores problemas de qualidade na distribuição. Muitos fatores podem influenciar direta ou indiretamente o consumo de água, porém, como o principal deles é o próprio padrão de vida dos consumidores, a série histórica de demandas exibe uma forte similaridade cíclica entre os dias de uma mesma semana e entre semanas adjacentes. Estas relações de semelhança foram aqui utilizadas para criar um preditor horário prático, simples e intuitivo baseado em uma metodologia de série temporal univariada com sazonalidade dupla. Um algoritmo desenvolvido recentemente denominado Busca Harmônica foi aplicado na estimativa dos parâmetros do modelo ARIMA usando dados históricos de demanda de água. O algoritmo de otimização de Busca Harmônica é baseado no processo da teoria musical de busca por um estado de perfeita harmonia. Uma função objetivo de mínimos quadrados foi formulada minimizando-se a série de erros aleatórios obtida. A ferramenta desenvolvida foi disponibilizada como “toolkit” de programação, o que permitiu sua integração como um aplicativo de análise e simulação online. Foram gerados resultados de previsão para dados de demandas em quatro setores de fornecimento para duas cidades brasileiras, sendo testado o compotamento do modelo perante as estações do ano e às mudanças regionais. Finalmente, algumas considerações sobre a aplicação de técnicas de detecção e correção de erros (outliers) são apresentadas.

Palavras-Chave: Série temporais; Sazonalidade dupla; Estimação de parâmetros

ABSTRACT

OLIVEIRA, P. J. A. de (2012). *Water demand forecasting on urban water distribution systems by seazonal arima models. Campo Grande, 2012. 57 p. Master Dissertation – Federal University of Mato Grosso do Sul, Brasil (in Portuguese).*

Accurate short term forecasts of water demand are significant information in decision making process of water resources management. They can be applied to design, future studies and quality problems in water distribution systems. Many factors can influence directly or indirectly the water demand, however, as the main one is the proper behavior of consumers, the series of demands presents a strong relationship between the adjacent days on a week and between adjacent weeks. These similarity relationships were used to create a practical, simple and intuitive forecasting method based on a univariate time series with double seasonality. A recently developed algorithm called Harmonic Search was applied to the estimation of the ARIMA model based on historical data on water demand. The optimization of Harmonic Search is based on the process of music theory to search for a state of perfect harmony. A least square objective function was formulated by minimizing the number of obtained random errors. The developed tool was available as a programming toolkit, allowing its integration with an analysis tool and online simulation. Forecasting results were generated for data demands of four supply sectors in two Brazilian cities allowing to verify the model behavior by seasons and regional changes. Finally, some considerations on the application of techniques for detecting and correcting outliers are presented.

Keywords: Time Series; Double Seazonality; Parameter Estimation

SUMÁRIO

AGRADECIMENTOS iii

RESUMO iiiv

ABSTRACT v

LISTA DE FIGURAS viii

LISTA DE TABELAS ix

LISTA DE SIGLAS E SIMBOLOS x

1 Introdução 1

2 Pesquisa Bibliográfica 3

2.1 Operação dos sistemas de distribuição de água 3

2.2 Operação otimizada e economia de energia 4

2.3 Como um algoritmo de previsão de demandas pode ajudar 5

2.4 Fatores que influenciam o consumo de água 7

2.5 Conceitos fundamentais de séries temporais 8

2.5.1 Modelo estocástico X Modelo determinístico 9

2.5.2 Estacionariedade 10

2.5.3 Operador de atraso (B) 10

2.5.4 Operador de auto-diferenças 11

2.5.5 Estacionariedade homogênea 11

2.5.6 Função de autocorrelação 12

2.5.7 Função de auto-correlação parcial 13

2.5.8 Ruído Branco 13

2.5.9 Método da decomposição 13

2.6 Características da série de demandas de água 14

2.7 Modelos ARIMA e SARIMA 16

2.8 Critérios para escolha de um algoritmo de previsão 20

3 Metodologia 22

3.1 Procedimento de aplicação dos modelos sarima 22

3.1.1 Identificação 22
3.1.2 Estimação 22
3.1.3 Diagnóstico 22
3.1.4 Previsão 23
3.2 Algoritmo de busca harmônica 23
3.2.1 Função Objetivo 24
3.2.2 Inicializar a memória harmônica (HM) 24
3.2.3 Improvisar uma nova harmonia a partir de HM 25
3.2.4 Atualizar a memória harmônica 25
3.2.5 Checar o critério de parada 25
3.3 Planejamento de experimentos 26
3.3.1 Método da Superfície de Resposta (RSM) 27
3.3.2 Planejamento fatorial 27
4 Resultados e Discussões 28
4.1 Estudo de caso 28
4.2 Tratamentos iniciais 29
4.3 Identificação do modelo 31
4.4 Estimação dos modelos 34
4.4.1 Otimização do Algoritmo de Busca 34
4.4.2 Análise dos modelos obtidos 37
4.5 Previsões Obtidas 39
5 Conclusões 42
6 Referências Bibliográficas 44

LISTA DE FIGURAS

Figura 2.1. Comportamento de série estacionária chamada passeio aleatório. 10

Figura 2.2. Aplicação de operador linear de diferenças de ordem unitária..... 12

Figura 2.3. Perfil típico de consumo para setores residenciais..... 14

Figura 3.1. Fluxograma do algoritmo de Busca Harmônica. 26

Figura 4.1. Mapa com classificação climática das cidades monitoradas. Fonte: Base Cartográfica IBGE, Ano 2001, Clima Adaptado A. STRAHLER. 28

Figura 4.2. Vazão observada no período de uma semana. 29

Figura 4.3. Falha em medição de vazão: possivelmente por manobra em reservatório. 30

Figura 4.4. Falha em medição de vazão: possivelmente por falha no medidor. 31

Figura 4.5. Falha em medição de vazão: possivelmente por mudança de setorização ou vazamento..... 31

Figura 4.6. FAC da série original (a), série diferenciada diaria (b) e diferenciada semanal (c).32

Figura 4.7. (a) função de auto-correlação da série diferenciada e (b) função de auto-correlação parcial da série diferenciada. 33

Figura 4.8. Coeficiente de correlação entre as funções de autocorrelação dos setores de fornecimento..... 34

Figura 4.9. Superfícies ajustadas para análise da influência dos fatores. 35

Figura 4.10. Resultados de previsão para horizonte de 24 horas em uma semana típica. 39

Figura 4.11. Avaliação do erro de previsão para diferentes horizontes para modelos com e sem retirada de valores outliers. 40

Figura 4.12. Avaliação do erro de previsão para diferentes horizontes nas diferentes estações do ano..... 41

LISTA DE TABELAS

Tabela 2.1. Regras resumo para a identificação de modelos ARMA. 18

Tabela 4.1. Intervalo dos fatores para a primeira avaliação. 35

Tabela 4.2. Resultados ordenados para o número de iterações (ótimo em cinza). 36

Tabela 4.3. Resumo dos coeficientes estimados. 38

Tabela 4.4. Resumo da comparação dos modelos. 38

LISTA DE SIGLAS E SIMBOLOS

Siglas

AIC	Critério de Informação de Akaike
AR	Modelo Auto-regressivo
ARIMA	Modelo Auto-regressivo Integrado de Médias Móveis
ARMA	Modelo Auto-regressivo de Médias Móveis
BIC	Critério de Informação Bayesiano
CB	Conjunto motor-bomba
DMA	Distrito de medição
FAC	Função de auto-correlação
FACP	Função de auto-correlação parcial
GA	Algoritmos genéticos
Hmcr	Taxa de escolha de valores da memória harmônica
HS	Algoritmo meta-heurístico de busca harmônica
MA	Médias Móveis
MAPE	Erro percentual absoluto médio
MV	Método de máxima verossimilhança
PAR	Taxa de deslocamento lateral
RMSE	Erro quadrático médio
RSM	Método da superfície de resposta
SAA	Sistema de abastecimento de água
SARIMA	Modelo Sazonal Auto-regressivo Integrado de Médias Móveis
SARMA	Modelo Sazonal Auto-regressivo de Médias Móveis
SCADA	Sistema de supervisão e controle
SDA	Sistema de distribuição de água
SMA	Modelo Sazonal de Médias Móveis

Símbolos

IT	Número de iterações
SQ	Soma quadrática
P1, P2	ordens dos coeficientes auto-regressivos sazonais

Q1, Q2	ordens dos coeficientes de médias móveis sazonais
S1	período diário
S2	período semanal
B	operador de atraso
p	ordem da parcela auto-regressiva
d	ordem das diferenças
q	ordem da parcela de médias móveis
P	ordem da parcela auto-regressiva sazonal
D	ordem das diferenças sazonais;
Q	ordem da parcela de médias móveis sazonal
s	tamanho da sazonalidade
∇^d	operador de diferenças
∇_s^D	operador de diferenças sazonal
a_t	parcela de erro do sistema
$\varphi_p(B)$	polinômio auto-regressivo de ordens *p*
$\theta_q(B)$	polinômio de médias móveis de ordem *q*
$\Phi_{P1}(B^{S1})$	polinômio auto-regressivo sazonal diário de ordem *P1*
$\Theta_{Q1}(B^{S1})$	polinômio de médias móveis sazonais diários de ordem *Q1*
$\Pi_{P2}(B^{S2})$	polinômio auto-regressivos sazonal semanal de ordem *P2*
$\Psi_{Q2}(B^{S2})$	polinômio de médias móveis sazonal semanal de ordem *Q2*
d_t	demanda consumida em um tempo t
ϕ	coeficiente auto-regressivo
θ	coeficiente de médias móveis
Φ	coeficiente auto-regressivo sazonal diário
Θ	coeficiente de médias móveis sazonal diário
Π	coeficiente auto-regressivo sazonal semanal
Ψ	coeficiente de médias móveis sazonal semanal

1

1 Introdução

Sistemas de abastecimento de água por todo o mundo têm sido sobrecarregados nos últimos anos devido ao rápido crescimento populacional e o aumento do consumo per capta de água. A fim de solucionar este problema, devem ser criados planos de ação que envolvam uma combinação das seguintes características: otimização dos sistemas de abastecimento através do seu controle e o desenvolvimento de estratégias de expansão de infraestrutura sustentável e de baixo custo (Adamowski, 2008).

Em ambas características a previsão de demanda de água é essencial. A otimização da operação em qualquer sistema de distribuição de água depende basicamente das previsões precisas das demandas horárias e do agendamento das operações de bombeamento durante os períodos de baixas tarifas de energia elétrica (Ghiassi *et al.*, 2008). Esta informação também é essencial na composição de estratégias de expansão a longo prazo, pois auxilia em questões de planejamento empresarial, tais como analisar a tendência do consumo de água, medir esforços de conservação, e a realização da previsão de receitas.

Pesquisas recentes em sistemas de abastecimento têm documentado os impactos do tempo de retenção em reservatórios na deterioração da qualidade da água. A demanda de água de todo o sistema é a diferença entre o bombeamento e o aumento do volume nos reservatórios. Entretanto, se a previsão da demanda de água está disponível, tanto o bombeamento como a reservação podem ser ajustados de acordo com a demanda. Isto não apenas beneficia minimizando os custos de bombeamento, mas também reduz o tempo de retenção nos reservatórios melhorando a qualidade da água.

A demanda de água é uma série altamente dominada pelos ciclos sazonais diários, semanais e anuais. Um modelo de séries temporais univariado baseado na série de dados históricos pode ser útil para previsão de demandas de curto prazo quando são retratados os vários ciclos periódicos e sazonais na especificação do modelo e na previsão (Caiado, 2010). Nexte contexto, Pulido-Calvo (2003) ressalta que a técnica de Box-Jenkins para modelos de séries temporais e as regressões lineares tem sido as técnicas mais utilizadas na previsão do

consumo de água, pois são relativamente simples de se desenvolver e se implementar. Shang *et al.* (2006) ainda demonstram que reduzindo-se o passo de tempo em um modelo de séries temporais pode-se reduzir a incerteza da previsão da demanda.

Assim, o objetivo deste trabalho é o desenvolvimento de modelos SARIMA para previsão do consumo de água em curto prazo. É apresentado um procedimento detalhado de modelagem de séries temporais de demanda de água, incluindo a identificação do modelo, estimação dos seus coeficientes, diagnóstico e realização de previsões para as próximas 24 horas com passo de tempo de 15 minutos, empregando uma nova metodologia de estimação de parâmetros.

2

2 Pesquisa Bibliográfica

2.1 Operação dos sistemas de distribuição de água

Em geral, os sistemas de abastecimento de água (SAA) das cidades de médio a grande porte apresentam complexidade elevada por contar com dezenas de conjuntos motor-bomba e reservatórios de distribuição que precisam ser operados simultaneamente. Muitas vezes também a decisão de qual operação será realizada deve ser tomada dentro de uma escala de alguns poucos minutos. A responsabilidade deste gerenciamento de curto e curtíssimo prazo, escala de horas e minutos respectivamente, de todo o SAA recai basicamente sobre a figura dos operadores.

Parte das tarefas de ligar e desligar os conjuntos motor-bomba é realizada automaticamente através de regras já programadas, baseadas no nível dos reservatórios buscando mantê-los cheios. Mesmo assim, os operadores tendem a manobrar o sistema seguindo a sua intuição e experiência no sistema específico, buscando tomar a ação mais adequada em cada momento (León *et al.*, 2000). Esta experiência passa por inúmeros fatores, desde a previsão do consumo de água para as próprias horas até o conhecimento técnico de elementos do sistema que podem se danificar em determinadas situações.

A técnica de previsão da demanda de água baseada na experiência dos operadores é chamada de técnica tradicional de previsão. Seguindo esta linha, a previsão se torna simplificada, pois nem todas as informações recebidas podem ser processadas. Esta limitação ocorre devido à elevada complexidade do sistema que está sendo operado. Apesar desta restrição, a técnica tradicional de previsão se torna útil devido à flexibilidade da ponderação de informações vindas de diversas fontes.

Esta flexibilidade permite a interpretação de eventos improváveis como desastres naturais e dias atípicos de consumo, através da inserção automática de quaisquer fatores que a experiência comprovou influenciar no fenômeno específico (Greenaway *et al.*, 2006). Em dias típicos, porém, os quais representam a maior parte do período, a previsão sob perspectiva

tradicional, vem associada a um excessivo grau de subjetividade, pois pode variar consideravelmente de um operador para outro e até mesmo para um mesmo operador com o passar dos dias.

Buscando-se minimizar a subjetividade e baseados no fato de que as demandas de água possuam ciclos diários pronunciados, tipicamente são preparados planos de operação com a antecedência de um dia (Kim *et al.*, 2001). Estes planos se baseiam tanto na experiência dos diversos operadores como no reflexo que a operação escolhida terá nas condições hidráulicas do sistema para o próximo dia, buscando minimizar os recursos a serem dispendidos pela empresa.

Neste sentido, de facilitar a avaliação do impacto das medidas a serem aplicadas no sistema, são muito úteis os modelos computacionais de simulação. Em resumo, os modelos de simulação, baseados nas condições físicas do sistema, aplicam sobre a condição atual de reservação, as futuras demandas de água permitindo a verificação da operação do sistema.

A modelagem computacional dos sistemas de abastecimento de água é uma tecnologia que tem sido largamente difundida e sua aplicação tem sido facilitada sobremaneira pela difusão de softwares livres como o Epanet (Rossman, 1999). Segundo Shang (2006), estes modelos têm sido amplamente empregados na análise e projeto de sistemas, porém, apesar da ampla aplicação de sistemas de supervisão, aquisição e controle de dados e a disponibilização destes dados praticamente em tempo real, os modelos não são atualizados na mesma escala. Isto reduz os modelos apenas a análises “off line” inutilizando assim a aplicação na operação dos sistemas.

2.2 Operação otimizada e economia de energia

Como consequência da excassez dos recursos energéticos e a necessária busca por novas fontes, tem-se justificado os recentes aumentos nos custos da energia. Estas mudanças têm exigido que as companhias de água otimizem suas operações de forma que mantenham ou até melhorem os padrões de qualidade do abastecimento, mantendo um valor de tarifa adequado. Cabe destacar que apenas a otimização das operações já pode resultar em economias substanciais dos custos operacionais totais para companhia de águas.

Minimizar os gastos com energia, já consagrados como a segunda ou até a primeira fonte de gastos para as empresas de saneamento, é um importante objetivo na otimização da operação dos SAA. Atingir este objetivo depende de dois fatores cruciais: previsões precisas

de demanda em tempo real, e a programação das operações de bombeamento durante os períodos de baixas tarifas da eletricidade (Ghiassi *et al.*, 2008).

Em outras palavras, a minimização do custo com energia é conseguida pelo conhecimento da estrutura de tarifação da energia, o que permite o agendamento dos níveis de produção e conseqüente gasto energético de forma que essa seja aumentada durante os períodos de menor tarifação, e diminuída quando os preços estão elevados. A aplicação deste tipo de rotina é dificultada principalmente pela coincidência do pico do preço de energia com o pico da série de demandas de água (Greenaway *et al.*, 2006). Dessa forma, nem sempre é possível o completo deslocamento da carga do horário de pico.

A redução do consumo em horário de pico é fundamental para toda a matriz energética nacional, pois, apesar de só acontecer em um período curto do dia, ela necessita estar preparada para suprir essa demanda. Essa redução pode ser conseguida pela utilização otimizada dos conjuntos motor-bomba. A operação otimizada pode ser feita por meio da escolha de bombas mais eficientes ou pelo agendamento do bombeamento de tal modo que a capacidade total dos reservatórios melhor aproveitada.

Nesse contexto, a cultura operacional de manter todos os reservatórios em algum nível pré-determinado próximo da sua capacidade máxima vem de encontro à minimização do consumo de energia. Ao invés disso, essa estratégia essencialmente resulta num aumento da freqüência de utilização dos conjuntos motor-bomba, conseqüentemente causando um aumento no consumo de energia e desnecessário desgaste dos CBs (Homwongs *et al.*, 1994).

2.3 Como um algoritmo de previsão de demandas pode ajudar

Ao realizar uma análise de sensibilidade dos SAA, Arandia-Perez e*t al.* (2009) afirmam que as demandas de água são os parâmetros que mais influenciam na resposta hidráulica da rede de distribuição de água. Este fato pode ser claramente entendido à medida em que observamos que as demandas de água influem diretamente na vazão a ser produzida bem como com as perdas de carga na rede.

Desse modo, nas últimas décadas tem-se motivado esforços para caracterizar a demanda de água em resoluções espaciais e temporais, para apoiar a necessidade de modelos de simulação cada vez mais precisos. Cabe destacar também que se os modelos estiverem operando ou sob padrões de demanda desatualizados, como é o caso da tão comum previsão determinística, jamais poderiam gerar resultados precisos para as condições hidráulicas da

rede, pois não representam as demandas do período a ser simulado. Neste contexto, o modelo de previsão de demandas torna-se potencialmente o fator mais crítico para uma boa estimação das demandas (Shang *et al.*, 2006).

A tarefa de se realizar a previsão de demanda não é simples, pois este tipo de previsão sempre será estocástica e baseada em uma série de parâmetros envolvidos e até mesmo a interação entre estes parâmetros. Além disso, a precisão da previsão é dificultada pela presença de ruídos, falhas e *outliers* nas séries de valores observados os quais devem ser tratados convenientemente a fim de se decidir quanta informação guardam sobre a variável original. Também o nível de incertezas aumenta consideravelmente quando se busca fazer previsões com horizonte de tempo maior.

Como já foi dito anteriormente, embora os operadores possuam vasta experiência sobre o sistema que operam, estes não são capazes de absorver e processar todas as informações recebidas de modo a gerar uma previsão de demanda com nível razoável de precisão para a maioria dos dias típicos. Neste ponto, o uso de métodos de previsão sistemáticos acarretam a minimização da aleatoriedade e subjetividade inseridos pelo operador humano, uma melhor descrição quantitativa do verdadeiro impacto dos fatores que realmente afetam o consumo de água fornecendo uma capacidade mais concreta de simulação de cenários (Greenaway *et al.*, 2006).

Para os modelos de simulação não ficarem desatualizados, os modelos de previsão de demanda bem como as condições de operação atuais devem ser atualizadas constantemente por dados armazenados em banco de dados históricos. Estas informações variam desde a própria demanda anterior ao período que se quer analisar, a atual capacidade de reservação e ao estado atual de bombeamento.

Dessa forma, ao se realizar uma simulação sobre um modelo de rede, utilizando-se as demandas previstas e com as condições iniciais atualizadas, forma-se os modelos de simulação em tempo real. Estes modelos são altamente desejáveis pois tem a capacidade de rapidamente sintetizar os dados de vazão, reservação e bombeamento do SCADA transformando em cenários futuros para a rede de abastecimento.

Como uma alternativa viável para se melhorar a eficiência dos sistemas, algoritmos de otimização podem ser aplicados sobre os modelos de simulação em tempo real de forma a gerar planos de operação para as próximas 24 horas. Estes modelos de operação podem ser muito úteis para auxiliar na atuação dos operadores ou até subtituí-los por algum tempo em

que não puderem estar presentes. Dessa forma, a previsão da demanda em tempo real, juntamente com os sistemas de supervisão e controle (SCADA), os modelos de simulação e os algoritmos de otimização, formam as bases para o controle computadorizado "on-line" do sistema de distribuição de água (Jowitt e Xu, 1992).

Este sistema composto, aliado à experiência dos operadores, gera uma maior segurança e confiabilidade das operações refletindo numa melhor programação das operações de bombeamento bem como a redução do tempo de detenção dos reservatórios através de uma melhor utilização dos mesmos. Assim sendo, a previsão da demanda de água é uma componente crucial no sucesso da operação de qualquer SAA.

2.4 Fatores que influenciam o consumo de água

Muitos fatores podem influenciar o consumo de água, dentre eles estão incluídos aspectos econômicos, hábitos culturais, o consumo em dias especiais como feriados, condições físicas da rede e condições climáticas locais. Os algoritmos de previsão de demandas objetivam sintetizar as relações mais importantes entre o consumo de água e os fatores que o influenciam e, posteriormente, usando o modelo estabelecido prever a demanda futura.

Em geral, os fatores que influenciam o consumo de água são complexos e, se relacionando de maneira não linear, tornam a previsão de demanda um processo desafiador. Isto dificulta a aplicação de simples modelos de regressão, por exemplo, à medida que eles necessitam de uma formulação pré-estabelecida. Também, como existe interação entre os fatores, muitas vezes são obtidas relações inconsistentes ou indiretas, o que dificulta a obtenção de um modelo mais geral e torna o modelo construído cada vez mais complexo e pouco intuitivo.

Além disso, a análise do consumo é limitada à informação que está disponível e pode ser quantificada. Ressalta-se também que para a relação de um fator poder ser utilizada por um algoritmo de previsão é preciso que seja avaliado o seu impacto em eventos anteriores e ainda que seja feita a previsão do seu acontecimento para o próximo período. Desta forma, fatores acidentais como faltas de energia, vazamentos, quebras de adutoras e incêndios, apesar de serem influentes na demanda, são fatores muito difíceis de quantificar e de se prever dificultando a sua inserção nos modelos (Liu *et al.*, 2009). Esses fatores contribuem para a

incidência de variações entre os padrões históricos e atuais de uso de água dificultando a atuação mesmo dos modelos de previsão mais refinados.

Ghiassi *et al.* (2008) reiteram que para modelos de previsão horária ou de curto prazo, pode-se considerar que as condições econômicas e sociais não sofram variações significativas. Assim, a relativa estabilidade própria destas variáveis ao longo do período analisado permite que os próprios dados de demanda de água implicitamente já representem a atual condição sócio-econômica da população, podendo-se eliminar a dependência dessas variáveis explicativas externas no modelo.

Quanto às condições climáticas, estas têm sido amplamente utilizadas como entradas para modelos estatísticos multivariados (regressão, funções de transferência, vetores autoregressivos, e redes neurais artificiais) para modelagem da demanda de água e previsão principalmente para previsão diária. Devido aos padrões de variação das condições climáticas além das próprias características dinâmicas de mudança da demanda de água são necessárias atualizações seqüênciais da previsão da demanda a fim de manter uma precisão aceitável para os períodos posteriores (Homwongs *et al.*, 1994).

2.5 Conceitos fundamentais de séries temporais

Segundo Box e Jenkins (1994) uma série temporal é qualquer conjunto de observações tomadas seqüencialmente no tempo. Uma característica fundamental das séries temporais é que, geralmente, as observações adjacentes são correlacionadas. A natureza desta dependência é de considerável interesse prático. A técnica análise de séries temporais analisa o grau de correlação.

As séries reais são contínuas no tempo, ou seja, em qualquer instante desejado podem ser tomadas medidas de sua grandeza. Por outro lado, sem perca de informações, os métodos discutidos a seguir são apropriados para séries onde as observações são igualmente espaçadas no tempo, como por exemplo, os valores obtidos via data-logger. Assim, define-se ao invés de relacionarmos a variável original a um instante exato de tempo na qual foi retirada, relaciona-se a uma quantidade inteira referente à quantidade de valores já tomados por esta variável. Assim surge a notação Zt onde Z representa a variável de estudo e t é quantidade de instantes de tempo no qual ela está sendo tomada. Este intervalo constante entre observações é denominado lag.

Obtida a série temporal Z(t) observada Moretin (2004) destaca as seguintes aplicações:

1. Investigar o mecanismo gerador da série temporal;
2. Fazer previsões de valores futuros da série de curto ou longo prazo;
3. Descrever apenas o comportamento da série, neste caso construção de gráficos, verificação da existência de tendências, ciclos e variações sazonais;
4. Procurar periodicidades relevantes nos dados;

Em todos os casos, modelos probabilísticos ou modelos estocásticos são construídos. Estes modelos devem ser simples e parcimoniosos (no sentido de que o número de parâmetros envolvidos deve ser o menor possível) e, se possível sua utilização não deve apresentar dificuldades às pessoas interessadas em manipulá-los.

2.5.1 Modelo estocástico X Modelo determinístico

Em geral a idéia principal ao se desenvolver um sistema é de se utilizar um modelo matemático para se descrever um fenômeno físico. Em particular, às vezes é possível derivar um modelo baseado em leis físicas, o qual nos permite calcular o valor exato de alguma variável dependente do tempo praticamente em qualquer instante. Este é o caso, por exemplo de um movimento uniforme ou uniformemente variado sem atrito. Um modelo assim desenvolvido chama-se determinístico. (Equação 2.1)

$$Z_t = a + b\,Z_{t-1} \tag{2.1}$$

Em um escala perfeita, não existe fenômeno inteiramente determinístico, porém em alguns casos esta aproximação é perfeitamente aceitável. Em outros fenômenos, no entanto, com uma mesma série de entradas conhecidas, as quais são de comprovada influência, podem ser geradas uma infinita quantidade de respostas diferentes. Dessa forma, a resposta não está perfeitamente determinada pelas entradas.

Como é sabido que não existem outras variáveis que influenciam o fenômeno, uma quantidade infinita de testes sucessivos podem ser realizados e, assim é verificado que existe uma tendência que a resposta se aproxime de um valor médio. Em outras palavras, a resposta pode ser explicada em termo de uma média e uma distribuição de probabilidades. Este tipo de processo ou fenômeno é chamado probabilístico ou estocástico.

Dessa forma, quando temos uma determinada observação, tomada de uma série temporal, ela não pode ser entendida como um resultado simples e sim como uma amostra de realização para uma população infinita de possíveis realizações como resultado da combinação dos fatores mais uma componente de erro aleatório. (Equação 2.2)

$$Z_t = a + b\,Z_{t-1} + \varepsilon_t \tag{2.2}$$

2.5.2 Estacionariedade

Uma das suposições mais freqüentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. A maior parte das séries reais são classificadas como não estacionárias, pois geralmente possuem algum nível de tendência ou alguma forma de não estacionariedade explosiva.

A Figura 2.1 mostra a série estacionária mais simples denominada passeio aleatório. No passeio aleatório a variável de estudo varia aleatoriamente ao redor de uma média constante. Este seria o caso da quantidade de grãos ensacados por uma máquina, por exemplo.

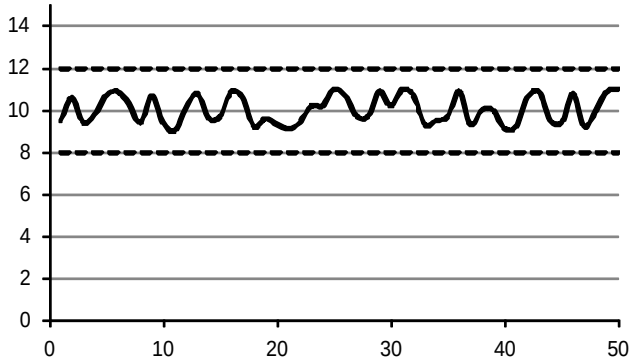

Figura 2.1. Comportamento de série estacionária chamada passeio aleatório.

2.5.3 Operador de atraso (B)

O operador de atraso é um dos operadores lineares mais utilizados em séries temporais e permite com que representemos simplificadamente uma variável deslocada no tempo. Ele é definido como na Equação 2.3.

$$B\,Z_t = Z_{t-1} \tag{2.3}$$

Como um operador linear, pode ser aplicado recursivamente de modo a obter o retrocesso de mais de um período no tempo como na Equação 2.4.

$$B(BZ_t) = B^2 Z_t = Z_{t-2} \tag{2.4}$$

Ou de maneira genérica, pode ser aplicado para uma ordem qualquer *m* como na Equação 2.5.

$$B^m Z_t = Z_{t-m} \tag{2.5}$$

2.5.4 Operador de auto-diferenças

O operador de auto-diferenças também é um operador linear e pode ser aplicado nas séries a fim de simplificar a sua posterior utilização pelo modelo preferido. A sua definição é mostrada na Equação 2.6. Como nas análises de séries temporais, as medidas são tomadas sempre sobre um mesmo intervalo de tempo, o operador de diferenças acaba sendo um indicativo da derivada discreta da série no período determinado.

$$\nabla Z_t = Z_t - Z_{t-1} \tag{2.6}$$

Este operador também pode ser aplicado para períodos maiores de tempo espefícicando-se seu índice como mostrado na Equação 2.7.

$$\nabla_m Z_t = Z_t - Z_{t-m} \tag{2.7}$$

2.5.5 Estacionariedade homogênea

O operador de auto-diferenças como apresentado anteriormente pode ser utilizado para gerar séries estacionárias a partir de séries não estacionárias. As séries que possuem esta particularidade são chamadas de séries não estacionárias homogêneas, pois apesar de apresentarem variação significativa no seu nível esta variação é homogênea no período considerado. A Figura 2.2 (a) mostra uma série não estacionária homogênea e a Figura 2.2 (b) mostra o resultado da aplicação de diferenças de ordem unitária sobre a série original.

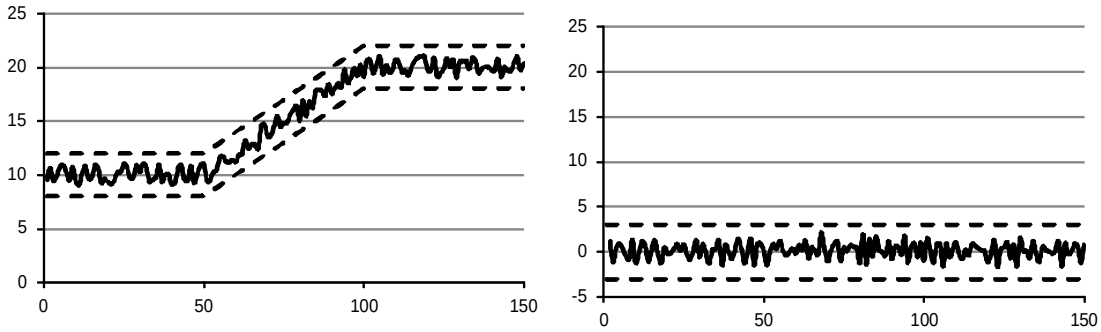

Figura 2.2. Aplicação de operador linear de diferenças de ordem unitária.

Como já foi dito, por se tratar de um operador linear, o operador de diferenças também pode ser aplicado recursivamente sobre a série original buscando-se atingir uma série estacionária. Na prática, no entanto, geralmente o número de diferenças necessárias suficientes é bem reduzido.

2.5.6 Função de autocorrelação

A fim de avaliar as relações existentes entre os valores de uma determinada variável Zt e seus valores adjacentes foi criada a função de autocorrelação FAC definida na Equação 2.8.

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} \tag{2.8}$$

Onde γ é a função de auto-covariância definida pela Equação 2.9.

$$\gamma_\tau = Cov\{X_t X_{t+\tau}\} = \frac{1}{N} \sum_{t=1}^{n-\tau} \left[(X_t - \overline{X})(X_{t+\tau} - \overline{X}) \right] \tag{2.9}$$

Esta função de auto-correlação desempenha papel importante na análise dos modelos de séries temporais e possui as características básicas de que a autocorrelação para lag = 0 seja igual a 1 o que significa que a série tem máxima auto-correlação consigo mesma e tende a diminuir até a zero quando se aumenta o lag até o infinito.

2.5.7 Função de auto-correlação parcial

Box, Jenkins e Reinsel (1994) propõem a utilização de uma outra função significativa ϕ da série denominada função de autocorrelação parcial (FACP) definida na Equação 2.10 matricialmente em função da FAC.

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots \rho_{k-2} \\ \vdots & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots \quad 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \cdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdots \\ \rho_k \end{bmatrix} \tag{2.10}$$

2.5.8 Ruído Branco

Diz-se que uma série ε_t é um ruído branco se esta for estritamente estacionária e não correlacionada. Isto implica em:

Condições de estacionariedade : $E\{\varepsilon_t\} = \mu_\varepsilon$ e $Var\{\varepsilon_t\} = \sigma_\varepsilon^2 \quad \forall t$

Condição de não auto-correlação: $Cov\{\varepsilon_t \varepsilon_{t+\tau}\} = \begin{cases} \sigma^2, se\ \tau = 0 \\ 0, se\ \tau \neq 0 \end{cases}$

2.5.9 Método da decomposição

O método da decomposição é uma das técnicas mais utilizadas em séries temporais para se realizar a compreensão das principais componentes presentes nos dados. Esta abordagem parte do pressuposto que a série temporal pode ser entendida como a composição de parcelas de tendência (T_t), variações sazonais (S_t) além de fatores aleatórios (a_t) representados de maneira genérica como um ruído branco conforme Equação 2.11.

$$Z_t = T_t + S_t + a_t \tag{2.11}$$

Tendência em uma série temporal é a mudança gradual observada por meio da variação dos valores da série ao longo do tempo e que se mantém ao se remover os componentes de cliclos, sazonalidade e fatores aleatórios. Este conceito normalmente está relacionado ao comportamento da série percebidas a longo prazo. A sazonalidade por sua vez, é um comportamento que acontece de maneira recorrente ao longo de um período definido uniforme. Este período em que a sazonalidade se repete pode ser facilmente percebido através da função de auto-correlação pois em valores múltiplos deste período sazonal esta função FAC apresentará um pico característico.

2.6 Características da série de demandas de água

Em geral, as vazões utilizadas pelos algoritmos de previsão são medidas por macro-medidores convenientemente distribuídos por todo o sistema. Estas vazões são um somatório de variadas parcelas de consumo dependendo da região e são constituídas principalmente por consumos residenciais, comerciais, industriais e perdas. Dependendo de qual a importância de cada uma destas parcelas no setor específico é que se definem as características únicas do padrão de demandas para a região.

Para um setor em que a maior parte do consumo seja residencial, objeto mais importante para os algoritmos de previsão, o consumo atinge o seu mínimo durante a madrugada em torno das 2h e 4h e se torna alto durante o dia. Em dias típicos, o consumo aumenta consideravelmente a partir das 5h, pois é o período referente ao deslocamento para o trabalho diário e atinge um pico entre as 9h e 11h com as atividades preparativas para o almoço. Em seguida, o consumo de água tende a diminuir e se eleva novamente por volta das 19h referente ao retorno da população do serviço.

Caso não haja uma alteração significativa de outros fatores intervenientes, o padrão típico mostrado na Figura 2.3 é mantido. Em geral, os padrões de demanda para dias de semana (de segunda-feira à sexta-feira) tendem a ser semelhantes. Além disso, em uma semana típica, os padrões de utilização de água durante o fim de semana (sábado e domingo) são significativamente diferentes de um padrão de dia típico de semana e semelhantes entre si (Homwongs *et al.*, 1994).

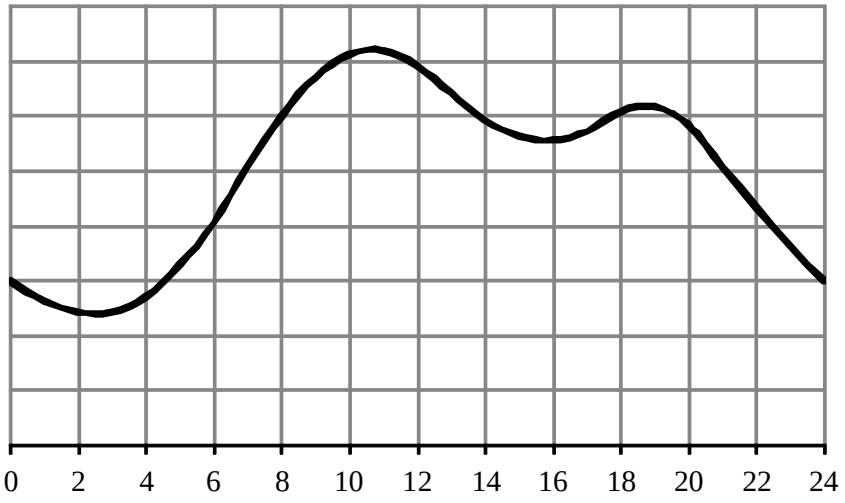

Figura 2.3. Perfil típico de consumo para setores residenciais

De modo geral, o padrão de demanda de água sobre um determinado período pode ser entendido como uma série temporal estocástica não-estacionária com características específicas. Por série estocástica, entende-se que, apesar da série de demandas de água ser influenciada por vários fatores, ela não pode ser perfeitamente explicada por meio destes. Desta forma, as demandas são tratadas mais convenientemente adotando-se sempre a presença de uma componente de erro não explicável.

A estacionariedade estrita é a característica de uma série temporal que garante que sua média e variância não se altera com o passar do tempo. Desta forma, ao se tomar qualquer origem no tempo as características da série se mantém. As séries de demanda de água são classificadas como não-estacionárias, pois podem conter vários efeitos, podendo incluir a presença de uma tendência na média, variância não constante, e descontinuidades (Jowitt e Xu, 1992). Por todas essas características, além do fato da demanda de água depender basicamente dos hábitos de uma população específica, Shang *et al*. (2006) afirmam que a demanda de água distribuída é o fator mais incerto e variável de um sistema de distribuição.

Como ponto em comum de todos os padrões de demanda, tem-se que a vazão observada para um determinado dia tem forte relação com os dias anteriores a ele observados podendo se dizer que a demanda de água possui forte sazonalidade diária. Da mesma forma, as vazões observadas para um determinado dia tem forte relação com para o mesmo dia só que na semana anterior podendo se dizer que a demanda de água possui também forte sazonalidade semanal.

De forma mais geral, pode-se dizer que a demanda de água é altamente dominada por ciclos sazonais diários, semanais e anuais (Caiado, 2010). Estas relações podem ser mensuradas em modelos de previsão horários, através de picos de auto-correlação da série de demandas principalmente separados de intervalos de 24 horas (1 dia) e 168 horas (1 semana).

Essas relações sazonais de semelhança formam as bases para o estabelecimento de qualquer previsor de demanda de curto prazo. Neste caso, avaliam-se as relações que os dias passados tiveram com seus anteriores podendo-se projetar essas relações para o futuro. Ressalta-se também que estas semelhanças sazonais em geral são bem expressivas e contínuas, pois são influenciadas pelos hábitos de vida dos usuários que geralmente se mantém.

Além da componente sazonal, outra componente a ser analisada é a tendência, a qual geralmente é afetada pela variação contínua das variáveis meteorológicas com as estações do

ano ou o crescimento populacional. Estes fatores causam não-estacionaridades aem longo prazo, pois tendem a aumentar ou diminuir gradativamente o consumo de água. Por gerarem não-estacionariedades apenas em longo prazo, para uma previsão de escala horária, estes efeitos podem ser facilmente negligenciados (Homwongs *et al.*, 1994).

Quanto às descontinuidades inerentes das séries de demandas, geralmente são causadas por uma manobra realizada no sistema, decorrente de mudança de setorização, re-calibração de algum medidor ou algum fator acidental expressivo. Quando um destes fatos ocorre, o modelo de previsão deve ser identificado novamente, pois podem ter se alterado as características básicas da série de demandas em estudo.

2.7 Modelos ARIMA e SARIMA

Abaixo são descritas as bases para o desenvolvimento do modelo de previsão utilizado cujo método é baseado na análise de séries temporais. Ghiassi *et al.* (2008) afirmam que este tipo de modelagem é a abordagem mais utilizada em problemas de previsão e se baseiam na direta identificação dos padrões presentes nos dados históricos de demanda de água. Mesmo sendo relativamente simples, este tipo de modelagem pode ser satisfatoriamente utilizada para predizer o consumo para o caso do controle de um sistema de abastecimento em tempo real (Jowitt e Xu, 1992).

Quando esta série é constituída de valores que são relacionados aos valores anteriores, como é o caso da série de demandas de água, podemos entender cada observação de demanda d_t em função de p observações de demanda anteriores.

Adotando-se a diferença de tempo entre cada observação como uma constante chamada lag tem-se que, por simplicidade, os valores anteriores podem ser representados basicamente por d_{t-1}, d_{t-2} e assim por diante. A ponderação destes valores de demanda anteriores é feita através de coeficientes auto-regressivos denominados ϕ_1, ϕ_2.

Este modelo é dito auto-regressivo (AR) de ordem p. Como esta relação não é determinística, sempre deve-se também admitir para o instante t um valor de erro (a_t), ou seja uma diferença entre o valor previsto e o verdadeiro (Equação 2.12).

$$d_t = \phi_1 d_{t-1} + \phi_2 d_{t-2} + ... + \phi_p d_{t-p} + a_t \tag{2.12}$$

Equações como a apresentada em 2.12 representam a observação presente em função das observações já passadas. Esta equação, portanto permite também que seja feita a extrapolação para momentos futuros a medida que pode-se inserir tempos cada vez maiores t+1, t+2 etc.

Quando se deseja descrever a Equação 2.12 para os próximos tempos nota-se que não utilizamos a informação de que foi obtido um erro a_t no tempo t o que representaria um desperdício de informação. De fato, a informação dos erros obtidos nos passos anteriores dá uma boa indicação de quais erros pode-se evitar para os próximos passos. Assim surge a idéia dos modelos de erro ou modelos de médias móveis (MA).

Num modelo de médias móveis puro, as demandas futuras seriam função apenas de q termos de erros obtidos na estimativa das demandas passadas a_{t-1}, a_{t-2} e assim por diante e da mesma maneira que o modelo auto-regressivo inclui-se um termo de erro para o instante de tempo t (a_t). A ponderação destes valores de erros anteriores é feita através de coeficientes de médias móveis denominados θ_1, θ_2. (Equação 2.13).

$$d_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \tag{2.13}$$

De modo geral a maioria dos processos é obtido através de um mixto de ambos os tipos de modelo auto-regressivo e de médias móveis formando o chamado modelo autoregressivo de médias móveis (ARMA) de ordens (p, q). (Equação 2.14).

$$d_t = \mu + \phi_1 d_{t-1} + \ldots + \phi_p d_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \tag{2.14}$$

Pode-se afirmar que estes modelos ARMA gerados são superiores aos tradicionais modelos de regressão, pois os modelos regressão se baseiam apenas na minimização do termo de erro quadrático nas equações usadas, enquanto que os modelos de médias móveis aproveitam esses erros na estimação da demanda futura em longo alcance evitando que aconteçam novamente (Aly e Wanakule, 2004).

A formulação de modelos ARMA inicialmente desenvolvida para modelos estacionários, podem ser generalizadas também para processos não estacionários homogêneos. O número necessário de auto-diferenças é denominado ordem de integração, indicada por “d”. O modelo resultante é dito auto-regressivo integrado de médias móveis (ARIMA) de ordem (p,d,q) (Equação 2.15). Os modelos ARIMA são alguns dos mais utilizados modelos para análise de séries temporais (Morettin e Tolloi, 2004).

$$\nabla^d Z_t = \frac{\theta(B)}{\phi(B)} a_t \tag{2.15}$$

Em que B é o operador de atraso e $BZ_t = Z_{t-1}$; p é a ordem da parcela auto-regressiva; d é a ordem das diferenças; q é a ordem da parcela de médias móveis; $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ é o operador auto-regressivo ; $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ é o operador de médias móveis; ∇^d é o operador de diferenças em que $\nabla^d Z_t = Z_t - Z_{t-d}$; e a_t é um ruído aleatório tipo branco.

A ordem das parcelas auto-regressivas e de médias-móveis pode ser percebida através da análise conjunta das funções de auto-correlação (FAC) e de autocorrelação parcial (FACP) analisando-se as regras da Tabela 2.1.

Tabela 2.1. Regras resumo para a identificação de modelos ARMA.

Processo	FAC	FACP
AR (p)	FAC decai de acordo com exponenciais e/ou senoides amortecidas, infinita em extensão	FACP finita, no sentido que ela apresenta um corte após o lag p
MA (q)	FAC finita, no sentido que ela apresenta um corte após o lag q	FACP decai de acordo com exponenciais e/ou senoides amortecidas, infinita em extensão
ARMA (p,q)	FAC infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o lag q-p	FACP infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o lag p-q

Como já foi visto, a série de demandas de água possui características sazonais, ou seja, suas observações separadas por um período *s* são similares. Se estes efeitos cíclicos forem retirados da série, uma aproximação ARIMA pode ser aplicada em seus resíduos, ou seja, um modelo ARIMA pode ser aplicado sobre os resíduos do modelo ARIMA original. Este modelo composto, resultante da incorporação de dados sazonais ao modelo ARIMA, é dito modelo ARIMA sazonal multiplicativo ou simplesmente SARIMA. De forma simplificada, o

modelo SARIMA é formado por dois modelos ARIMA conjugados, um modelo sazonal de ordens (P,D,Q) e outro não sazonal de ordens (p,d,q) conforme mostrado na (Equação 2.16).

$$\nabla_s^D \nabla^d d_t = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)} a_t \tag{2.16}$$

Em que P é a ordem da parcela auto-regressiva sazonal; D é a ordem das diferenças sazonais; Q é a ordem da parcela de médias móveis sazonal; s é o tamanho do ciclo; $\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$ é o operador auto-regressivo sazonal; $\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs}$ é o operador de médias móveis sazonal; ∇_s^D é o operador de diferenças sazonal; a_t é um ruído branco e os demais parâmetros já foram definidos.

Para a série de demandas de água, adotamos como componentes sazonais principais a sazonalidade diária (lag=24h) e semanal (lag=168h), portanto precisa-se adotar um modelo que comporte ambas as sazonalidades. Sem nenhum problema adicional, segundo Box e Jenkins (1994) podemos adicionar uma segunda sazonalidade formando assim o modelo SARIMA duplo mostrado na (Equação 2.17).

$$\nabla_{S2}^{D2} \nabla_{S1}^{D1} \nabla^d d_t = \frac{\theta_q(B)\,\Theta_{Q1}(B^{S1})\,\Psi_{Q2}(B^{S2})}{\phi_p(B)\,\Phi_{P1}(B^{S1})\,\Pi_{P2}(B^{S2})} a_t \tag{2.17}$$

Este modelo univariado de séries temporais duplo sazonal, por ser baseado nas séries históricas de dados e, portanto ter a capacidade de acomodar os vários ciclos sazonais do processo subjacente de geração da série, pode ser bem útil para previsões de demanda de curto prazo. Comparando-se com o também muito utilizado processo caixa preta de redes neurais, Alvisi *et al.* (2007) afirma que embora a rede neural artificial tenha proporcionado bons resultados, foram encontradas dificuldades com a transição de um dia para o outro, especialmente entre dias úteis e fins de semana o que poderia ser evitado ao se adotar uma abordagem baseada em ciclos sazonais, semanais e diários.

Quanto à não inclusão de fatores climáticos no modelo de séries temporais isso acarreta um defeito caso estes sejam desprezados completamente em curto prazo, pois podem influenciar as demandas principalmente para previsão diária. No entanto, caso não haja uma mudança brusca no clima, Alvisi *et al.* (2007) afirma que a própria inclusão de componentes sazonais pode em grande medida superar essas limitações.

Esse fato pode ser explicado porque o efeito das componentes climáticas acabam sendo incluídos indiretamente por meio dos termos auto-regressivos de dados de demandas passados o que resultou no estudo de Jowitt e Xu (1992) uma não indicação de qualquer dependência dos fatores climáticos para a maioria das áreas de estudo além do que já tinha sido representado pelos termos auto-regressivos.

Além disso, vários autores (Kim *et al.*, 2001; Aly e Wanakule, 2004) consideram que apesar dos fatores climáticos influírem significativamente na previsão da demanda de água, se apenas um parâmetro de entrada for utilizado pode-se utilizar as demandas dos dias anteriores como parâmetro de entrada, o que demostra que a memória da série temporal é muito mais pronunciada que a própria dependência do clima.

Vale reiterar também que o uso de informações exógenas climáticas apesar de poder beneficiar alguns modelos de previsão de demanda de água, exige valores previstos precisos dessas variáveis para o alcançe de previsão. Este é também um problema desafiador adicional pois as previsões climáticas são muitas vezes imprecisas, introduzindo erros adicionais no modelo (Ghiassi *et al.*, 2008).

2.8 Critérios para escolha de um algoritmo de previsão

Ao se selecionar um algoritmo de previsão, o primeiro critério, e mais óbvio, a ser adotado seria a precisão obtida na aplicação do método. Apenas este indicador, porém não é suficiente, pois assim desprezar-se-iam os aspectos relativos à aplicação do método dentro da empresa de saneamento. Este fator pode ser facilmente percebido pelo fato dos modelos mais precisos necessitarem, geralmente, da inclusão de um maior número de fatores e de interpretação mais aguçada das situações por parte dos operadores, o que nem sempre é possível.

Em resumo, Zhang (2007) destaca que ao se selecionar uma ferramenta para previsão, as concessionárias devem equilibrar a necessidade de precisão com muitos fatores orientados à operação, como a aceitação dos usuários, baixo custo de manutenção, rápida resposta, etc. Além disso, é importante para a ferramenta de previsão ter a flexibilidade de ser adaptável para atender às necessidades individuais de cada companhia além de ter a capacidade de extrair registros históricos e compartilhar essas informações obtidas entre os operadores.

Detacam-se os fatores de aplicação juntamente ao usuário como a sua própria aceitação e a rápida resposta, pois os operadores quando se deparam com uma aplicação

pouco intuitiva, ou que acabe resultando em valores incorretos ou até mesmo absurdos, acabam abandonando a proposta e se voltando aos modelos tradicionais. Da mesma forma, o operador ao se deparar com um modelo tipo caixa preta, por não conseguir entender a aplicação acaba por abandoná-la o que de qualquer forma é indesejável.

Os modelos também precisam ser flexíveis (León *et al.*, 2000), pois como já mencionado, é necessário sempre ter um modelo diferente para cada sistema e alterar os modelos existentes, caso tenhamos uma mudança de setorização ou topologia do sistema. Dessa forma, um modelo que exija de grande esforço para mudança de estrutura ou ainda um modelo que não possa ser alterado diretamente e de forma rápida pelos operadores, com as mudanças do sistema, se torna inaplicável.

Além disso, Jowitt e Xu (1992) destacam fatores computacionais, tais como a exigência em termos de requisitos de armazenamento e tempo de execução, a robustez do modelo, a facilidade de inicialização em caso de desligamento do sistema e a facilidade da comunicação com os bancos de dados existentes nas empresas, como fatores limitantes da aplicação de previsão de demandas.

3

3 Metodologia

3.1 Procedimento de aplicação dos modelos sarima

Ao se estudar uma série temporal procura-se descobrir qual é o comportamento apresentado pela série da variável, sendo este expresso pela ordem dos termos do modelo além dos valores dos seus coeficientes. Deve-se então fazer uma busca pelo modelo que melhor represente o comportamento da série, ou seja, um modelo que se ajuste aos valores observados e seja parcimonioso. Segundo Box e Jenkins (1994), a construção do modelo e sua aplicação devem seguir um ciclo iterativo composto de quatro etapas:

3.1.1 Identificação

Neste estágio são definidos os valores de p, d, q, P, D, Q. Primeiramente faz-se uma análise para decidir as ordens de diferenciação d e diferenciação sazonal D buscando atingir a estacionaridade da série original. Em seguida, calcula-se a função de auto-correlação (FAC) e função de auto-correlação parcial (FACP) da série já diferenciada buscando-se identificar as ordens dos termos auto-regressivos e de médias móveis p, q, P, Q.

3.1.2 Estimação

Neste estágio os coeficientes do polinômio auto-regressivo e de médias móveis, sazonal e não sazonal, são calculados. O processo de estimação de um modelo é realizado através do método de máxima verossimilhança (MV) onde ocorre minimização dos valores dos quadrados dos ruídos aleatórios previstos.

3.1.3 Diagnóstico

Nesta etapa, busca-se confirmar a hipótese assumida de que os resíduos geradores do modelo são do tipo ruído branco. Faz-se o cálculo da correlação entre os próprios resíduos e entre os resíduos e a série original. Caso os resíduos sejam auto-correlacionados, a dinâmica

da série não pode ser explicada suficientemente pelos coeficientes do modelo e novos termos de médias móveis devem ser sugeridos. Caso os resíduos sejam relacionados com a série original, novos termos auto-regressivos devem ser adicionados.

3.1.4 Previsão

A última etapa consiste em realizar a previsão propriamente dita. A melhor previsão será aquela que apresentar os menores valores para: AIC, BIC e erro percentual absoluto médio (MAPE).

Como já foi dito o padrão de demanda de água e, conseqüentemente, as relações subjacentes entre os padrões de consumo dependem de vários fatores como aspectos econômicos, hábitos culturais, condições climáticas do dia, etc o que resulta em fraca correlação ente os padrões de regiões distintas.

Dessa forma, Babel e Shinde (2011) ressaltam a necessidade do desenvolvimento de modelos de previsão de demanda de água específicos para cada cidade e Jowitt e Xu (1992) para cada zona de distribuição de uma mesma cidade. Vale destacar que esses modelos podem até possuir mesma estrutura, ou seja, mesmo número de coeficientes, porém, sempre deverá ser realizada a etapa de estimação dos coeficientes especificamente para cada área baseados nos consumos medidos da área.

3.2 Algoritmo de busca harmônica

O algoritmo meta-heurístico de Busca Harmônica (HS) foi originalmente proposto por Geem *et al*. (2001) e é inspirado no processo de busca pela harmonia musical perfeita. Este algoritmo compara a busca por uma perfeita harmonia pelos músicos com a busca por soluções ótimas determinadas por uma função objetivo.

O algoritmo HS considera vários vetores solução simultaneamente e o seu procedimento é similar ao algoritmo genético (GA). Entretanto, a maior diferença entre o GA e o algoritmo HS é que o segundo gera um novo vetor a partir de todo o conjunto de vetores existentes, considerando que o primeiro gera um novo vetor a partir de apenas dois dos vetores existentes. Além disso, as iterações no algoritmo HS são muito mais rápidas do que nos algoritmos genéticos (Lee e Geem 2005).

O algoritmo HS é simples em conceito, possui poucos parâmetros, e é fácil de implementar. Os passos dos procedimentos de busca harmônica são apresentados na Figura 3.1. Estes passos são descritos nas próximas quatro subseções.

3.2.1 Função Objetivo

As previsões de modelos auto-regressivos puros são funções lineares dos seus coeficientes assim como funções lineares dos dados passados. Por isso, termos auto-regressivos são casos especiais de modelos de regressão linear por isso eles podem ser ajustados por procedimentos de regressão múltipla por algoritmos comuns de mínimos quadrados. As previsões de termos de medias móveis, por outro lado, são funções lineares dos valores passados, mas são funções não lineares dos seus coeficientes. Assim, estes modelos requerem a utilização de um algoritmo de estimação não linear.

Em modelos de medias móveis e modelos mistos auto-regressivos de medias moveis, os termos de erro não são conhecidos a menos que o modelo já esteja ajustado. Eles necessitam ser calculados sequencialmente dadas as estimativas atuais dos próprios parâmetros, período a período, em um procedimento recursivo e inverso chamado "backward forecasting" (BF). Este modelo pode ser avaliado pela função de máxima verossimilhança estimada no procedimento BF como uma função objetivo.

3.2.2 Inicializar a memória harmônica (HM)

Neste passo, a matriz *HM* (Equação 3.1) é preenchida com tantos modelos gerados aleatoriamente quanto for o tamanho da memória harmônica *hms* e os seus correspondentes valores de função objetivo. A memória harmônica *HM* é um local onde todos os vetores soluções e as correspondentes funções objetivo são armazenadas. Cada coluna representa um modelo $X = (\phi,\Phi,\theta_1,\theta_2,\theta_3,\Theta,\Psi)$ e $f(X)$ é a respectiva função objetivo para o modelo genérico X. Modelos inviáveis que violem as condições de estacionaridade não são incluídos em *HM*, nestes casos outro modelo que esteja na área de solução viável é gerado.

$$HM = \begin{bmatrix} \varphi^1 & \Phi^1 & \theta_1^1 & \theta_2^1 & \theta_3^1 & \Theta^1 & \Psi^1 \\ \varphi^2 & \Phi^2 & \theta_1^2 & \theta_2^2 & \theta_3^2 & \Theta^2 & \Psi^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi^{hms} & \Phi^{hms} & \theta_1^{hms} & \theta_2^{hms} & \theta_3^{hms} & \Theta^{hms} & \Psi^{hms} \end{bmatrix} \Rightarrow \begin{bmatrix} f(X^1) \\ f(X^2) \\ \vdots \\ f(X^{hms}) \end{bmatrix} \tag{3.1}$$

3.2.3 Improvisar uma nova harmonia a partir de HM

Neste passo, uma nova harmonia é gerada baseada em três regras as quais são consideração da memória, ajuste lateral e geração aleatória. Na consideração da memória, o valor do primeiro coeficiente ϕ para a nova harmonia é escolhido a partir de uma posição aleatória da memória harmônica ou a partir de um conjunto de valores possíveis. O parâmetro *hmcr* representa a taxa de escolha de um valor a partir dos valores históricos armazenados em *HM* enquanto que (1-*hmcr*) é a taxa de se selecionar aleatoriamente um valor do conjunto de valores possíveis. O parâmetro *hmcr* previne que o processo de busca convirja para ótimos locais pois ele permite a busca em regiões ainda não percorridas.

Após isso, o coeficiente obtido pela consideração da memória é examinado para se determinar se ele deve ser ajustado lateralmente. O parâmetro *par* representa a taxa de ajuste lateral dos valores enquanto que (1-*par*) é a taxa de não se fazer nada. Com a taxa de *par*/2 o coeficiente ϕ é ajustado um valor para cima e com uma taxa de *par*/2 o coeficiente ϕ é ajustado um valor para baixo. A taxa de ajuste lateral é essencial para melhorar as soluções já encontradas pelo *hmcr* e promove a busca local. Os valores dos outros coeficientes ($\Phi,\theta_1,\theta_2,\theta_3,\Theta,\Psi$) são selecionados da mesma maneira.

3.2.4 Atualizar a memória harmônica

Todos os modelos e o valor da função objetivo em *HM* são ordenados do melhor para o pior e a nova harmonia é comparada com o vetor com a pior função objetivo obtida neste passo. Se o novo modelo ($\phi,\Phi,\theta_1,\theta_2,\theta_3,\Theta,\Psi$) gerado é melhor do que o pior em *HM*, o novo modelo é incluído em *HM*, e a pior harmonia existente é excluída de *HM*.

3.2.5 Checar o critério de parada

Os passos das duas seções anteriores são repetidos até que o critério de parada seja satisfeito. Neste estudo, dois critérios de parada foram adotados para HS. O primeiro deles termina o algoritmo quando um número predeterminado de buscas (adotado como cem mil) é atingido. O segundo critério termina o processo antes do número máximo de iterações, se um modelo melhor do que o atual na memória harmônica não seja encontrado durante um número definido de sucessivas buscas em HS (adotado como vinte mil).

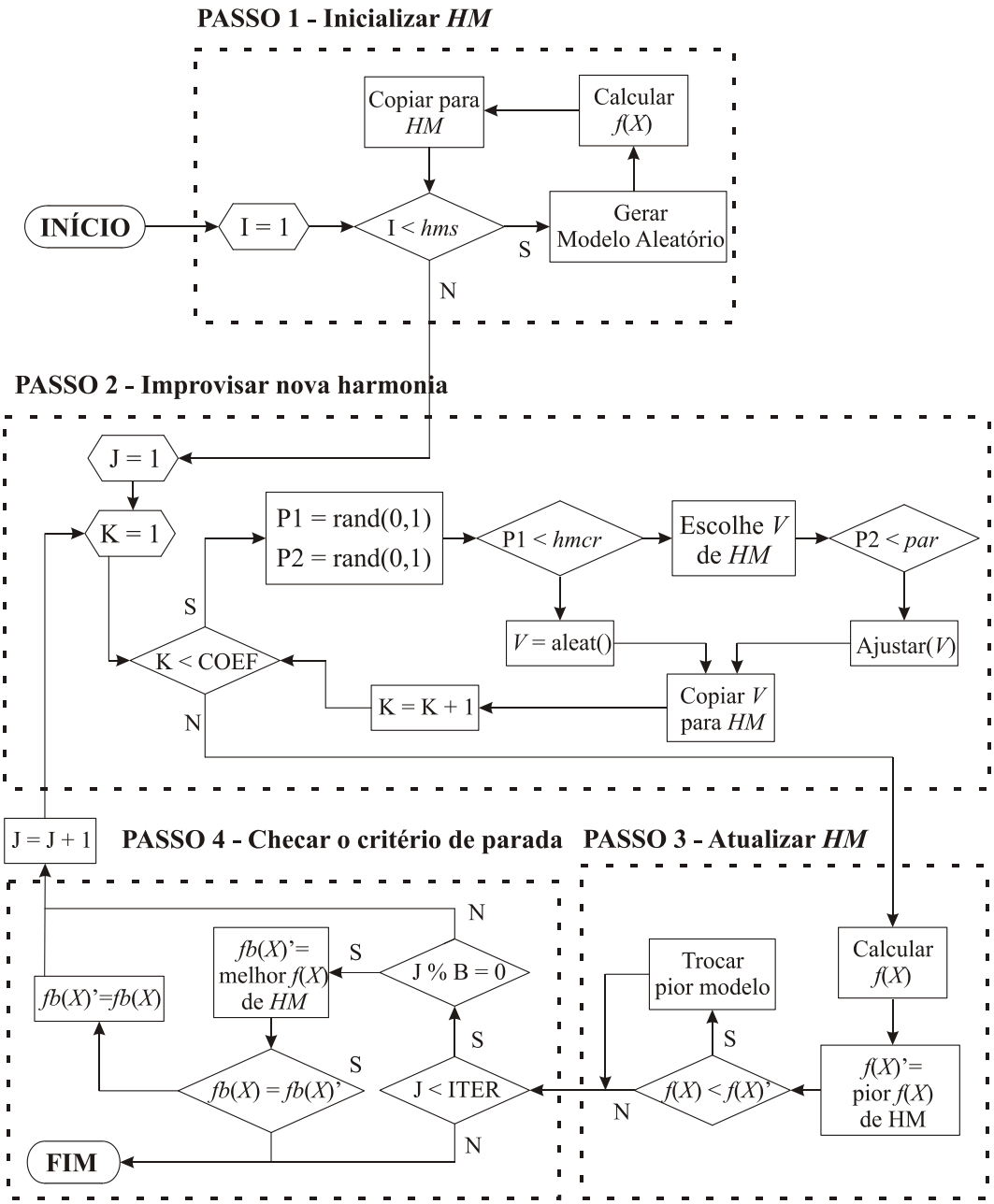

Figura 3.1. Fluxograma do algoritmo de Busca Harmônica.

3.3 Planejamento de experimentos

O planejamento de experimentos é baseado em modernas análises estatísticas. No planejamento experimental, uma variável de resposta pode ser definida a fim de analisar o desempenho desta em função dos parâmetros do planejamento. A convergência da otimização da busca harmônica foi realizada alterando-se os parâmetros originais de tamanho da memória harmônica (HM), taxa de escolha (*hmcr*) e taxa de ajuste lateral (*par*). Esta busca foi feita

seguindo o planejamento fatorial de experimentos e a metodologia de superfície de resposta, originalmente desenvolvida por Box e Wilson (1951). O número de iterações para se atingir o modelo otimizado foi analisado como o indicador mais importante de convergência.

3.3.1 Método da Superfície de Resposta (RSM)

A principal ideia do RSM é utilizar uma sequencia de planejamentos experimentais a fim de encontrar as condições ótimas dos fatores que possam otimizar o fator resposta. Para determinar um ótimo é necessário que uma função polinomial contendo termos quadráticos seja ajustada, assim as superfícies de resposta são utilizadas para determinar o ótimo. A função polinomial da Equação 3.2 descreve um modelo quadrático com três variáveis:

$$F(x,y,z) = b_0 + b_1 x + b_2 y + b_3 z + b_4 xy + b_5 xz + b_6 yz + b_7 x^2 + b_8 y^2 + b_9 z^2 \qquad (3.2)$$

Introduzindo-se estes termos no modelo, é possível determinar as relações não lineares entre o número de iterações e as constantes da busca harmônica. O planejamento de Box reduz o custo da experimentação de modo que um modelo quadrático possa ser ajustado, o que leva a uma tão procurada direção de subida. Esta direção pode ser bem estimada pelas equações diferenciais da Equação 3.3 as quais indicam a estimativa para o ótimo global da função:

$$\frac{\partial F(x,y,z)}{\partial x} = 0 \quad \frac{\partial F(x,y,z)}{\partial y} = 0 \quad \frac{\partial F(x,y,z)}{\partial z} = 0 \qquad (3.3)$$

3.3.2 Planejamento fatorial

A abordagem correta para se lidar com vários fatores é conduzir um experimento fatorial. Esta é uma estratégia experimental na qual os fatores são variados ao mesmo tempo, ao invés de um de cada vez (Montgomery 2001). Os experimentos são realizados sobre todos os níveis possíveis de combinação de todos os fatores. O número de experimentos realizados deve ser no mínimo n^k, com repetições a n níveis e k fatores.

A vantagem dos experimentos fatoriais é que pode-se considerar o efeito principal do fator (o efeito do fator sozinho) e o efeito da interação entre os fatores (o efeito entre os fatores). Este é um método de busca conveniente que pode ser utilizado para encontrar o fator principal quando se tem muitos fatores envolvidos no começo do experimento.

4

4 Resultados e Discussões

4.1 Estudo de caso

Os dados coletados referem-se ao consumo de água observado em quatro setores de fornecimento de duas cidades brasileiras (Figura 4.1). A primeira cidade (CID1), localiza-se no centro da grande área tropical da qual foram tomados dados de dois setores de fornecimento (SF1 e SF2), já a segunda cidade (CID2) localiza-se no centro da grande área de clima subtropical úmido da qual também foram tomados dois setores de fornecimento chamados (SF3 e SF4).

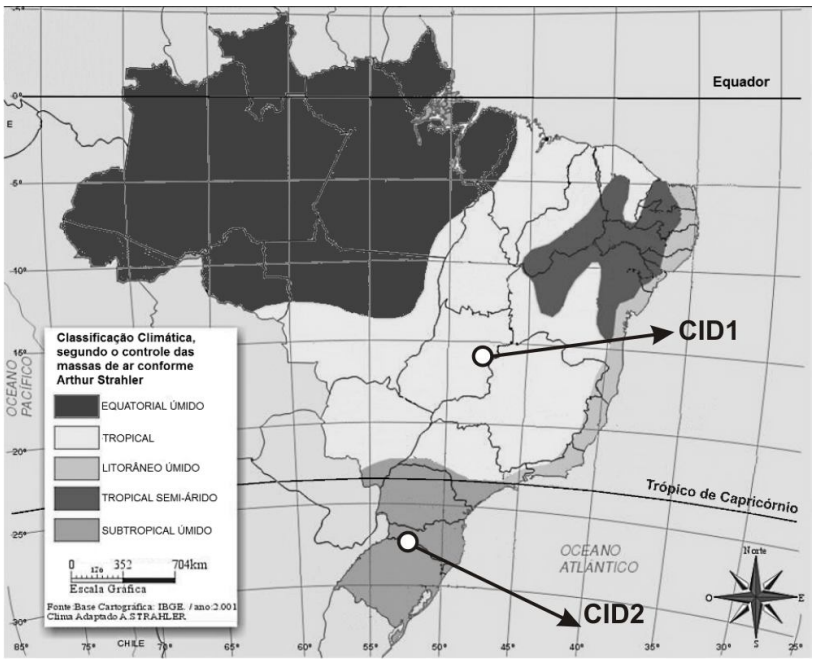

Figura 4.1. Mapa com classificação climática das cidades monitoradas. Fonte: Base Cartográfica IBGE, Ano 2001, Clima Adaptado A. STRAHLER.

É comprovada a influência das condições climáticas nos algoritmos de previsão de demanda, razão pela qual têm sido amplamente utilizadas como entradas para os mais variados modelos estatísticos multivariados (regressão, funções de transferência, vetores

autoregressivos, e redes neurais artificiais) para modelagem da demanda de água e previsão, principalmente para previsão diária (Homwongs et al., 1994). Assim, foi considerado um período de um ano de medidas para cada setor buscando a caracterização completa das variações do comportamento do modelo que porventura viessem a acontecer com a mudança das estações do ano bem como a variação do seu comportamento entre setores de um mesmo município.

Assim, para entender estas relações para o modelo específico, as séries de demandas dos quatro SF monitorados foram divididas em quatro períodos representativos das quatro estações do ano. Como o período mínimo de dados necessário para a obtenção do modelo ARIMA sazonal utilizado é de dois meses, foi separado, para cada estação do ano e em cada setor de fornecimento, o período correspondente aos dois primeiros meses de cada estação os quais foram estudados separadamente.

4.2 Tratamentos iniciais

As séries foram analisadas graficamente no intuito de investigar o seu comportamento temporal. A Figura 4.2 apresenta uma amostra de uma semana de dados retirada das vazões observadas, onde se observa a presença marcante da componente sazonal diária. Considerando o padrão apresentado, pode-se assumir que os valores podem ser modelados como uma série temporal periódica utilizando-se um modelo SARIMA (p,d,q)(P,D,Q) com sazonalidade diária (Shang et al, 2006).

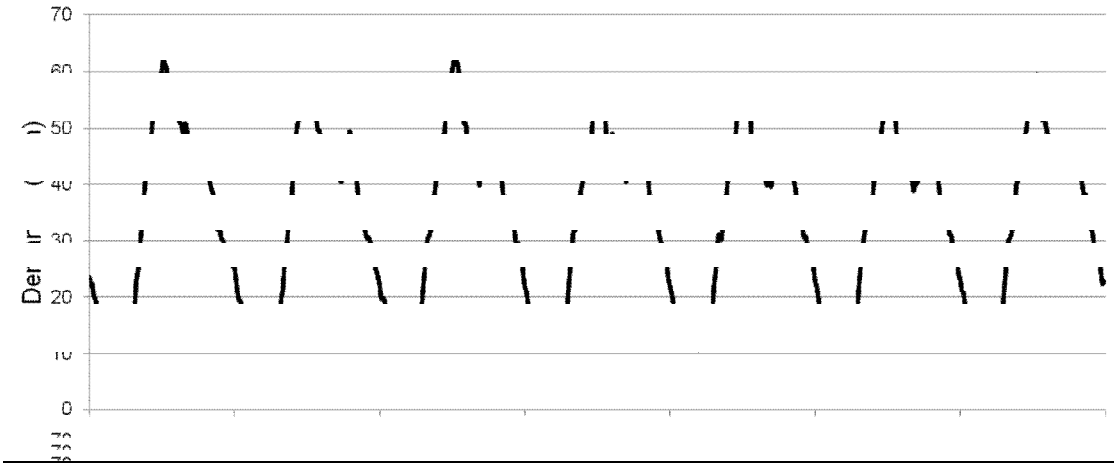

Figura 4.2. Vazão observada no período de uma semana.

Após a análise visual, as séries foram tratadas convenientemente, transformando-as em séries com intervalo de tempo entre pontos observados de 15 minutos, totalizando uma média de 5760 pontos para cada série. Cada série foi analisada separadamente para identificar e corrigir possíveis falhas de medição e manobras atípicas de rede. Tanto o processo de identificação das falhas como o processo de correção (preenchimento) foi baseado em um modelo ARIMA sazonal unitário provisório SARIMA(1,0,1)(1,1,1)(1,0,1) - sazonalidades diária e semanal - considerado não tão distante do modelo mais adequado para todas as séries. Assim, as observações que se distanciassem significativamente (nível de confiança de 95%) dos valores esperados pelo modelo SARIMA para o período respectivo eram consideradas falhas (outliers) e em seguida eram substituídas pelos valores esperados.

As Figuras 4.3, 4.4 e 4.5 mostram três tipos característicos de outliers tipicamente encontrados nas séries de demandas. A Figura 4.3 representa possivelmente uma manobra na saída do reservatório, pois inicialmente tem-se um súbita interrupção do abastecimento seguida de vazões elevadas para compensação. A Figura 4.4 mostra uma falha característica de um medidor em manutenção, que apresenta um intervalo atípico com valores constantes, após o qual não se percebe um acréscimo significativo das vazões que representaria uma compensação pelo desabastecimento temporário.

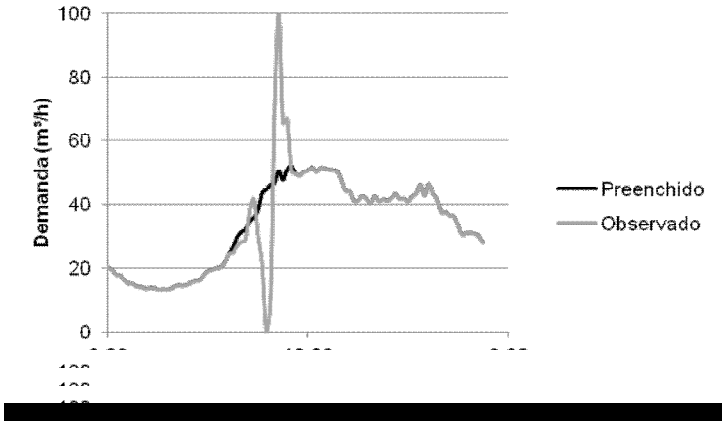

Figura 4.3. Falha em medição de vazão: possivelmente por manobra em reservatório.

A Figura 4.5 exemplifica uma mudança repentina de setorização ou um vazamento visível, pois o consumo aumenta significativamente por um período curto e logo volta a sua normalidade. Vale a pena ressaltar que os outliers não são, em todos os casos, erros de medição, porém em todos os casos devem ser corrigidos para utilização do algoritmo de previsão, pois representam comportamentos atípicos. Caso não sejam corrigidos, esses valores

de medição que fogem ao padrão habitual do local serão entendidos como padrões normais e portanto propagados nas previsões futuras.

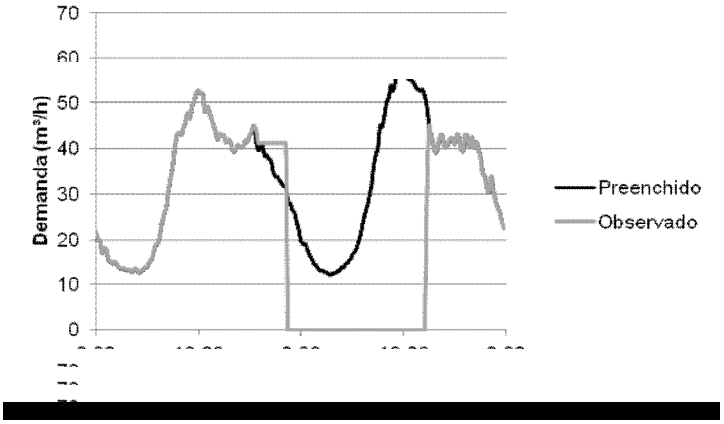

Figura 4.4. Falha em medição de vazão: possivelmente por falha no medidor.

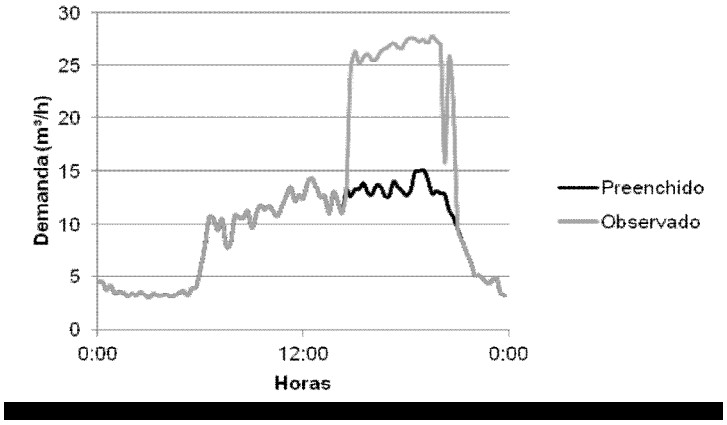

Figura 4.5. Falha em medição de vazão: possivelmente por mudança de setorização ou vazamento.

4.3 Identificação do modelo

Partindo para a etapa de identificação do modelo, fez-se o cálculo das funções de auto-correlação da série original, da série diferenciada diariamente e diferenciada sazonalmente (Figura 4.6). Segundo Box and Jenkins (1994) assume-se que o nível de diferenciação d, necessário para atingir a estacionariedade, é atingido quando a função de auto-correlação decai rapidamente para zero.

A série original (4.6 a) de demandas não apresenta este comportamento, retornando a valores altos de FAC em diferenças no tempo (lags) sazonais, assim deve-se escolher entre um grau de diferenciação sazonal para atingir a estacionaridade. Os dois tipos de diferenças

(a)

(4.6 b) e (4.6 c) satisfazem a condição de estacionaridade, porém a diferenciação semanal (4.6 c) reduz significativamente as auto-correlações diárias mais significativas o que prejudicaria o modelo. Assim, escolhe-se a diferenciação sazonal diária (4.6 b) com ordens de diferenças d=0 e D1=1 e D2=0.

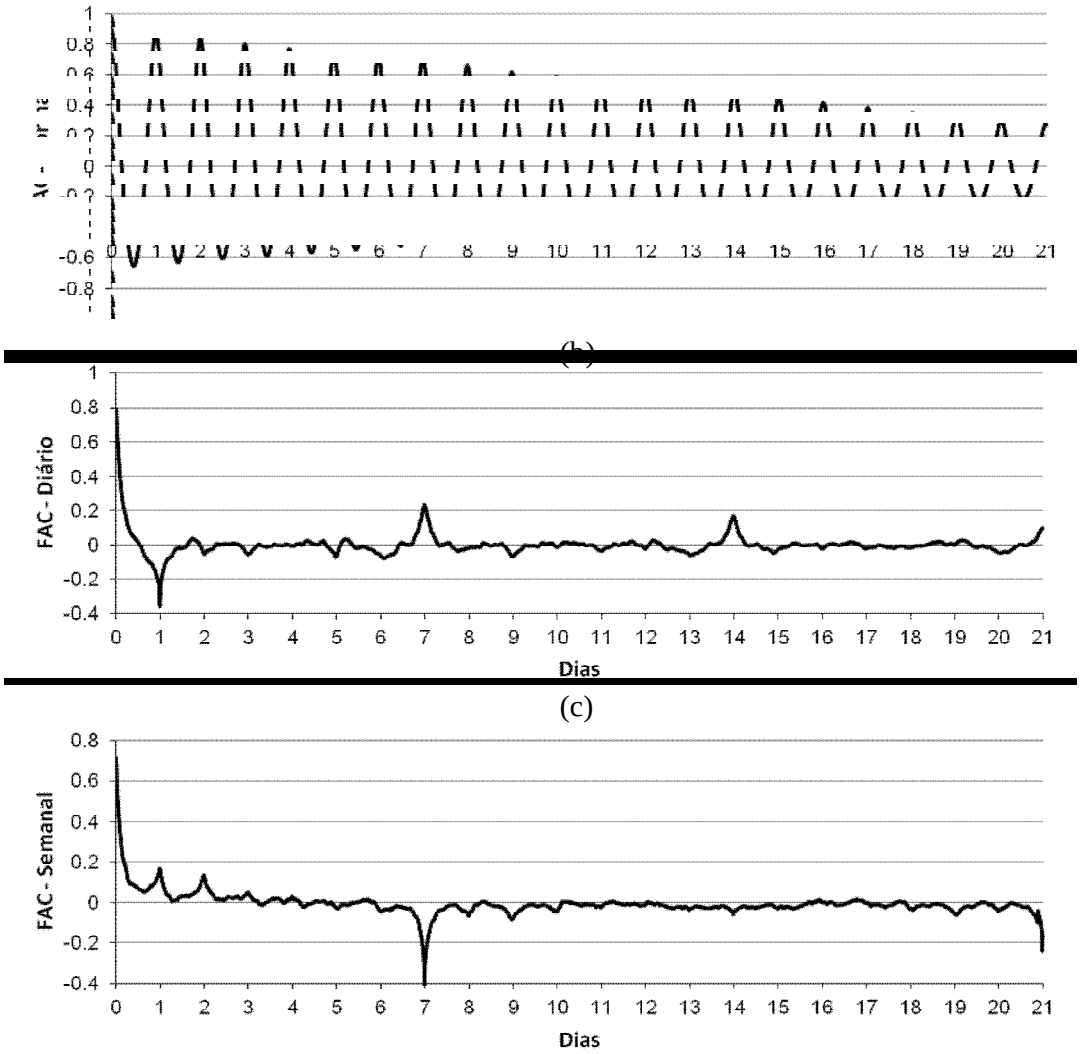

Figura 4.6. FAC da série original (a), série diferenciada diaria (b) e diferenciada semanal (c).

Partindo-se da utilização de diferenças diárias no modelo, as auto-correlações amostrais e as auto-correlações parciais da série diferenciada diária foram examinadas a fim de identificar as ordens restantes do modelo p, q, P1, Q1, P2, e Q2. As Figuras 4.7 (a) e 4.7 (b) apresentam, respectivamente, as autocorrelações e autocorrelações parciais médias para as séries de estudo. Estas função foram representadas simplificadamente apenas pelas suas três primeiras componentes nas lags normal, sazonal diária (S1) e sazonal semanal (S2).

Segundo a metodologia proposta, visualizam-se as ordens dos coeficientes a partir de um decaimento abrupto nas funções, as quais devem ser analisadas conjuntamente. A função de autocorrelações FAC fornece indicação para os coeficientes de médias móveis, enquanto a função de autocorrelações parcial FACP é representativa da indicação de coeficientes autorregressivos.

Analisando-se a Figura 4.7 para a componente normal percebe-se que a função FAC apresenta um decaimento suave, ao passo que a função FACP apresenta um decaimento abrupto após o lag 1, o que nos sugere a investigação de um modelo autorregressivo de ordem 1. Para a componente sazonal diária (S1) percebe-se um comportamento oposto - decaimento abrupto após o lag 1S1 para a função FAC e suave para a função FACP - que nos sugere a investigação de um modelo de médias móveis de ordem 1.

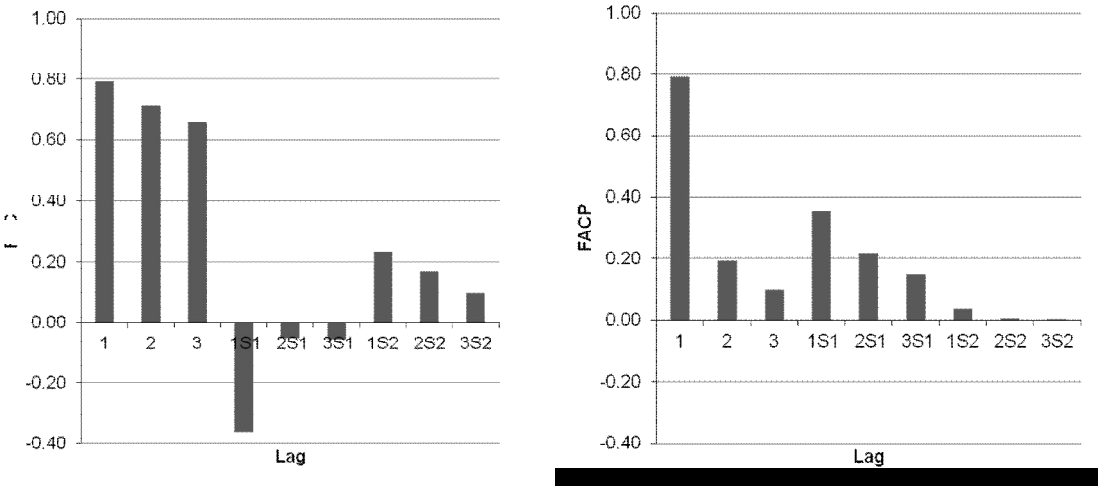

Figura 4.7. (a) função de auto-correlação da série diferenciada e (b) função de auto-correlação parcial da série diferenciada.

Para a componente sazonal semanal (S2), percebe-se um comportamento em que ambas funções possuem decaimentos suaves, o que sugere um modelo misto autoregressivo e de médias móveis. Shang et al. (2006) utilizaram um modelo SARIMA$(1,0,1)(0,1,1)_{24}$ para a modelagem horária da demanda de água com apenas um ajuste sazonal. Neste trabalho, um intervalo de tempo entre medidas foi tomado como 15 minutos, podemos dizer que a primeira identificação dos modelos nos sugere um modelo sazonal SARIMA $(1,0,0)(0,1,0)_{S1}(1,0,1)_{S2}$.

Juntamente com a análise do comportamento de autocorrelação médio das séries, foi realizada uma investigação para verificar se havia relações significativas entre as características de dependência temporal dos setores de fornecimento. Para tanto foi calculado

o coeficiente de correlação entre as autocorrelações de cada setor de fornecimento (Figura 4.8). Como esperado, foi obtida uma relação bem alta entre setores de fornecimento da mesma cidade, SF1 e SF2 (0.98) e SF3 e SF4 (0.99) o que demonstra que, por simplicidade, caso não haja dados suficientes em algum setores de fornecimento para determinação da ordem dos coeficientes do modelo, pode-se utilizar de estrutura semelhante para compor um outro setor de fornecimento.

	SF1			
SF1	1	SF2		
SF2	0.98	1	SF3	
SF3	0.90	0.94	1	SF4
SF4	0.88	0.92	0.99	1

Figura 4.8. Coeficiente de correlação entre as funções de autocorrelação dos setores de fornecimento.

4.4 Estimação dos modelos

A partir do modelo geral identificado, foi seguido com o procedimento iterativo de Box and Jenkins (1994), com ciclos consecutivos de estimação, diagnóstico e reidentificação dos modelos até que conseguiu-se chegar em modelos parcimoniosos mais adequados para cada série individualmente.

4.4.1 Otimização do Algoritmo de Busca

A estimação dos coeficientes do modelo foi feita através de um algoritmo de busca por harmonias originalmente desenvolvido por Geem (2011). Esta meta-heurística foi empregada para minimização da função objetivo de erros mínimos quadrados entre os valores observados e os valores da série ajustada pelo modelo com ordem fixa. Buscando-se evitar a obtenção de falsos modelos ótimos globais foi feita a otimização dos parâmetros característicos do algoritmo de busca por harmonias segundo um planejamento fatorial.

Lee e Geem (2005) recomendam o intervalo de valores dos parâmetros entre 0.7 e 0.95 para *hmcr*, 0.2 e 0.5 para *par*, e 10 e 50 para *hms* para se produzir um bom desempenho do algoritmo de busca HS. Estes valores são baseados empiricamente.

A função polinomial da Equação 3.2 contém nove parâmetros não conhecidos a serem determinados. Para diferentes modelos, diferentes tipos de planejamentos experimentais são necessários. Neste estudo, considerando o número de fatores, o número total de experimentos

possíveis, o número mínimo de experimentos necessário e o tempo disponível, foi decidido por realizar um planejamento fatorial 3^3.

Baseado nisso, a Tabela 4.1 apresenta os intervalos das variáveis utilizadas para a primeira avaliação do método da superfície de resposta. Para cada experimento dez avaliações foram feitas. Se um bloco de experimentos não atingisse 100% de modelos otimizados esta avaliação era desprezada.

Tabela 4.1. Intervalo dos fatores para a primeira avaliação.

Nível	HM	HMCR	PAR
-1	25	80	20
0	50	90	30
+1	100	95	40

O primeiro planejamento (3^3) foi utilizado para detectar os fatores ou variáveis independentes que tiveram o maior impacto no número de iterações. O modelo de segunda ordem pode ser ajustado como uma superfície tri-dimensional representando as interações em função de três fatores: tamanho da memória harmônica, taxa de consideração da harmonia e taxa de deslocamento lateral para o intervalo do experimento considerado. A Figura 4.9 mostra as superfícies ajustadas duas a duas para os três parâmetros em função do número total de interações em cada caso.

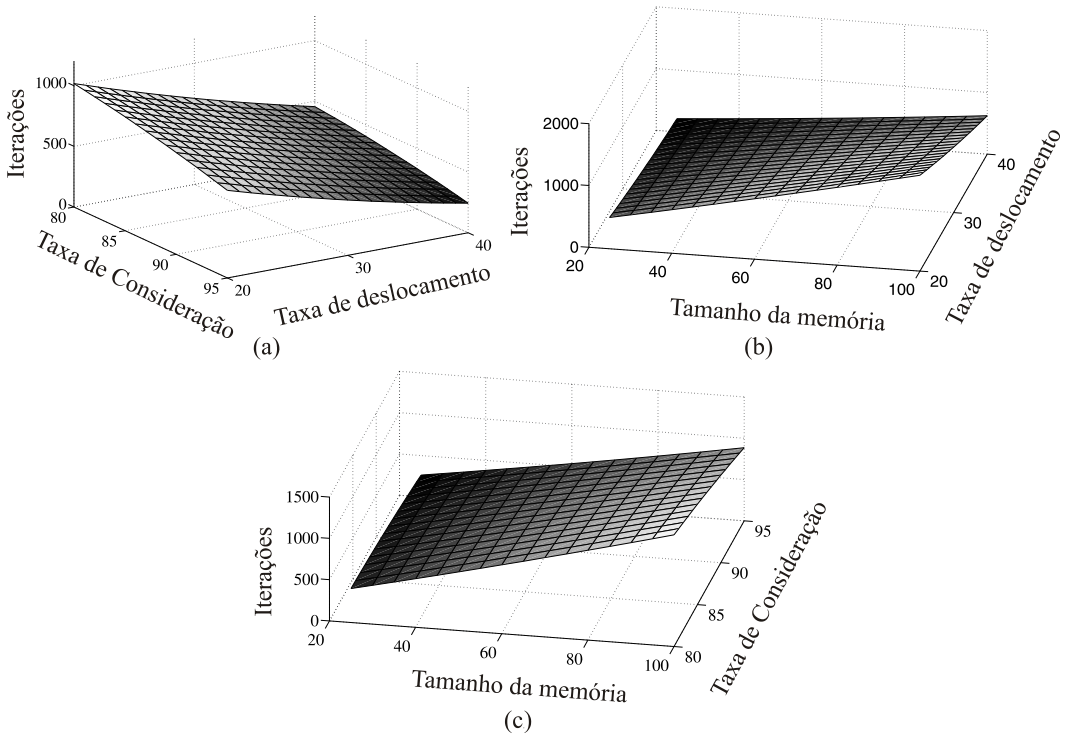

Figura 4.9. Superfícies ajustadas para análise da influência dos fatores.

Os resultados experimentais foram ajustados para o modelo completo de segunda ordem em três variáveis (Equação 4.1). Os parâmetros do modelo de segunda ordem foram determinados por regressão múltipla:

$$IT = -1425 + 46.5x + 55.5y - 45.4z - 0.28xy - 0.42xz + 0.28yz + 0.01x^2 - 0.38y^2 + 0.27z^2 \quad (4.1)$$

onde *IT* = número de iterações; *x* = tamanho da memória harmônica; *y* = taxa de consideração da memória; *z* = taxa de deslocamento lateral. Esta equação gerou um valor ótimo de *hms* = 17; *hmcr* = 86; *par* = 53.

A superfície de resposta correspondente ao modelo de segunda ordem indicou que o fator principal é o tamanho da memória harmônica o qual tem um efeito positivo. Sobre a taxa de deslocamento lateral, para pequenos valores de memória harmônica, o número de iterações aumenta com o seu aumento ao contrário do que acontece para altos valores de memória harmônica. A taxa de consideração da memória teve o menor efeito e, no range considerado, tem um efeito aproximadamente linear e negativo. Um valor mínimo de iterações é, portanto, conseguido com menores níveis de memória harmônica, níveis mais altos de taxa de deslocamento lateral e taxa de consideração da memória.

Baseado nestas indicações foi escolhido outra iteração fatorial em dois níveis. Foi escolhido variar o tamanho da memória harmônica em 10 e 25; a taxa de escolha da memória em 90 e 95 e a taxa de ajuste lateral em 40 e 55. A Tabela 4.2 mostra os resultados ordenados para o número de iterações:

Tabela 4.2. Resultados ordenados para o número de iterações (ótimo em cinza).

Posição	HM	HMCR	PAR	Iterações
1	10	95	40	7,100
2	10	95	55	7,600
3	10	90	40	8,200
4	10	90	55	8,700
5	25	90	55	12,500
6	25	95	55	13,900
7	25	95	40	15,100
8	25	90	40	18,400

4.4.2 Análise dos modelos obtidos

Utilizando os parâmetros otimizados (Tabela 4.2), o algoritmo harmônico foi utilizado para estimar os coeficientes dos modelos ARIMA sazonal para os quatro setores do estudo de caso. Após cada etapa de estimação, a adequação do modelo era avaliada utilizando as funções de autocorrelação dos resíduos (FACR) e a função de correlação cruzada (FCC) entre a série de resíduos e a série inicial. Respectivamente, caso existam picos nestas funções, estas indicam a necessidade da inserção de mais um coeficiente de médias móveis ou de um coeficiente autorregressivo na sazonalidade correspondente.

Quanto à variação dos modelos em escala espacial, Babel e Shinde (2011) ressaltam a necessidade do desenvolvimento de modelos de previsão de demanda de água específicos para cada cidade e Jowitt e Xu (1992) para cada zona de distribuição de uma mesma cidade. Percebe-se, pela Tabela 4.3, que em geral os modelos para uma mesma cidade em uma mesma estação do ano tiveram comportamentos próximos, o que não ocorreu para setores de cidades diferentes - o que valida parcialmente os resultados obtidos por Babel e Shinde (2011).

Quanto as questões climáticas, Homwongs et al. (1994) destacam que são necessárias atualizações sequenciais da previsão da demanda a fim de manter uma precisão aceitável para os períodos posteriores. Na Tabela 4.3 pode-se observar que as condições climáticas mostraram ter maior influência sobre os modelos gerados, até mais do que as relações regionais. Destaca-se o fato de que os modelos para ambas as cidades mostraram aproveitar as relações semanais para as estações de outono e inverno e não as incluíram nas estações de verão e primavera, o que é facilmente explicado pela maior variação climática destas duas últimas estações.

Por fim, analisou-se o caso da utilização de um modelo unificado para todos os setores de fornecimento em todas as estações, SARIMA(1,0,3)(0,1,1)(1,0,1). Quando da aplicação deste modelo temos os valores a direita destacados na Tabela 4.4 em asterisco para erro quadrático RMSE e critério de Akaide AIC. Percebe-se que, apesar de na maioria dos casos o critério AIC, ter aumentado indicando um excesso de parâmetros, os valores de erro de previsão diminuiram. Esta escolha é justificável, pois, ao se adotar um modelo único para todo o ano, evitam-se trabalhos desnecessários de troca de modelos, ficando-se a adequação dos modelos restrita apenas a nova estimação de seus coeficientes.

Tabela 4.3. Resumo dos coeficientes estimados.

Estação	Cidade	Setor	Modelo	Coeficientes						
				f0	t01	t02	f1	t1	f2	t2
Verão	CID1	SEF1	$(1,0,0)(0,1,1)_{S1}(0,0,0)_{S2}$	0.92				0.86		
		SEF2	$(1,0,0)(0,1,1)_{S1}(0,0,0)_{S2}$	0.89				0.93		
	CID2	SEF3	$(1,0,1)(0,1,1)_{S1}(0,0,0)_{S2}$	0.82	0.30			0.94		
		SEF4	$(1,0,1)(0,1,1)_{S1}(0,0,0)_{S2}$	0.90	0.24			0.97		
Outono	CID1	SEF1	$(1,0,1)(0,1,1)_{S1}(1,0,1)_{S2}$	0.93	0.39			0.95	0.92	0.78
		SEF2	$(1,0,1)(0,1,1)_{S1}(1,0,1)_{S2}$	0.86	0.25			0.91	0.94	0.81
	CID2	SEF3	$(1,0,1)(0,1,1)_{S1}(1,0,1)_{S2}$	0.94	0.58			0.91	0.90	0.84
		SEF4	$(1,0,1)(0,1,1)_{S1}(1,0,1)_{S2}$	0.86	0.39			0.94	0.90	0.80
Inverno	CID1	SEF1	$(1,0,2)(0,1,1)_{S1}(1,0,1)_{S2}$	0.92	0.37	0.20		0.94	1.00	0.76
		SEF2	$(1,0,2)(0,1,1)_{S1}(1,0,1)_{S2}$	0.96	0.38	0.16		0.94	0.95	0.74
	CID2	SEF3	$(1,0,1)(0,1,1)_{S1}(1,0,1)_{S2}$	0.81	0.42			0.94	0.92	0.83
		SEF4	$(1,0,2)(0,1,1)_{S1}(1,0,1)_{S2}$	0.94	0.51	0.16		0.92	0.90	0.80
Primavera	CID1	SEF1	$(1,0,0)(0,1,1)_{S1}(0,0,0)_{S2}$	0.93				0.97		
		SEF2	$(1,0,1)(0,1,1)_{S1}(0,0,0)_{S2}$	0.94	0.24			0.96		
	CID2	SEF3	$(1,0,1)(0,1,1)_{S1}(0,0,0)_{S2}$	0.94	0.54			0.95		
		SEF4	$(1,0,1)(0,1,1)_{S1}(0,0,0)_{S2}$	0.89	0.28			0.95		

Tabela 4.4. Resumo da comparação dos modelos.

Estação	Cidade	Setor	RMSE	AIC	RMSE*	AIC*
Verão	CID1	SEF1	0.688	3.86	0.683	3.93
		SEF2	0.868	4.32	0.859	4.39
	CID2	SEF3	0.605	3.62	0.600	3.67
		SEF4	0.484	3.18	0.482	3.23
Outono	CID1	SEF1	0.448	3.06	0.033	3.24
		SEF2	0.573	3.55	0.035	3.72
	CID2	SEF3	0.420	2.94	0.032	3.12
		SEF4	0.379	2.73	0.031	2.90
Inverno	CID1	SEF1	0.504	3.32	0.504	3.32
		SEF2	0.546	3.48	0.546	3.48
	CID2	SEF3	0.531	3.40	0.530	3.42
		SEF4	0.441	3.05	0.441	3.05
Primavera	CID1	SEF1	0.607	3.61	0.600	3.67
		SEF2	0.718	3.96	0.704	3.99
	CID2	SEF3	0.487	3.19	0.483	3.24
		SEF4	0.470	3.11	0.459	3.13

4.5 Previsões Obtidas

A Figura 4.10 mostra a aplicação do modelo obtido para a previsão com horizonte de 24 horas sobre a série de demandas de uma semana com características de erro médias. Mais detidamente, pode-se perceber que nesta semana o modelo se adaptou muito bem na maior parte dos dias, apesar do padrão de demandas variar significativamente de um dia para outro resultando em valores de erro MAPE próximos de 11%.

No segundo dia de previsão, porém, vemos um dia atípico, onde os valores de erro tendem a aumentar consideravelmente resultando em um valor de MAPE próximo a 20%. Mesmo neste dia, entretanto, o padrão dos valores previstos é semelhante ao dos valores observados, o que resultaria numa boa previsão em um horizonte menor (de algumas horas).

Nota-se que o modelo demonstra ser robusto no ajuste de dias típicos, principalmente onde as características climáticas são mantidas, além de tender a ajustar-se também razoavelmente em dias atípicos. Esta característica de adaptação rápida permite realizar o controle do sistema até mesmo em dias atípicos, desde que para isso haja macromedição online das demandas. Quanto à análise de erros, nota-se que são não correlacionados entre si, nem tampouco relacionados com a série original, o que demonstra que o modelo foi adequadamente identificado.

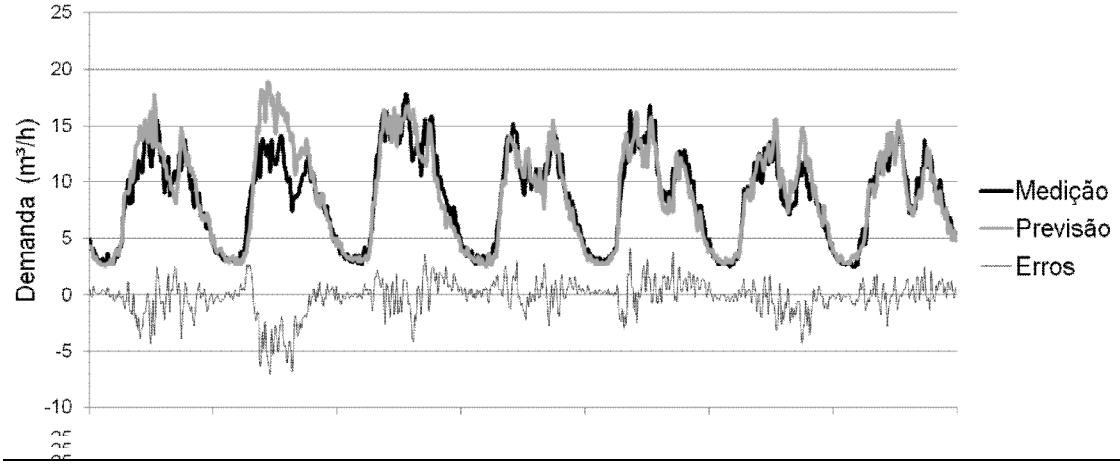

Figura 4.10. Resultados de previsão para horizonte de 24 horas em uma semana típica.

A Figura 4.11 mostra a evolução da precisão média das previsões realizadas com o crescimento do horizonte de previsão para o modelo ARIMA adotado, para um modelo determinístico de comparação determinado a partir da médias dos valores medidos nas

semanas passadas e do mesmo modelo ARIMA aplicado sobre a série original sem a retirada de falhas e outliers.

Foram gerados resultados pontuais para horizontes crescentes de 1h, 3h, 6h, 12h, 24h e 48h observando-se que sempre o erro médio crescia com o aumento do horizonte tendendo, mas sempre melhor do que uma previsão determinística. O valor médio de MAPE para um horizonte de até 24h foi próximo de 10.0%, o que indica que o modelo pode ser dito de alta precisão (Chang *et al.*, 2008). Além disso nota-se um prejuízo significativo em todos os horizontes de previsão, cerca de 2%, ao não se adotar a correção de falhas e outliers. Destaca-se que, para horizontes acima de 48h, o modelo aplicado sobre séries com falhas e outliers pode ser até pior do que uma previsão determinística, o que pode ser facilmente evitado.

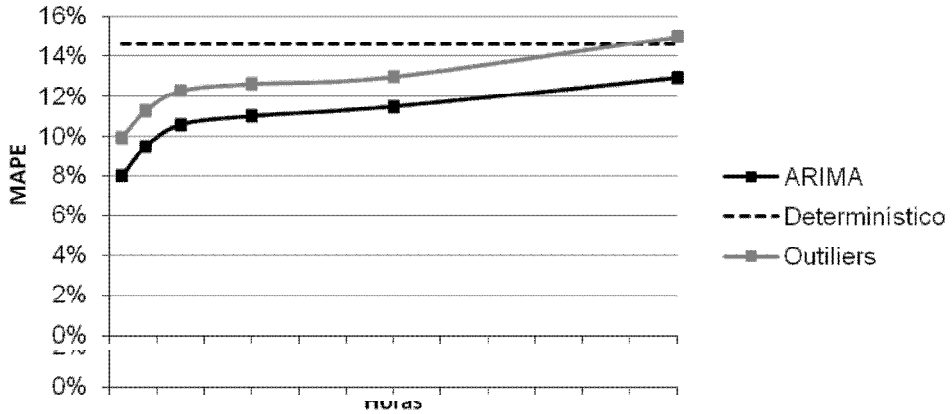

Figura 4.11. Avaliação do erro de previsão para diferentes horizontes para modelos com e sem retirada de valores outliers.

A Figura 4.12 mostra os valores médios de erro para cada estação do ano segundo se aumenta o horizonte de previsão. Nota-se que o comportamento para o erro é semelhante nas estações de Verão e Primavera bem como nas estações de Outono e Inverno. Este resultado vai ao encontro do que havia sido exposto na Tabela 4.4, pois há uma maior variação climática das estações de Verão e Primavera, o que dificulta a aplicação das componentes semanais.

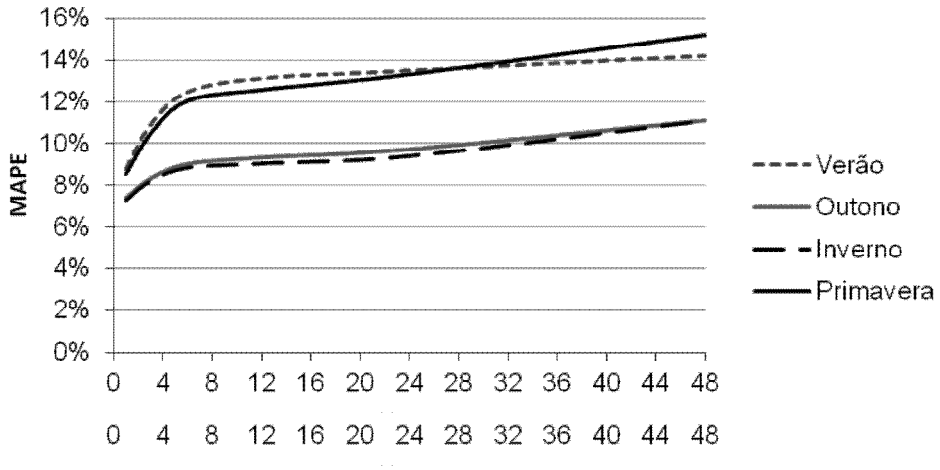

Figura 4.12. Avaliação do erro de previsão para diferentes horizontes nas diferentes estações do ano.

5

5 Conclusões

Neste trabalho, um modelo de previsão de demandas de água foi apresentado. O modelo foi baseado em análises de séries temporais e tem a vantagem utilizar apenas valores históricos da demanda de água como variáveis explicativas. Para estimar os parâmetros do modelo ARIMA foi aplicado um algoritmo recentemente desenvolvido chamado busca por harmonias. A convergência do algoritmo foi otimizada utilizando-se planejamento fatorial e o método de superfície de resposta.

Medidas reais de consumo de água residenciais de quatro setores de fornecimento em duas cidades brasileiras foram utilizadas para validar a metodologia de previsão proposta. Dessa forma cabe ressaltar que as conclusões tomadas são restritas as regiões analisadas, porém podem ser generalizadas com um certo nível de rigor.

Os resultados de simulação mostraram uma boa correspondência das previsões com as demandas de água medidas principalmente em dias típicos onde as características climáticas são mantidas. Isto demonstrou que o modelo ARIMA duplo sazonal adotado tem uma boa habilidade de compreender a estrutura de série temporal das demandas de água resultando em modelos de elevada precisão para a previsão um dia a frente. Além disso, o modelo tendeu a ajustar-se também razoavelmente em dias atípicos mostrando uma característica de adaptação rápida o que permitiria realizar o controle do sistema até mesmo em dias atípicos.

O valor médio do erro médio percentual MAPE para um horizonte de até 24h foi próximo de 10.0%, o que indicou que o modelo pode ser considerado de alta precisão. Além disso, nota-se um prejuízo significativo em todos os horizontes de previsão, cerca de 2%, ao não se adotar a correção de falhas e outliers o que demonstra a importância da execução deste tipo de procedimento.

Foi obtida uma alta relação entre setores de fornecimento da mesma cidade, porém foi verificado que existe uma variação significativa do comportamento do modelo para estações do ano distintas. Este resultado sugere a importância da realização do procedimento de identificação em cada localidade bem como o processo de estimação de parâmetros

periodicamente. Também pode-se observar que as condições climáticas mostraram ter maior influência sobre os modelos gerados do que as relações regionais.

Por fim, analisou-se o caso da utilização de um modelo unificado para todos os setores de fornecimento em todas as estações. Percebe-se que, apesar de na da maioria dos casos o critério AIC, ter aumentado indicando um excesso de parâmetros, os valores de erro de previsão diminuiram. Esta escolha se tornou justificável, pois, ao se adotar um modelo único para todo o ano, evitam-se trabalhos desnecessários de troca de modelos, ficando a adequação dos modelos restrita apenas à nova estimação de seus coeficientes.

6

6 Referências Bibliográficas

ADAMOWSKI, J.F. Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks. **Journal of Water Resources Planning and Management**, v.134, n.2, p.119-128, 2008.

ALVISI, S.; FRANCHINI, M.; MARINELLI, A. A short-term, patternbased model for water-demand forecasting. **Journal of Hydroinformatics**, v.9, n.1, p.39–50, 2007.

ALY, A.; WANAKULE, N. Short-term forecasting for urban water consumption. **Journal of Water Resources Planning and Management**, v.130, n.5, p.405–410, 2004.

ARANDIA-PEREZ, E.; UBER, J.G.; SHANG, F.; BOCCELLI, D.L.; JANKE, R.; HARTMAN, D.; LEE, Y. Preliminary Spatial-Temporal Statistical Analysis of Hourly Water Demand at Household Level. In: WORLD ENVIRONMENTAL & WATER RESOURCES CONGRESS, 2009, Kansas. **Anais….** Reston: ASCE, 2009. v.632, p.73-88.

BABEL, M.S.; SHINDE, V. Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok. **Journal of Water Resource Management**, v.25, n.6, p.1653–1676, 2011.

BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. **Time series analysis: forecasting and control**. Englewood Cliffs: Prince-Hall, 1994.

BOX, G.E.P.; WILSON, K.B. On the Experimental Attainment of Optimum Conditions. **Journal of the Royal Statistical Society**, v.13, n.1, p.1–45, 1951.

CAIADO, J. Performance of combined double seasonal univariate time series models for forecasting water demand. **Journal of Hydrologic Engineering**, 15(3), 215–222. (2010)

CHANG, K.; GAO, J.; CHEN, M.; YUAN, Y. Urban Daily Water Demand Short-term Forecasts Based on the Chaos Theory. In: INTERNATIONAL CONFERENCE ON NATURAL COMPUTATION, 4., 2008, Jinan. **Anais….** Washington: IEEE, 2008. v.4, p.696-699.

GEEM, Z.W.; KIM, J.H.; LOGANATHAN, G.V. A new heuristic optimization algorithm: harmony search. **Journal of Simulation**, v.76, n.2, p.60–68, 2011.

GHIASSI, M.; ZIMBRA, D.K.; SAIDANE, H. Urban Water Demand Forecasting with a Dynamic Artificial Neural Network Model. **Journal of Water Resources Planning and Management**, v.134, n.2, p.138–146, 2008.

GREENAWAY, G.; GUANLAO, R.; BAYDA, N.; ZHANG, K. Water Distribution Systems Demand Forecasting with Pattern Recognition. In: ANNUAL WATER DISTRIBUTION SYSTEMS ANALYSIS SYMPOSIUM, 8., 2006, Cincinnati. **Anais….** Reston: ASCE, 2006. v.126, p.1-16.

HOMWONGS, C.; SASTRI, T.; FOSTER, J.W. Adaptive forecasting of hourly municipal water-consumption. **Journal of Water Resources Planning and Management**, v.120, n.6, p.888–904, 1994.

JOWITT, P.W.; XU, C. Demand forecasting for water distribution systems. **Journal of Civil Engineering Systems**, v.9, n.2, p.105–121, 1992.

KIM, J.H.; KWANG, S.H. e Shin, H. S. A Neuro-Genetic Approach for Daily Water Demand Forecasting. **Journal of Civil Engineering**, v.5, n.3, p.281-288, 2001.

LEE, K.S.; GEEM, Z.W. A new meta-heuristic algorithm for continuous enginerring optimization: harmony search theory and practice. **Journal of Computational Methods in Applied Mechanical Engineering**, v.194, n.36-38, p.3902–3933, 2005.

LEON, C.; MARTIN, S.; JOSE, M.E.; LUQUE, J. EXPLORE-Hybrid Expert System for Water Networks Management. **Journal of Water Resources Planning and Management**, v.126, n.2, p.65–74, 2000.

LIU, H.; DENG, T.; ZHANG, H. Research on forecasting method of urban water demand based on fuzzy theory. In: INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS AND KNOWLEDGE DISCOVERY, 6., 2009, Tianjin. **Anais….** Washington: IEEE, 2009. v.6, p.389-395.

MONTGOMERY, D.C. **Design and Analysis of Experiments**. New York: Wiley, 2001.

MORETTIN, P.; TOLOI, C.M. **Análise de Séries Temporais**. São Paulo: Edgard Blucher, 2004.

PULIDO-CALVO, I.; ROLDÁN, J.; LÓPEZ-LUQUE, R.; GUTIÉRREZ-ESTRADA, J.C. Demand Forecasting for Irrigation Water Distribution Systems. **Journal of Irrigation and Drainage Engineering**, v.129, n.6, p.422-431, 2003.

SHANG, F.; UBER, J.G.; VAN BLOEMEN WAANDERS, B.G.; BOCCELLI, D.; JANKE, R. Real Time Water Demand Estimation in Water Distribution System. In: ANNUAL WATER DISTRIBUTION SYSTEMS ANALYSIS SYMPOSIUM, 8., 2006, Cincinnati. **Anais….** Reston: ASCE, 2006. v.8, p.95-108.

ZHANG, Q.J. Water Demand and Supply Management from Utilities Perspective. In: WORLD ENVIRONMENTAL AND WATER RESOURCES CONGRESS, 2007, Tampa. **Anais….** Reston: ASCE, 2007. v.483, p.1-10.