IVALTEMIR BARROS CARRIJO

Extração de regras operacionais ótimas de sistemas de
distribuição de água através de algoritmos genéticos
multiobjetivo e aprendizado de máquina

Tese apresentada à Escola de Engenharia de São Carlos
da Universidade de São Paulo, como parte dos requisitos
para obtenção do Título de Doutor em Hidráulica e
Saneamento.

Área de Concentração: Hidráulica e Saneamento
Orientadora: Profa. Assoc. Luisa Fernanda Ribeiro Reis

São Carlos
2004

FOLHA DE JULGAMENTO

Candidato: Bacharel **IVALTEMIR BARROS CARRIJO**

Tese defendida e julgada em 10-12-2004 perante a Comissão Julgadora:

Profa. Associada **LUISA FERNANDA RIBEIRO REIS (Orientadora)**
(Escola de Engenharia de São Carlos/USP)

Aprovado

Prof. Titular **FAZAL HUSSAIN CHAUDHRY**
(Escola de Engenharia de São Carlos/USP)

APROVADO

Profa. Dra. **MARIA CAROLINA MONARD**
(Instituto de Ciências Matemáticas e de Computação/USP)

Aprovado

Prof. Dr. **ANTONIO MAROZZI RIGHETTO**
(Escola de Engenharia de São Carlos/USP)

APROVADO

Prof. Dr. **EDEVAR LUIZOTTO JUNIOR**
(Universidade Estadual de Campinas/UNICAMP)

APROVADO

Profa. Titular **MARIA DO CARMO CALLJURI**
Coordenadora do Programa de Pós-Graduação em
Engenharia (Hidráulica e Saneamento) e
Presidente da Comissão de Pós-Graduação

Ao meu pai Pedro, mãe Galega (Perolina) e irmãos, Ilma, Dete, Bete e Carrijo, pelo amor, carinho, compreensão, amizade e, acima de tudo, por compartilhar os bons e maus momentos desde o início desta caminhada (a vida).

AGRADECIMENTOS ESPECIAIS

- Diretoria da Saneamento de Goiás S.A. SANEAGO – 2002-2005.
- *Centre for Water Systems* da Universidade de Exeter, Inglaterra.
- Professora Luisa Fernanda Ribeiro Reis.
- Professora Maria Carolina Monard.
- Professores Godfrey Walters, Dragan Savic e Soon Thiam e Doutores Edward Keedwell, Artem Babayan, Davi Prasad, Zoran Kapelan e (Ja)Kobus van Zyl e Sra. Sarah Adam.
- Dr. Peter Batista Cheung.
- Colemar, Ismar, Tião e Ismene.
- Vita de Sá.

AGRADECIMENTOS

- A Deus.
- À minha família.
- Ao CNPq pelo apoio financeiro concedido ao Projeto Temático (CT-HIDRO 01/2001).
- Aos Professores Fazal Hussain Chaudhry, Antonio Righetto, Edevar Luvizotto Jr e Edmundo Koelle.
- Aos especiais colegas da SANEAGO: João Guimarães, Mercia, Helio, Marilene, Wanessa Castro, José Ubaldo, José Grande e Augustinho.
- Aos colegas Sjobom, Edson, Eduardo, Eliana e Carlos.
- Aos colegas do Labsin, especialmente, Fernando, Alexandre, Zé Edu, Daniel e Klebber.
- Aos especiais funcionários do Departamento de Hidráulica e Saneamento, Sá, Pavi, Rose, Sergião, André, Flávia e Fernanda.
- Aos amigos de sempre de Sanca: Lara, Karina, Monique, Cyrão, Gilberto(Jeba), Luis (toco), Sonia, Quaresma, Marcelo di Julio, Fernandão e toda a galera da Confraria dos Pernetas.
- Aos amigos de guerra de Gyn: Camélia, Gordo, Monteiro, Cidinha, Álvaro, Jauner, César, Luiz, Nivaldo, Betão, Agostinha e Papagaio.
- À professora de inglês e grande amiga Beatrice Ballain.
- A todos os professores, funcionários e colegas do Departamento de Hidráulica e Saneamento da EESC-USP.

RESUMO

CARRIJO, I.B. (2004). *Extração de regras operacionais ótimas de sistemas de distribuição de água através de algoritmos genéticos multiobjetivo e aprendizado de máquina*. São Carlos, 2004. 217p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2004.

A operação eficiente do sistema é uma ferramenta fundamental para que sua vida útil se prolongue o máximo possível, garantindo o perfeito atendimento aos consumidores, além de manter os custos com energia elétrica e manutenção dentro de padrões aceitáveis. Para uma eficiente operação, é fundamental o conhecimento do sistema, pois, através deste, com ferramentas como modelos de simulação hidráulica, otimização e definição de regras, é possível fornecer ao operador condições de operacionalidade das unidades do sistema de forma racional, não dependendo exclusivamente de sua experiência pessoal, mantendo a confiabilidade do mesmo. Neste trabalho é desenvolvido um modelo computacional direcionado ao controle operacional ótimo de sistemas de macro distribuição de água potável, utilizando um simulador hidráulico, um algoritmo de otimização, considerando dois objetivos (custos de energia elétrica e benefícios hidráulicos) e um algoritmo de aprendizado para extração de regras operacionais para o sistema. Os estudos foram aplicados no sistema de macro distribuição da cidade de Goiânia. Os resultados demonstraram que podem ser produzidas estratégias operacionais satisfatórias para o sistema em substituição ao julgamento pessoal do operador.

Palavras-chave: otimização multiobjetivo; algoritmos genéticos; operação ótima; aprendizado de máquina.

ABSTRACT

CARRIJO, I.B. (2004). *Extraction of optimal operation rules of the water distribution systems using multiobjective genetic algorithms and machine learning*. São carlos, 2004. 217p. Thesis (Doctorate) – São Carlos School of Engineering, University of São Paulo, São Carlos, 2004.

The efficient operation of a system is a fundamental tool to postpone the system's service life as much as possible, thus ensuring a good service to the consumer while keeping electrical energy and maintenance costs at acceptable levels. Efficient operation requires knowledge of the system, for this knowledge, supported by tools such as models for hydraulic simulation, optimization, and definition of rules, provides the operator with proper conditions for the rational operating of the system's units without depending exclusively on personal experience while maintaining the system's reliability. In this work is developed a computational model for the optimal operation control of macro water distribution systems using a hydraulic simulator, an optimization algorithm, and a learn algorithm to extract operational rules (strategies) for the system. These studies are to be based on the macro system of the city of Goiânia, in Brazil. The results show that solutions for satisfactory operation can be quickly produced as a substitute to the personal judgment of the operator.

Key words: multiobjective optimization; genetic algorithms; optimal operation; machine learning.

LISTA DE FIGURAS

Figura 1.1 – Organização (Layout) desta Tese	4
Figura 3.1 – Conceito de soluções Pareto ótimas	49
Figura 3.2 – Metodologia do algoritmo genético Standard	57
Figura 3.3 – Algoritmo genético típico	58
Figura 3.4 – Esquemas de uma recombinação simples	64
Figura 3.5 – Exemplo de mutação	66
Figura 3.6 – Procedimento geral do NSGA (Cheung, 2004).....	72
Figura 3.7 – Procedimento geral para o SPEA (Cheung, 2004).....	75
Figura 3.8 – Procedimento geral do NSGA II (Cheung, 2004)	77
Figura 4.1 – Hierarquia do aprendizado	90
Figura 4.2 – Parte da árvore de decisão induzida por C4.5 Para o conjunto de exemplos Cleveland heart disease (BLAKE E MERZ, 1998)	91
Figura 5.1 – Fluxograma do modelo proposto	107
Figura 5.2 – Soluções - espaço bidimensional objetivo 1 x objetivo 2	111
Figura 5.3 – Soluções - espaço bidimensional objetivo 1 x objetivo 3	111
Figura 5.4 – Soluções – espaço bidimensional objetivo 3 x objetivo 2	112
Figura 5.5 – Esquema de macro-adição de Goiânia	116
Figura 5.6 – Goiânia – curvas das demandas horárias por áreas de reservação – Saneago S.A.(2004)	117
Figura 5.7 – Frentes Pareto - algoritmos multiobjetivo	126
Figura 5.8 – Frentes Pareto – operador mutação – recombinação BLX-a	127
Figura 5.9 – Frentes Pareto – operador mutação – recombinação uniforme	128
Figura 5.10 – Frentes Pareto – operador mutação – recombinação linear	128
Figura 5.11 – Frentes Pareto – recombinação uniforme – mutações uniforme e não uniforme	129
Figura 5.12 – Frentes Pareto para diversos números de gerações-algoritmo SPEA	129
Figura 6.1 – Goiânia – algoritmo multiobjetivo SPEA – frente Pareto	139
Figura 6.2 – Goiânia – frente Pareto - soluções ótimas e boas – melhores soluções AM	143
Figura 6.3 – Goiânia – Níveis de água nos reservatórios – solução AM1	143
Figura 6.4 – Goiânia – Níveis de água nos reservatórios – solução AM2	144
Figura 6.5 – Goiânia – Pressões nos nós de demanda – solução AM1	146
Figura 6.6 – Goiânia – Pressões nos nós de demanda – solução AM2	146
Figura 6.7 – Número de bombas em funcionamento – soluções AM1 e AM2	147
Figura 6.8 – Curva típica de nível de água – reservatório Amazônia	148
Figura 6.9 – Estratégia de funcionamento típico das bombas na estação elevatória SENAC	148
Figura 6.10 – Cronograma de funcionamento diário das bombas – solução AM1	149
Figura 6.11 – Cronograma de funcionamento diário das válvulas – solução AM1	149
Figura 6.12 – Cronograma de funcionamento das bombas – solução AM2	150
Figura 6.13 – Cronograma de funcionamento das válvulas - Solução AM2	150
Figura 6.14 – Soluções FO_1 x FO_2 e frente Pareto -utilização do OP	151
Figura 6.15 – Soluções FO_1 x FO_3 e frente Pareto - utilização do OP	152
Figura 6.16 – Soluções FO_2 x FO_3 e frente Pareto - utilização do OP	152
Figura 6.17 – Estratégia operacional selecionada – espaço FO_1 x FO_2 - OP	153

Figura 6.18 – Estratégia operacional selecionada – espaço FO_1 x FO_3 – OP	153
Figura 6.19 – Goiânia – Níveis de água nos reservatórios – solução OP	154
Figura 6.20 – Goiânia – Pressões nos nós de demanda – solução OP	155
Figura 6.21 – Número de bombas em funcionamento – Solução OP	155
Figura 6.22 – Cronograma de funcionamento das bombas – solução OP	156
Figura 6.23 – Cronograma de funcionamento das válvulas – solução OP	156
Figura 6.24 – Cenário de demanda 1 (0.75*demandas de referência)	157
Figura 6.25 – Cenário de demanda 3 (1.2*demandas de referência)	158
Figura 6.26 – Cenário de demanda 4 (1.4*demandas referência)	158
Figura 6.27 – Cenário de demanda 5 (1.5*demandas de referência)	159
Figura 6.28 – Frente Pareto e melhores soluções AM – cenário de demanda 1	161
Figura 6.29 – Frente Pareto e melhores soluções AM – cenário de demanda 3	162
Figura 6.30 – Frente Pareto e melhores soluções AM – cenário de demanda 4	162
Figura 6.31 – Frente Pareto e melhores soluções AM – cenário de demanda 5	163
Figura 6.32 – Cenários de demanda – estratégias selecionadas pelo AM	163
Figura 6.33 – Níveis de água nos reservatórios – estratégia geral – cenário 1	164
Figura 6.34 – Níveis de água nos reservatórios – estratégia geral – cenário 3	165
Figura 6.35 – Níveis de água nos reservatórios – estratégia geral – cenário 4	165
Figura 6.36 – Níveis de água nos reservatórios – estratégia geral – cenário 5	166
Figura 6.37 – Frente Pareto – níveis iniciais médios nos reservatórios - soluções AM	168
Figura 6.38 – Níveis de água nos reservatórios – níveis iniciais médios de água	168
Figura 6.39 – Pressões nos nós de demanda – níveis iniciais médios de água	169
Figura 6.40 – Número de bombas em funcionamento – níveis iniciais médios de água	169
Figura 6.41 – Frente Pareto – demandas máximas constantes - soluções AM	172
Figura 6.42 – Níveis de água nos reservatórios – demandas máximas constantes	172
Figura 6.43 – Pressões nos nós de demanda – demandas máximas constantes	173
Figura 6.44 – Número de bombas em funcionamento – demandas máximas constantes	173

LISTA DE TABELAS

Tabela 4.1 – Conjunto de dados no formato atributo-valor	92
Tabela 4.2 – Matriz de confusão de um classificador	96
Tabela 4.3 – Matriz de confusão de um classificador ideal	97
Tabela 4.4 – Matriz de confusão para problemas de classificação binária	98
Tabela 4.5 – Matriz de contingência para a regra $L \rightarrow R$	98
Tabela 4.6 – Definições de cobertura da regra <i>if L then R</i>	101
Tabela 5.1 – Valores numéricos das três funções objetivo	110
Tabela 5.2 – Aplicação do método OP nos dados da tabela 5.1	114
Tabela 5.3 – Goiânia – dados de demanda	118
Tabela 5.4 – Goiânia – dados dos reservatórios	118
Tabela 5.5 – Goiânia - dados das bombas da estação Senac	118
Tabela 5.6 – Goiânia – dados das válvulas de controle	118
Tabela 5.7 – Goiânia – dados dos nós	119
Tabela 5.8 – Goiânia – dados dos trechos	120
Tabela 5.9 – Goiânia – frente Pareto – descrição dos atributos	135
Tabela 6.1 – Goiânia – frente Pareto – descrição do conjunto de dados	139
Tabela 6.2 – Regras selecionadas e suas matrizes de contingência	140
Tabela 6.3 – Medidas genéricas para regras R1 e R2	141
Tabela 6.4 – Matriz de confusão	142
Tabela 6.5 – Goiânia – resultados do See5 – condição operacional	142
Tabela 6.6 – Comparação de custos das soluções	149
Tabela 6.7 – Descrição dos conjuntos de dados-AM	159
Tabela 6.8 – Matriz de confusão – cenário 1	159
Tabela 6.9 – Matriz de confusão – cenário 3	160
Tabela 6.10 – Matriz de confusão – cenário 4	160
Tabela 6.11 – Matriz de confusão – cenário 5	160
Tabela 6.12 – Goiânia – Resultados do See5 – condição operacional.....	160
Tabela 6.13 – Descrição do conjunto de dados – cenários de demandas-geral	164
Tabela 6.14 – Matriz de confusão – cenários de demandas – geral	164
Tabela 6.15 – Goiânia – Resultados do See5 – condição operacional.....	164
Tabela 6.16 – Descrição do conjunto de dados – níveis iniciais médios	167
Tabela 6.17 – Matriz de confusão – níveis iniciais médios	167
Tabela 6.18 – Goiânia – Resultados do See5 – condição operacional.....	167
Tabela 6.19 – demandas máximas constantes	171
Tabela 6.20 – Descrição do conjunto de dados – demandas constantes máximas	171
Tabela 6.21 – Matriz de confusão – demandas constantes máximas	171
Tabela 6.12 – Goiânia – Resultados do See5 – condição operacional.....	171

LISTA DE ABREVIATURAS E SIGLAS

AG	Algoritmo Genético
AM	Aprendizado de Máquina
OP	Ordem de Preferência
SANEAGO	Saneamento de Goiás S.A.
MOC	Método das Características
VCA	Válvula de Controle Automático
MHSDP	Modelos de Simulação Hidráulica Dirigidos pela Pressão
LM	Método de <i>Levenberg – Marquardt</i>
SD	<i>Sampling Design</i>
SCADA	<i>Supervisory Control and Data Acquisition</i>
SIG	Sistema de Informações Geográficas
CLP	Controlador Lógico Programável
WADSOP	<i>Water Distribution System Optimization Program</i>
IA	Inteligência Artificial
RNA	Redes Neurais Artificiais
AE	Algoritmos Evolucionários
SGA	<i>Simple Genetic Algorithm</i>
Pcross	Probabilidade de Recombinação
Pm	Probabilidade de Mutação
VEGA	<i>Vector Evaluated Genetic Algorithm</i>
MOGA	<i>Multi-objective Genetic Algorithm</i>
NSGA	<i>Non-dominated Sorting Genetic Algorithm</i>
NPGA	<i>Niched-Pareto Genetic Algorithm</i>
PAES	<i>Pareto Archived Evolution Strategy</i>
PESA	<i>Pareto Envelope-based Selection Algorithm</i>
SPEA	<i>Strength Pareto Evolutionary Algorithm</i>
MOMGA	<i>Multiobjective Messy Genetic Algorithm</i>
COMOGA	<i>Constrained Optimization by Multi-objective Genetic Algorithm</i>
mGA	<i>Messy Genetic Algorithm</i>
SMGA	<i>Structured Messy Genetic Algorithm</i>
IB	<i>Instance-based</i>
NB	<i>Naive Bayes</i>
MD	Mineração de Dados
KDD	<i>Knowledge Discovery in Databases</i>
DM	<i>Data Mining</i>

LISTA DE SÍMBOLOS

Acc	-	Acurácia do classificador (AM)
atr	-	Número de atributos (AM)
B	-	Bombas (variável de decisão – AG)
BH	-	Benefício hidráulico (AG – FO_2)
C	-	Classes preditas (AM)
cov	-	Cobertura (AM)
cu	-	Custo unitário (AG – FO_1)
D	-	Fator de demanda máxima (AG – FO_1)
E	-	Dados de treinamento (AM)
Err	-	Taxa de erro do classificador (AM)
ErrA	-	Erro aparente (AM)
Err10cv	-	Erro verdadeiro (AM)
F	-	Frentes não dominadas (AG); Número de exemplos erroneamente classificados (AM)
f	-	Função de aptidão (AG); Vetor função (AG)
g	-	Restrições de desigualdade (AG)
H	-	Carga hidráulica (AG – FO_1)
h	-	Restrições de igualdade (AG)
K	-	Índice de condição operacional (AM e OP)
L	-	Condição (AM)
M	-	Número de objetivos (AG); Atributos (AM); Matriz (AM)
Mean	-	Média – <i>cross validation</i> (AM)
N	-	Dimensão da população (AG); Exemplos (AM)
Nat	-	Nível de água atual (AG – FO_2)
Nmín	-	Nível de mínimo de água (AG – FO_2)
nov	-	Novidade (AM)
nrel	-	Confiabilidade negativa (AM)
Nreq	-	Nível de água requerido (AG – FO_2)
ns	-	Número de soluções na frente Pareto (AG)
P	-	Probabilidade (AM)
Pat	-	Pressão atual no nó de demanda (AG – FO_2)
PEP	-	Dimensão do conjunto da população externa (AG)
Pmín	-	Pressão mínima no nó de demanda (AG – FO_2)
POP	-	Dimensão do conjunto da população corrente (AG)
prel	-	Confiabilidade positiva (AM)
Preq	-	Pressão requerida no nó de demanda (AG – FO_2)
Pt	-	Conjunto de soluções
Q	-	População filha (AG); Vazão (AG – FO_1)
Qdem	-	Vazão horária nos nós de demanda
R	-	População combinada (AG); Conclusão (AM)
r	-	Número aleatório

S	-	Vetor solução (AG); População de cromossomos (AG)
sat	-	Satisfação (AM)
Sd	-	Desvio padrão (AM)
sens	-	Sensitividade (AM)
spec	-	Especificidade (AM)
sup	-	Suporte (AM)
T	-	Número de exemplos corretamente classificados (AM)
t	-	Geração atual (AG); Hora do dia (AG)
tacc	-	Precisão total (AM)
V	-	Válvulas (variável de decisão – AG)
x	-	Variáveis de decisão (AG)
x^*	-	Vetor de otimização (AG)
X^L	-	Limite inferior da variável de decisão (AG)
X^U	-	Limite superior da variável de decisão (AG)
Y	-	Classe (AM)
# Ex	-	Número de exemplos (Matriz de confusão – AM)
# MR	-	Número médio de regras induzidas (Matriz de confusão – AM)
# R	-	Número de regras induzidas (Matriz de confusão – AM)
a	-	Parâmetro do operador de recombinação BLX- a
?	-	Coeficiente do benefício hidráulico (AG – FO_2)
?	-	Parâmetro de funcionamento (AM e OP)
?	-	Rendimento do conjunto motor-bomba (AG – FO_1)
< >	-	Símbolos

SUMÁRIO

RESUMO	v
<i>ABSTRACT</i>	vi
Lista de Figuras	vii
Lista de Tabelas	ix
Lista de Abreviaturas e Siglas.....	x
Lista de Símbolos.....	xi
Sumário	xiii
1 INTRODUÇÃO	1
1.1 Objetivos	3
1.2 Organização da tese	4
2 APLICAÇÃO DE MODELOS MATEMÁTICOS EM SISTEMAS DE ABASTECIMENTO DE ÁGUA POTÁVEL	7
2.1 Simulação hidráulica	7
2.2 Calibração de sistemas de distribuição de água	13
2.3 Operação de sistemas de distribuição de água	18
2.4 Otimização de sistemas de abastecimento de água	24
2.4.1 Métodos de otimização	36
2.4.1.1 Métodos de programação matemática	36
2.4.1.1.1 Programação linear	36
2.4.1.1.2 Programação não linear	36
2.4.1.2 Programação dinâmica	36
2.4.1.3 Métodos heurísticos	37
2.4.1.3.1 Algoritmos genéticos	38
2.4.1.3.2 Lógica <i>fuzzy</i>	38
2.4.1.3.3 Otimização heurística	38
2.5 O uso de aprendizado de máquina (AM) em sistemas de abastecimento de água	39
3 ALGORITMOS EVOLUCIONÁRIOS	42

3.1	Introdução	42
3.2	Origens do pensamento evolutivo	44
3.3	Conceitos básicos de genética	45
3.4	Aplicação de algoritmos evolucionários em problemas de otimização multiobjetivo	47
3.4.1	Breve histórico	47
3.4.2	Implementação dos algoritmos evolucionários a problemas multiobjetivos	48
3.5	Algoritmos genéticos (AGs)	51
3.5.1	Terminologia	53
3.5.2	Visualização geral dos AGs	54
3.5.3	O algoritmo	56
3.5.4	Representação das soluções	58
3.5.5	Função de aptidão	59
3.5.6	Operadores genéticos	59
3.5.6.1	Seleção	59
3.5.6.2	Recombinação (<i>Crossover</i>)	61
3.5.6.3	Mutação.....	65
3.5.7	Crítérios de parada	67
3.5.8	Problemas de convergência	67
3.6	Otimização multiobjetivo utilizando algoritmos genéticos	67
3.6.1	Uso de funções de agregação (<i>aggregating approaches</i>)	69
3.6.2	Uso de técnicas não elitistas e não baseadas na teoria de Pareto.....	69
3.6.2.1	<i>Vector evaluated genetic algorithm – VEGA</i>	69
3.6.2.2	<i>Lexicographic ordering</i>	70
3.6.2.3	Método <i>e – constrained</i>	70
3.6.3	Uso de técnicas não elitistas baseadas na teoria de Pareto	70
3.6.3.1	<i>Multiple objective genetic algorithm – MOGA</i>	70
3.6.3.2	<i>Nondominated Sorting Genetic Algorithm – NSGA</i>	71
3.6.3.3	<i>Niched-Pareto Genetic Algorithm – NPGA</i>	71
3.6.4	Uso de técnicas elitistas baseadas na teoria de Pareto	71
3.6.4.1	<i>Pareto Archived Evolution Strategy – PAES</i>	71
3.6.4.2	<i>Pareto Envelope-based Selection Algorithm – PESA</i>	71
3.6.4.3	<i>Strength Pareto Evolutionary Algorithm – SPEA</i>	73

3.6.4.4	<i>Non-dominated Sorting Genetic Algorithm – NSGA II</i>	74
3.6.4.5	<i>Multiobjective Messy Genetic Algorithm – MOMGA</i>	76
3.7	O Uso de algoritmos genéticos na análise de sistemas de abastecimento público de água	76
4	APRENDIZADO DE MÁQUINA (AM)	85
4.1	Generalidades	85
4.2	Hierarquia do aprendizado	89
4.3	Aprendizado supervisionado: Definições	91
4.4	Indutores	101
4.5	Finalização	102
5	MODELO COMPUTACIONAL	103
5.1	Generalidades	103
5.2	Simulação hidráulica	106
5.3	Otimização operacional de sistemas de abastecimento de água potável	108
5.4	Extração de regras utilizando o algoritmo See5	109
5.5	Método de ordem de preferência (OP)	109
5.6	Descrição do problema em estudo	113
5.6.1	Sistema de abastecimento de água de Goiânia	114
5.6.2	Definição das funções objetivo	116
5.7	Implementação dos métodos evolucionários multiobjetivo para o problema de operação ótima de sistemas de distribuição de água	124
5.7.1	Representação das soluções	124
5.7.2	Definição do método evolucionário multiobjetivo e dos operadores genéticos e critério de parada	125
5.7.2.1	Definição do método evolucionário multiobjetivo	125
5.7.2.2	Definição dos operadores genéticos e critério de parada	126
5.8	Arquivo de dados para o sistema See5	130
5.9	Modelo computacional proposto	134
5.10	Verificação complementar com a utilização da rotina de ordem de preferência-OP	136
6	RESULTADOS E ANÁLISES	137

6.1.1	Resultados e análises considerando os parâmetros típicos fornecidos pela SANEAGO	138
6.2	Resultados e análises utilizando processo da ordem de preferência (OP).....	150
6.3	Resultados e análises considerando diferentes cenários de demandas.....	156
6.4	Resultados e análises considerando níveis iniciais médios nos reservatórios.	167
6.5	Resultados e análises considerando demandas máximas constantes.....	170
7	CONCLUSÕES E ESTUDOS FUTUROS	175
8	BIBLIOGRAFIA.....	181
8.1	Artigos apresentados pelo autor	181
8.2	Referências bibliográficas	182
	ANEXO A – Saída de dados do <i>See5</i> para a simulação relatada no item 6.1.....	199

Capítulo**1****INTRODUÇÃO**

*All men by nature desire knowledge.**- Aristotle*

Empresas concessionárias de serviços de água são criadas para construir, operar e manter sistemas de abastecimento público de água. A função básica dessas empresas é obter água de uma ou mais fontes, tratá-la para que atinja um padrão de qualidade apropriado ao consumo e distribuí-la ao longo da rede, sempre que houver solicitação por parte dos usuários. Tais empresas devem, portanto, avaliar a operação dos principais componentes do sistema de abastecimento no que diz respeito à qualidade ambiental do local de captação, adução da água bruta, reservação desta água, tratamento, reservação e distribuição final da água potável, assim como as unidades associadas. Devido às interações existentes, os reservatórios de distribuição são avaliados em conjunto com o sistema de distribuição final e os reservatórios de água bruta conjuntamente com as fontes de captação.

O principal objetivo de um sistema de distribuição de água é suprir os usuários de água em quantidade suficiente às suas necessidades, com pressões adequadas sob diversas condições de solicitação, definidas a partir de padrões de demandas nodais. O sistema pode estar sujeito a diferentes condições de consumo: demandas de incêndio em diferentes nós, picos de demandas diárias, uma série de padrões variando ao longo do dia ou críticas situações que ocorrem quando um ou mais tubos da rede são danificados. Assim, um sistema pode ser considerado confiável à medida que o seu abastecimento se processa de maneira satisfatória sob as diversas condições operacionais possíveis de ocorrer.

A operação de sistemas de abastecimento de água é uma tarefa bastante complexa, pois envolve vários aspectos distintos tais como a necessidade de garantia da

confiabilidade no atendimento dos serviços, economia no uso de equipamentos (energia e manutenção) e o planejamento de investimento para expansões futuras.

Embora a operação de um sistema de abastecimento de água seja entendida como uma mera seqüência de comandos exercidos sobre os equipamentos, que têm como objetivo o atendimento da demanda, na realidade, o problema é muito mais amplo, envolvendo aspectos de planejamento, controle e supervisão, serviços de infraestrutura de apoio e de atendimento ao usuário, todos considerados simultaneamente e interdependentes entre si.

A equipe de planejamento da operação geralmente define as regras (fixas ou variáveis) de controle dos sistemas, baseada em informações e experiências anteriores e no conhecimento do estado atual do sistema. As regras são transmitidas à equipe responsável pelo controle do sistema, que por sua vez, implementa tais regras e retorna os resultados das operações efetuadas ao setor de planejamento, para avaliação do desempenho e as adequações necessárias.

Para a obtenção de uma melhor regra de operação, três condições básicas são necessárias: o conhecimento do sistema, a definição clara dos objetivos a serem alcançados e a disponibilidade de modelos de análise hidráulica e otimização.

A definição dos objetivos é uma das tarefas mais árduas do processo, na medida em que vários aspectos não técnicos estão envolvidos. Após a implantação do sistema de abastecimento, por um determinado período, o sistema conta com uma capacidade ociosa que deve ser utilizada através de uma política de incentivo ao consumo. Entretanto, este procedimento deve ser adotado com cautela para que não ocorram desperdícios por parte dos consumidores. Nas ocasiões em que há oferta e demandas equilibradas, o objetivo da operação pode ser voltado à minimização de custos operacionais, de forma a maximizar os benefícios líquidos. Quando o sistema apresenta demanda superior à capacidade de produção, o objetivo passa a ser o de minimizar os prejuízos causados pela escassez do recurso oferecido.

Uma vez definidos os objetivos da operação, devem ser fixadas as políticas de operação. A definição das regras de operação exige que dois requisitos sejam obedecidos: a informação atualizada sobre o estado do sistema e a disponibilidade de modelos de simulação, calibração e de otimização que, acoplados, definem uma política de operação que procure atingir um objetivo pré-fixado.

O controle e a supervisão da operação são os responsáveis diretos pelos comandos de alterações da operação do sistema. Este controle pode ser efetuado de

forma simples, através dos elementos isolados, de forma local, ou de forma mais complexa, através da operação automática centralizada sem a presença de operadores locais.

Geralmente, quando a operação do sistema não é automática, os operadores procuram atender a demanda visando o custo mínimo e mantendo as condições de segurança do sistema. Os critérios de operação não seguem uma regra pré-fixada variando de acordo com a experiência e o conhecimento de cada operador. Quando a operação é automática, o operador passa a atuar como supervisor, corrigindo os desvios da situação real em relação àquela prevista pelos modelos. Quando ocorre um evento acidental não previsto na simulação, o operador passa a ser novamente o decisor das operações.

O acompanhamento constante dos operadores no controle da operação do sistema permite avaliar a eficiência das regras operacionais impostas pelos modelos e detectar elementos do sistema que se comportam de forma diferente da modelada.

Várias técnicas de otimização como programação linear, não linear e dinâmica e algoritmos genéticos (AGs) têm sido aplicadas a problemas de otimização operacional, com variados graus de sucesso. AGs têm se mostrado um dos mais eficientes métodos de otimização. Vários trabalhos na literatura apresentam a eficácia dos AGs em otimização operacional de sistemas de distribuição de água (GOLDBERG e KUO, 1987; SCHWAB et al., 1996; SAVIC e WALTERS, 1997; KLEMPPOUS et al., 1997 e RIGHETTO, 2002).

1.3 Objetivos

O objetivo desta pesquisa é desenvolver um modelo aplicável à otimização da operação de sistemas de distribuição de água para abastecimento público. Para tanto, deve-se, numa primeira etapa, proceder à análise e identificação dos objetivos mais promissores do sistema. Em seguida, deve-se avaliar e adotar técnicas de otimização que possibilitem um tratamento mais abrangente para um problema de natureza multiobjetivo. Finalmente, deve-se selecionar a(s) melhor (es) alternativa (s) dentre aquelas apontadas como resposta da fase anterior.

A revisão do “estado da arte” enfocará, portanto, as aplicações de modelos matemáticos em sistemas de abastecimento público de água, os algoritmos evolucionários como ferramentas que vêm se mostrando adequados ao problema em

estudo e o algoritmo de aprendizado de máquina, como ferramenta recente com potencial aplicação para os problemas de engenharia de uma forma geral.

O desenvolvimento do modelo proposto visa obter uma ferramenta útil e prática que minimize as intervenções subjetivas por parte dos operadores, que normalmente executam a tarefa de acordo com o grau de experiência pessoal adquirida no exercício de suas atividades diárias.

1.4 Organização da tese

Esta tese é organizada em sete capítulos, conforme o fluxograma apresentado através da figura 1.1.

Figura 1.1 – Organização (*Layout*) desta Tese

A primeira parte do trabalho (capítulos 2, 3 e 4) contém a revisão bibliográfica de modelos hidráulicos de sistemas de distribuição de água, otimização operacional e aprendizado de máquina.

A segunda parte (capítulos 5 e 6) apresenta o desenvolvimento de um modelo hidráulico de simulação, otimização operacional e algoritmo de aprendizado de máquina utilizado na extração de regras operacionais ótimas para operação de um sistema de distribuição de água potável, além dos resultados e análises. O modelo hidráulico de otimização e o algoritmo de aprendizado de máquina são desenvolvidos separadamente e, posteriormente, implementados em conjunto para a aplicação em uma parte do sistema de macro distribuição da cidade de Goiânia. Para verificação e avaliação dos resultados obtidos da aplicação do modelo proposto utilizando o AM, foi implementado um método multicriterial, denominado ordem de preferência – OP, para a escolha das melhores soluções da frente Pareto. Finalmente, as principais conclusões e sugestões, para estudos futuros, são apresentadas no capítulo 7.

Maiores detalhes do conteúdo de cada capítulo são apresentados a seguir.

- **Capítulo 2** - Neste capítulo é apresentada uma revisão bibliográfica da aplicação de modelos matemáticos em sistemas de abastecimento público de água. O capítulo é dividido em cinco subitens objetivando mostrar as diversas metodologias apresentadas pela literatura para este fim. Mesmo não sendo tema e objetivo do trabalho, apresenta-se um item sobre calibração de sistemas de distribuição de água, pela importância do tema, além de pesquisas na área do trabalho, como operação de sistemas de distribuição de água e métodos utilizados na otimização dessa operação.

- **Capítulo 3** – O objetivo principal deste capítulo é apresentar de forma abrangente a teoria e as diversas aplicações dos algoritmos evolucionários. São mostrados, a origem, os conceitos, as terminologias e, especialmente, os aspectos gerais das aplicações dos algoritmos genéticos multiobjetivo, que são os utilizados na elaboração do modelo de otimização proposto neste trabalho.

- **Capítulo 4** – São apresentados, neste capítulo, aspectos gerais sobre o algoritmo de aprendizado de máquina (*machine learning*). Conceitos, aplicações e formas de apresentação e avaliação de resultados são os aspectos mais importantes levantados.

- **Capítulo 5** – Neste capítulo, é apresentado o modelo computacional proposto pelo trabalho. É mostrada, primeiramente, a metodologia para o desenvolvimento do

modelo de otimização. São relatados os diversos aspectos considerados nos módulos do modelo, quais sejam, o simulador hidráulico, o módulo de otimização e o módulo de extração de regras operacionais. Neste capítulo também é introduzida a teoria do processo de ordem de preferência-OP (*Preference Ordering Routine-POR*), utilizado no trabalho, para avaliar os resultados obtidos através da aplicação do modelo proposto no sistema de macro distribuição de Goiânia. Além disso, é descrito o problema em estudo, com a definição das funções objetivo, modo de representação das soluções, definição do método multiobjetivo e operadores genéticos e, finalmente, do modelo geral proposto pelo trabalho.

- **Capítulo 6** – Neste capítulo são apresentados os resultados das avaliações implementadas. O capítulo é dividido em cinco subitens que descrevem as diferentes simulações implementadas, com o objetivo de avaliar o modelo desenvolvido para diversas situações operacionais e de projeto do sistema em estudo. Primeiramente, são avaliados os resultados usando o modelo de otimização e extração de regras desenvolvido utilizando os dados padrões do sistema de distribuição, fornecidos pela Saneamento de Goiás S.A. - SANEAGO; posteriormente, estes dados são mantidos, porém, é utilizado o método de ordem de preferência e não o algoritmo de aprendizado de máquina, como módulo de extração de regras; na terceira avaliação, é implementado o modelo desenvolvido utilizando diferentes cenários de demandas com o objetivo de avaliar a possibilidade de obtenção de uma única estratégia operacional para diferentes cenários; no quarto subitem, apresenta-se a avaliação do comportamento do modelo utilizando níveis de água iniciais médios nos reservatórios, no início do período de simulação (zero hora); por último, avalia-se o modelo proposto, considerando demandas diárias máximas constantes na área de influência de cada reservatório de distribuição.

- **Capítulo 7** – Neste capítulo são apresentadas as principais conclusões e são sugeridos alguns tópicos para estudos futuros na área do trabalho.

APLICAÇÃO DE MODELOS MATEMÁTICOS EM SISTEMAS DE ABASTECIMENTO DE ÁGUA POTÁVEL

*The most beautiful thing we can experience is the
mysterious. It is the source of all true art and
Science.*

- Albert Einstein

Pretende-se desenvolver um modelo de otimização para a operação de sistemas de distribuição de água, orientado para uma melhor visualização e aplicação, por parte do operador, dos procedimentos relativos às regras ótimas que possibilitem melhor desempenho do sistema. Para tanto, é imprescindível um conhecimento relacionado à simulação hidráulica, calibração, operação e otimização de sistemas de distribuição de água, além do conhecimento de técnicas capazes de fornecer regras para uma melhor operação das unidades que compõem estes sistemas. Estes aspectos, entre outros, serão abordados a seguir.

2.1 Simulação hidráulica

O termo simulação geralmente refere-se à representação do comportamento de um sistema físico de interesse. No caso específico dos sistemas de abastecimento de água, modelos matemáticos que representam as leis físicas que regem o fenômeno do escoamento de fluidos sob pressão são utilizados para fins de simulação. Nesse caso, as simulações visam reproduzir a dinâmica de um sistema existente ou proposto, e são

comumente utilizadas na fase de projeto, para o dimensionamento das unidades e, após o início de operação, para avaliação de seu desempenho.

Os modelos de simulação de sistemas de distribuição de água consistem do equacionamento das principais variáveis envolvidas no processo: vazões nos trechos de uma rede e pressões nos nós, entre trechos consecutivos. Podem ser utilizados na operação em tempo real, associados ou não a modelos otimizadores, de forma a fornecer uma resposta do sistema a uma dada manobra.

A análise da simulação e operação pode ser realizada admitindo-se regime permanente de escoamento ou variações das vazões e pressões ao longo do tempo. Neste segundo caso, o procedimento auxilia a operação do sistema, pois verifica se as vazões e as pressões são mantidas adequadas em todos os nós e se a reservação equilibra as distribuições da adução e da distribuição.

Geralmente, os modelos que consideram o escoamento em regime permanente, utilizam a equação da continuidade nos nós e a equação da energia ao longo dos trechos. O sistema de equações assim formado é resolvido pelo método de linearizações sucessivas de sistemas de equações, pelo método de Hardy-Cross ou Newton-Raphson. Davis e Jeppson (1979), Wood (1980), Shawcross (1985), Tanner (1985), Bhawe (1988), Hamberg e Shamir (1988) foram pesquisadores que desenvolveram modelos utilizando estes métodos.

Estes autores utilizaram basicamente dois procedimentos. No primeiro, calculam-se as vazões nos trechos e as pressões nos nós em um dado intervalo de cálculo, admitindo-se regime permanente. Ao final do intervalo, os níveis dos reservatórios são atualizados em função dos consumos do período e das vazões de adução obtidas no cálculo. No segundo procedimento, admite-se o escoamento em regime transitório, utilizando as teorias de coluna rígida ou de coluna elástica. Este segundo procedimento é mais complexo em termos de equacionamento e resolução, porém menos restritivo para aplicação em redes de distribuição em funcionamento. Ribeiro (1985) e Onizuka (1986) foram alguns dos pesquisadores que utilizaram este procedimento para avaliação da operação de sistemas de distribuição de água.

Com o objetivo principal de corrigir falhas decorrentes de equívocos e simplificações na fase de planejamento e projetos de redes de distribuição de água, é comum a instalação das chamadas válvulas de controle automático (VCAs) em sistemas em operação. Há vários tipos e modelos disponíveis no mercado, mas as mais usuais são as redutoras e as controladoras de vazão. Prescott e Ulanicki (2003) ressaltaram que a

literatura tem apresentado o desenvolvimento de vários modelos que representam o procedimento dinâmico operacional de vários componentes do sistema, porém estes modelos em geral não consideram a presença de válvulas redutoras de pressão. Os autores propuseram o desenvolvimento de um modelo para esta consideração, composto de quatro parâmetros dinâmicos: dois fenomenológicos, um procedimental e um linear. Os modelos variaram em complexidade, mas foram similares em termos de performance. Após a realização de alguns testes, os autores concluíram que os modelos fenomenológicos eram derivados de leis físicas e forneceram uma excelente, porém complexa, representação das válvulas redutoras. O modelo procedimental era mais simples e aplicável para a maioria das condições reais de funcionamento. O modelo linear não considerava a utilização das válvulas agulhas na válvula redutora e, portanto, tinha um uso limitado.

Com o intuito de modelar as válvulas de controle automático (VCA) em conjunto com outros elementos constituintes de redes de distribuição de água, Poll (1999) desenvolveu um modelo simulador para a análise do desempenho deste tipo de válvulas. No modelo, o autor considerou a rede hidráulica em qualquer situação operacional (permanente ou transitória). O sistema hidráulico automático de regulação da VCA, constituído pela válvula piloto, operada com o fluxo do líquido circulante na rede, foi analisado em conjunto com toda a rede hidráulica. O modelo simulador desenvolvido resultou da união de dois modelos matemáticos, um representando o comportamento físico da VCA e outro que representava a rede completa. A VCA foi simulada através do conjunto de equações diferenciais ordinárias que modelavam a dinâmica de seus principais componentes internos, incluindo a válvula piloto. A rede foi simulada utilizando o método das características (MOC), que era o algoritmo numérico utilizado para a resolução das equações básicas da continuidade e da quantidade de movimento, em condições transitórias do escoamento. Esta mesma consideração foi apresentada por Brunone e Morelli (1999), porém, avaliando apenas o transitório provocado na rede pela abertura e fechamento deste tipo de válvula. Righetto (2002) também propôs a utilização do MOC para o cálculo do escoamento permanente de uma rede sem VCAs. Como condição inicial do método numérico iterativo, o autor partiu de uma situação simples, hidrostática, para a rede. Apesar do transitório causado pelas aberturas bruscas nos nós de consumo, as quais impunham as vazões de demanda, o cálculo numérico com o MOC manteve-se estável, convergindo rapidamente para a solução desejada. Segundo o autor, a principal vantagem do uso do MOC é que as

equações são todas lineares e explícitas, sendo possível calcular, a cada passo do processo iterativo, os valores das cargas piezométricas nodais.

A avaliação da presença de válvulas automáticas nas redes de distribuição de água também fez parte dos estudos apresentados por Simpson (1999). O objetivo do trabalho era avaliar o desempenho de modelos simuladores comerciais quando introduzidas válvulas redutoras e sustentadoras de pressão e controladoras de vazão no sistema de distribuição de água. Da análise desenvolvida foi relatado que, em alguns casos, os modelos não convergiam para uma solução adequada. O autor utilizou o simulador hidráulico *EPANET* e outros modelos computacionais comerciais como simuladores e, como exemplo aplicativo, alguns esquemas hipotéticos onde as válvulas eram localizadas em pontos diferenciados. O autor não mencionou qual a versão do *EPANET* foi utilizada, assim como não fez referência aos outros pacotes comerciais. Além disso, algumas configurações propostas não são usuais em termos de sistemas reais. No estudo, concluiu-se que, para alguns casos, o *EPANET* não convergia de forma satisfatória.

Os modelos de simulação são ferramentas eficazes e necessárias não só na fase de planejamento e projetos, mas, principalmente, na fase de operação dos sistemas de distribuição de água, principalmente nos dias atuais em que a maioria das cidades já conta com sistemas em operação e, em geral, projeta-se a expansão do sistema existente e não um novo sistema. Zahed Filho (1990) e Souza (1994) ressaltaram que os modelos de simulação podem ser utilizados na operação em tempo real, de forma a verificar a resposta do sistema a uma dada manobra. A simulação por períodos extensivos ajuda na operação, verificando se as pressões nos nós e nos pontos de reserva apresentam valores adequados. O autor considerou que os modelos em períodos extensivos deveriam ser utilizados admitindo-se uma curva de consumo conhecida ou adotada. Estes modelos de simulação poderiam ser utilizados em estudos de dimensionamento ou de otimização de regras operacionais. Vários modelos de simulação para períodos extensivos foram desenvolvidos com a finalidade de auxiliar no planejamento e projeto de sistemas existentes, assim como na sua operação em tempo real (RAO e BREE, 1977; SOUSA e SILVA, 1987; SIMONOVIC, 1992; RAHAL, 1994; DANDY et al., 1997; KRITPIPHAT et al., 1998).

Também com o objetivo de avaliar a performance de um sistema de distribuição de água em operação diante de alguns riscos, Zongxue et al. (1998) apresentou um modelo de simulação combinado a um modelo de riscos que incorporava previsão de

demandas. O modelo foi aplicado no sistema de abastecimento de água da cidade de Fukuoka no Japão. No trabalho, o risco de obtenção de uma boa performance foi investigado para três alternativas: a operação do sistema existente quando a disponibilidade de água do rio Chikugo não era suficiente para suprir a necessidade de atendimento das demandas, redução de água distribuída em diversas porcentagens e, por último, o aumento do abastecimento de água quando era implementada a dessalinização da água do mar. O modelo era composto de três componentes: demanda, simulação e risco. Como o funcionamento do reservatório determinava o atendimento das demandas e a diminuição de riscos, o modelo de simulação basicamente previu a operação adequada desta unidade, prevendo diferentes análises de riscos para o sistema.

Um dos maiores entraves para a avaliação da operação de um sistema de distribuição é a quantificação das perdas e previsão das demandas. Apesar do rápido desenvolvimento de modelos avaliadores de sistemas de água, estes dois problemas ainda continuam sendo tratados com certa ressalva. Obradovic (2000) ressaltou que ambos os parâmetros são dependentes da pressão e que esta relação ainda não é conhecida. Ele propôs um melhoramento no modelo de simulação, caso esta relação passasse a ser conhecida. Porém, o trabalho mostrou apenas a necessidade de maiores pesquisas nesta área, essencialmente no que diz respeito a fatores internos e externos ao sistema, que influenciam na demanda e nas perdas. Recentemente, Soares et al. (2002), Soares (2003) e Soares et al. (2003) realizaram análises de desempenho de modelos de simulação hidráulica dirigidos pela pressão (MSHDP), admitindo que tanto as demandas como os vazamentos poderiam e deveriam ser explicitados como função das pressões. Os autores utilizaram modelos já desenvolvidos de demandas e perdas dependentes da pressão e propuseram um procedimento iterativo. Após as análises realizadas em uma rede de distribuição hipotética, concluiu-se pela viabilidade do emprego de modelos de simulação hidráulica dirigidos pela pressão (MSHDP), para a análise de redes.

Esta preocupação em considerar as perdas do sistema na formulação de um modelo de simulação hidráulica foi demonstrada por Hernández et al. (1999). Utilizando o simulador hidráulico *EPANET*, os autores desenvolveram um modelo denominado *HIPERWATER* que buscava atingir dois objetivos: minimização de perdas e melhoria no tempo de simulação. O modelo foi implementado através de computação de alta performance (através de computação paralela) e de um modelo de perdas com um módulo paralelo de otimização visando a redução de perdas através do controle de

pressão no sistema. O modelo de minimização de perdas não representou grandes contribuições em termos de eficácia hidráulica, porém ficou demonstrado que, para este tipo de implementação, a computação paralela representava uma eficiente ferramenta em termos de tempo de processamento.

A influência das perdas de água nos custos finais da operação de um sistema de distribuição de água foi estudada por Colombo e Karney (2002). Segundo os autores estes custos dependem de vários fatores que incluem o regime de demanda, a distribuição espacial das perdas e a complexidade do sistema. Eles utilizaram o simulador hidráulico *EPANET* e concluíram que, em geral, os custos de energia têm uma relação polinomial de segunda ordem com as perdas. Eles citaram que além dos custos, a qualidade da água distribuída poderia ser alterada em função das perdas.

A confiabilidade na operação de um sistema de abastecimento de água é fundamental para garantir o perfeito atendimento dos consumidores. Porém, quantificar este componente num modelo hidráulico não é tarefa fácil pela complexidade de suas variáveis. A quantidade de água que um sistema de distribuição deve entregar com uma pressão adequada é, segundo Tanyimboh et al. (2001), um dos principais fatores na determinação da performance e confiabilidade do sistema. Os autores utilizaram um modelo de simulação que relacionava a pressão nodal com a demanda e determinaram a confiabilidade do sistema para condições críticas de atendimento dessas demandas. A confiabilidade hidráulica do sistema também foi a preocupação de Ostfeld et al. (2002) que apresentaram um modelo de simulação estocástico, baseado no simulador hidráulico *EPANET*, quantificando três medidas de confiabilidade: a fração de volume distribuído, a fração de demanda atendida e a fração da qualidade de água distribuída. Os resultados mostraram que é possível quantificar a confiabilidade para um sistema específico considerando estas variáveis, porém, não de forma generalizada.

A confiabilidade associada à disponibilidade do sistema foi considerada no modelo de simulação proposto por Shistine et al. (2002). Os autores utilizaram um modelo de confiabilidade existente, baseado no método de mínimo corte, como interface para um modelo de simulação em regime permanente que implicitamente resolvia as equações da continuidade e energia. Os resultados do modelo de simulação eram usados no modelo de confiabilidade para definir os conjuntos de mínimos cortes e determinar os valores da confiabilidade nodal e do sistema. A disponibilidade foi definida como a porção do tempo em que o sistema atendia adequadamente as demandas. O modelo foi testado em uma rede que, devido às altas pressões, apresentava

vazamentos constantes. Os resultados mostraram a viabilidade do modelo para a identificação dos pontos de falhas na rede, devido o não atendimento das demandas.

Os sistemas de distribuição de água apresentam, geralmente, uma configuração complexa e de difícil compreensão e modelação. As várias unidades como reservatórios, estações de bombeamento, válvulas, redes de distribuição primária e secundária, são interligados, porém, não de forma padronizada, o que dificulta a simulação hidráulica para avaliação da operação do sistema ou mesmo para projetos de futuras expansões. Os cadastros técnicos em geral são defasados e não apresentam informações confiáveis do sistema. Preocupados com esta questão, Anderson e Al-Jamal (1995) sugeriram uma metodologia para simplificação de sistemas hidráulicos complicados. Eles propuseram a criação de um pequeno sistema hidráulico que imitaria o procedimento de um grande e complicado sistema. A metodologia utilizada, era baseada em programação não linear para minimizar a medida da diferença entre os sistemas original e simplificado. O modelo desenvolvido foi aplicado a uma parte do sistema de distribuição de água de Londres. Os resultados mostraram que era possível utilizar a metodologia proposta em grandes sistemas de distribuição de água. Apenas para o caso de válvulas de retenção instaladas na rede o modelo não apresentou bons resultados.

A falta de um critério único, em pormenores, para modelos de simulação de redes de água como definição da topologia, inicialização das variáveis de estado (estimativa inicial de vazões e cargas) e critérios de convergência, foi mencionado por Luvizotto Júnior (1995). Segundo o autor, este fator dificultava a comparação entre os modelos propostos por vários autores. Visando resolver este problema, ele propôs, para o caso de regime permanente, a elaboração de rotinas de simulação com base no modelo elástico, compostas por três módulos: módulo da topologia, módulo do modelo matemático e o módulo das condições de contorno. O autor afirmou que, com aplicação desta nova metodologia, a comparação entre os modelos e análise crítica entre eles se tornaram mais factíveis e viáveis.

2.2 Calibração de sistemas de distribuição de água

Os modelos de simulação hidráulica são largamente usados por planejadores, consultores e muitos outros profissionais envolvidos em análises, projetos, operação ou manutenção de sistemas de distribuição de água. Para tornar os modelos de simulação

de sistemas úteis é fundamental calibrá-los antes de utilizá-los (WALSKI, 1983, ORMSBEE et al., 1989 e RIGHETTO, 2001).

Mesmo utilizando dados coletados no sistema em operação para implementação e teste do modelo de simulação, o modelador não deve assumir que o modelo seja uma representação matemática fiel do sistema. O modelo de simulação hidráulica simplesmente resolve as equações da continuidade e energia usando os dados coletados. Desta forma, a qualidade dos resultados dependerá da qualidade dos dados de campo. A acurácia de um modelo hidráulico depende da qualidade dos resultados obtidos através da calibração. Assim, antes de utilizar os resultados obtidos pelo modelo desenvolvido, na tomada de decisões para operação das unidades do sistema, é fundamental a calibração deste. O processo de calibração é necessário pelas seguintes razões (WALSKI, 2001):

- confiança – resultados fornecidos pelo modelo computacional são frequentemente utilizados para tomadas de decisão na operação ou melhorias nos projetos hidráulicos. A calibração produz informações sobre as condições reais do sistema, aumentando a confiança e fornecendo ao responsável pela operação ou projeto, previsões sobre os procedimentos para melhoria do sistema.

- entendimento – o processo de calibração de um modelo hidráulico fornece excelentes informações dos procedimentos e performances do sistema. Com um melhor entendimento do sistema, o modelador terá uma idéia dos possíveis impactos das mudanças operacionais e das futuras expansões do sistema.

O objetivo desta pesquisa não é desenvolver um modelo para calibração de sistemas. Porém, a citação de alguns estudos nesta área tem como meta elucidar a relação entre este tema e aqueles propostos por este trabalho.

Calibração é um processo que consiste na comparação de valores de vazões e pressões previstas e observadas, para condições de operação conhecidas e que ajusta os dados de entrada de um modelo, para melhorar a concordância entre tais valores observados e previstos. Este conceito foi apresentado por Walski (1983) juntamente com uma técnica de calibração de redes de distribuição de água que dispensava o processo de tentativa e erro. Basicamente, eram observadas vazões e pressões nos nós (hidrantes) para duas condições de escoamento distintas e, através de dois coeficientes, as vazões e pressões eram corrigidas. Para o autor, um modelo de simulação pode ser considerado calibrado, se houver a previsão do comportamento do sistema para inúmeras e distintas condições operacionais e diversos usos da água.

Os maiores problemas na resolução dos modelos de simulação, residem na restrição das dimensões do sistema, na representação realística, na distribuição de vazões nos nós ao longo da rede e na determinação dos coeficientes de rugosidade dos trechos, por medidas de campo ou por calibração. Zahed Filho (1990) ressaltou que a calibração poderia ser realizada através de algoritmos matemáticos que estimam os parâmetros, de forma que os valores medidos no campo coincidam com os valores calculados no modelo de simulação. Segundo o autor, o conhecimento dos coeficientes de rugosidade dos trechos, através de medidas de campo, é um processo bastante difícil e oneroso, uma vez que apresenta uma característica dinâmica e, por conta disso, necessita de um levantamento abrangente e contínuo sobre todo o sistema.

Uma técnica de calibração explícita foi desenvolvida por Ormsbee e Wood (1986). De acordo com essa técnica, as equações utilizadas para caracterização hidráulica de uma rede eram determinadas em função dos ajustes das perdas de carga, que satisfaziam uma ou mais condições de pressões e vazões, medidas para uma dada condição operacional de demandas. Para os autores, um modelo de calibração deveria compreender sete passos básicos: identificação do uso pretendido para o modelo; determinação das estimativas iniciais dos parâmetros; coleta dos dados para calibração; avaliação dos resultados; execução da macro-calibração; realização da análise de sensibilidade e execução da micro-calibração.

Utilizando como parâmetros: pressão em determinados nós, consumo de água, produção da estação de tratamento de água, vazão em trechos selecionados e o nível de água em todos os reservatórios, Siwon (1998) apresentou uma metodologia destinada à calibração de sistemas de distribuição de água, baseada nas medidas realizadas num período de sete dias. Os dados de entrada eram relacionados com a topologia, com a representação esquemática dos nós e, com o comprimento, o diâmetro e o material dos trechos. O autor afirmou que as dificuldades na obtenção de resultados satisfatórios na calibração eram freqüentemente causadas por erros na estimativa inicial dos valores dos parâmetros, pela ausência de correlação entre o uso da água e a pressão necessária, bem como, pelos erros nas medidas realizadas no campo.

A principal finalidade do processo de calibração de uma rede de distribuição de água é ajustar os dados de entrada do modelo, visando reduzir o desvio entre os valores observados e prognosticados. Para prognosticar com eficiência o comportamento real de um sistema de distribuição de água, um modelo deve ser calibrado utilizando dados observados da rede. Cheung e Souza (2001) ressaltaram que as técnicas para calibração

estão divididas em duas categorias: aquelas que ajustam apenas os coeficientes de perda de carga, e as que ajustam os coeficientes de perda de carga e as demandas nos nós. Segundo eles, os coeficientes de rugosidade das tubulações e as demandas nos nós devem ser inicialmente estimados e ajustados para que os valores preditos se aproximem dos valores reais. Os autores implementaram um método clássico de calibração em um modelo de rede teórica e investigaram a influência do zoneamento de uma rede na acurácia dos resultados produzidos, concluindo que o aumento do número de zonas diminuía significativamente as discrepâncias entre os valores calibrados e os reais.

Amostragens de campo, em geral, são realizadas para apoiar atividades de modelação, especificamente para estimar os parâmetros de um modelo matemático, ou, mais precisamente, para calibrar o modelo. Neste caso, uma questão relevante para procedimentos de amostragens de campo é como maximizar a confiança nos valores de parâmetros estimados, dado um nível de trabalho amostrado. Bush e Uber (1998) abordaram esta questão, utilizando idéias estabelecidas em estimativa de parâmetro e teoria de projeto de amostragens. Eles propuseram métodos gerais, baseados em sensibilidade, para classificar os locais e tipos de medidas, com o objetivo de estimar os parâmetros para um modelo para sistema de distribuição de água. Estes métodos propostos eram sub ótimos, embora práticos, e eram aplicados com o objetivo de selecionar bons traçadores e locais de medição de pressão, para a estimativa do coeficiente de rugosidade do tubo. Os autores sugeriram que, quando comparadas com medições de pressão, as medições de traçador podiam ser informativas para calibração de parâmetros hidráulicos do sistema. Entretanto, devia-se tomar cuidado ao selecionar sua localização. Segundo os autores, utilizando os métodos propostos, uma seleção, tanto de medição do traçador como de pressão, melhorava a confiança na estimativa quando comparada àquela obtida usando medida de pressão ou traçador separadamente.

No sentido de dar suporte às atividades de operação, manutenção e reabilitação de sistemas de distribuição de água, é fundamental a previsão do comportamento hidráulico do sistema via computador. Entretanto, previsões realísticas dependem da etapa de calibração. Silva et al. (2002) propuseram um modelo de calibração de redes de distribuição de água em termos das rugosidades absolutas, de trechos dotados de singularidades e consideração de perdas por vazamentos. Os autores utilizaram algoritmos genéticos e verificaram o modelo em um setor da rede de distribuição da

cidade de São Carlos, SP. As respostas do modelo de calibração mostraram-se consistentes.

Kapelan et al. (2002) apresentaram um modelo para calibração de sistemas de distribuição de água utilizando um método híbrido baseado em algoritmos genéticos. Os autores aplicaram este novo método na calibração de um sistema de distribuição hipotético e compararam os resultados com os obtidos quando aplicados métodos de busca local como o de Levenberg-Marquardt (LM) ou de busca global como algoritmos genéticos. Foi utilizada uma forma simples de hibridização dos métodos AG e LM, em que, primeiramente era implementada a busca por AG e, posteriormente, com a melhor solução obtida era implementada a busca através do método LM. Os autores ressaltaram que a maior vantagem do método proposto, quando comparado com o AG era a eficácia computacional.

Com o objetivo de determinar localizações ótimas para os pontos de coleta de dados que seriam utilizados em modelos de calibração de um modelo hidráulico para sistemas de distribuição de água, Kapelan et al. (2003) apresentaram uma metodologia para *Sampling Design (SD)* formulado como um problema de otimização com dois objetivos: maximização da acurácia do modelo calibrado pela minimização das incertezas relevantes e minimização dos custos totais do *sampling design*. O problema de ótimo *SD* foi resolvido utilizando algoritmo genético multiobjetivo baseado em ranqueamento Pareto.

Ressaltando que, para grandes sistemas de abastecimento de água, é usual dividir as zonas de consumo para diversos reservatórios de distribuição, Martínez et al. (2003) apresentaram um modelo de calibração dirigido por válvulas com o objetivo de mostrar que era possível obter uma boa performance do sistema utilizando válvulas para a divisão destas zonas. De acordo com os autores, as dificuldades da calibração neste tipo de sistema aumentavam em função da grande quantidade de válvulas a serem reguladas. As principais vantagens eram a flexibilidade no controle das pressões e a redistribuição de vazões de acordo com as necessidades, assim como a redução da idade da água no sistema, porém, em contrapartida, estas vantagens adicionavam um alto grau de dificuldade na operação do sistema e, conseqüentemente, em sua calibração. Os autores aplicaram o modelo em uma parte do sistema de abastecimento de água da cidade de Valência, na Espanha. O sistema era composto de dois reservatórios elevados e 40 válvulas reguladoras de pressão. Os dados foram obtidos através do sistema *SCADA (Supervisory Control and Data Acquisition)* e do cadastro técnico em meio digital

através do sistema de informações geográficas (SIG). Foram utilizados algoritmos genéticos *standard* como processo de otimização e o *EPANET2* como simulador hidráulico. De acordo com os autores, os resultados finais demonstraram um eficiente modelo de sistema de água calibrado utilizando uma base de dados oriunda do SIG, podendo ser utilizado em procedimentos de tomadas de decisões para operação de sistemas.

2.3 Operação de sistemas de distribuição de água

Sistemas de abastecimento de água são grandes e complexos, com vários reservatórios, bombas e válvulas, vários quilômetros de redes de distribuição e milhares de consumidores. Os processos de captação, tratamento, transporte e distribuição envolvem vários especialistas e profissionais. No procedimento operacional, dados com a descrição do estado do sistema são capturados, processados e transmitidos para a gerência operacional. É, então, efetuada uma previsão para os procedimentos futuros. Obradovic e Lonsdale (1998) citaram três exemplos que podem ser utilizados para um gerenciamento adequado de sistemas de abastecimento de água, dependendo do prazo:

- gerenciamento de longo prazo – planejamento de novas fontes superficiais ou profundas, para captação e abastecimento futuros.
- gerenciamento de médio prazo – mudanças na operação do sistema para adequação aos horários de tarifas de energia elétrica de baixo custo.
- gerenciamento de curto prazo – redirecionamento de pessoal da área de manutenção para serviços específicos de redução de perdas no sistema.

Vários sistemas de abastecimento de água são caracterizados pelo crescimento elevado do consumo, em níveis que extrapolam as previsões e a capacidade de oferta da água. Este aspecto torna o custo da água elevado e faz com que o sistema passe a trabalhar em condições adversas, bastante distintas daquelas previstas em projeto. Luvizoto Júnior (1995) ressaltou que a falta de dados operacionais e as deficiências no atendimento enfatizam a necessidade da utilização de um controle operacional mais rígido para o sistema.

As formas de se operar um sistema de abastecimento de água são muito variadas e dependem de vários fatores, tais como a dimensão e complexidade do sistema, a experiência dos operadores, a disponibilidade de equipamentos apropriados para comunicação, para comando de estruturas de controle e para cálculos e a disponibilidade de modelos matemáticos para análises das informações. Outros aspectos

igualmente importantes são a segurança, a confiabilidade e os custos operacionais. Além destes, Gargano e Pianese (2000) acrescentaram outros aspectos. O consumo (demanda) que varia de forma aleatória, possíveis falhas ou remoção do serviço de um ou mais componentes eletromecânicos do sistema (tubulações, bombas, válvulas, etc.), quantidade de água disponível nos tanques para suprir o aumento diário ou semanal das demandas e a qualidade da água a ser distribuída ao consumidor, foram citados pelos autores.

O controle das várias unidades integrantes de um sistema de distribuição de água pode ser feito isoladamente ou através de uma central (controle local). O controle local, de acordo com Tardelli (1987), pode ser feito de três formas diferentes. Na primeira, manual, o operador maneja a estrutura de controle em função do estado local do equipamento e anota os dados operacionais e de consumo. A segunda forma é a automática não programada como, por exemplo, nos casos de reservatórios controlados por bóias e elevatórias comandadas por pressostatos. Nesses casos, há a necessidade de parâmetros definidos pelos planejadores para acionamento dos dispositivos de controle. A terceira forma é a automática programada que difere da anterior pela complexidade possível de regras operacionais, que podem incluir critérios mais flexíveis, além de limites máximos e mínimos adotáveis. Nessa forma, o comando do controle é efetuado através de redes ou de controladores lógicos programáveis (CLPs) que contêm microprocessadores capazes de adquirir e analisar dados e comandar um processo de emissão desses. Esse tipo de automação pode ser utilizado da mesma forma em controles centralizados.

No controle centralizado, as decisões são emitidas por um centro de controle específico. Neste caso, a forma mais simples é aquela em que o operador de uma unidade do sistema obedece a comandos do centro de controle e transmite os dados operacionais a este centro. Para uma eficiente resposta desses procedimentos, é imprescindível uma boa infra-estrutura de comunicação, factível em sistemas de grande porte, apesar de dificultar a análise de dados, uma vez que, geralmente, não são criados bancos de dados para consultas posteriores.

Para incrementar a confiabilidade e a segurança operacionais, pode-se prescindir do operador local, passando o sistema a ser controlado a partir do centro de controle. Em uma primeira modalidade, a decisão é tomada pelo operador central. A esse tipo de controle denomina-se *SCADA (Supervisory Control and Data Acquisition)*. As unidades do sistema devem ser controladas remotamente a partir do centro.

O estágio mais avançado de operação de um sistema de grande porte exige a mudança de uma forma *SCADA* para uma outra, denominada controle automático por computador, que exige algoritmos de otimização da operação.

O operador é indispensável, tanto na fase de transição de uma operação tipo *SCADA* para outra automática, quanto após sua implementação. A mudança deve ser feita gradualmente, com base nas análises de mudanças operacionais executadas *off-line* e apresentadas aos operadores para serem avaliadas. Somente após a comprovação da eficiência e segurança de um algoritmo desenvolvido para a operação de um sistema, é que se deve automatizá-lo.

A automatização do sistema pode atingir níveis diferenciados em função de sua própria necessidade, de sua importância e do risco de falhas admissível. Entretanto, mesmo quando o objetivo for um sistema de controle totalmente automático, será necessária a presença de operador, cuja função será a de censor das regras definidas pelo computador e de responsável pela operação do sistema em situações anômalas ou em casos de falha dos computadores de controle.

O controle automático de sistemas urbanos de distribuição de água, com a utilização de CLP, também foi proposto por Ericsson (1989). O autor ressaltou que através deste processo é viável controlar, com certo grau de confiabilidade, sistemas constituídos de um único reservatório de distribuição. Porém, para operar sistemas com vários reservatórios, mantendo o equilíbrio entre o consumo variável num extremo e suprimento relativamente constante no outro extremo do processo, é indispensável reunir grande quantidade de informações, analisá-las e agir em tempo, evitando o desequilíbrio que se traduz em deficiências no atendimento ao usuário. Desta maneira, um sistema deve possuir sensores transmitindo informações que, sendo processadas, permitam atuar no processo antes que o desequilíbrio ocorra. O desenvolvimento de estações remotas inteligentes, instaladas nos centros de distribuição, possibilita a concepção de sistemas de automação, produzindo assim um controle automático. Através da adoção deste princípio, pode-se automatizar sistemas de distribuição de grande porte, estabelecendo vários níveis de decisão e permitindo o aumento da eficiência no setor de distribuição de água.

A relativa facilidade de programação torna o CLP extremamente flexível no caso de aplicação e/ou alteração da lógica operacional, fator muito importante em termos de projeto, operação e manutenção. Kravosac et al. (1989) ressaltaram que, quanto mais complexo o controle a ser efetuado, mais indicado é o seu uso e considerou que a

implantação de um equipamento confiável para o controle operacional tem como objetivo, a otimização do processo, além do aumento da eficiência dos equipamentos e das instalações para que seja obtido um bom padrão de regularidade no abastecimento de água. Estes procedimentos evitam desperdícios de energia elétrica e de água, assim como a falta deles pode acarretar uma diminuição da vida útil dos equipamentos, causada por operações indevidas, revertendo em decréscimo dos custos operacionais e de manutenção.

O grau de automatização é variável e dependente do porte do sistema e de suas necessidades. Os sistemas mais complexos tendem a evoluir de uma operação baseada em *SCADA* para uma operação automática, de forma a liberar os operadores para tarefas de supervisão, visando incrementar a confiabilidade da operação (ZAHED FILHO, 1990).

O gerenciamento da distribuição é uma forma de organizar a estrutura técnico-administrativa, essencial para a melhoria da operação do sistema. Esta foi a proposta de Freitas (1989). O autor ressaltou que a operação de um sistema poderia ser definida como um conjunto de procedimentos que, através da supervisão e atuação, possibilitassem o funcionamento do mesmo, atendendo os objetivos definidos no projeto.

O gerenciamento da distribuição tem como objetivo básico, o atendimento do consumidor com eficiência e confiabilidade. As atividades ligadas ao atendimento são as que compõem a operação propriamente dita, de caráter permanente e ininterrupto, visando manter o sistema em funcionamento e atendendo as demandas. Estas atividades envolvem, dentre outras, o controle do abastecimento, a manutenção corretiva, o atendimento ao público, a geração de dados operacionais e relatórios de operação. A confiabilidade mecânica é o principal suporte da operação, pois envolve a inspeção e a manutenção preventiva das instalações, a análise dos dados operacionais e a análise do desempenho.

Obtenção de informações sobre os procedimentos operacionais, sobre a metodologia de cobrança de tarifa elétrica, sobre a variabilidade da topologia e das demandas, consiste num dos maiores problemas para uma operação adequada de sistemas de água. Preocupados com este aspecto, León et al. (2000) desenvolveram um sistema especialista híbrido chamado *EXPLORE* que foi desenvolvido para auxiliar no gerenciamento do sistema de abastecimento de água de Sevilha. O *EXPLORE* empregava a demanda de água prevista para obtenção de pontos de funcionamento

diário das bombas, reduzindo o custo da operação de bombeamento. Os autores ressaltaram que as informações repassadas pelos operadores do sistema serviam para um refinamento do processo operacional através da simulação de novas estratégias.

O desenvolvimento de um sistema especialista crítico, com a participação efetiva dos operadores do sistema pode evitar que várias ações propostas sejam inadequadas ou de difícil aplicação. Sheperd e Ortolano (1996) propuseram um sistema especialista que tinha como principal objetivo fornecer apoio à decisão para a operação de um sistema de abastecimento de água de grande porte. Os autores ressaltaram que a metodologia implementada aumentou a capacidade de um sistema especialista convencional pelo refinamento de um plano elaborado pelo operador. Este sistema avaliava o plano de operação e fornecia medidas auxiliares para melhoria do sistema além de alternativas flexíveis para operação.

Do ponto de vista técnico, o principal objetivo de um sistema de abastecimento é fornecer água potável em quantidade e pressões adequadas. Para atendimento deste objetivo, vários parâmetros operacionais devem ser determinados e avaliados. As decisões de operação são baseadas no estado atual e restrições operacionais do sistema.

A distribuição de água potável requer decisões contínuas na manipulação dos componentes físicos do sistema, permitindo o controle sob condições extremamente dinâmicas. Fatores como férias, variação das condições climáticas e sócio-econômicas de diferentes áreas da cidade e a presença de áreas industriais, podem afetar significativamente o consumo e, conseqüentemente, a distribuição de água. Além disso, restrições técnicas e considerações econômicas devem ser avaliadas.

As decisões, em geral, dependem da experiência dos operadores e do uso do sistema *SCADA* para obtenção de valores para as variáveis de estado do sistema, principalmente vazões e pressões. Angel et al. (1999) introduziram um modelo de lógica *fuzzy* para um sistema urbano de água constituído de várias estações de bombeamento, tanques e reservatórios, avaliando estes fatores. O modelo considerava a filosofia operacional e as contingências do sistema sob condições operacionais diversas. O modelo consistia de um sistema especialista *fuzzy* e foi utilizado para diagnóstico e previsões da operação das unidades do sistema.

O sistema de informações geográficas (SIG) é uma excelente ferramenta complementar para o gerenciamento e operação de sistemas urbanos de abastecimento de água. Porém, seu uso tem sido restrito pela falta de informação histórica dos sistemas. A maioria dos cadastros de sistemas de água não conta com informações

técnico-comerciais digitalizadas e acopladas a uma base de dados. Em função disto, a utilização é bastante limitada, particularmente, para propósitos operacionais.

Recentes avanços tecnológicos tratam o SIG como um verdadeiro modelador espacial em que, a conectividade com a rede, as características hidráulicas do sistema e suas relações com os consumidores são superiores a outras ferramentas disponíveis para o mesmo fim. Vários autores trataram deste assunto e propuseram a utilização desta ferramenta como forma de auxiliar na operação de sistemas de distribuição de água (SCHINDLER e GARRARD, 1999; MAKSIMOVIC e CARMI, 1999; BURROWS et al., 1999; WESTPHAL et al., 2003).

Um aspecto que já foi ressaltado anteriormente e que é fundamental como rotina na operação de um sistema de abastecimento de água é a confiabilidade. Mays (1989) definiu dois tipos de confiabilidade que devem ser consideradas neste contexto. A confiabilidade mecânica e a hidráulica. A mecânica é a capacidade que os componentes do sistema têm de atender as necessidades dos consumidores continuamente e, por longo período de operação, sem necessidade de freqüentes reparos ou modificações de componentes ou sub-componentes. O cálculo adequado desta confiabilidade requer conhecimento da composição e do funcionamento dos componentes básicos. No caso de sistemas de distribuição de água, os dados básicos sobre falhas e informações sobre reparos são escassos, o que dificulta o cálculo matemático da confiabilidade mecânica. A confiabilidade hidráulica é a capacidade que os componentes do sistema têm para fornecer serviço com aceitável nível de interrupções, mesmo em condições anormais. A avaliação da confiabilidade hidráulica recai diretamente sobre a função básica de um sistema de distribuição de água, qual seja: a distribuição de uma quantidade específica de água, para um local adequado, num intervalo de tempo previamente definido, à pressão desejável. A confiabilidade hidráulica é uma medida da performance de um sistema de distribuição. O desempenho hidráulico do sistema de distribuição depende, em grande parte, dos seguintes fatores: interação entre as unidades de adução, reservação, bombeamento e reguladores de vazão e pressão; confiabilidade/disponibilidade de componentes individuais do sistema e variação espacial e temporal de demandas.

A implementação de um modelo para avaliar a confiabilidade de um sistema de distribuição, reconhecendo suas incertezas, foi proposta por Xu e Goulter (1998). Neste modelo, as análises hidráulicas determinavam o comportamento da rede em termos de pressão nos nós, vazão individual dos componentes na rede e o nível de armazenamento

dos reservatórios. Os resultados das aplicações indicaram que o modelo era flexível e eficiente na determinação da confiabilidade, constituindo em avanço na integração entre variação da demanda, variação do nível de água nos reservatórios e a ocorrência de falhas no sistema de abastecimento de água.

Com o objetivo de avaliar a confiabilidade hidráulica em sistemas de distribuição de água, tanto na fase de projeto quanto de operação, Gargano e Pianese (2000) apresentaram uma metodologia baseada na análise estatística de índices de performance hidráulica, derivados de um grande número de simulações e de vários cenários de demandas de sistemas de distribuição de água e/ou condições operacionais. Os autores assumiram que o índice de confiabilidade hidráulica era a probabilidade de que, sob uma dada condição operacional, o índice de performance hidráulica estaria acima de um certo valor pré-fixado. Projeto e operação adequados para um sistema de distribuição de água dependem de um grande número de fatores, podendo ser citados, dentre outros, a demanda, que varia aleatoriamente com o tempo e espaço, a possível falha ou remoção de um equipamento eletro-mecânico do sistema (tubos, válvulas, bombas), a quantidade de água armazenada nos tanques para atender a variação diária e semanal da demanda, além da qualidade da água distribuída entre os usuários do sistema. No trabalho, a confiabilidade foi definida como a capacidade que o sistema possui de distribuir uma quantidade de água requerida pelos usuários, sob as diversas condições operacionais a que o sistema pode ser submetido.

Uma nova metodologia, para a interpretação de resultados da análise hidráulica convencional dirigida pela demanda, como uma aproximação para simulação dirigida pela pressão e, para calcular a confiabilidade em sistemas com uma única fonte de captação, foi proposta por Tanyimboh et al. (2001). Os autores definiram confiabilidade como sendo o valor, dentro de um intervalo de tempo determinado, da relação entre vazão distribuída e requerida. Segundo os autores, a quantidade de água que o sistema poderia distribuir com uma pressão adequada, era um dos principais fatores na determinação da performance e confiabilidade. Eles ressaltaram que a relação entre demanda e pressão nodais deveria ser incorporada em qualquer medida real de confiabilidade do sistema.

2.4 Otimização de sistemas de abastecimento de água

Os modelos de simulação hidráulica representam ferramentas poderosas para a determinação das características hidráulicas de um sistema de abastecimento de água.

Estes modelos podem ser usados para a determinação das principais características do sistema (pressões, operação das estações elevatórias, níveis de água nos reservatórios, etc.), porém, não podem determinar o ótimo funcionamento ou os custos mínimos de operação e manutenção (MAYS e YEOH-KOUNG, 1992).

A principal função de um sistema de distribuição de água é atender as demandas com pressões adequadas. A principal função do projetista e do operador é determinar um custo mínimo para operação e manutenção do sistema de forma a atender adequadamente o usuário. Em geral, o problema é determinar o custo inicial que pode incluir investimentos em tubulações, reservatórios, válvulas, bombas e os custos com energia elétrica das estações elevatórias. De forma geral, o problema de otimização de sistemas de distribuição de água potável pode ser descrito da seguinte forma:

Minimizar: custos de investimentos + custos de energia elétrica + impostos

Sujeito a:

- restrições hidráulicas
- atendimento das demandas de água
- pressões adequadas

A literatura reporta diversos modelos de otimização aplicados à operação de sistemas de abastecimento de água expressos na forma linear (SHAMIR, 1974; ALPEROVITS e SHAMIR, 1977; QUINDRY et al., 1981; MORGAN e GOULTER, 1985; CARRIJO, 1989).

Como visto, os problemas de otimização apresentam uma estrutura clássica, em que se procura maximizar ou minimizar uma função objetivo, respeitando um conjunto de uma ou mais equações ou inequações conhecidas como restrições. As funções objetivo e as restrições envolvem as variáveis de decisão, de estado e os parâmetros de processamento. Elas representam uma forma de medir o desempenho do modelo especificado através dos valores das variáveis de decisão, as quais definem como o sistema será operado. As restrições para problemas de sistemas de abastecimento de água podem ser divididas em três grupos (ORMSBEE e LANSEY, 1994):

- limitações físicas dos sistemas (capacidade dos reservatórios, capacidade dos mananciais, configuração das bombas, etc.);
- leis físicas (conservação de massa nos nós da rede, conservação de energia nos anéis, etc.); e
- requisitos externos (atendimento da demanda, manutenção dos níveis de pressão aceitáveis, etc.).

Um sistema de distribuição de água é composto de vários elementos que, em conjunto, determinam a operação confiável ou não das estações de bombeamento, adutoras, reservatórios, válvulas de controle e redes de distribuição.

Com o objetivo de otimizar a rede de distribuição de água da cidade de Goiânia, Carrijo (1989) apresentou um modelo computacional para dimensionamento econômico de redes de distribuição, utilizando programação linear inteira. A otimização do sistema foi obtida através de soluções sub-ótimas. Segundo o autor, foram obtidas configurações com custos inferiores a 60%, quando comparadas com aquelas especificadas no projeto original.

Várias técnicas têm sido apresentadas pela literatura para a otimização de sistemas de distribuição de água. Righetto (2002) propôs a utilização dos algoritmos genéticos para a otimização de regras operacionais, para um sistema de distribuição de água composto de reservatórios, estações de bombeamento, válvulas e rede de distribuição. O autor sugeriu a adoção de parâmetros de performance ótimos, relacionados com o consumo mínimo de energia elétrica, máxima confiabilidade, máximo atendimento da demanda e a minimização das perdas do sistema através da adoção de pressões adequadas nos nós de demanda. Para a implementação do método e atendimento dos objetivos propostos, o autor sugeriu um modelo hidráulico e de otimização combinados, baseado no método das características, nos algoritmos genéticos com a utilização de ranqueamento para seleção das soluções e na definição de parâmetros para obtenção da regra operacional mais adequada para a operação do sistema hipotético proposto.

Métodos evolucionários, como algoritmos genéticos, têm sido usados com grande frequência em otimização de projetos e operação de sistemas de distribuição. Um algoritmo de otimização denominado *Ant Colony Optimization Algorithms* (ACOAs) foi proposto por Maier et al. (2003). ACOAs são métodos evolucionários baseados no procedimento utilizado pelas formigas para busca de alimentos. Os autores propuseram a utilização do referido método na implementação de um modelo de otimização de projetos de sistemas de distribuição. Os resultados obtidos com a aplicação dos ACOAs foram comparados com os obtidos com a aplicação dos algoritmos genéticos. Os ACOAs mostraram-se, de acordo com os autores, mais eficientes, tanto em termos de tempo de processamento computacional quanto em termos da habilidade para a obtenção de soluções próximas das ótimas globais.

Com o objetivo de resolver um problema de otimização multicriterial, na determinação de uma estratégia ótima para o controle de um sistema em tempo real, Bhattacharya et al. (2003) propuseram a utilização de técnicas de aprendizado de máquina (*machine learning*) como uma rede neural artificial (*artificial neural network-ANN*) e aprendizado reforçado (*reinforcement learning-RL*). Segundo os autores, a programação não linear e/ou a dinâmica poderiam ser usadas para a obtenção da solução, porém, para sistemas complexos, esta técnica consumiria um tempo computacional extremamente elevado. Na proposição, o *RL* foi usado para diminuir o erro do componente baseado no *ANN*. O modelo foi testado no controle ótimo da operação de um complexo sistema de abastecimento de água da Holanda e os resultados, de acordo com os autores, foram bastante satisfatórios.

Um método meta-heurístico para resolução de problemas de otimização discreta na otimização de projetos de sistemas de distribuição de água, denominado *shuffled frog leaping algorithm (SFLA)*, foi proposto por Eusuff e Lansey (2003). O modelo foi utilizado na determinação de dimensões ótimas de tubulações para expansões de sistemas de distribuição de água. Segundo os autores, o modelo admitia a troca de informações entre as buscas locais, orientando para a obtenção de um ótimo global. Foi desenvolvido um modelo computacional denominado *SFLANET* que conectava a técnica *SFLA* com o simulador hidráulico *EPANET*. O modelo foi aplicado a alguns sistemas exemplos utilizados pela literatura e apresentou resultados promissores, quando comparados aos de técnicas tradicionalmente utilizadas para o mesmo fim.

Utilizando um modelo denominado *Structured Messy Genetic Algorithm (SMGA)*, Walters et al. (1999) propuseram a otimização de projetos de sistemas de distribuição de água existentes, considerando adição de novas tubulações, recondicionamento e substituição de tubulações existentes, locação e dimensionamento de novos reservatórios, e estações elevatórias. O modelo foi aplicado ao sistema teórico denominado *Anytown* e apresentou resultados bastante promissores quando comparados com os obtidos da aplicação de outras técnicas, apresentadas pela literatura, no mesmo sistema.

Com o objetivo de aprimorar a utilização dos algoritmos genéticos na otimização operacional de sistemas de distribuição de água, Zyl et al. (2004) propuseram a utilização de um método híbrido que combinava o algoritmo genético com uma estratégia de busca *hillclimber*. Segundo os autores, as estratégias *hillclimber* complementavam os algoritmos genéticos aumentando a eficiência na busca de um

ótimo local. Foram investigadas duas estratégias diferentes: o método de Hooke e Jeeves e o de Fibonacci. O método híbrido provou ser melhor que o algoritmo genético puro na busca rápida da solução ótima, tanto para um sistema teórico simples quanto para um sistema real complexo.

A complexidade da otimização de um sistema de distribuição de água reside na dinâmica e quantidade de parâmetros e variáveis envolvidos. Vários estudos têm sido desenvolvidos nas últimas décadas visando adequar estas dificuldades à realidade operacional dos sistemas. Para facilidade de aplicação do processo, vários pesquisadores têm implementado modelos de otimização para componentes isolados do sistema (estações de bombeamento, redes de distribuição, válvulas, reservatórios), considerando a variação das demandas ao longo do dia e os diferentes custos de energia elétrica, buscando uma confiabilidade adequada, em termos de atendimento ao consumidor. Há autores que preferem considerar as unidades do sistema (bombas, válvulas, reservatórios, entre outros) conjuntamente no modelo de otimização, tratando o problema com múltiplos objetivos. Estes objetivos podem ser: a confiabilidade hidráulica e mecânica, os custos de operação e manutenção e os benefícios hidráulicos. Em geral, grande parte dos sistemas de abastecimento de água já está implantada e em plena operação por alguns anos ou até décadas. Em função disso, as unidades componentes apresentam problemas ou falhas nos procedimentos operacionais que podem acarretar o não atendimento das necessidades dos usuários. Em função disso, a otimização da reabilitação dessas unidades tem sido um dos temas apresentados na literatura especializada em desenvolvimento de ferramentas para a melhoria dos sistemas.

Dentre as unidades de um sistema de abastecimento de água, as estações de bombeamento são as que apresentam maior complexidade com relação às questões de custos, confiabilidade, benefícios hidráulicos e flexibilidade. Em função disso, vários autores têm preferido desenvolver modelos de otimização considerando exclusivamente o funcionamento desta unidade.

Bombas aparentemente operadas com eficácia são similares àquelas que desperdiçam energia elétrica. O problema maior é identificar quais bombas são ineficientes e porquê. A melhor abordagem depende da utilidade e do orçamento. Walski (1993) enumerou uma série de causas que podem levar uma bomba a operar de forma a não atender uma condição de custo ótimo:

- seleção incorreta das bombas;

- limitação na capacidade de transmissão-distribuição do sistema;
- inadequação na operação do tanque hidropneumático;
- incompatibilidade dos equipamentos de telemetria;
- incapacidade de controle automático ou remoto de bombas e válvulas;
- inexistência de uma política adequada de tarifa de energia elétrica;
- não verificação da demanda ou da potência;
- equívocos por parte do operador; e
- não atendimento das condições ótimas na definição das estratégias de controle.

O desenvolvimento de uma metodologia para otimizar a operação de bombas em um sistema de abastecimento de água, através da diminuição dos custos de energia elétrica, foi apresentado por Ormsbee et al. (1989). Os autores propuseram uma metodologia envolvendo duas fases básicas: o desenvolvimento de uma trajetória ótima do nível de água no reservatório e desenvolvimento de uma política operacional ótima da bomba para encontrar uma trajetória (curva) ótima global. A trajetória ótima do nível de água no reservatório foi determinada utilizando programação dinâmica, enquanto a política associada à bomba foi determinada através da utilização de enumeração explícita. O modelo foi aplicado a um setor da rede de distribuição de Washington D.C. e possibilitou uma diminuição no custo anual de energia elétrica de 6,7%, quando comparado ao custo previsto em projeto. Utilizando uma abordagem similar, porém, considerando curvas diferentes para diversas combinações de conjuntos motor-bomba, Lansey e Awumah (1994) propuseram uma análise do problema via programação dinâmica, com restrição no número de bombas em funcionamento. Diferentemente de Ormsbee et al (1989), Lansey e Awumah (1994) desenvolveram curvas para cada combinação de bombas, possibilitando uma maior versatilidade com relação à operação do sistema.

Nas últimas décadas, os responsáveis pela administração dos serviços de saneamento têm buscado novas tecnologias no sentido de aprimorar a operação diária dos sistemas de distribuição de água, diminuindo os custos operacionais. Seguindo esta premissa, Ormsbee e Reddy (1995) enfatizaram que um dos componentes de maior custo nos sistemas é a estação de bombeamento. Visando a diminuição dos custos de operação deste componente do sistema, os autores propuseram uma política operacional para uma estação de bombeamento. Utilizando o sistema *SCADA* e um *software* de apoio de controle operacional, o modelo definia regras para a operação diária dos conjuntos moto-bombas de uma determinada estação. A política ótima de bombeamento

foi definida como aquela que resultava no menor custo operacional total para um dado conjunto de condições limites e restrições hidráulicas do sistema. Os referidos autores desenvolveram uma heurística não-linear para obtenção das políticas de bombeamento de baixo custo. Como não foram apresentados resultados da aplicação da técnica, em um sistema real de distribuição de água, não é possível avaliar a validade do modelo proposto.

Considerando que os custos da energia elétrica em sistemas de distribuição de água representavam a maior parte dos custos totais relativos às despesas operacionais, Jowitt e Germanopoulos (1992) implementaram um modelo baseado em programação linear para determinação do custo ótimo (mínimo) em estações de bombeamento com base em 24 horas de funcionamento. Os autores consideraram os custos relativos à energia consumida e à demanda. Foram enumeradas as restrições operacionais, as características hidráulicas do sistema, o perfil de demanda e a política tarifária local de energia elétrica. O modelo foi aplicado a um pequeno sistema de uma cidade da Inglaterra, porém os resultados com relação à otimização não foram apresentados, pois a preocupação maior dos autores era com relação ao tempo de processamento. Também com a preocupação de reduzir o tempo de processamento do modelo de otimização, Pezeshk e Helweg (1996) propuseram um método, chamado algoritmo de busca por ajuste, para otimizar a operação de uma estação de bombeamento de água. Este algoritmo era um modelo discreto de otimização por busca, que selecionava uma das bombas para ser ligada ou desligada, usando uma combinação de coeficientes de influência e leituras de pressão na rede. Os autores ressaltaram que atingiram valores ótimos após três ou quatro iterações e orientaram para a utilização do modelo no controle de sistemas em tempo real.

Com o objetivo de otimizar a operação de estações de bombeamento em sistemas de distribuição de água, através da redução dos custos de energia elétrica, Brion e Mays (1991), Yu et al. (1994), Pezeshk et al. (1994) e Kadar e Bariudin (2001), McCormick e Powell (2003), implementaram modelos utilizando programação não linear, devido ao formato do problema, em termos do número das variáveis de decisão e de suas características não-lineares.

Com a proposta de gerar uma regra geral para a operação de estações de bombeamento em tempo real, considerando dificuldades como complexa política tarifária, variação de demanda, vazões de bombeamento discretas entre outras restrições hidráulicas, Nitivattananon et al. (1996) propuseram um modelo de otimização

progressiva. O modelo dividia o sistema em vários subsistemas no tempo e espaço, planejando intervalos dentro dos períodos operacionais. A otimização progressiva era aplicada para resolver o modelo de programação dinâmica. As vazões das bombas eram discretizadas e rearranjadas através de métodos heurísticos, com o objetivo de reduzir o período de tempo em que as bombas estavam em operação. O modelo foi aplicado ao sistema de abastecimento de água da cidade de Pittsburgh e a solução ótima apresentou uma redução de 20% nos custos de operação das estações de bombeamento, quando comparados com os custos reais.

Utilizando a qualidade da água como parâmetro, Sakarya e Mays (2000) desenvolveram uma metodologia para a determinação da operação ótima de bombas de um sistema de distribuição de água. Na metodologia proposta, a solução do problema de otimização era obtida utilizando uma interface entre o simulador hidráulico e de qualidade *EPANET* e um código de otimização não linear denominado *GRG2*. As variáveis de estado eram incorporadas como restrições na função objetivo usando método de penalidade Lagrangeano. Foram consideradas três funções objetivo no problema: minimização dos desvios entre as concentrações atuais e desejadas de um constituinte em particular, minimização do tempo total de operação das bombas e minimização dos custos totais de energia. O modelo foi aplicado em um sistema de distribuição de água hipotético. Segundo os autores, foi possível encontrar uma política ótima de funcionamento das bombas, com considerações de qualidade da água, atendendo os três objetivos.

A relação entre a operação das bombas na estação elevatória e os níveis dos reservatórios para atendimento das demandas diárias, tem sido a preocupação de vários pesquisadores nos procedimentos de otimização da operação de um sistema de água. Ko et al. (1997) utilizaram uma técnica de análise multiobjetiva para determinar a operação ótima de bombas e reservatórios em sistemas de distribuição. Os autores definiram três objetivos distintos: econômico, em termos de minimização dos custos de energia elétrica na estação de bombeamento, estabilidade na operação das bombas e confiabilidade no atendimento de demandas que variam aleatoriamente. Estes três objetivos foram avaliados, baseados no tempo total em que as bombas estavam ou não em operação, no custo total requerido de energia e no nível mínimo de água requerido nos reservatórios. Um algoritmo de programação dinâmica foi aplicado para encontrar a solução ótima para o problema multiobjetivo discretizado. A metodologia foi testada no sistema de distribuição de água da cidade de Kumi na república da Korea. Segundo os

autores, os resultados da otimização multiobjetivo indicaram um potencial melhoramento da operação do sistema quando comparada às operações históricas. Considerando, da mesma forma, a relação entre estações de bombeamento e reservatórios, Kazantzis et al. (2002) propuseram uma metodologia para a otimização da operação de uma estação elevatória considerando dois objetivos: maximização do bombeamento no período de tarifa mínima e minimização das cargas das bombas com a racionalização dos níveis dos reservatórios. Foi desenvolvida uma ferramenta de otimização utilizando algoritmos genéticos para encontrar os controles ótimos para uma faixa de demandas diárias. A nova metodologia foi testada em parte do sistema de abastecimento de água de Murray Bridge na Austrália. Segundo os autores, os resultados apresentaram uma redução entre 15 e 20% nos custos relativos a energia elétrica do sistema em questão.

Tendo como preocupação a redução de perdas de água num sistema de distribuição de água, Jowitt e Xu (1990) propuseram um algoritmo para determinação do ponto ótimo de funcionamento e localização de válvulas controladoras de vazão. Equações não lineares, relacionando carga nos nós e vazão nos trechos, foram adicionadas a equações que expressavam a dependência entre a perda por vazamento e pressão e que modelavam os efeitos da ação da válvula no processo. O modelo foi testado num sistema de distribuição de água em operação e demonstrou eficiência com relação à minimização de perdas e, por consequência, com relação à diminuição dos custos relativos à troca de tubulações devido à diminuição de pressão. Vale ressaltar, que a redução de pressão em locais específicos da rede deve ser feita utilizando válvulas redutoras de pressão e não controladores de vazão, como foi sugerido pelos autores.

Objetivando solucionar o problema de localização ótima de válvulas de controle de pressão em um sistema de abastecimento de água, assim como o número de válvulas a implantar, Reis et al. (1997) utilizaram algoritmos genéticos para obter a máxima redução de vazamentos, para dados níveis de água em reservatórios e demandas nos nós. O vazamento, considerado explicitamente na formulação, é minimizado para obter as aberturas ótimas das válvulas para uma dada localização destas, e, então, a redução de vazamento é maximizada pelo uso de algoritmo genético para localização ótima destas válvulas. Um estudo do efeito da variação da demanda na localização ótima da válvula indicou combinações distintas do local da válvula para os diferentes padrões espaciais de nós de demanda e para diferentes demandas totais. De acordo com o estudo

apresentado, uma das vantagens do uso de algoritmos genéticos na formulação é a variedade de soluções envolvendo um número pequeno de válvulas.

Em termos de otimização de sistemas de abastecimento de água, a literatura apresenta vários trabalhos nos quais é considerado o sistema como um todo, não adotando uma metodologia específica para unidades isoladas. Nesta linha de verificação, termos como confiabilidade do sistema, custos operacionais, benefícios, flexibilidade, reabilitação, entre outros, são considerados nos objetivos definidos para o processo de otimização. Trabalhos propondo metodologias de otimização para a confiabilidade de sistemas de água foram apresentados por Su et al. (1987), Ormsbee e Kessler (1990), Xu e Goulter (1999), Shinstine et al. (2002), Tolson et al. (2004) e Prasad e Park (2004), considerando a maximização de benefícios hidráulicos para o sistema foram sugeridos por Walters et al. (1999), Walski (2001) e Wu et al. (2002) e considerando a reabilitação de sistemas de distribuição de água, podem ser encontrados em trabalhos propostos por Cheung et al. (2003a), Cheung et al. (2003b) e Cheung (2004).

O sistema de distribuição de água é um elemento componente dos sistemas públicos de abastecimento composto de adutoras, elevatórias, reservatórios, válvulas e redes de distribuição e é implantado para satisfazer demandas em pontos determinados (nós), atendendo uma determinada pressão requerida. Cunha e Sousa (1999) implementaram um modelo de simulação hidráulica e otimização para minimização dos custos de uma rede de distribuição de água, envolvendo variáveis relativas a projeto e operação. Os autores utilizaram uma heurística baseada em *simulated annealing*. Este algoritmo utilizava buscas aleatórias que admitiam, tanto na teoria como na prática, a determinação do ótimo global de uma determinada função. Seguindo esta mesma linha, porém utilizando o protótipo de um sistema de apoio à decisão, denominado *WADSOP* (*Water Distribution System Optimization Program*), Taher e Labadie (1996) propuseram um modelo que otimizava o projeto e analisava o sistema, considerando a variação da demanda. O modelo propunha uma interface com o sistema de informações geográficas (SIG) para implementação da base de dados.

Embora os modelos de otimização possam auxiliar na escolha dos diâmetros dos tubos, são necessárias análises críticas e experiências para definição de uma solução efetivamente de baixo custo. Esta foi a conclusão de Walski et al. (1987) que apresentaram uma comparação entre os vários métodos de otimização para um exemplo específico, numa conferência ocorrida em Buffalo, New York (Junho/1985). Apesar das

diferenças nas soluções obtidas, todos os métodos utilizados chegaram a respostas com variações de 10% de custo.

Souza (1998) implementou modelos hidráulicos da coluna rígida e da coluna elástica, utilizando os métodos de incidência e das características, para otimizar a operação de um determinado sistema de distribuição de água. O autor utilizou a teoria do regulador quadrático linear para derivar uma expansão para abertura ótima de uma válvula redutora de pressão, com o objetivo de conduzir o sistema à condição desejada de equilíbrio na presença de distúrbios internos e externos. A eficácia da técnica foi demonstrada através da aplicação em uma rede exemplo com válvulas de controle.

Um modelo para otimizar a operação de um sistema de abastecimento de água, utilizando otimização progressiva, num método de programação dinâmica iterativa, foi implementado por Zessler e Shamir (1989). Dados, demandas para um período de 24 horas, condições iniciais e finais dos reservatórios, propriedades hidráulicas de todos os elementos do sistema e os custos de energia elétrica ao longo do dia, encontra(m)-se o(s) ponto(s) ótimo(s) de trabalho do(s) conjunto(s) moto-bomba(s). Os autores afirmaram que o processo convergia para um ótimo global (sob certas condições definidas nas funções) quando valores ótimos não mudavam de uma geração para outra.

Com o objetivo de desenvolver uma ferramenta suporte para a obtenção de políticas operacionais ótimas em sistemas de distribuição de água para o período estendido de 24 horas, Almeida e Barbosa (2002.a) desenvolveram um modelo hidráulico de otimização, estruturado na forma de um problema de programação não-linear inteira mista de grande escala, cuja solução foi obtida pela associação do algoritmo do gradiente reduzido generalizado com um algoritmo de programação inteira, o *branch and bound*. O modelo proposto foi utilizado para avaliar as influências das condições iniciais de reservação de um sistema de distribuição de água com relação ao consumo de energia elétrica de uma estação de bombeamento tipo *booster*. Os resultados indicaram economias da ordem de 36% no consumo de energia elétrica, quando se alteravam as condições iniciais dos reservatórios do sistema. Os autores sugeriram que o modelo era capaz de fornecer subsídios para as tomadas de decisão com relação à operação de sistemas de abastecimento de água.

Francato (2002) apresentou um modelo que utilizava uma metodologia baseada no método da teoria linear para a otimização da operação de um sistema de distribuição de água. Para a abordagem multiobjetivo do problema, o autor sugeriu a utilização do método dos pesos. Devido ao grande número de variáveis e da necessidade de geração

de um conjunto de soluções não dominadas, o autor optou pela utilização de programação linear para a otimização.

Um sistema de distribuição de água deve ser operado na base diária com o menor gasto de energia elétrica e de manutenção, fornecendo para cada nó da rede de distribuição a demanda de consumo à pressão desejada. Além disso, deve ser confiável no sentido de permitir o suprimento de água em condições anormais ocasionadas por avarias em componentes do sistema. Com o objetivo de atender a estes requisitos, Righetto (2002) propôs uma metodologia para encontrar a operação ótima de sistemas de distribuição de água, envolvendo parâmetros relacionados com o consumo de energia elétrica, confiabilidade operacional, satisfação quanto ao atendimento da demanda e controle das pressões nos nós. O modelo de otimização foi baseado em algoritmo genético. O autor utilizou cinco índices para avaliar a performance das regras operacionais ao longo de 24 horas de funcionamento do sistema: índice de consumo de energia elétrica, índice de nível de água de reservatório, índice de atendimento de demanda, índice de adequação da pressão média e índice de mudanças operacionais. O modelo foi aplicado a um sistema hipotético simples, com o objetivo de ilustrar o procedimento para a determinação das regras operacionais desejadas, apresentando eficácia nesta aplicação.

O principal objetivo de um sistema de abastecimento de água é satisfazer a demanda dos consumidores. Para que isto ocorra, deve haver um equilíbrio entre a adução e os diversos componentes do sistema, como estações de bombeamento, válvulas e reservatórios. Na maioria dos modelos de otimização para sistemas de abastecimento de água, a minimização dos custos é o critério básico adotado. Klempous et al. (1997) propôs um modelo baseado em três níveis:

- o primeiro era o controle direto das unidades de bombeamento, ajuste das válvulas de controle, cargas e vazões de um sistema. Nesse nível, baseado num algoritmo de controle, era determinado o número de bombas em funcionamento, assim como a posição desejada de regulação de válvulas. Os valores recomendados de carga e vazão eram fornecidos como parâmetros do algoritmo de controle, a partir do segundo nível do modelo.

- o segundo, determinava os valores dos parâmetros descritos anteriormente (carga e vazão) que garantiam a implementação da demanda aos usuários. Os valores desses parâmetros eram obtidos da minimização dos custos de energia elétrica. Como

resultado deste nível, obtém-se um gráfico ilustrando a cooperação entre estações de bombeamento e reservatórios.

- o terceiro nível determinava um gráfico ótimo para enchimento dos reservatórios. Este era baseado no histograma previsto de demanda dos consumidores. O algoritmo garantia o atendimento da demanda e a minimização dos custos de energia, considerando a variação dos preços de energia elétrica.

Walski (1995) ressaltou que uma importante razão para a não aceitação e utilização dos modelos de otimização, por parte dos operadores de sistemas, é o fato destes modelos não levarem em consideração vários aspectos como planejamento, projetos ótimos de concepção, reabilitação do sistema existente, setorização da rede, relação entre níveis de água nos reservatórios e demandas, seleção e operação de bombas e operação de válvulas reguladoras. O autor enfatizou que cada tipo de problema leva a específicas avaliações e considerações que devem ser usadas na preparação do modelo.

Depois de mais de duas décadas de pesquisas na área da otimização de sistemas de abastecimento de água, os modelos continuam mostrando-se ferramentas inadequadas aos técnicos que trabalham na operação. O mais importante objetivo, quando se trata de planejar a operação de um sistema, é maximizar os benefícios onde o conjunto de soluções ótimas é a diferença entre os custos e os benefícios. Os modelos de otimização tendem a reduzir a capacidade do sistema (reduzindo os benefícios), visando atender a condição de custo mínimo. Os pesquisadores envolvidos com modelos deveriam, primeiramente, conhecer um sistema de distribuição de água real, com todos os seus componentes e formas diferenciadas de operação, para, posteriormente, empregar diversas técnicas desenvolvidas para operação (WALSKI, 2001).

2.4.1 Métodos de otimização

Diferentes métodos de otimização têm sido aplicados em sistemas de abastecimento público de água. Os principais são: métodos de programação matemática, programação dinâmica e métodos heurísticos (ZYL, 2001).

2.4.1.1 Métodos de programação matemática

2.4.1.1.1 Programação linear

Programação linear é largamente usada como técnica de otimização operacional de sistemas de distribuição de água. Geralmente é utilizada para resolução de um tipo específico de problema, onde as relações entre as variáveis, tanto nas restrições como nas funções objetivo são lineares.

Sob certas condições, problemas não lineares podem ser linearizados e resolvidos por iteração ou procedimentos de aproximação usando programação linear. Para que seja obtido um resultado satisfatório na aplicação, é necessária a utilização de funções custo convexas, e funções benefício côncavas.

2.4.1.1.2 Programação não linear

Programação não linear não tem sido utilizada com muita frequência em otimização de sistemas de distribuição de água. Isto ocorre, particularmente, devido o fato do processo de otimização ser usualmente lento e depender de um tempo computacional longo quando comparado com outros métodos.

A programação não linear pode, entretanto, oferecer uma formulação matemática mais geral que a programação dinâmica e pode fornecer uma fundamentação para análise por outros métodos. A programação não linear inclui a programação quadrática, a programação geométrica e, num caso especial, a programação separável que pode ser usada iterativamente como um programa mestre ou como um subprograma na aplicação em grandes sistemas.

2.4.1.2 Programação dinâmica

Programação dinâmica é um procedimento utilizado para otimização de processos de decisão multi-estágio e bastante aplicado em sistemas de distribuição de água. A popularidade e o sucesso desta técnica podem ser atribuídos ao fato de que as características não lineares e estatísticas, que caracterizam a maior parte dos sistemas de água, podem ser traduzidas na formulação da programação dinâmica (YEH, 1985).

Para aplicação da programação dinâmica na otimização de sistemas de distribuição de água, o problema deve ser decomposto em um determinado número de estágios e analisado, de um estágio para o próximo, para todos os estados operacionais possíveis. Na finalização dos resultados, aquele que apresentar o valor mais econômico é selecionado como solução do problema.

2.4.1.3 Métodos heurísticos

2.4.1.3.1 Algoritmos genéticos

De acordo com Goldberg e Kuo (1987), tanto a programação matemática como os métodos gradientes têm aplicação restrita quando considerados tamanho e tipo de problemas encontrados em sistemas de abastecimento público de água. Os autores sugeriram como método de otimização, para estes casos, os algoritmos genéticos, por apresentarem as seguintes vantagens:

- têm uma orientação mais global que muitos outros métodos usados em engenharia;
- são eficientes e robustos;
- têm grande aplicação em várias áreas da engenharia;
- trabalham com o código dos parâmetros e não com os parâmetros propriamente ditos;
- realizam a busca em uma população de pontos e não em um único ponto;
- utilizam informação apenas da função objetivo, não necessitando de derivadas ou de outras informações; e,
- usam regras de transição probabilísticas e não determinísticas na identificação de soluções alternativas.

2.4.1.3.2 Lógica fuzzy

Angel et al. (1999) desenvolveram um sistema *expert* baseado em lógica *fuzzy* para um decisor da operação de um sistema de distribuição de água. O sistema usava o conhecimento derivado da experiência de operadores e dados do sistema *SCADA*. Eram coletados dados, em diferentes intervalos de tempo, da variação dos níveis de água nos reservatórios e estes eram usados no sistema *expert* para desenvolver decisões operacionais na manutenção dos níveis ótimos.

2.4.1.3.3 Otimização heurística

É possível desenvolver métodos de otimização operacional heurísticos, baseados em custos e performances de diferentes pontos de funcionamento de elementos de um sistema hidráulico. Ormsbee e Reddy (1995) desenvolveram um modelo utilizando esta idéia. Eles primeiramente implementaram combinações viáveis de bombas com base num custo padrão de operação para cada combinação. As variáveis de decisão consistiam de uma variável simples (correspondente a uma combinação de bomba) para cada estação de bombeamento, em cada intervalo de controle. Um modelo de simulação

do sistema foi acoplado ao método para verificação da acurácia do processo. Os resultados mostraram que o método poderia fornecer um conjunto de soluções factíveis em termos de custo e eficiência.

Um outro método heurístico foi proposto por Pezeshk e Helweg (1996). Eles usaram um método de otimização de busca adaptado para otimizar o funcionamento de bombas de um sistema com combinações simples e tarifa de energia elétrica fixa. Primeiramente uma boa combinação de bombas era selecionada. Esta combinação era testada num simulador hidráulico e as vazões e pressões no sistema eram comparadas com as requeridas. Se esta situação fosse inadequada para a demanda, bombas adicionais eram selecionadas e inseridas no sistema. Um coeficiente de influência era determinado para cada ponto monitorado e estação de bombeamento. Este coeficiente fornecia uma indicação do impacto na variação de pressão no ponto monitorado, quando a estação de bombeamento estava em operação. A estação de bombeamento com o maior coeficiente, era selecionada para o sistema. O processo era repetido até que a seleção da bomba satisfizesse a pressão requerida para o sistema.

2.5 O uso de aprendizado de máquina (AM) em sistemas de abastecimento de água

Os custos econômico e social associados com a ruptura de tubulações e consequentes perdas numa rede de distribuição de água, estão crescendo rapidamente e em proporções inaceitáveis. Os riscos de rupturas de redes de distribuição dependem de inúmeros fatores que são extremamente difíceis de se caracterizar. Babovic et al. (2002) propuseram o uso de algoritmos de aprendizado de máquina para determinar os riscos de rupturas de tubulações numa rede de distribuição de água. Análises dos dados de um banco contendo informações de eventos de rupturas ocorridas poderiam ser usadas para estabelecer um modelo de risco como uma função das características associadas à tubulação rompida (idade, diâmetro, material, etc.), tipo de solo em que a tubulação foi implantada, fatores climáticos, tipo de tráfego, etc. Os autores afirmam que com o modelo de risco disponível era possível planejar estratégias ótimas de reabilitação antes da ocorrência de rupturas.

Bessler (1998) analisou o desempenho do aprendizado de máquina, como processo de identificação de sistema para o desenvolvimento de um conjunto de regras de controle em sistemas de abastecimento de água. Para determinar a qualidade dos conjuntos de regras obtidos, o autor desenvolveu um programa de simulação para

avaliar o comportamento do sistema, usando as regras de controle identificadas. Para a otimização foi utilizado o algoritmo *RELAX*, o programa *See5* (QUINLAN, 1993) para o aprendizado de máquina e, para a simulação, um programa desenvolvido pelo autor denominado *SIMU.FOR*. O modelo foi aplicado a um sistema de múltiplos reservatórios de Roadford, Sudoeste da Inglaterra. Os resultados demonstraram a viabilidade da aplicação do modelo para a extração de regras em sistemas com poucas unidades operacionais.

Bessler et al. (2001) relataram que o algoritmo de aprendizado de máquina poderia ser utilizado para a previsão de falhas em sistemas de abastecimento de água, auxiliando os operadores na tomada de decisões em longo prazo, sobre as estratégias de investimento. Os autores utilizaram, como técnica para aplicação do algoritmo, o classificador *C 5.0* (QUINLAN, 1993) e, para o refinamento e otimização das regras fornecidas pelo *C 5.0*, um *software* denominado *rule refiner*, desenvolvido na Universidade de Exeter, Inglaterra. O objetivo do trabalho era refinar as regras extraídas, permitindo a combinação e a alteração das regras obtidas de diferentes origens, como, por exemplo, peritos humanos, análises de árvores de decisão, indução de regras, entre outros. A metodologia combinava o conhecimento induzido de máquina com o conhecimento do perito humano.

Utilizando AM, Bessler et al (2003) desenvolveram uma política operacional geral para sistemas de abastecimento de água. Para definição da política de operação usando esta abordagem, foram modelados e otimizados um sistema de água com múltiplos reservatórios e outro com reservatório único, considerando uma série de vazões históricas. Os resultados otimizados definiram a melhor performance possível para os sistemas e foram usados como dados de entrada no processo. O algoritmo de AM gerou um conjunto de regras de controle que forneciam a melhor política de operação para a série histórica de afluições naturais. Os resultados destas simulações foram comparados com os obtidos usando políticas de operação derivadas de regressão linear. Foi efetuada uma outra comparação entre regras de operação, utilizando, além dos resultados obtidos pela aplicação do AM e regressão linear, as regras disponíveis na *U.K. Environment Agency (South West)*. Os autores ressaltaram que os resultados obtidos através do AM estão muito próximos dos resultados otimizados obtidos através da regressão linear. Eles alertaram, porém, que idealmente seria melhor não depender completamente de regras de controle geradas por AM, mas melhorar e combinar os

conjuntos de regras que incorporam o conhecimento e a experiência de gerentes de operação de reservatórios na direção de uma operação ótima.

ALGORITMOS EVOLUCIONÁRIOS

All truths are easy to understand once they are discovered; the point is to discover them.

- Galileo Galilei

3.1 Introdução

O desenvolvimento da inteligência artificial (IA) pode ser dividido em quatro períodos: o mais antigo, chamado período sub-simbólico, ocorreu entre 1950 e 1965 e apresentava como principal característica, a representação do conhecimento de forma não simbólica, mas numérica. Como principais ferramentas representantes desta época tem-se as Redes Neurais Artificiais (RNA) e as primeiras idéias sobre algoritmos evolucionários (AE). Seguiu-se a esta fase, o chamado período simbólico que ocorreu por volta de 1962 até 1975. Este período caracterizou-se pelo desenvolvimento dos algoritmos de representações simbólicas como a Lógica de Predicados e Redes Semânticas. No período de conhecimento intensivo (de 1976 a 1988), a ênfase recaiu na grande quantidade de conhecimentos incorporados nos sistemas de aprendizagem. Atualmente, os pesquisadores de IA estudam, nas mais diversas áreas do conhecimento, a possibilidade de tornar o computador tão confiável quanto o homem (*experts*) no desenvolvimento de determinadas tarefas.

Os algoritmos evolucionários surgiram, rudimentarmente, no período sub-simbólico da IA. Estes métodos procuravam determinar uma boa solução para problemas com vastos espaços de busca. Estes espaços eram muito grandes para uma enumeração completa, principalmente devido aos recursos computacionais da época.

Os algoritmos evolucionários pertencem ao ramo da IA classificado como aprendizagem por indução: a observação, a descoberta, a pesquisa por regularidades e regras gerais feitas por exploração.

Este tipo de técnica forma uma classe de algoritmos de pesquisa probabilística e de otimização baseados no modelo de evolução orgânica, onde a natureza é a fonte de inspiração. Hoje, os principais representantes deste paradigma computacional, e que foram desenvolvidos independentemente, são conhecidos como: estratégias evolutivas, programação evolucionária e algoritmos genéticos.

Back (1996) ressaltou que os algoritmos evolucionários formam um campo de pesquisa interdisciplinar, pois envolvem disciplinas como a biologia, inteligência artificial, otimização numérica entre outras.

Para modelar a evolução orgânica, é necessária uma grande variedade de conceitos, tais como: genótipo, fenótipo, cromossomo, seleção natural, maximização, aprendizagem, etc., pois existe uma inter-relação entre algoritmos e biologia, pesquisa aleatória e algoritmos evolucionários, otimização e inteligência artificial.

De acordo com Pessis-Pasternak (1993), a ênfase inicial da IA era baseada na inteligência individual, guiada por tópicos de redes neurais e conhecimento simbólico. No cruzamento das neurociências e da informática, buscava-se compreender o pensamento e o comportamento humano a fim de reproduzi-los artificialmente. Posteriormente, a ênfase foi deslocada para propriedades de aprendizagem de populações de indivíduos onde havia o benefício da alta diversidade de seu material genético: a população é inicializada e caminha para melhores regiões do espaço de busca por meio de processos probabilístico-aleatórios (recombinação, mutação e seleção). O ambiente devolve a informação de qualidade (*fitness*) e o processo de seleção favorece os indivíduos de maior qualidade para sobrevivência e reprodução sobre os de pior qualidade. O mecanismo de recombinação permite a mistura de informação dos pais para seus descendentes e a mutação introduz inovações neste processo. As mudanças nas populações que são consideradas evolutivas são aquelas herdáveis através de material genético de uma geração para outra.

Os algoritmos evolutivos modelam, portanto, um processo de aprendizagem coletivo dentro de uma população de indivíduos, onde cada um representa não somente um ponto no espaço de soluções potencial de um dado problema, mas um depósito temporal de conhecimento sobre as leis do ambiente (MEDAWAR e MEDAWAR, 1978).

A programação genética é uma das diversas técnicas para simulação computacional da evolução. Algoritmos evolucionários imitam aspectos da evolução natural, seleção natural e reprodução diferencial. Estes algoritmos são implementados através da definição de um objetivo na forma de um critério de qualidade e usa este objetivo para medir e comparar soluções candidatas no refinamento de um conjunto de estruturas de dados. Se houver sucesso, um algoritmo evolucionário retornará um ótimo ou um próximo do ótimo individual depois de algumas iterações (BANZHAF et al., 1997).

3.2 Origens do pensamento evolutivo

Idéias antigas são lentamente abandonadas, pois são mais que categorias e formas lógicas abstratas. São hábitos, predisposições, atitudes de aversão e preferência profundamente enraizadas. Além disso, persiste a convicção - embora a história mostre que se trata de uma alucinação - de que todas as questões que a mente humana formulou podem ser respondidas em termos de duas alternativas que as próprias questões apresentam. Entretanto, na verdade, o progresso intelectual ocorre normalmente através do abandono completo das questões, juntamente com as alternativas que elas pressupõem, um abandono que resulta de sua vitalidade enfraquecida e de uma mudança do interesse mais urgente. Nós não resolvemos os problemas, passamos por cima deles. Velhas questões são resolvidas pelo desaparecimento, volatilização, enquanto os novos problemas, correspondentes às atitudes de iniciativa e preferência modificadas, tomam seu lugar. Sem dúvida, a revolução científica teve seu clímax em *A Origem das Espécies*, que é a maior dissolvente das velhas questões, a maior iniciadora de novos métodos, novas intenções, novos problemas dentro do pensamento contemporâneo.

Esta conclusão, de 1910, do filósofo John Dewey, ainda é aceita atualmente, apesar de já terem decorrido quase 150 anos da publicação do livro de Darwin, *A Origem das Espécies*.

Medawar e Medawar (1978) relataram que embora a teoria de Darwin tenha prevalecido e seja aceita como verdadeira até hoje, ele não foi o primeiro a advogar uma teoria evolutiva. As origens do pensamento evolutivo, onde a idéia de evolução é apresentada pela primeira vez, em oposição ao criacionismo, são atribuídas a Jean-Baptiste de Lamarck (1744-1829). A teoria de Lamarck, conhecida até hoje como *Lamarckismo*, propunha o conceito de “herança de características adquiridas” segundo a

qual admite-se que a informação pode ser gravada no organismo por suas necessidades ou por influência do exterior.

Barcellos (2000) ressaltou que, embora errôneas, as idéias de Lamarck foram o estopim para que, em meados do século XIX, idéias sobre evolução começassem a ser difundidas e discutidas.

Charles Robert Darwin (1809-1882) iniciou sua carreira como naturalista em 1831. Em 1844 ele escreveu, mas não publicou, um ensaio sobre seleção natural e, em 1856, começou a trabalhar numa grande obra, *Natural Selection* que, entretanto, não foi acabada. Em junho de 1858 ele recebeu um manuscrito sobre a tendência das variedades se afastarem indefinidamente a partir do tipo natural, escrito pelo jovem naturalista Alfred Russel Wallace (1823-1913), que concebeu, independente de Darwin, o mecanismo de seleção natural a partir de suas viagens para América do Sul e o arquipélago malaio.

Darwin, a conselho de amigos, publicou o resumo de seu grande livro, em 24 de novembro de 1859, sob o título de *A Origem das Espécies por Meio da Seleção Natural*. Este livro contém duas teses separadas: a que todos os organismos descendem, com modificações, de ancestrais comuns, e que o principal agente de modificação é a ação da seleção natural sobre a variação individual. Darwin foi o primeiro a mostrar a realidade histórica da evolução.

3.3 Conceitos básicos de genética

O germe fertilizado de um dos animais superiores, sujeito como é, a tão vasta série de mudanças, desde a célula germinativa até a velhice, talvez seja o objeto mais maravilhoso da natureza. É provável que dificilmente qualquer tipo de mudança que afete um dos pais não deixe marca alguma no material germinativo. Por outro lado, pela doutrina da reversão, o ovo torna-se um objeto muito mais maravilhoso, pois, além das mudanças visíveis que vai sofrendo, precisamos crer que ele contenha uma infinidade de caracteres invisíveis apropriados para ambos os sexos, para os dois lados do corpo e para a extensa linhagem de machos e fêmeas ancestrais, separada do presente por centenas ou mesmo milhares de gerações; e estes caracteres, como aqueles escritos em papel com tinta invisível, permanecem prontos para desenvolverem-se sempre que a organização for perturbada por certas condições conhecidas ou desconhecidas. (*A Variação dos Animais e Plantas sob Domesticação*, Charles Darwin, 1868).

Neste trecho do livro, percebe-se que Darwin estava tentando desenvolver uma teoria para a transmissão de caracteres. Observa-se, também, que Darwin equivocou-se ao constatar que mudanças que afetam os pais não deixam marca no “germe” mas, por outro lado, ele percebeu a existência de uma variação oculta e a distinção crucial entre genótipo e fenótipo.

Genótipo é o projeto de um organismo, o conjunto de instruções recebidas dos pais. O fenótipo é a manifestação, numa série de etapas do desenvolvimento, da interação dessas instruções com fatores físicos e químicos que permite a realização do projeto do organismo.

Dois dos mais importantes princípios da hereditariedade devem ser ressaltados: o de que o fluxo de informação do genótipo para o fenótipo é unidirecional, e o de que as unidades hereditárias transmissíveis mantêm sua identidade, de geração em geração. O estudo da evolução sempre esteve ligado ao estudo da transmissão de características de um indivíduo ao outro.

Todos os seres vivos são formados a partir de um conjunto de instruções que estão contidas no núcleo de todas as suas células. As estruturas que armazenam as informações de como será o indivíduo recebem o nome de cromossomos. O cromossomo pode ser dividido em segmentos longitudinais chamados genes, que são considerados a unidade básica do cromossomo e, juntos, carregam todas as características que o organismo pode ter.

A recombinação cromossômica (*crossover*) é o fenômeno que ocorre durante a meiose, onde os cromossomos homólogos se pareiam antes de se segregarem para gametas diferentes. O que ocorre, basicamente, é que uma parte de um cromossomo pode ser trocada por outra parte do cromossomo homólogo, fazendo com que alelos, que antes pertenciam ao mesmo cromossomo, passem a pertencer a cromossomos distintos. Este fenômeno constitui a chamada Terceira Lei da Herança (de Morgan): “Genes situados no mesmo cromossomo tendem a se manter unidos de uma geração à seguinte, só se separando pelo processo da permuta, cuja frequência reflete, até certo ponto, as relações espaciais entre aqueles genes”.

A hereditariedade é uma força conservadora que confere estabilidade a sistemas biológicos. Contudo, nenhum mecanismo composto de moléculas e sujeito ao impacto do mundo físico pode ser perfeito. Erros na cópia produzem seqüências alteradas de DNA – mutações – que são perpetuadas. Mutação é um termo vago, que é freqüentemente definido como uma mudança na seqüência de pares de base de um gene,

mas às vezes o termo é usado de maneira mais ampla, de modo a incluir mudanças no número e estrutura dos cromossomos. A mutação representa a matéria prima da evolução, ou seja, sem ela a vida humana nunca passaria de uma protobactéria.

3.4 Aplicação de algoritmos evolucionários em problemas de otimização multiobjetivo

O problema de otimização multiobjetivo, também chamado de otimização multicriterial ou multiperformance, pode ser definido como uma investigação que tem como objetivo encontrar um vetor de variáveis de decisão, que satisfaça restrições e otimize um vetor em que os elementos representam as funções objetivo. Estas funções formam uma descrição matemática dos critérios de performance que, usualmente, estão em conflito.

Com várias funções objetivo, há uma mudança no conceito de ótimo, pois, neste tipo de problema, tenta-se encontrar bons compromissos (*trade-offs*) e não uma simples solução como nos casos de otimização global.

3.4.1 Breve histórico

Em 1951 Tjalling C. Koopmans editou um livro chamado *Activity Analysis of Production and Allocation* onde o conceito de vetor “eficiente” foi usado pela primeira vez num conceituado estudo.

Os fundamentos matemáticos da otimização multiobjetivo tiveram origem no período compreendido entre 1895 e 1906. Durante este período, Georg Cantor e Felix Hausdorff sugeriram os fundamentos básicos dos espaços dimensionais infinitos. Entretanto, foi o conceito do problema de máximo vetor, introduzido por Harold W. Kuhn e Albert W. Tucker em 1951, que demonstrou o enfoque matemático da otimização multiobjetivo.

A teoria da otimização multiobjetivo não obteve muita evolução durante os anos 50. Somente nos anos 60 que os seus fundamentos foram consolidados e passaram a fazer parte integrante dos estudos de pesquisadores das ciências matemáticas. A aplicação da técnica, fora do domínio das ciências econômicas, foi apresentada pela primeira vez por Koopmans em 1951 com a teoria da produção e com o trabalho de Marglin sobre planejamento de recursos hídricos, em 1967. O primeiro trabalho aplicado à engenharia, citado pela literatura, foi um artigo de Zadeh em meados dos anos 60. Entretanto, o uso da otimização multiobjetivo tornou-se generalizada a partir

dos anos 70. Atualmente, existem mais de 30 técnicas de programação matemática para otimização multiobjetivo (COELLO, 1996).

3.4.2 Implementação dos algoritmos evolucionários a problemas multiobjetivos

Na otimização com objetivo único, o algoritmo tenta obter a melhor estratégia operacional ou o melhor projeto que, geralmente, é o máximo ou o mínimo global. Com base numa concepção muito diferente, um típico método multiobjetivo procura um conjunto de soluções que são superiores às outras soluções do espaço de busca.

Muitos problemas de engenharia têm vários objetivos (em geral conflitantes) a serem solucionados. No passado, muitos desses problemas eram freqüentemente tratados e resolvidos como problemas de objetivo único. No entanto, existem diferenças fundamentais entre os princípios de aplicação de algoritmos de otimização simples e multiobjetivos. No problema de otimização de um único objetivo, a tarefa é encontrar uma solução que otimize uma função objetivo específica. Simplesmente estendendo esta idéia para otimização multiobjetivo, pode-se considerar que a tarefa seria encontrar uma solução ótima correspondente a cada função objetivo, o que não é correto.

O problema de otimização multiobjetivo pode ser definido como o problema que busca um vetor de variáveis de decisão que satisfaça às restrições e otimize uma função cujos elementos representam as funções objetivo. Estas funções formam uma descrição matemática dos critérios de performance que estão geralmente em conflito mútuo. Assim, otimizar significa encontrar as soluções que forneçam o melhor compromisso entre os diversos objetivos para o decisor.

O problema geral de otimização multiobjetivo pode ser definido como: encontrar um vetor $x^* = [x_1^*, x_2^*, \dots, x_n^*]^T$ que satisfaça as m restrições de desigualdade, $g_i(x) \geq 0$, $i = 1, 2, \dots, m$ e as p restrições de igualdade, $h_i(x) = 0$, $i = 1, 2, \dots, p$ e otimize um vetor função, $f(x) = [f_1(x), f_2(x), \dots, f_k(x)]^T$.

Há uma mudança no conceito de ótimo quando existem várias funções objetivo, pois, nesse caso, tenta-se encontrar bons compromissos (negociações) ao invés de uma simples solução como na otimização global. O entendimento de ótimo mais comumente utilizado na literatura é o proposto por Francis Ysidro Edgeworth, em 1881. Este conceito foi posteriormente generalizado por Vilfredo Pareto, em 1896. Alguns autores o chamam de Edgeworth-Pareto-ótimo, porém é mais comumente conhecido como Pareto-ótimo (COELLO, 1996).

Diz-se que um vetor de variáveis de decisão $x^* \in F$ é Pareto-ótimo se não existe um outro $x \in F$ tal que $f_i(x) \leq f_i(x^*)$ para todo $i = 1, \dots, k$ e $f_j(x) < f_j(x^*)$ para pelo menos um j . Em outras palavras, esta definição diz que x^* é Pareto-ótimo se não existe vetor factível de variáveis de decisão $x \in F$ que decresça algum critério sem causar um simultâneo acréscimo em pelo menos um outro critério, no caso da maximização. Este conceito fornece um conjunto de soluções ao invés de uma simples solução, chamado de conjunto Pareto-ótimo. Os vetores x^* correspondentes às soluções incluídas no conjunto Pareto-ótimo são chamados não-dominados. Plotando num gráfico, as funções objetivo, cujos vetores não-dominados estão no conjunto Pareto-ótimo, obtém-se a chamada frente Pareto. A figura 3.1 ilustra este aspecto com um hipotético problema exemplo. A figura considera dois objetivos a serem minimizados. O ponto A representa a solução com o menor valor para o objetivo 1 porém com o maior valor do objetivo 2.

Figura 3.1 – Conceito de soluções Pareto ótimas

Por outro lado o ponto B apresenta o menor valor para o objetivo 2 mas o maior valor para o objetivo 1. Se os dois objetivos forem importantes para a solução do problema não é possível dizer qual o melhor. Uma solução é melhor que a outra em um objetivo e pior em outro. De fato, existem muitas soluções, como por exemplo a solução D, que também está inserida no conjunto ótimo Pareto e não é possível definir uma hierarquia entre as soluções A, B e D. Todas as soluções interceptadas pela linha que liga a solução A à solução B são conhecidas como soluções ótimas de Pareto.

Avaliando a figura é possível verificar que existem soluções não Pareto ótimas, como o caso do ponto C. Se for comparada, a solução C com a solução A, novamente

não é possível definir qual é a melhor em termos dos dois objetivos. Neste caso, pode-se pensar que esta solução(C) é ótima Pareto. Porém, como existe uma solução D no espaço de busca, que é melhor que a solução C para ambos os objetivos, esta solução não é considerada ótima Pareto. Soluções como a C são chamadas de soluções dominadas ou inferiores.

A maior parte das técnicas multiobjetivo para delimitação da região Pareto utiliza o conceito de dominância. Segundo este conceito, todas as soluções são comparadas duas a duas sendo verificada a dominância ou não de uma sobre a outra.

Segundo Deb (2001), o conceito de dominância pode ser descrito da seguinte forma: uma solução x_1 domina uma solução x_2 , se:

- a solução x_1 não é pior que a solução x_2 segundo nenhum dos m objetivos; e
- a solução x_1 é estritamente melhor do que a solução x_2 segundo pelo menos um objetivo.

Tendo como ponto de partida este conceito, pode-se listar três possíveis critérios de dominância entre duas soluções x_1 e x_2 :

$$\begin{array}{ll} x_1 \succ x_2 & x_1 \text{ domina } x_2 \\ x_1 \prec x_2 & x_2 \text{ domina } x_1 \\ x_1 \sim x_2 & x_1 \text{ e } x_2 \text{ são indiferentes} \end{array}$$

Seja uma série de soluções P , o conjunto de soluções não-dominadas P' é composto por todas as soluções que não são dominadas por nenhum elemento de P . Quando a série P é representada por toda a região factível, o conjunto não-dominado resultante, P' , é chamado conjunto Pareto ótimo (DEB, 2001).

De acordo com Fonseca e Fleming (1995) as técnicas evolucionárias para otimização multiobjetivo podem ser classificadas em:

- métodos que consideram os objetivos separadamente: estes métodos levam em conta um objetivo de cada vez no processo evolutivo, ou seja, a cada passo (ou iteração) a escolha do indivíduo para reprodução é feita considerando um objetivo diferente;
- métodos que utilizam técnicas de agregação dos objetivos: estas técnicas se baseiam no método tradicional dos pesos, em que o problema multiobjetivo é substituído por uma função simples, no entanto, os pesos dessa função são variados aleatoriamente a cada geração; e

- métodos baseados no conceito de dominância Pareto: utilizam o conceito de dominância Pareto proposto por Goldberg (1989) na ordenação das soluções.

Pode-se utilizar uma classificação simples objetivando a visualização dos diferentes processos atualmente utilizados na otimização com algoritmos evolucionários multiobjetivo:

- técnicas da primeira geração ou não elitistas, não baseadas na teoria de Pareto (*Aggregating approaches*, *Vector Evaluated Genetic Algorithm-VEGA*, *Lexicographic ordering*, *The ϵ -constrained Method*) e baseadas na teoria de Pareto (*Multiobjective Genetic Algorithm- MOGA*, *Non-dominated Sorting Genetic Algorithm-NSGA* e *Niched-Pareto Genetic Algorithm-NPGA* e *NPGA II*);e

- técnicas da segunda geração ou elitistas – estão inseridos os métodos: *Pareto Archived Evolution Strategy - PAES*, *Pareto Envelope-based Selection Algorithm- PESA* e *PESA II*, *Strength Pareto Evolutionary Algorithm-SPEA* e *SPEAII*, *Non-dominated Sorting Genetic Algorithm-NSGA II* e *Multiobjective Messy Genetic Algorithm-MOMGA* e *MOMGA II*.

Estes métodos serão detalhados no item 3.6.

3.5 Algoritmos genéticos (AGs)

Os AGs foram inspirados em alguns processos observados na evolução natural. Os mecanismos que regem a evolução natural não são completamente conhecidos, mas algumas de suas características são claras e amplamente aplicadas no processo de aplicação dos AGs.

De acordo com Davis (1991), muitos dos interessados em algoritmos genéticos são pessoas determinadas a resolver problemas reais, que não são pesquisadores em AGs, mas que se sentem desconfortáveis com as notações formais que são utilizadas nos resultados de pesquisas utilizadas em sistemas de distribuição de água. Segundo o autor, esta é uma boa oportunidade para produzir um trabalho, orientado a explicar, em linguagem não matemática, o que são algoritmos genéticos e como estes podem ser aplicados.

Os AGs são métodos de otimização por busca inspirados nos mecanismos de evolução de populações de seres vivos. Estes algoritmos seguem o princípio da seleção natural e sobrevivência dos mais aptos, declarado em 1859, pelo naturalista e fisiologista inglês Charles Darwin.

Segundo Barcellos (2000), tais algoritmos podem ser vistos como uma representação matemática-algorítmica das teorias de Darwin e da genética, chamada nova sintaxe da teoria da evolução, cujos principais postulados podem ser resumidos da seguinte forma:

- a evolução é um processo que opera sobre os cromossomos do organismo e não sobre o organismo que os carrega. Desta maneira, o que ocorrer com o organismo, durante sua vida, não irá se refletir sobre seus cromossomos. O inverso, entretanto, não é verdadeiro, pois os cromossomos do organismo são o projeto e terão reflexo direto sobre todas as características desse organismo (o indivíduo é a decodificação de seus cromossomos).

- a seleção natural é o elo entre os cromossomos e o desempenho que suas estruturas decodificam. O processo de seleção natural faz com que aqueles cromossomos representativos dos organismos mais bem adaptados ao seu meio ambiente sobrevivam e reproduzam mais do que aqueles menos adaptados.

- a reprodução é o operador através do qual a evolução se caracteriza. Mutações podem causar diferenças entre os cromossomos dos pais e os de seus filhos. Além disso, processos de recombinação (*crossover*) podem fazer com que os cromossomos dos filhos sejam bastante diferentes dos de seus pais, uma vez que eles combinam materiais cromossômicos de dois genitores.

Na década de 1970, estes postulados intrigaram John Holland que acreditava que, incorporando estes princípios em um programa de computador, pudesse resolver, por simulação, problemas complexos, justamente como ocorria na natureza. Quando Holland começou a estudar estes algoritmos, eles ainda não tinham nome. Como os estudos começaram a demonstrar o potencial do método, foi necessário batizá-lo. Pelo fato de originar de estudos da genética, Holland chamou-os de algoritmos genéticos (AGs). Os AGs são considerados, como as redes neurais e *simulated annealing*, algoritmos baseados no potencial metafórico do mundo natural (DAVIS, 1991).

Holland iniciou o seu trabalho manipulando cadeias de 0's e 1's, para representar os cromossomos, e cada organismo constituindo uma tentativa de solução do problema.

O algoritmo desenvolvido por Holland resolvia problemas complexos de uma maneira muito simples. Como na natureza, o algoritmo ignorava o tipo do problema a ser resolvido. Uma simples função de ajuste fazia o papel da medida de adaptação dos organismos (cromossomos) ao meio ambiente. Assim, os cromossomos com uma

adaptação melhor, medida por esta função, tinham maior oportunidade de reprodução do que aqueles com uma má adequação, imitando o processo evolucionário da natureza.

Para incorporar as características da teoria da evolução natural em seus algoritmos genéticos, Holland iniciou considerando os mecanismos que ligam um AG ao problema que está sendo resolvido. Há dois mecanismos: um de codificação das soluções para o problema em cromossomos e uma função de avaliação que retorna uma medida de valor médio dos cromossomos no contexto do problema. As técnicas para codificação das soluções podem variar de problema para problema e de AG para AG. Em vários trabalhos, a codificação é feita utilizando uma cadeia de bits, mas há outros tipos de técnicas de codificação. De acordo com Davis (1991), provavelmente nenhuma técnica tem um desempenho melhor para todo tipo de problema. Uma questão importante é que diversos fatores devem ser considerados na seleção de uma técnica de representação no contexto de um problema real.

O algoritmo genético proposto por Holland era composto por três partes principais:

- módulo de avaliação – contém uma função de avaliação que mede o valor de todos os cromossomos no contexto do problema.
- módulo de população – contém uma população de cromossomos e técnicas para criar e manipular esta população.
- módulo de reprodução – contém técnicas para criação de novos cromossomos durante a reprodução.

De acordo com a literatura, o primeiro trabalho publicado sobre AG foi em 1975, nos Estados Unidos, com o título: *Adaptation in Natural and Artificial Systems* de autoria de John Holland. Este livro foi o marco para o desenvolvimento e aplicação dos AGs na resolução de vários problemas na área de engenharia.

3.5.1 Terminologia

Como já foi visto no item anterior, um AG é a metáfora dos fenômenos que ocorrem na natureza, o que explica o fato dos algoritmos genéticos possuírem muitos termos originados da biologia. Apresenta-se, a seguir, uma lista com os principais termos encontrados sobre o assunto na literatura (GOLDBERG, 1989; MICHALEWICZ e ATTIA, 1994; GALVÃO et al., 1999):

- **cromossomo e genoma**: na biologia, genoma é o conjunto completo de genes de um organismo e pode ter vários cromossomos. Nos AGs, os dois representam a

estrutura de dados que codifica uma solução para um problema. Representam um simples ponto no espaço de busca; ou seja, uma possível solução para o problema;

- **gene**: na biologia, é a unidade de hereditariedade que é transmitida pelo cromossomo e que controla as características do organismo. Nos AGs, é uma característica codificada no cromossomo, ou seja, um elemento do vetor que representa o cromossomo. O gene representa uma variável de decisão do problema;

- **indivíduo** – um simples membro da população. Nos AGs, um indivíduo é representado pelo cromossomo e o respectivo valor de aptidão;

- **genótipo** – na biologia, representa a composição genética contida no genoma. Nos AGs, representa a informação contida no cromossomo ou genoma;

- **fenótipo** – nos AGs, representa o objeto, estrutura ou organismo construído a partir das informações do genótipo. É o cromossomo decodificado;

- **alelo** – na biologia, representa uma das formas alternativas de um gene. Nos AGs, representa os valores que as variáveis de decisão do problema podem assumir.

- **epistasia** – interação entre genes do cromossomo, isto é, quando um valor de gene influencia o valor do outro gene; e

- **elitismo** – é comum no processo de AGs que o melhor membro de uma população falhe na produção de um descendente (*off spring*) na próxima geração. DeJong (1975) propôs a estratégia do elitismo, que se baseia na transferência do melhor cromossomo de uma geração para outra, sem alterações, já que o melhor cromossomo pode ser perdido devido à aplicação dos operadores de reprodução. A estratégia do elitismo pode aumentar a velocidade de dominação de uma população devido a um super indivíduo, mas no balanço, esta estratégia melhora a performance do AG. A solução, no AG com elitismo, é encontrada de forma mais rápida que no AG não elitista.

3.5.2 Visualização geral dos AGs

Os AGs são técnicas de busca estocásticas que imitam, matematicamente, os mecanismos de evolução natural, compreendendo os processos de seleção e genética das populações, em analogia com o processo biológico de sobrevivência dos indivíduos.

Na resolução de problemas através da utilização de AGs, dois aspectos devem ser considerados: a maneira de codificação da solução na forma de *string* e a função de aptidão a adotar.

A técnica de codificar soluções pode variar de acordo com as características do problema e tipo de AG. No trabalho original de Holland, a codificação foi feita usando cadeias de *bits*. Segundo Goldberg (1989), este tipo de representação é a que resulta melhor desempenho, pois a base binária apresenta maior número de esquemas por *bit* de informação.

A função de aptidão é a maneira de se fazer a ligação entre o código representativo de uma solução e quão boa essa solução é para o problema em análise. Ela toma como entrada o cromossomo, que é uma tentativa de solução para o problema, e devolve um número real, informando o desempenho deste cromossomo no problema; este número representa o seu grau de adaptabilidade, que informa quão próxima essa solução está da ótima. Dessa forma, a função de aptidão desempenha, no AG, papel análogo ao que o meio ambiente realiza sobre os seres vivos: avalia o grau de adaptação do organismo (cromossomo) ao ambiente e este valor é usado no processo de seleção para reprodução.

Nos AGs, os cromossomos reúnem unidades chamadas genes, que representam as variáveis de decisão do problema em estudo. Cada gene controla a herança de uma ou mais características do indivíduo a que pertence. O código genético de um gene é denominado genótipo e o conjunto decodificado de variáveis de decisão é o fenótipo do cromossomo. Cada cromossomo representa uma solução potencial para o problema de otimização e as populações criadas sucessivamente são chamadas gerações. Tais gerações são criadas mediante a seleção de indivíduos de uma população para serem pais, cujos filhos darão origem à próxima geração.

Para criar uma nova população, cromossomos chamados descendentes são formados pela união de dois cromossomos pais da geração atual, através do operador de recombinação (*crossover*) e em seguida modificados, através do operador mutação (*mutation*). Assim, novas populações são formadas sucessivamente, mantendo as soluções (cromossomos) mais aptas e descartando as menos qualificadas, para que seja mantido o tamanho da população inicial, até a convergência para uma solução “ótima”.

Dentre as vantagens dos AGs sobre os métodos de busca direta convencionais, podem ser ressaltadas:

- trabalham com o código dos parâmetros e não com os parâmetros propriamente ditos;
- a busca é realizada a partir de uma população de pontos e não de um único ponto;

- utilizam informação apenas da função objetivo, não necessitando de derivadas ou de outras informações;
- são de fácil implementação, não sendo necessários conhecimentos matemáticos complexos;
- são capazes de otimizar um grande número de variáveis, trabalhando com funções objetivo com superfícies de resposta complexas, reduzindo a incidência de mínimos ou máximos locais;
- fornecem uma gama de parâmetros ótimos e não uma simples solução;
- usam regras de transição probabilísticas e não determinísticas na identificação de soluções alternativas; e
- mesmo quando o número de variáveis de decisão é muito grande, o algoritmo é capaz de “varrer” o espaço de soluções de maneira bastante eficaz.

Apesar dessas vantagens, os AGs são lentos e não garantem que a solução ótima seja atingida. A figura 3.2 mostra um fluxograma com os passos básicos de um algoritmo genético *standard*.

3.5.3 O algoritmo

O algoritmo proposto por Holland, é conhecido na literatura como *Simple Genetic Algorithm* ou *Standard Genetic Algorithm* ou, simplesmente, *SGA*. Pode-se descrever o algoritmo, sucintamente, em seis passos (DAVIS, 1991):

- a- inicie uma população de tamanho N com soluções geradas aleatoriamente;
- b- avalie a função de aptidão para cada solução desta população;
- c- crie novos cromossomos através de recombinação de soluções (pais) selecionadas desta população. Aplique recombinação e mutação sobre tais soluções;
- d- elimine membros da antiga população, de modo a ter espaço para inserir as soluções filhas, mantendo a população com o mesmo número N de soluções;
- e- aplique a função de aptidão nestes cromossomos e insira-os na população; e
- f- se a solução ideal for encontrada ou, um número pré-estabelecido de gerações for desenvolvido, retorne a solução com a melhor aptidão. Caso contrário, volte ao passo c.

Figura 3.2 – Metodologia do algoritmo genético *standard*

Esta simulação do processo evolutivo deverá produzir, à medida que as gerações forem se sucedendo, soluções cada vez melhores, isto é, com melhor valor da função de aptidão, de maneira que, no final, obtém-se uma solução (cromossomo) com alto grau de adaptação ao problema proposto.

De maneira geral, a implementação de AGs para resolver um problema específico requer:

- escolha da sistemática de representação das variáveis de decisão que representam soluções potenciais para o problema;
- geração de uma população inicial de soluções alternativas, geralmente aleatórias;

- definição de um módulo de avaliação do desempenho das soluções através da função objetivo ou de aptidão;
- especificação dos operadores genéticos de recombinação e mutação; e
- escolha dos parâmetros de AGs tais como, probabilidade de recombinação e de mutação , tamanho da população e número de gerações ou outro critério de convergência.

A figura 3.3 ilustra o procedimento de um AG típico.

3.5.4 Representação das soluções

Cada solução alternativa (cromossomo) para o problema em estudo é representada através de uma cadeia (*string*) de características necessárias à sua completa definição e que constituem as variáveis de decisão do problema. Tais características ou códigos podem ser valores binários ou reais.

Seja $S(t)$ a população de cromossomos na geração t .

$t \leftarrow 0$

inicializar $S(t)$

avaliar $S(t)$

enquanto o critério de parada não for satisfeito, faça

$t \leftarrow t + 1$

selecionar $S(t)$ a partir de $S(t-1)$

aplicar recombinação(*crossover*) sobre $S(t)$

aplicar mutação sobre $S(t)$

avaliar $S(t)$

fim

Figura 3.3 – Algoritmo genético típico – Galvão et al. (1999)

A representação binária é a tradicional, sendo fácil de utilizar e manipular, além de simples de analisar teoricamente. No entanto, se um problema tem parâmetros contínuos e deseja-se trabalhar com uma boa precisão, haverá a necessidade do armazenamento de cromossomos longos na memória, os quais precisam ser decodificados para avaliação da função de aptidão. Este aspecto pode fazer com que o algoritmo tenha uma convergência vagarosa.

De acordo com Goldberg (1989), o melhor método para trabalhar com AG é codificar os parâmetros de uma possível solução do problema em uma cadeia de bits de 0's e 1's. Segundo o autor, isto permite uma manipulação fácil e eficiente dos operadores genéticos sobre os cromossomos.

Por outro lado, a representação real gera cromossomos menores e é compreendida mais naturalmente pelo ser humano quando comparada à cadeia de bits.

A melhor estratégia de representação, binária ou real, tem sido tema de várias pesquisas e muitas delas têm demonstrado a maior utilização da representação real (MICHALEWICZ e ATTIA, 1994; HAUPT e HAUPT, 1998).

3.5.5 Função de aptidão

A função de aptidão tem o objetivo de trabalhar como interface entre o AG e o problema em estudo. É usada para avaliar a solução para o uso posterior pelos operadores de seleção e reprodução. Esta função admite como entrada uma cadeia de bits e, como saída, um valor real.

Deve-se ressaltar que é necessário dispor de algum procedimento para avaliar a distância que um cromossomo está da possível solução ideal. Este processo pode ou não ser exato. Nos casos de grande complexidade, alguma heurística poderá ser utilizada.

A função de aptidão se diz normalizada quando seus valores estão sempre dentro do intervalo $[0,1]$. Isto significa que um cromossomo com valor de aptidão 1 (o maior valor) é a melhor solução possível. Portanto, este cromossomo é a resposta do problema. No caso oposto, um valor 0 (zero) representaria a pior solução.

Não é necessário, contudo, que a função seja normalizada, pois, basicamente, é uma função de comparação de cromossomos e, o que interessa, na verdade, são os valores relativos entre um cromossomo e os outros. Para facilitar o algoritmo de seleção, é necessário que a função de aptidão nunca apresente valores negativos.

3.5.6 Operadores genéticos

3.5.6.1 Seleção

O processo de seleção ocorre após a avaliação da função de aptidão das soluções. Ela desempenha o papel da seleção natural na evolução, selecionando, para sobreviver e reproduzir, os organismos melhor adaptados ao meio, no caso, as soluções com melhor valor na função de aptidão.

A maneira pela qual os cromossomos são selecionados para reprodução, pode variar, dependendo do método utilizado. Entretanto, é certo que os cromossomos melhor adaptados terão, necessariamente, uma probabilidade maior de sobrevivência e reprodução que os de baixa função de aptidão.

Os tipos de seleção mais usuais são roda da roleta, torneio, ordenamento, uniforme e amostragem determinística. .

O modo pelo qual Holland propôs o processo de seleção é conhecido como método de seleção por roda da roleta. De acordo com tal algoritmo, são calculadas as probabilidades de seleção para cada solução de acordo com a razão de sua aptidão e da aptidão da população acumulada. Em seguida, um número aleatório é gerado no intervalo de zero até o valor da aptidão acumulada e, então, a solução selecionada sofre a ação dos operadores recombinação e mutação que serão descritos na sequência. Vale observar que a aptidão pode assumir valores negativos e o algoritmo roda da roleta não funciona com aptidões negativas. Além disso, pode gerar também problemas como convergência prematura.

Sucintamente o processo pode ser descrito da seguinte forma:

Se f_i é o valor da função de adequação do i -ésimo cromossomo e f_m a soma dos valores da função de aptidão sobre todos os cromossomos, então, de acordo com esse método, cada cromossomo receberá um setor da roleta imaginária, de ângulo f_i/f_m . Ou seja, um setor com ângulo proporcional ao grau de adaptabilidade do cromossomo.

O próximo passo é selecionar um cromossomo com probabilidade igual à do setor por ele representado, isto é, a probabilidade de selecionar um dado cromossomo, para reprodução, é proporcional à sua função de aptidão. A seleção é realizada conforme descrição a seguir (Barcellos, 2000):

- gera-se um número aleatório entre 0 e 1;
- soma-se o valor do setor alocado de cada cromossomo. Continua-se somando enquanto esta soma não ultrapassar o número aleatório gerado;
- o cromossomo cujo setor fizer esta soma ultrapassar o número aleatório gerado será o cromossomo escolhido; observe-se que isto sempre ocorrerá, pois a soma de todos os setores é 1.

Este processo é semelhante a uma roda de roleta, na qual os setores vão se sucedendo, sob um pino marcador, até que a roda pare com o pino sobre um setor qualquer. É claro que, quanto maior a largura do setor, maior a probabilidade deste setor se encontrar sob o pino marcador.

A cada dois cromossomos escolhidos, são gerados dois novos cromossomos (clones). O processo de seleção termina quando o número de novos cromossomos atinge o mesmo valor da população anterior. Quando isto ocorre, a antiga população é substituída por esta nova geração.

É possível dispensar o algoritmo roda da roleta e utilizar seleção por torneio. Neste caso, são escolhidos, aleatoriamente e com probabilidades iguais, n cromossomos da população e o cromossomo com maior valor de aptidão é selecionado para a população intermediária. O processo repete-se até preencher a população intermediária. Galvão et. al. (1999) ressaltaram que um valor usual para n é três.

Quando o algoritmo troca toda a antiga população pela nova, ele é chamado AG Geracional (*Generational GA*). Deve-se ressaltar que existem outras variantes, com relação à maneira de substituir a população antiga, como, por exemplo, trocando apenas uma fração da antiga população pela nova, caso em que recebe o nome de *Steady-State GA*.

Com a mesma probabilidade de seleção para todos os cromossomos da população, a seleção uniforme escolhe aleatoriamente os cromossomos a serem recombinados e mutados. A seleção uniforme não leva em consideração a aptidão do cromossomo.

Na seleção por amostragem determinística, as probabilidades de seleção são calculadas como na roda da roleta. Em seguida, cada probabilidade é multiplicada pelo número de cromossomos da população e, então, cada cromossomo é selecionado de acordo com a parte inteira do valor calculado e a população ordenada de acordo com a parte decimal deste mesmo valor. O restante da população é selecionado a partir do topo da lista ordenada. A seguir, selecionam-se uniformemente os cromossomos que sofrerão a ação dos operadores genéticos (SOARES, 2003).

3.5.6.2 Recombinação (*Crossover*)

Em biologia, aprende-se que a evolução é um processo baseado na seleção natural e mutação aleatória. Mas esta visão de evolução é incompleta. Há outros processos que alteram os cromossomos durante a reprodução e, alguns biólogos evolucionistas, acreditam que eles são tão importantes quanto a mutação. Um deles é a chamada recombinação. Na natureza, a recombinação ocorre quando dois pais trocam parte de seus cromossomos correspondentes.

A recombinação, segundo Davis (1991), é um componente extremamente importante do algoritmo genético. Vários pesquisadores que trabalham com AGs acreditam que se o operador recombinação for extraído deste algoritmo o resultado não mais será de um AG. Segundo eles, isto não acontece com o operador mutação. De fato, muitos pesquisadores acreditam que o uso de um operador recombinação distingue AG de todos os outros algoritmos de otimização.

A recombinação ou cruzamento é o mecanismo de criação de indivíduos (novas soluções) filhos através da recombinação de indivíduos pais (soluções atuais).

Geralmente este processo é realizado através de simples permuta ou da combinação de características correspondentes entre soluções pais. O primeiro esquema mostra-se mais apropriado para o caso em que as soluções são representadas através de código binário, enquanto o segundo se aplica mais ao caso de representação real das soluções.

A recombinação é aplicada com uma dada probabilidade a cada par de cromossomos selecionados. Na prática, esta probabilidade, denominada de taxa de recombinação (Pcross), varia entre 60% e 100%. Não ocorrendo a recombinação, os filhos serão iguais aos pais (isto permite que algumas soluções sejam preservadas). Isto pode ser implementado, gerando números pseudo-aleatórios no intervalo [0,1]. Assim, a recombinação só é aplicada se o número gerado for menor que a taxa de recombinação.

De forma simples, através deste processo dois cromossomos (A e B) trocam genes da seguinte forma: um ponto do cromossomo (*locus*) é escolhido aleatoriamente. Ambos cromossomos são quebrados neste mesmo ponto. A primeira parte do cromossomo A é concatenada à segunda parte do cromossomo B, formando um novo cromossomo. O mesmo ocorre com as outras partes: a primeira parte do cromossomo B é ligada à segunda parte do cromossomo A, formando um segundo cromossomo. Este processo pode ser resumido da seguinte forma:

- gera-se um número aleatório entre 0 e 1;
- se este número for menor ou igual à probabilidade de recombinação (Pcross), então a recombinação irá ocorrer, caso contrário não haverá recombinação;
- escolhe-se, aleatoriamente, o ponto em que haverá a quebra;
- quebram-se os dois cromossomos neste ponto; e
- formam-se dois novos cromossomos combinando-se a primeira parte do cromossomo A com a segunda parte do cromossomo B, e a primeira parte do cromossomo B com a segunda parte do cromossomo A.

Pode-se também quebrar os cromossomos (possíveis soluções) em dois ou mais pontos. Quando o cromossomo é quebrado em um ou dois pontos a recombinação recebe o nome de simples ou tradicional. A figura 3.4 apresenta um esquema do processo simples.

A recombinação pode ser realizada de diversas maneiras. Wardlaw e Sharif (1999) apresentaram três diferentes formas, chamadas operadores convencionais de *um ponto*, *dois pontos* e *uniforme*.

A recombinação ocorre entre dois cromossomos selecionados com uma probabilidade pré-fixada, normalmente na faixa entre 0,6 e 1,0 (os cromossomos selecionados têm esta probabilidade de serem utilizados na recombinação). A *recombinação de um ponto* prevê a criação de cromossomos a partir da troca de genes de determinada posição escolhida aleatoriamente até o comprimento total do *string* (cadeia de códigos representativos de determinada solução). Na *recombinação de dois pontos*, o material genético entre duas posições escolhidas aleatoriamente ao longo do comprimento do cromossomo é trocado. A recombinação uniforme opera nos genes individuais dos cromossomos selecionados ao invés de atuar nos blocos de material genético e cada gene é considerado um por um (não blocos) tanto na recombinação quanto na troca.

Michalewicz (1992) descreveu o método de recombinação uniforme como um operador que modifica as variáveis de decisão individuais de um vetor solução. O autor indicou a utilização desse tipo de operador para a representação binária.

Um importante aspecto da recombinação aplicada a um problema multivariado, em código binário, é que ela deve ocorrer apenas nos limites dos genes porque cada gene consiste de alelos, ou bits, e a recombinação pode dividir os genes e perder alguns alelos. Isto não é um problema para representações reais. No código real o gene consiste de um único alelo e ele próprio é o valor do parâmetro.

O operador recombinação pode combinar características, através da simples troca de valores, mais adequada à representação binária, ou realizar operação aritmética sobre os valores das variáveis de decisão, mais adequada à representação real. Apresentam-se, a seguir, duas das possíveis formas deste tipo de operador, para aplicação em problemas com representação real.

Pais:	(recombinação após o oitavo bit)	Filhos:
1 1 1 1 1 1 1 1 1 1 1	-----	1 1 1 1 1 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0	-----	0 0 0 0 0 0 0 1 1 1
Pais:	(recombinação após o segundo bit)	Filhos:
1 0 1 1 0 1 0 0 0 1	-----	1 0 0 1 1 0 0 0 1 1
1 1 0 1 1 0 0 0 1 1	-----	1 1 1 1 0 1 0 0 0 1

Figura 3.4 – Esquemas de uma recombinação simples

Recombinação Linear: Uma das primeiras implementações desse tipo de operador foi apresentada por Wright (1991). De acordo com esta implementação, um operador de recombinação linear criaria três soluções, $0.5(x_i^{(1,t)} + x_i^{(2,t)})$, $(1.5x_i^{(1,t)} - 0.5x_i^{(2,t)})$ e $(-0.5x_i^{(1,t)} + 1.5x_i^{(2,t)})$ a partir de duas soluções pais $x_i^{(1,t)}$ e $x_i^{(2,t)}$ da geração t , sendo as duas melhores soluções escolhidas como filhas. x_i é a variável de decisão, i é o índice da variável de decisão, t é o índice da geração e 1 e 2 são índices dos vetores pais 1 e 2, respectivamente.

Recombinação BLX (Blend Crossover): Eshelman e Schaffer (1993) sugeriram uma recombinação mistura (*Blend Crossover*) ou BLX-a.. Para duas soluções pais $x_i^{(1,t)}$ e $x_i^{(2,t)}$ na geração t assumindo que $x_i^{(1,t)} < x_i^{(2,t)}$, o operador BLX-a toma aleatoriamente uma solução na faixa $\left[x_i^{(1,t)} - a(x_i^{(2,t)} - x_i^{(1,t)}), x_i^{(2,t)} + a(x_i^{(2,t)} - x_i^{(1,t)}) \right]$. Se u_i é um número

aleatório entre 0 e 1, o valor da variável de decisão correspondente na solução descendente é avaliada pela seguinte expressão:

$$x_i^{(1,t+1)} = (1 - \gamma_i) x_i^{(1,t)} + \gamma_i x_i^{(2,t)} \quad (3.1)$$

onde $\gamma_i = (1 + 2\alpha)u_i - \alpha$. Se $\alpha = 0$, esta recombinação cria uma solução aleatória na faixa $(x_i^{(1,t)}, x_i^{(2,t)})$. Deb (2001) sugeriu o valor 0,5 para α .

É possível encontrar na literatura (MICHALEWICZ e ATTIA, 1994; GEN e CHENG, 1997; DEB, 2001) outras diversas formas para aplicação do operador recombinação a problemas com representação binária ou real.

3.5.6.3 Mutação

A mutação é um processo que possibilita a introdução de características genéticas novas às soluções existentes, através da eventual substituição de genes dos indivíduos. Geralmente é realizada bit a bit (representação binária) ou variável por variável (representação real) possibilitando que cada bit ou gene seja alterado com uma pequena taxa de mutação (Pm) entre 0,001 e 0,1 (REIS e AKUTSU, 2002).

A mutação é um importante processo que permite que um novo material genético seja introduzido na população durante o processo iterativo. A tarefa do operador mutação é “perturbar” cada solução da população pai para criar uma nova população de soluções descendentes.

Mutação Limite: Michalewicz (1992) descreveu um operador mutação denominado limite (eq. 3.2). O autor ressaltou que este operador substitui o valor de uma variável de decisão, selecionada aleatoriamente, por outra cujo valor é obtido aleatoriamente $\left[X_i^{(L)} \text{ ou } X_i^{(U)} \right]$, onde $X_i^{(L)}$ e $X_i^{(U)}$ são os limites inferiores e superiores das variáveis de decisão, respectivamente. Assim, tem-se:

$$x_i^{(t+1)} = \begin{cases} X_i^{(L)} & \text{se } r < 0,5 \\ X_i^{(U)} & \text{se } r > 0,5 \end{cases} \quad (3.2)$$

onde x_i^{t+1} é a variável de decisão obtida e r é um número aleatório pertencente ao intervalo $[0,1]$.

Mutação Aleatória: Outra mutação que pode ser usada é a aleatória. Deb (2001) ressaltou que este é o esquema de mutação mais simples para criar uma solução aleatória pertencente ao espaço de busca. A eq. (3.3) mostra a simplicidade do processo.

$$x_i^{(t+1)} = r_i (X_i^{(U)} - X_i^{(L)}) \quad (3.3)$$

onde r_i é um número aleatório no intervalo $[0,1]$.

Mutação Não-Uniforme: Depende da geração atual (pai) e do número máximo de gerações permitido para criar uma nova solução.

Neste caso, a probabilidade de criar uma solução próxima da solução pai é maior que criar uma população próxima dela mesma. De acordo com esse operador de mutação, soluções descendentes são criadas de acordo com a equação 3.4.

$$y_i^{(t+1)} = X_i^{(1,t+1)} + \tau (X_i^{(U)} - X_i^{(L)}) \left(1 - r_i \left(1 - \frac{t}{t_{\max}} \right)^b \right) \quad (3.4)$$

onde τ admite um valor booleano, -1 ou 1, cada um com a probabilidade de 0,5, $t_{\max}$ é o número máximo de gerações permitido, t representa a geração atual e b é um parâmetro definido pelo usuário. Michalewicz (1992) sugeriu $b=6$.

O operador de mutação pode ser aplicado antes ou depois do operador de recombinação e, diferentemente desse, irá operar sobre cada solução.

O operador mutação percorre todos os bits da solução e, para cada bit, gera um evento com probabilidade P_m . Se este evento ocorrer o valor do bit é trocado.

A mutação melhora a diversidade das soluções na população. Por outro lado, pode destruir a informação contida no cromossomo. Em função disso, a probabilidade de mutação deve ter um valor pequeno, porém suficiente para assegurar a diversidade. Como a probabilidade de mutação é pequena, poucos cromossomos são afetados. Um exemplo de mutação é mostrado na figura 3.5, onde se observa que apenas o valor do terceiro bit do segundo cromossomo foi alterado, considerando a probabilidade de mutação 0,1.

Número	Cromossomo Antigo	Números Aleatórios	Cromossomo Novo
1	0 1 0 1	0,12;0,70;0,45;0,92	0 1 0 1
2	1 1 0 1	0,92;0,13; 0,06 ;0,23	0 1 1 1

Figura 3.5 – Exemplo de mutação

Assim como no caso do operador recombinação, vários autores demonstraram outras diversas formas de utilização do operador mutação (FOGEL et al. 1992; HANSEN e OSTERMEIER, 1996; DEB e GOYAL, 1996).

3.5.7 Critérios de parada

Após a aplicação dos operadores de recombinação e mutação, são implementadas as avaliações pertinentes ao processo, até que o critério de parada seja satisfeito. A literatura descreve vários critérios de parada para os AGs, dos quais podem-se destacar: a realização de um número de gerações (avaliações) pré-determinado; iterar até que não ocorra melhoramento significativo no cromossomo de maior aptidão por um determinado número de gerações ou até que a variabilidade dos valores de aptidão relativos às soluções da população possa ser considerado desprezível.

3.5.8 Problemas de convergência

A convergência prematura é um problema que os usuários de AGs devem evitar. Ocorre quando surgem soluções de valor de aptidão elevado (mas não ótimo), e as soluções realmente ótimas ainda não estão presentes na população. Tais cromossomos geram um número excessivo de filhos que dominam a população, uma vez que a mesma é finita. Estes cromossomos espalham seus genes por toda a população, enquanto outros genes desaparecem (*genetic drift*). Como consequência, o algoritmo converge para um máximo ou mínimo local.

Este problema pode ser minimizado limitando-se o número de filhos por cromossomos. Esta limitação pode ser realizada através dos diferentes critérios de seleção. Uma outra forma é manter a diversidade dos cromossomos na população, através do aumento da taxa de mutação (mais genes são criados) ou evitando a inserção de filhos duplicados na população.

A solução procurada será o melhor indivíduo, ou seja, aquele que tem a maior aptidão da população final, ou de uma população dentro do processo.

4.6 Otimização multiobjetivo utilizando algoritmos genéticos

Em problemas de otimização, é possível que a função objetivo corresponda a um determinado parâmetro que se deseja minimizar ou maximizar. Nestes casos, os algoritmos genéticos costumam apresentar bons resultados utilizando uma determinada função de avaliação (função objetivo), considerando uma otimização de objetivo único, onde se deseja minimizar ou maximizar o objetivo. Porém, em outras ocasiões, pode-se ter um problema de otimização, com dois ou mais objetivos a serem minimizados ou maximizados simultaneamente. Para estes casos, é difícil combinar em uma única

função objetivo todos os parâmetros a otimizar. Para problemas com estas características deve-se utilizar uma técnica de otimização multiobjetivo.

Para estes casos, vários estudos apresentados pela literatura sugerem técnicas tradicionais de otimização e de busca de soluções, porém, nos últimos anos, vários trabalhos têm sugerido a utilização de algoritmos genéticos.

Em problemas envolvendo um único objetivo busca-se a melhor solução (que apresenta o menor ou maior valor de sua função objetivo). Na otimização multiobjetivo, a noção de ótimo não é tão óbvia, já que se deve compatibilizar as diferenças existentes entre os objetivos. Nestes casos, é útil o conceito de ótimo de Pareto que sugere a obtenção não de uma única solução, mas uma família de soluções que são chamadas de não dominadas.

Goldberg (1989) relatou que a noção de busca genética em problemas multicriteriais foi sugerida pela primeira vez nos anos 60. Rosenberg, trabalhando com caracterização química de células, sugeriu o uso de propriedades múltiplas em sua simulação de características genéticas e químicas de uma população de organismos unicelulares. Naquela implementação, havia apenas uma única propriedade, e não várias que caracterizassem uma função multiobjetiva, mas foi o ponto de partida para que pesquisadores começassem a se interessar pelo assunto.

Algoritmos genéticos requerem informações escalares para sua implementação, o que significa que, pretendendo aplicá-los em problemas multicriteriais é fundamental que os vetores objetivos sejam escalares. Este procedimento nem sempre é possível pela dificuldade de obtenção de um critério global baseado na formulação do problema. Na ausência de informações, os objetivos tendem a assumir o mesmo grau de importância para o problema. Em caso contrário, pode-se combiná-los de acordo com a informação disponível, provavelmente assumindo mais importância para determinados objetivos. Otimizar uma combinação de objetivos tem a vantagem de produzir uma solução de compromisso único, não requerendo iteração com o decisor. O problema é que se a solução ótima não for aceita, tanto porque a função usada excluiu aspectos do problema que eram desconhecidos antes da otimização, ou porque foram escolhidos coeficientes inadequados da função de combinação, deverão ser processadas implementações complementares até que uma solução factível seja encontrada.

Listam-se a seguir, de forma sucinta, as principais técnicas desenvolvidas para aplicação de algoritmos genéticos em problemas de engenharia. Informações mais

detalhadas sobre cada técnica podem ser encontradas em Goldberg (1989), Coello (1996), Deb (2001).

3.6.1 Uso de funções de agregação (*aggregating approaches*)

Fonseca e Fleming (1993) relataram que várias técnicas têm sido desenvolvidas para combinar as diversas funções objetivo em diferentes linhas. Uma das metodologias utilizadas para este fim, envolve o uso de funções de agregação. A literatura apresenta várias opções de métodos neste sentido. As principais são descritas sucintamente na sequência.

- (a) **Soma dos pesos:** neste método, o vetor ideal não necessita ser conhecido, mas um conjunto de pesos e fatores de escala têm de ser fornecidos pelo usuário. Este método normalmente produzirá somente um ótimo (máx-mín) e não uma frente Pareto.
- (b) **Redução para um objetivo único:** a idéia central desta técnica é baseada no método *e – constrained*. O vetor ideal não é requerido, mas em alguns casos é muito difícil transformar os objetivos em restrições. Este método normalmente produz somente um ótimo (mín-máx) e não um ótimo Pareto ou uma frente Pareto.
- (c) **Funções de penalidade:** baseado no método *e – constrained* . Assim como no método anterior, não é necessário o vetor ideal e também é produzido somente um ótimo (mín-máx). Neste método a definição dos valores das penalidades não é uma tarefa fácil.

3.6.2 Uso de técnicas não elitistas e não baseadas na teoria de Pareto

Devido às dificuldades de implementação da técnica de agregação para alguns tipos de problema, muitos pesquisadores têm preferido o desenvolvimento de técnicas alternativas baseadas em ranqueamento. A seguir são listadas as mais populares técnicas não baseadas na teoria Pareto, apresentadas pela literatura.

3.6.2.1 *Vector evaluated genetic algorithm - VEGA*

Esta técnica foi usada como uma extensão do algoritmo genético simples (*SGA*), porém, diferiu deste apenas na forma em que o operador seleção foi implementado. Este operador foi modificado considerando que em cada geração, um número de sub-populações é gerada pela performance da seleção proporcional de acordo com cada

função objetivo. Neste método, o vetor ideal não é requerido e a frente Pareto pode ser gerada. Entretanto, como esta metodologia seleciona indivíduos baseados em suas características de dimensão (função) simples, em problemas com objetivos altamente conflitantes, os bons compromissos (*trade-offs*) serão eliminados. O método utiliza subpopulações que otimizam cada objetivo separadamente. Em certos tipos de problema, serão encontrados tanto o ótimo (mín-máx) como o Pareto ótimo.

3.6.2.2 Lexicographic ordering

Neste método, os objetivos são ranqueados em função do grau de importância. A solução ótima é obtida minimizando as funções objetivo, iniciando pela mais importante e prosseguindo de acordo com a importância dos objetivos. É possível selecionar aleatoriamente um único objetivo para otimizar as funções.

3.6.2.3 Método *e* – constrained

Este método é baseado na otimização de uma função objetivo (preferencial ou primária), considerando os outros objetivos como restrições com níveis permissíveis chamados e_i . Desta forma, é minimizada uma função objetivo simples (objetivo único) com o objetivo mais relevante, sujeita a restrições adicionais das outras funções objetivo. Os níveis e_i são, então, alterados para gerarem o conjunto Pareto ótimo.

Informações mais detalhadas a respeito de cada método descrito nos itens 3.6.1 e 3.6.2 podem ser encontradas em Coello, 1996.

3.6.3 Uso de técnicas não elitistas baseadas na teoria de Pareto

3.6.3.1 Multiple objective genetic algorithm – MOGA

Este método é uma variação dos descritos anteriormente. Fonseca e Fleming (1993) introduziram um algoritmo genético multiobjetivo que usava a classificação não-dominada de uma população de algoritmo genético. Os pesquisadores foram os primeiros a sugerir um algoritmo genético multiobjetivo que explicitamente servia para enfatizar soluções não dominadas e, simultaneamente, manter a diversidade dessas soluções. A técnica consistia de um esquema em que o ranqueamento de um certo indivíduo correspondia ao número de indivíduos na população que domina este

indivíduo. Esta técnica difere do algoritmo genético padrão pelo fato da cadeia (*fitness*) ser alocada para cada solução na população.

3.6.3.2 *Nondominated Sorting Genetic Algorithm – NSGA*

O *NSGA* foi proposto por Srinivas e Deb (1994) com o objetivo de se trabalhar o conceito de dominância e a manutenção da diversidade da população simultaneamente, seguindo a idéia inicial de Goldberg (1989). Este método é similar ao algoritmo *MOGA*, entretanto, o *NSGA* difere da aproximação proposta por Fonseca e Fleming (1993) na maneira como classifica os vetores solução não dominados e na forma com que o método de compartilhamento é considerado (espaço dos parâmetros) e na garantia de diversidade da população. Os operadores genéticos são aplicados como nos AGs convencionais. O fluxograma da figura 3.6 apresenta o procedimento geral do *NSGA*. Cheung (2004) apresentou o detalhamento das etapas do procedimento iterativo do *NSGA*.

3.6.3.3 *Niched-Pareto Genetic Algorithm – NPGA*

Horn et al. (1994) propuseram um esquema de seleção por torneio baseado na dominância Pareto. No processo, dois indivíduos são escolhidos aleatoriamente e comparados com um sub-conjunto da população (usualmente, 10% da população). Quando ambos competidores são dominados ou não dominados, o resultado do torneio é decidido através da divisão da cadeia no domínio dos objetivos.

3.6.4 Uso de técnicas elitistas baseadas na teoria de Pareto

3.6.4.1 *Pareto Archived Evolution Strategy – PAES*

Introduzido por Knowles e Corne (2000) este método usa uma estratégia de evolução em conjunto com um arquivo externo que registra todos os vetores não dominados, previamente encontrados. Esta técnica usa um *grid* adaptativo para manter a diversidade.

3.6.4.2 *Pareto Envelope-based Selection Algorithm – PESA*

Proposto por Corne et al. (2000), esta técnica usa uma pequena população interna e uma grande população externa (ou secundária). O *PESA* utiliza a mesma

divisão de *hypergrid* do espaço da função objetivo adotada pelo *PAES* para manter a diversidade. Entretanto, este mecanismo de seleção é baseado numa medida de agrupa-

Figura 3.6 – Procedimento geral do *NSGA* (Cheung, 2004)

mento usada pela *hypergrid* mencionada anteriormente. Esta mesma medida de agrupamento é usada para decidir quais soluções serão introduzidas na população externa, isto é, o arquivo de vetores não dominados encontrados no processo evolucionário.

3.6.4.3 *Strength Pareto Evolutionary Algorithm – SPEA*

O *SPEA* foi proposto por Zitzler e Thiele (1998) com o objetivo de introduzir o elitismo nos algoritmos evolucionários multiobjetivo. As melhores soluções (soluções não dominadas) de cada iteração, denominadas elite, são armazenadas em um conjunto externo (temporário). Em cada iteração do algoritmo, as soluções externas são comparadas com as soluções pertencentes ao conjunto corrente, de forma que as melhores soluções sejam preservadas.

Existem várias similaridades entre o *SPEA* e outros algoritmos evolucionários multiobjetivo. Dentre as principais, podem ser citadas a garantia de vetores solução representantes da frente não dominada entre as soluções consideradas e a utilização do conceito de dominância Pareto para atribuir valores de aptidão baseados nesse conceito. Entretanto, o *SPEA* difere dos demais métodos, em relação a alguns aspectos: atribuição do valor aptidão de um vetor solução é determinado entre os conjuntos externo e corrente, e não entre vetores pertencentes à mesma população, todos os vetores solução no conjunto externo participam da seleção e a utilização de um novo método de manutenção da diversidade da população. A figura 3.7 apresenta um fluxograma do procedimento geral do *SPEA*. Do fluxograma da figura 3.7, pode-se dizer que o *SPEA* cria um conjunto inicial aleatório P_0 de dimensão POP e um conjunto externo vazio P_e com uma dimensão máxima PEP. Na iteração t , as soluções não dominadas da população P_t são copiadas para a população externa P_e . Com isso, acontece a união entre as soluções não dominadas (população elite) da iteração corrente e da iteração prévia, e, em seguida, identificam-se e excluem-se as soluções não dominadas desse tipo unificado. O que permanece nesse conjunto externo são as melhores soluções elite, encontradas durante o processo iterativo. Entretanto, à medida que o processo evolui, em alguns problemas, o número de soluções contidas nesse conjunto é considerado elevado.

Como forma de restringir o aumento exagerado de soluções não dominadas no conjunto externo, uma dimensão (PEP) é fixada. Assim, quando a dimensão da

população externa é menor que PEP, toda a população elite é encontrada no conjunto. Entretanto, quando uma dimensão excede PEP, nem toda a população elite pode estar na população. Em função disso, Zitzler e Thiele (1998), propuseram uma técnica para reduzir as soluções da população externa denominada algoritmo de agrupamento (*clustering algorithm*). O detalhamento das etapas do procedimento iterativo do *SPEA* pode ser visto em Cheung (2004).

3.6.4.4 Non-dominated Sorting Genetic Algorithm – NSGA II

Deb et al. (2002) propuseram uma nova versão do *Nondominated Sorting Genetic Algorithm (NSGA)*, chamado *NSGA II*, que é mais eficiente (computacionalmente falando), usa o elitismo e um operador de comparação de vários indivíduos que fornece diversidade sem especificação de parâmetros adicionais.

De acordo com Cheung (2004), em se tratando de complexidade computacional, o *NSGA II* contém um algoritmo de classificação não dominada com menor complexidade computacional (MN^2) frente ao *MOGA*, *NSGA* e *SPEA*, que apresentam uma complexidade aproximada na ordem de MN^3 , sendo M o número de objetivos e N a dimensão da população. Já o elitismo no *NSGA II*, é garantido através da preservação das melhores soluções (frentes não dominadas) na população posterior. Em geral, o procedimento desenvolve-se da seguinte forma, uma população derivada (filha) da população corrente (pai) é criada utilizando a seleção por torneio de multidão seguida da recombinação e mutação. Após este procedimento, esses dois conjuntos de mesma dimensão (pai e filha) são unidos em uma única população e aplica-se a classificação não dominada em frentes de dominância. Inicia-se, então, o processo de preenchimento de uma nova população, colocando as melhores frentes não dominadas nesse novo conjunto, garantindo dessa forma o elitismo.

Ressalta-se que nem todas as frentes podem ser incluídas na nova população. Assim, os autores propuseram um método, que combina as frentes não incluídas no conjunto, para compor os últimos espaços da população. Este método é denominado distância de multidão. Com a utilização dessa técnica para alocação de soluções na população, é garantida a diversidade da população e eliminação da subjetividade de introdução do parâmetro de compartilhamento utilizado nos métodos não elitistas, pois diferentes soluções podem ter chances de pertencer à nova população.

Figura 3.7 – Procedimento geral para o SPEA (Cheung, 2004)

A figura 3.8 apresenta o procedimento geral do *NSGA II*. Avaliando o fluxograma da figura 3.8, percebe-se que o *NSGA II* inicia seu processo através de um conjunto inicial aleatório P_0 de dimensão POP. Através desse conjunto inicial obtém-se uma população derivada Q_0 de dimensão POP denominada população filha, utilizando os operadores de seleção, recombinação e mutação. Esses dois conjuntos são unidos e formam um conjunto R_0 de dimensão 2POP. A partir desse passo, inicia-se o processo de classificação não dominada e formação da nova população P_t . Essa nova população

é preenchida com soluções de diversas frentes não dominadas, que garantem a diversidade da população. Cheung (2004) apresentou o detalhamento das etapas de iteração para o *NSGA II*.

3.6.4.5 Multiobjective Messy Genetic Algorithm – MOMGA

Esta técnica foi proposta por Veldhuizen e Lamont (2000). É um procedimento desenvolvido com o objetivo de estender o *Messy Genetic Algorithm* para resolução de problemas de otimização multiobjetivo. O *MOMGA* consiste de três fases. Na primeira fase, o *MOMGA* produz todos os blocos de construção de um certo tamanho especificado através de um processo determinístico chamado de inicialização enumerativa parcial. Na segunda fase, o método executa seleção por torneio na população e reduz o tamanho da população, caso seja necessário. Na última fase, o *Messy GA* trabalha construindo a população final, através do uso do operador de recombinação para corte e escolha.

3.7 O Uso de algoritmos genéticos na análise de sistemas de abastecimento público de água

Para avaliação e otimização de sistemas de abastecimento de água, vários algoritmos e técnicas já foram propostos. Os algoritmos genéticos estão, atualmente, entre os mais utilizados para este fim.

A pesquisa para identificação do projeto ótimo de um sistema de distribuição de água pode ser uma tarefa bastante onerosa. Um grande número de procedimentos pode ser usado para resolução deste problema, incluindo técnicas baseadas em programação linear, programação não linear e algoritmos de enumeração completa. Entretanto, freqüentes simplificações e hipóteses são exigidas, dada à sua natureza complexa. Murphy e Simpson (1992) definiram a solução para o problema de otimização de redes de distribuição de custo mínimo em termos da combinação de diâmetros de tubos que, para um dado conjunto de demandas nos nós da rede, satisfaça restrições de pressões mínimas e máximas nos nós. Segundo os autores, a otimização poderia ser utilizada para auxiliar os projetistas na busca de combinações com menores custos. Eles afirmaram que, para o caso de grandes sistemas, é praticamente impossível identificar o ótimo global em função da grande quantidade de alternativas.

Figura 3.8 – Procedimento geral do *NSGA II* (Cheung, 2004)

Nestes casos, deve-se considerar uma solução ótima, assim como soluções sub-ótimas e, usando um julgamento de especialista (*expert*), selecionar a melhor solução.

Vários trabalhos da literatura empregam AGs em modelos de redes de distribuição de água. Walters e Lohbeck (1993) implementaram um modelo para o traçado ótimo de uma rede de distribuição de água, utilizando AGs. No trabalho, os autores compararam os resultados obtidos via AGs, com dois tipos de representação

(binária e real inteira) das variáveis de decisão, com os obtidos por programação dinâmica. Os autores concluíram que, para redes de distribuição de água de pequena dimensão, os AGs atingiram os mesmos resultados que a programação dinâmica, com menor esforço computacional. Já para redes de grandes dimensões, os AGs apresentaram resultados melhores. A representação real inteira apresentou melhores resultados à medida que se consideraram redes com maior grau de conectividade (maior número de trechos interligados).

Com o objetivo de comparar os AGs com outras técnicas utilizadas na otimização de sistemas de adução de água, Simpson et al. (1994) apresentaram os resultados da otimização, aplicando AGs, técnicas de enumeração completa e otimização não linear. Os autores sugeriram que a técnica de enumeração completa deveria ser usada em sistemas com poucos tubos. Segundo os pesquisadores, a otimização não linear é uma técnica bastante efetiva quando aplicada à expansão de pequenos sistemas de abastecimento de água. No caso dos AGs, múltiplas soluções próximas da ótima são geradas, sendo que uma delas poderia ser escolhida como ótima, com base em outra medida não quantificável. Os referidos autores estabeleceram faixas de valores para parâmetros como tamanho da população (entre 30 e 200), probabilidades de recombinação (entre 0,7 e 1,0) e de mutação (0,01 a 0,05). Estabeleceram, ainda, que a probabilidade de mutação deveria ser maior que o inverso do número de cromossomos da população e menor que o inverso do comprimento dos cromossomos. Os AGs apresentaram resultados favoráveis para aplicação a sistemas sem restrições.

Uma melhoria na formulação para algoritmos genéticos aplicados na otimização de redes de distribuição de água foi a proposta de Dandy et al. (1996). Os autores sugeriram a utilização de diferentes parâmetros para os algoritmos genéticos, daqueles usualmente utilizados. Eles propuseram a codificação denominada *Gray* (CARUANA e SCHAFFER, 1988) ao invés de binária e mutação adjacente ou *creep* (COOMBS e DAVIS, 1987), além do emprego de escalonamento exponencial da função de aptidão. O modelo foi aplicado a uma parte do sistema de água da cidade Nova York e os resultados comparados com AG tradicional ou simples e com diferentes técnicas de otimização: programação linear, não linear e dinâmica e enumeração completa. Segundo os autores, os resultados mostraram a superioridade da aplicação do novo modelo para este tipo de problema.

Utilizando como método de busca os AGs e comparando-os com métodos de tentativa e erro, Savic e Walters (1997b) implementaram uma metodologia de otimização para calibração de redes de distribuição de água. Os autores citaram várias aplicações dos AGs, além de enumerar suas vantagens quando utilizados para o propósito de calibração. Um dos principais pontos relatados pelos autores foi a grande quantidade de parâmetros hidráulicos da rede que puderam ser calibrados utilizando tal modelo, ou seja, calibração não só de coeficientes de rugosidade e demandas, mas de cotas topográficas, diâmetros e outros parâmetros da rede.

Preocupados em otimizar projetos de expansão de sistemas de distribuição de água, Savic e Walters (1997a) utilizaram os AGs na forma de um modelo computacional denominado *GANET*. Foi usada codificação *Gray* que, segundo eles, melhorava o desempenho do AG quando comparado com a representação binária. O simulador hidráulico utilizado foi o *EPANET 2*. Utilizou-se a mutação aleatória com taxa de 0,03 e a recombinação uniforme com taxa aleatória variando entre 0,6 e 1,0. Para testar a eficiência do *GANET* na identificação de bons projetos de sistemas de distribuição de água, os autores aplicaram o modelo a três sistemas exemplos apresentados pela literatura utilizando outros métodos. Os resultados do modelo proposto apresentaram uma eficiência superior àquela apresentada pela literatura.

O dimensionamento de redes de distribuição de água consiste basicamente da determinação dos diâmetros da tubulação. Vários pesquisadores têm buscado, nas últimas décadas, uma metodologia capaz de dimensionar tais sistemas considerando diversos aspectos e incertezas, característicos desse tipo de projeto. Formiga e Chaudhry (2002) propuseram o emprego de um método multiobjetivo baseado em algoritmos evolucionários para geração de uma série de soluções não-dominadas. As funções objetivo propostas foram: o custo de implantação do sistema, as perdas por vazamentos na rede e a capacidade de atendimento das vazões. Os autores utilizaram, para simulação hidráulica, um modelo híbrido baseado nos métodos da teoria linear e Newton-Raphson. Os parâmetros dos AGs foram: seleção por torneio de 2 indivíduos, taxa de mutação 0,1 e recombinação 0,7. O modelo foi aplicado a uma rede hipotética composta de 9 trechos e 8 nós e abastecida por dois reservatórios. Segundo os autores, embora tenha mostrado uma boa capacidade para encontrar a região não-dominada, a metodologia utilizada perdeu, ao longo de suas gerações, algumas boas soluções, sugerindo a utilização de técnicas elitistas de forma a conservá-las ao longo do processo.

Com a preocupação de desenvolver projetos otimizados para sistemas de distribuição de água, Formiga et al. (2003) propuseram a aplicação de técnicas de decisão multiobjetivo, usando algoritmos evolucionários, para gerar uma série de soluções não-dominadas. As três funções objetivo utilizadas, incluíam custos de investimento, entropia e taxa de atendimento de demandas do sistema. Os autores utilizaram o método *NSGA II (Nondominated Sorting Genetic Algorithm)*, com recombinação tipo BLX-a, com probabilidade 1,0 (100%) e mutação aleatória com probabilidade de 0,05. O modelo foi aplicado a uma rede hipotética e apresentou resultados melhores, em termos de otimização das funções objetivo, que de outros métodos apresentados na literatura para a mesma rede de distribuição. Os autores ressaltaram que, para obter resultados satisfatórios usando métodos evolucionários multiobjetivos em sistemas de distribuição de água, é muito importante a escolha apropriada do tamanho das populações iniciais e o número de gerações, de forma a atingir uma frente Pareto estável.

Procurando avaliar perdas através da determinação dos parâmetros do modelo hidráulico pressão x vazamento, Caliman et al. (2001) desenvolveram um modelo utilizando AGs. A determinação dos parâmetros foi realizada através de modelo inverso de minimização das diferenças, entre valores de vazão e pressão observados em campo e previstos, através de modelo hidráulico de simulação com AG padrão. Os autores utilizaram recombinação aleatória com probabilidade de 1,0 e mutação uniforme com probabilidade de 0,05. O modelo foi aplicado a um sub setor da rede de água da cidade de São Carlos, SP, que contava com um índice elevado de perdas. Observou-se que, embora o modelo realizasse ajuste de parâmetros, com base na minimização dos desvios entre valores simulados de vazão e pressão, os valores de vazões obtidos eram muito próximos, o mesmo não ocorrendo para as pressões. Os autores explicaram que o resultado era função da estimativa das rugosidades, utilizadas em conformidade com o material empregado, sem considerar singularidades na rede.

Silva et al. (2002) desenvolveram um modelo de calibração de rede de distribuição de água considerando perdas por vazamento, utilizando AGs. Os autores propuseram a calibração em termos das rugosidades absolutas e trechos dotados de singularidades (registros). Foram utilizados dois tipos de AGs, geracionais elitistas e *steady state*, aplicados a um sub-setor da rede de água da cidade de São Carlos, SP. Constatou-se a vantagem do uso dos AGs *steady state* com relação ao tempo

computacional. Apesar de fornecer boas respostas, estas foram menos robustas que as identificadas por AGs geracionais.

Surry et al. (2001) propuseram uma nova técnica para manipulação de restrições em algoritmos evolucionários e demonstram sua eficácia para sistemas reais. A técnica, chamada de *COMOGA* (*Constrained Optimisation by Multi-objective Genetic Algorithms*), empregava duas idéias fundamentais de otimização multiobjetivo e as combinava com um algoritmo genético adaptativo. A idéia era produzir um modelo com o mesmo campo de aplicabilidade dos métodos com função de penalidades, porém com poucos parâmetros de controle e grande eficiência. Os autores, primeiramente, compararam a heurística previamente utilizada na prática, pela companhia de gás britânica (*British Gas*), com uma abordagem genética usando uma função de penalidade cuidadosamente ajustada. O algoritmo genético encontrou um sistema factível com um custo 4% menor que uma solução previamente conhecida. Os autores repetiram a aplicação do método *COMOGA* tratando cada restrição, explícita ou implícita, como um critério separado na formulação do problema multiobjetivo. Segundo os autores, os mesmos 4% de melhoramento sobre a heurística foi encontrado com *COMOGA*, com número similar de avaliações e consistência. No estudo, foi usada a representação inteira para as variáveis de decisão, mutação aleatória com probabilidade de 0,025 combinada com mutação *creep* à taxa de 0,05 e probabilidade de recombinação 1,0.

Com o objetivo principal de aplicar técnicas de otimização na reabilitação de redes de distribuição de água, Wu e Simpson (1996) propuseram um novo tipo de algoritmo genético denominado *Messy Genetic Algorithm* (*mGA*) para aplicação em estudos que apresentavam difíceis resoluções pelo AG padrão. No trabalho proposto, para um dado traçado de rede e demandas previstas de água nos nós, o custo total de todos os componentes do sistema (principalmente tubulações) foi minimizado pela busca da combinação ótima dos diâmetros dos tubos e ações de reabilitação. O processo de busca era sujeito a dois tipos de restrições. O primeiro envolvia um componente de limitação de diâmetro dos tubos (mínimo e máximo), o segundo incluía requisitos hidráulicos como pressões mínimas nos nós de demanda. O principal operador usado no AG padrão, a recombinação, não poderia ser usado no *mGA*. Usaram-se dois operadores *messy*, corte (*cut*) e encaixe/junção (*splice*) no lugar da tradicional recombinação. O corte divide o cromossomo em dois, enquanto a junção une dois cromossomos para formar um individual. Implementou-se um pacote computacional denominado *mGANET*

para otimização de projetos e/ou reabilitação de redes de distribuição de água integrando o *mGA* ao *EPANET*.

Objetivando encontrar soluções ótimas para reabilitação, substituição e expansão de redes de distribuição, Halhal et al. (1997) propuseram um AG diferente, denominado *structured messy genetic algorithms (SMGAs)*. Este tipo de algoritmo genético difere do padrão, por utilizar comprimentos variáveis das soluções em cada geração, codificação com números inteiros e seleção com escalonamento, que consiste na comparação da “distância” de um indivíduo com outros dois indivíduos através da função de aptidão. Utilizando-se de uma otimização multiobjetivo, os autores compararam o novo *SMGA* com o AG geracional, concluindo que o *SMGA* produzia melhores resultados, principalmente na seleção de soluções envolvendo pequenos conjuntos de componentes em grandes sistemas de distribuição de água.

Utilizando esta mesma ferramenta, *SMGA*, Walters et al. (1999) propuseram expandir o processo anteriormente proposto para ser utilizado na otimização de reabilitação da rede de distribuição, também para otimização de estações de bombeamento e tanques. O modelo para otimização de projetos de expansão incluía a adição de novos tubos, recondicionamento e substituição de tubulação existente, dimensionamento e localização de novos tanques e previsão de novas bombas no processo de decisão. O problema foi resolvido como multiobjetivo, ou seja, minimização dos custos e maximização dos benefícios hidráulicos. O modelo foi aplicado a um sistema hipotético denominado *Anytown* e duas soluções completamente viáveis foram encontradas, apresentando valores entre 4% e 5% mais baratas que aquelas publicadas por Halhal et al. (1997).

Reconhecendo a natureza multiobjetivo dos processos de decisão para reabilitação de sistemas de abastecimento de água, Cheung et al. (2003) apresentaram um estudo comparativo de dois métodos evolucionários multiobjetivo aplicados à resolução do problema de reabilitação de um sistema de distribuição de água: *multiobjective genetic algorithm (MOGA)* e *strenght Pareto evolutionary algorithm (SPEA)*. Os métodos foram aplicados a um sistema hipotético e os resultados demonstraram que o *SPEA* apresentava uma performance superior ao *MOGA*, em termos de qualidade de frentes Pareto produzidas e tempo de processamento necessário.

Políticas de operação de múltiplos reservatórios são usualmente definidas pelas regras que especificam os volumes de armazenamento desejados de um reservatório individual ou a liberação desejada, baseadas no período do ano e no volume total

armazenado em todos os reservatórios. Oliveira e Loucks (1997) propuseram o uso de algoritmos genéticos *standard* para identificar estas políticas de operação de reservatórios múltiplos. Os AGs foram usados com vetores reais, contendo informação necessária para definir, tanto a taxa de liberação como a de armazenamento do reservatório individual, como função do total armazenado por ano. Para gerar sucessivos conjuntos, de possíveis políticas operacionais, foram utilizados o elitismo, a recombinação aritmética com probabilidade de 1,0 e a mutação uniforme com probabilidade de 0,05.

A implementação dos AGs para um sistema de quatro reservatórios, operados dentro de um horizonte de planejamento de 12 períodos, foi apresentada por Wardlaw e Sharif (1999). Os autores utilizaram representação real, seleção por torneio, recombinação uniforme com probabilidade de 0,7 e mutação uniforme modificada com probabilidade de um gene por cromossomo. Os autores ressaltaram que é muito importante a maneira como os genes são agrupados no cromossomo. É preferível agrupar as liberações de vazões por estágio, para que o cromossomo contenha 12 grupos de 4 genes, representando a liberação de cada reservatório em um determinado estágio, para que o material genético mais diretamente relacionado seja mantido próximo.

Visando a redução de perdas por vazamento em redes de distribuição de água através da localização e abertura ótimas de válvulas controladoras, Reis et al. (1997) utilizaram AGs como método de busca. O objetivo era determinar o controle ótimo da válvula, em termos da minimização das perdas, para uma dada localização no sistema. Os autores ressaltaram que uma das vantagens da aplicação do AG para este tipo de problema é que são produzidas várias opções economicamente equivalentes, envolvendo pequeno número de válvulas.

Bombas de rotação variada têm sido aplicadas com mais frequência em sistemas de abastecimento de água, não só para reduzir os custos de energia elétrica, como também para reduzir as perdas no sistema. Lingireddy e Wood (1998) ressaltaram os benefícios hidráulicos e econômicos resultantes do uso desse tipo de equipamento, para a melhoria da operação do sistema. Os autores utilizaram métodos de otimização baseados em AGs e apresentaram um processo de cálculo direto das velocidades da bomba para condições de operação pré-estabelecidas, das velocidades ótimas da bomba para uma operação mais eficiente em termos de gasto de energia elétrica além de uma avaliação do controle de transientes para estas variações. O objetivo principal era

minimizar os custos de bombeamento mantendo uma superfície piezométrica sobre o sistema num valor mínimo, visando a minimização das perdas.

Procurando regras operacionais ótimas para um sistema de distribuição de água, Righetto (2002) utilizou AG como ferramenta de busca. O autor testou o modelo em um caso contendo tomada de decisões quanto ao acionamento de bombas e controle de válvulas reguladoras de vazão. A estratégia do processo de busca foi atribuir prioridade ao atendimento das demandas nodais. Durante as dez primeiras gerações de soluções factíveis, foi dada prioridade ao índice de armazenamento de água nos reservatórios e, em seguida, igual prioridade para todos os índices. Os resultados obtidos mostraram-se favoráveis para o exemplo hipotético estudado.

Os modelos de otimização são vitais para a avaliação e gerenciamento de sistemas de infra-estruturas hidráulicas. Wu et al. (2001) desenvolveram um sistema de suporte à decisão, para uso no gerenciamento de sistema de adução de água. O *software* desenvolvido integrava um sistema de informações geográficas (SIG), utilizado para o gerenciamento e análise de dados espaciais, considerando a teoria da otimização, com uma interface com o *software* gráfico *AutoCAD*. A interface gráfica interativa, oferecia informações estruturadas e gerenciamento dos dados espaciais, otimização dos sistemas e apresentação de resultados. O modelo de otimização utilizava uma variação eficiente dos AGs para calibrar o sistema, usando dados de campo, e apresentar uma configuração otimizada. Os autores utilizaram o *software* *H₂ONET* para fornecer uma relação entre o SIG e os AGs.

Os AGs foram aplicados por Ritzel e Eheart (1994) para um problema multiobjetivo de minimização da poluição de águas subterrâneas. O problema era encontrar o conjunto de soluções ótimas entre a confiabilidade hidráulica e o custo de operação, imaginando-se um conjunto de contaminantes hidráulicos. As variáveis de decisão eram, a quantidade de poços profundos, a quantidade de bombas e os locais para perfuração. Os autores utilizaram duas variações para os AGs multiobjetivo: um AG *vector-evaluated (VEGA)* e um AG Pareto. Os autores utilizaram uma população de 300 indivíduos com 300 gerações, uma probabilidade de mutação de 50% por *string* e de recombinação de 0,8. Alertaram que, aumentando o tamanho da população e o número de gerações, soluções com valores significativamente melhores eram obtidos.

APRENDIZADO DE MÁQUINA (AM)

*Our minds possess by nature an insatiable desire
to know the truth.*

- Marcus Tullius Cicero

4.1 Generalidades

Após a identificação das diversas soluções que compõem a frente Pareto, há a necessidade de escolher um número menor de alternativas com o objetivo de estabelecer as regras operacionais mais promissoras, em termos dos objetivos pré-definidos para o sistema em estudo. Quando, na definição do problema, há múltiplos critérios que governam a decisão operacional, diversas soluções de compatibilização dos objetivos são possíveis na composição da frente Pareto. Deve-se, portanto, adotar ou definir uma metodologia para escolher a solução mais eficiente dentre aquelas alternativas pertencentes ao conjunto ótimo Pareto. Nessa fase, a opinião de especialistas é de fundamental importância para que as soluções não dominadas possam ser classificadas. Alguns trabalhos da literatura vêm sugerindo os métodos de análise multicriterial como ferramentas para estabelecer ordem de prioridade das alternativas. Recentemente, Bessler et al. (2003) sugeriram que o algoritmo de aprendizado de máquina possa ser utilizado com o propósito de seleção de alternativas pré-selecionadas via otimização multiobjetivo. Uma outra opção foi apresentada por Das (1999) que propôs a utilização de uma técnica denominada ordem de preferência para a escolha de soluções pertencentes a uma determinada frente Pareto em problemas com mais de dois objetivos.

Neste trabalho, pretende-se analisar o desempenho de um algoritmo de aprendizado de máquina (AM) e implementá-lo para a extração de regras operacionais do sistema de macro adução de água de Goiânia. Além disso, pretende-se implementar a técnica de ordem de preferência para verificação e comparação dos resultados obtidos pela aplicação do AM.

A capacidade de aprender é essencial para um comportamento inteligente. A habilidade dos seres humanos de fazer generalizações a partir de alguns fatos, ou descobrir padrões em coleções de observações, é realizada por meio do aprendizado. Aprendizado, geralmente, é uma forma de inferência lógica que permite que conclusões sejam obtidas a partir de premissas ou exemplos particulares. Aprendizado de máquina é uma área de pesquisa em inteligência artificial cujo objetivo é o desenvolvimento de métodos computacionais que possam adquirir novos conhecimentos, novas habilidades e novos meios de organizar o conhecimento já existente (MITCHELL, 1998). Várias abordagens de aprendizado podem ser utilizadas para que um sistema computacional possa aprender e melhorar seu desempenho por meio de observações. Uma das estratégias mais utilizadas para aprender novos conceitos é o aprendizado por indução. Indução permite que conclusões genéricas sejam obtidas a partir de conjuntos de fatos ou observações (exemplos) particulares. A indução é caracterizada como o raciocínio que parte do específico para o geral, do particular para o universal, da parte para o todo. O conhecimento generalizado extrapola aquele contido nos fatos e ele pode ou não preservar a verdade. Na realidade, a conclusão obtida pode exceder a informação contida nas premissas, sendo a inferência indutiva um dos principais meios de criar novos conhecimentos e prever eventos futuros (MARTINS, 2003). Foi através da indução que Arquimedes descobriu a primeira lei da hidrostática e o princípio da alavanca, Kepler descobriu as leis do movimento planetário e Darwin descobriu as leis da seleção natural das espécies.

A indução da primeira lei da hidrostática ocorreu quando o rei de Siracusa, uma colônia da antiga Grécia, solicitou que Arquimedes descobrisse se um ourives havia adulterado a coroa real com prata, sendo que deveria conter apenas ouro. Arquimedes, ao entrar em uma banheira, notou o vazamento de água e percebeu que havia encontrado a solução, que também é conhecida como o princípio de Arquimedes: quando um corpo é parcial ou completamente imerso em um fluido, a perda aparente de peso é igual ao peso do fluido deslocado por ele. Dessa forma, Arquimedes induziu corretamente um princípio que é aplicado a outros corpos. Como o ouro é mais denso

que a prata, um dado peso em ouro representa um volume menor do que um mesmo peso em prata. Portanto ele precisava pesar a coroa imersa na água e pesar a água deslocada (BARANAUSKAS, 2001).

Apesar da indução ser o recurso mais utilizado pelo cérebro humano para derivar conhecimento novo, ela deve ser utilizada com cautela, pois se o número de exemplos for insuficiente, ou se os exemplos não forem bem escolhidos, as hipóteses obtidas podem ser de pouco valor.

Diversos algoritmos de AM utilizam o aprendizado indutivo para realizar generalizações e especializações na indução de uma hipótese, ou conhecimento, a partir de exemplos. Nesse caso, o algoritmo de aprendizado é um sistema computacional que toma decisões baseadas em experiências acumuladas contidas em casos resolvidos com sucesso (WEISS e KULIKOWSKI, 1991).

Basicamente, o aprendizado de máquina indutivo pode ser dividido em aprendizado supervisionado e aprendizado não supervisionado. No aprendizado supervisionado é fornecido ao algoritmo de aprendizado (indutor), um conjunto de exemplos de treinamento para os quais é conhecido o rótulo da classe. Normalmente, cada exemplo é descrito por um vetor de atributos e pelo rótulo da classe associada. O objetivo do algoritmo de indução é construir um classificador (hipótese) que possa determinar corretamente a classe de novos exemplos ainda não rotulados, ou seja, exemplos que não tenham o rótulo da classe. Este tipo de aprendizado trabalha com conjuntos de exemplos pré-classificados, isto é, rotulados com o atributo classe. O objetivo é construir um modelo preditivo que consiste em rotular novos exemplos que não possuem o atributo classe associado. Assim, a partir dos exemplos rotulados, denominados exemplos de treinamento, um algoritmo de aprendizado é utilizado para induzir padrões presentes em cada classe, gerando uma descrição da classe. Essas descrições são utilizadas para rotular novos exemplos. Para rótulos de classe discretos, esse problema é conhecido como classificação e para rótulos de classe contínuos, como regressão.

No aprendizado não supervisionado, o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, utilizando alguma medida de similaridade, formando agrupamentos ou *clusters*. Neste tipo de aprendizado, o problema é agrupar os exemplos não rotulados que apresentem um mesmo padrão de acordo com um critério pré-estabelecido. Diferentemente do que ocorre no aprendizado supervisionado, os resultados de um processo de *clustering* não

fornece nenhum tipo de explicação ou descrição. É necessário, nestes casos, uma interpretação dos grupos encontrados para extrair o significado conceitual de cada um. Em função dessas limitações, opta-se pela utilização do aprendizado de máquina supervisionado.

De acordo com Michalski et al. (1998), os sistemas de aprendizado são classificados em duas grandes categorias:

- sistemas caixa-preta que desenvolvem sua própria representação do conceito, isto é, sua representação interna pode não ser facilmente interpretada por humanos e não fornecem nem esclarecimento nem explicação do processo de reconhecimento;
- sistemas orientados a conhecimento que objetivam a criação de estruturas simbólicas compreensíveis por humanos.

Os sistemas baseados em conhecimento e sistemas especialistas solucionam problemas que, normalmente, requerem habilidades humanas para resolvê-los. Um requisito importante nos sistemas especialistas, que nem sempre é necessário para outras atividades de aprendizado, é a habilidade de explicar suas decisões. Para tanto, os sistemas especialistas possuem uma grande base de conhecimento cuja construção é geralmente atribuída a um engenheiro de conhecimento. De acordo com Rezende e Pugliese (1998), a construção manual de bases de conhecimento ou aquisição de conhecimento explícito é considerada um gargalo no desenvolvimento de sistemas especialistas.

Uma justificativa fundamental para pesquisas em aprendizado de máquina é a redução considerável do tempo gasto no processo de aquisição de conhecimento. Uma maneira de diminuir o tempo necessário para adquirir conhecimento é através do uso de exemplos reais, ou definidos por especialistas em um domínio de conhecimento específico, apresentando-os aos algoritmos de AM. Este processo é denominado de aquisição de conhecimento implícito.

Em várias situações, o processo de aprendizado automático produz resultados superiores, quando comparados com os resultados obtidos por especialistas. Como exemplo, considere o projeto de um controlador automático para um sistema complexo. Uma forma de construir um controlador consiste em definir um modelo preciso da dinâmica do sistema e utilizar uma variedade de métodos formais (incluindo métodos de planejamento em Inteligência Artificial) para projetar um controlador que tenha certas propriedades garantidas. Uma forma alternativa consiste em simplesmente aprender o mapeamento correto de um estado do sistema para uma ação correta. Para sistemas

muito complexos, tais como aeronaves, o desenvolvimento de um modelo detalhado pode ser inviável, o que leva a considerar a forma alternativa.

Kononenko (1993) salientou que regras de diagnóstico geradas automaticamente têm desempenho e precisão sensivelmente melhores do que aquelas definidas por especialistas.

4.2 Hierarquia do aprendizado

De acordo com Mitchell (1998), o aprendizado de máquina pesquisa métodos computacionais relacionados à aquisição automática de novos conhecimentos, novas habilidades e novas formas de organizar o conhecimento já existente.

Considerando a complexidade da inferência realizada sobre as informações disponíveis em casos prévios, dentre outros, existem diversas estratégias de AM: hábito, instrução, dedução, analogia e indução. A estratégia de aprendizado por hábito é a mais simples, enquanto que as por analogia e indução são as mais complexas, pois o aprendiz necessita de um maior esforço para o aprendizado.

Como visto no item anterior, o aprendizado indutivo pode ser supervisionado ou não supervisionado. A figura 4.1 mostra a hierarquia do aprendizado de máquina descrita anteriormente, na qual os nós sombreados levam ao aprendizado supervisionado utilizando classificação.

Para exemplificar o processo de classificação em termos práticos, suponha que se queira aprender uma forma para prever se um paciente tem problemas cardíacos. Para tanto é necessário verificar os históricos dos pacientes nos quais seriam encontrados registros contendo atributos, tais como idade, sexo, dor no peito, nível de colesterol, taxa máxima de batimentos cardíacos, a presença de angina induzida por exercícios, entre outros. Presume-se que cada registro histórico tenha sido rotulado por um especialista médico como um paciente saudável ou doente. O conjunto de exemplos composto por históricos de pacientes é então fornecido como entrada para um algoritmo de indução. A saída resultante, ou seja, a hipótese induzida, normalmente consiste em algumas regras que permitem classificar novos pacientes, isto é, que permitem determinar se um novo paciente apresenta ou não problemas cardíacos.

Figura 4.1 – Hierarquia do aprendizado

Na figura 4.2 é apresentada uma árvore de decisão induzida a partir de dados reais provenientes do conjunto de exemplos *cleveland* – *Cleveland heart disease* (BLAKE E MERZ, 1998). Esta árvore pode ser utilizada para classificar novos pacientes. Começando pela raiz da árvore, repetidamente segue-se o ramo de acordo com o atributo testado até que um nó folha seja encontrado, o qual rotula o paciente como saudável (healthy) ou doente (sick).

Baranauskas (2001) citou que um fator importante a ser considerado é que os classificadores, caso sejam analisados por seres humanos, devem fornecer uma descrição mais compacta do conceito embutido nos dados. Assim sendo, supondo um conjunto de n exemplos, é esperado que o classificador induzido forneça uma descrição menor do que n . Caso contrário, os exemplos descreveriam melhor e de forma mais compacta a si próprios, do que o próprio classificador.

Figura 4.2 – Parte da árvore de decisão induzida por C4.5 para o conjunto de exemplos *Cleveland heart disease* (BLAKE E MERZ, 1998)

4.3 Aprendizado supervisionado: Definições

Apresentam-se, a seguir, algumas definições de termos mais utilizados na literatura de AM:

- **indutor ou classificador** - consiste em extrair um bom classificador a partir de um conjunto de exemplos rotulados. A saída do indutor pode ser usada para classificar exemplos novos com a meta de prever corretamente o rótulo de cada um.

- **exemplo** - é um vetor de valores de atributos. Um exemplo descreve o objeto de interesse, tal como um paciente, dados médicos sobre uma determinada doença, histórico de clientes de uma dada companhia ou informações operacionais de sistemas de abastecimento de água (valores das variáveis de decisão).

- **atributo** - descreve alguma característica ou aspecto de um exemplo. Normalmente, há pelo menos dois tipos de atributos: discreto ou nominal, quando não existe uma ordem entre os valores (ex.: cor: vermelho, verde, azul) e contínuo, quando existe uma ordem linear nos valores (ex.: peso ? R, um número real).

Um ponto importante a ser considerado é a escolha de atributos com boa capacidade preditiva. De acordo com Weiss e Kulikowski (1991), independente do

método a ser empregado, os conceitos que podem ser aprendidos são dependentes dos dados e da qualidade dos atributos. Por exemplo, para a tarefa de determinar se uma pessoa está ou não com gripe, pode-se escolher atributos com baixo poder preditivo, tais como cor do cabelo, cor do olho, modelo do carro, número de filhos ou atributos com alto poder preditivo, tais como temperatura, resistência da pele e exame do pulmão.

- **classe** - no aprendizado supervisionado todo exemplo possui um atributo especial, o rótulo ou classe, que descreve o fenômeno de interesse, isto é, a meta que se deseja aprender para poder fazer previsões a respeito.

- **conjunto de exemplos** – usualmente, um conjunto de exemplos é dividido em dois subconjuntos disjuntos: o conjunto de treinamento que é usado para o aprendizado do conceito e o conjunto de teste usado para medir o grau de efetividade do conceito aprendido. A Tabela 4.1 ilustra o formato de uma tabela atributo-valor, um dos formatos mais utilizados pelos algoritmos de aprendizado. Nessa tabela está representado um conjunto de dados com N exemplos e M atributos. Nesse formato, as colunas ($A_1, \dots, A_M$) representam os diferentes atributos e as linhas ($E_1, \dots, E_N$) os diferentes exemplos.

Tabela 4.1 – Conjunto de dados no formato atributo-valor

	A_1	A_2	$\dots$	A_M	Y
E_1	a_{11}	a_{12}	$\dots$	a_{1M}	y_1
E_2	a_{21}	a_{22}	$\dots$	a_{2M}	y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
E_N	a_{N1}	a_{N2}	$\dots$	a_{NM}	y_N

Na tabela 4.1, a linha i refere-se ao i -ésimo exemplo ($i = 1, 2, \dots, N$) e cada valor a_{ij} refere-se ao valor do j -ésimo ($j = 1, 2, \dots, M$) atributo A_j do exemplo E_i . Y é o atributo que assume os valores do atributo classe de cada exemplo E_i , para o aprendizado supervisionado.

- **classificador** – dado um conjunto de exemplos de treinamento, um indutor gera como saída um classificador (também denominado hipótese ou descrição de

conceito) de forma que, dado um novo exemplo, ele possa prever precisamente sua classe.

Formalmente, em classificação, um exemplo é um par $(x_i, f(x_i))$ onde x é a entrada e $f(x)$ é a saída. A tarefa de um indutor é, dado um conjunto de exemplos, induzir uma função h que aproxima f , normalmente desconhecida. Neste caso, h é chamada uma hipótese sobre a função objetivo f , ou seja, $h(x_i) \sim f(x_i)$.

- **bias** – Todo indutor deve adotar algumas suposições de forma a generalizar os exemplos de treinamento. Estas suposições são denominadas de *bias* de AM. Já que a verdadeira função f é desconhecida, existem muitas escolhas possíveis para h , mas sem conhecimento adicional não é possível escolher entre as hipóteses. Qualquer preferência de uma hipótese sobre outra, além da simples consistência com os exemplos, é denominada *bias* de aprendizado (RUSSEL e NORVIG, 1995). Devido ao fato de que quase sempre existe um número grande de hipóteses consistentes, todos os indutores possuem alguma forma de *bias*.

- **under e overfitting** – Weiss e Indurkha (1998) relataram que, ao induzir uma hipótese, a partir dos dados disponíveis, é possível que esta seja muito específica para o conjunto de treinamento utilizado. Como o conjunto de treinamento é apenas uma amostra de todos os exemplos, é possível induzir hipóteses que melhorem seu desempenho no conjunto de treinamento, enquanto pioram o desempenho para exemplos diferentes daqueles pertencentes ao conjunto de treinamento. Nesta situação, o erro para um conjunto de teste independente leva a um desempenho ruim da hipótese. Neste caso, diz-se que a hipótese ajusta-se em excesso ao conjunto de treinamento ou que houve um *overfitting*. De outra forma, é possível induzir hipóteses que apresentem uma melhora de desempenho muito pequena para o conjunto de treinamento, assim como para o conjunto de teste. Neste caso, diz-se que a hipótese ajusta-se muito pouco ao conjunto de treinamento ou que houve *underfitting*.

- **overtuning** – é o ajuste de um algoritmo. Pode ocorrer quando o implementador ajusta um algoritmo de aprendizado ou os seus parâmetros, para otimizar o seu desempenho para todos os exemplos disponíveis. Pode ser uma técnica prejudicial quando todos os exemplos disponíveis são usados para o desenvolvimento ou ajuste do algoritmo.

- **poda** – é uma técnica para lidar com *overfitting* quando a hipótese induzida é expressa por regras ou árvores de decisão. A ideia geral consiste em lidar com o

problema de *overfitting* através do aprendizado de uma hipótese bem genérica a partir do conjunto de treinamento de forma a melhorar o desempenho para exemplos não vistos. Há, basicamente, dois métodos de poda:

- **regra** – uma regra assume a forma

if L then R

ou na forma simbólica

L ? R

Normalmente, as partes esquerda L e direita R da regra são complexos, sem atributos comuns entre eles, ou seja, atributos (L) n atributos = 0 . A parte esquerda é denominada condição, premissa, cauda ou corpo da regra, e a parte direita é denominada conclusão ou cabeça da regra.

- **erro e precisão** – com o resultado do aprendizado, a hipótese é a compreensão de um conceito pelo sistema. Tal conceito é obtido pela observação de exemplos no conjunto de treinamento. Entretanto, a principal meta do aprendizado supervisionado é generalizar esse conceito de forma a predizê-lo em exemplos não utilizados pelo indutor, ou seja, aqueles que não fazem parte do conjunto de treinamento. Tais medidas são denominadas erro ou precisão aparente, mas elas não representam uma boa estimativa do erro para novos exemplos. Uma medida muito utilizada em problemas de classificação é a taxa de erro do classificador h, denotada por $\text{err}(h)$. A taxa de erro pode ser obtida utilizando a seguinte equação:

$$\text{err}(h) = \frac{1}{N} \sum_{i=1}^N \| y_i \neq h(\bar{x}_i) \| \quad (4.1)$$

Esta expressão compara o rótulo real de cada exemplo com o rótulo predito pelo classificador. A expressão $\| y_i \neq h(\bar{x}_i) \|$ retorna 1, caso $y_i \neq h(\bar{x}_i)$, e zero em caso contrário. N é o número de exemplos. O complemento da taxa de erro, denominado acurácia do classificador, denotado por $\text{acc}(h)$ é dado pela equação 4.2.

$$\text{acc}(h) = 1 - \text{err}(h) \quad (4.2)$$

Freqüentemente, o cálculo do erro e da acurácia de uma hipótese é realizado utilizando técnicas de amostragem. Técnicas de amostragem caracterizam-se por dividir o conjunto de dados em várias partições, a serem utilizadas como conjunto de treinamento e teste. Algumas dessas técnicas são descritas a seguir.

- **amostragem aleatória** – segundo essa técnica de amostragem, diferentes partições de treinamento-teste são escolhidas aleatoriamente. Para cada partição de treinamento é induzida uma hipótese que tem então seu erro calculado a partir da partição de teste correspondente. O erro final é a média dos erros de todas as hipóteses induzidas.

- ***k-fold cross-validation*** – nesse caso os exemplos são aleatoriamente divididos em k partições mutuamente exclusivas de tamanho aproximadamente igual a $\frac{N}{k}$ exemplos, sendo N o número total de exemplos. Utilizam-se então $(k-1)$ partições para treinamento e a hipótese induzida é testada na partição restante. Esse processo é repetido k vezes, considerando a cada vez uma partição diferente para teste. O erro final é a média dos erros calculados com cada uma das k partições. Ressalta-se que a maioria dos programas de aprendizado de máquina que realizam *cross-validation* efetuam os cálculos da média ($mean(A)$) e o desvio padrão($sd(A)$) utilizados na verificação da qualidade do algoritmo.

- ***leave-one-out*** – esta técnica é um caso especial de ***k-fold cross validation***. Nesse caso, o valor de k é igual ao número total de exemplos N . Assim, a cada iteração, um dos exemplos é utilizado para testar a hipótese induzida a partir dos $(N-1)$ exemplos restantes. O erro final é a soma dos erros em cada teste dividido por N . A técnica *leave-one-out* é usualmente utilizada no caso do conjunto de dados ser pequeno.

- **prevalência de classe** – um ponto muito importante em aprendizado de máquina refere-se ao não balanceamento de classes em um conjunto de exemplos. Por exemplo, suponhamos um conjunto de exemplos com a seguinte descrição de classes e a proporção de exemplos em cada classe: $(C_1, C_2, C_3) = (99,00\%, 0,25\%, 0,75\%)$, com prevalência da classe C_1 . Um classificador simples que classifique sempre novos exemplos como pertencentes à classe majoritária C_1 teria uma precisão de 99,00%. Isto pode ser indesejável quando classes minoritárias são aquelas que possuem uma informação muito importante, por exemplo, supondo C_1 : paciente normal, C_2 : paciente com doença A e C_3 : paciente com doença B.

De acordo com Kubat et al. (1998) é desejável utilizar uma medida de desempenho diferente da precisão, quando se trabalha com conjunto de exemplos não balanceados. Isto, pelo fato de que a maioria dos sistemas de aprendizado é projetada para otimizar a precisão. Dessa forma, normalmente os algoritmos apresentam um

desempenho ruim se o conjunto de treinamento encontra-se fortemente não balanceamento, pois os classificadores induzidos tendem a ser altamente precisos para os exemplos da classe majoritária mas, freqüentemente, classificam incorretamente exemplos das classes minoritárias.

- **classificador simbólico** – é uma hipótese cuja linguagem de descrição pode ser transformada em um conjunto de regras, por exemplo, indutor de regras ou árvores de decisão.

- **matriz de confusão** – a matriz de confusão de uma hipótese h oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas *versus* as classificações preditas para cada classe, sobre um conjunto de exemplos T . Como pode ser visto na tabela 4.2, os resultados são totalizados em duas dimensões: classes verdadeiras e classes preditas, para k classes diferentes $\{C_1, C_2, \dots, C_k\}$. Cada elemento $M(C_i, C_j)$ da matriz, $i, j = 1, 2, \dots, k$ representa o número de exemplos de T que realmente pertencem à classe C_i , mas foram classificados como sendo da classe C_j .

Tabela 4.2 – Matriz de confusão de um classificador

Classe	predita C_1	predita C_2		predita C_k
verdadeira C_1	$M(C_1, C_1)$	$M(C_1, C_2)$		$M(C_1, C_k)$
verdadeira C_2	$M(C_2, C_1)$	$M(C_2, C_2)$		$M(C_2, C_k)$
.	.	.	.	.
.	.	.	.	.
verdadeira C_k	$M(C_k, C_1)$	$M(C_k, C_2)$		$M(C_k, C_k)$

Cada elemento $M(C_i, C_j)$ da matriz pode ser calculado pela equação 4.3.

$$M(C_i, C_j) = \left\| \sum_{\{ \forall (x,y) \in T : y = C_i \}} h(x) = C_j \right\| \quad (4.3)$$

O número de acertos, para cada classe, se localiza na diagonal principal $M(C_i, C_i)$ da matriz. Os demais elementos $M(C_i, C_j)$, para $i \neq j$, representam os erros para um

tipo particular de classificação. Na tabela 4.3 é apresentada uma matriz de confusão de um classificador ideal. Pode-se notar que todos os elementos $M(C_i, C_j)$ para $i \neq j$ são iguais a zero uma vez que o classificador não comete erros.

Considerando problemas de classificação binária, ou seja, problemas de classificação com apenas duas classes (geralmente rotuladas como “+” e “-”), as escolhas são estruturadas para prever a ocorrência ou não ocorrência de um simples evento ou hipótese. Quando apenas duas classes são consideradas, os dois erros possíveis são denominados falsos positivos e falsos negativos.

A tabela 4.4 ilustra a matriz de confusão para problemas de classificação com duas classes onde T_p é o número de exemplos corretamente classificados como exemplos positivos, F_p é o número de exemplos erroneamente classificados como positivos, T_n é o número de exemplos corretamente classificados como exemplos negativos, F_n é o número de exemplos erroneamente classificados como negativos e $n = (T_p + F_p + T_n + F_n)$ é o número total de exemplos.

Tabela 4.3 – Matriz de confusão de um classificador ideal

Classe	predita C_1	predita C_2		predita C_k
verdadeira C_1	$M(C_1, C_1)$	0		0
verdadeira C_2	0	$M(C_2, C_2)$		0
.	.	.	.	.
.	.	.	.	.
verdadeira C_k	0	0		$M(C_k, C_k)$

Considerando a tabela 4.4, quatro situações podem ocorrer:

1. O exemplo pertence à classe C_+ e é predito pelo classificador como pertencente à classe C_+ . Neste caso, o exemplo é um verdadeiro positivo.
2. O exemplo pertence à classe C_- e é predito pelo classificador como pertencente à classe C_- . Neste caso, o exemplo é um verdadeiro negativo.

Tabela 4.4 – Matriz de confusão para problemas de classificação binária

Classe	preditos como C_+	preditos como C_-	precisão da classe	precisão total
C_+	verdadeiros positivos T_p	falsos negativos F_n	$\frac{T_p}{T_p + F_n}$	$\frac{T_p + T_n}{n}$
C_-	falsos positivos F_p	verdadeiros negativos T_n	$\frac{T_n}{F_p + T_n}$	

3. O exemplo pertence à classe C_- e é predito pelo classificador como pertencente à classe C_+ . Neste caso, o exemplo é um falso positivo.

4. O exemplo pertence à classe C_+ e é predito pelo classificador como pertencente à classe C_- . Neste caso, o exemplo é um falso negativo.

- **matriz de contingência** – a matriz de confusão é aplicada ao classificador visto como uma caixa-preta, ou seja, o classificador pode ou não ser simbólico para se calcular a matriz. Para a matriz de contingência, que é calculada para cada regra, é exigido que o classificador seja simbólico.

Considerando cada regra no formato $L \rightarrow R$, sua correspondente matriz de contingência é mostrada na tabela 4.5. Nesta tabela, L denota o conjunto de exemplos para os quais a condição da regra é verdadeira e seu complemento $\bar{L}$ representa o conjunto de exemplos para os quais a condição da regra é falsa e, da mesma forma, para R e $\bar{R}$. LR corresponde ao conjunto de exemplos $L \cap R$ no qual ambos L e R são verdadeiros, $L\bar{R}$ denota o conjunto de exemplos $L \cap \bar{R}$ no qual L é verdadeiro e R é falso, e assim por diante.

Tabela 4.5 – Matriz de contingência para a regra $L \rightarrow R$

	L	$\bar{L}$	
R	lr	$\bar{l}r$	r
$\bar{R}$	$l\bar{r}$	$\bar{l}\bar{r}$	$\bar{r}$
	l	$\bar{l}$	n

Da tabela 4.5, temos:

- lr = número de exemplos para os quais L é verdade e R é verdade;
- $\bar{l}r$ = número de exemplos para os quais L é falso e R é verdade;
- $l\bar{r}$ = número de exemplos para os quais L é verdade e R é falso;
- $\bar{l}\bar{r}$ = número de exemplos para os quais L é falso e R é falso;
- r = número de exemplos para os quais R é verdade;
- $\bar{r}$ = número de exemplos para os quais R é falso;
- l = número de exemplos para os quais L é verdade;
- $\bar{l}$ = número de exemplos para os quais L é falso; e
- n = número total de exemplos.

Seja x a cardinalidade do conjunto X , ou seja $x = |X|$. Na tabela 4.4, l denota a cardinalidade do conjunto L , ou seja, $l = |L|$. Em outras palavras, l denota o número de exemplos para os quais a condição da regra é verdade. Da mesma forma, $r = |R|$ denota o número de exemplos para os quais a conclusão da regra é verdade. Assim, $l\bar{r} = |\overline{LR}|$ denota o número de exemplos para os quais a condição é verdadeira e a conclusão é falsa, ou seja, é o número de exemplos erroneamente cobertos pela regra. A cardinalidade n denota o número total de exemplos utilizados para determinar as demais cardinalidades expressas na matriz de contingência.

Associada a cardinalidade x , a frequência relativa f_x é utilizada como uma estimativa da probabilidade $P(X)$, ou seja $P(X) = f_x = \frac{x}{n}$. Por exemplo, considerando a cardinalidade lr , a probabilidade $P(LR)$ pode ser determinada da seguinte forma:

$$P(LR) = f_{lr} = \frac{lr}{r} \quad (4.4)$$

De forma semelhante podem ser determinados os valores das probabilidades $P(\bar{L}R)$, $P(L\bar{R})$ e $P(\bar{L}\bar{R})$. Conhecidas essas probabilidades, os valores de $P(L)$, $P(\bar{L})$, $P(R)$ e $P(\bar{R})$ podem ser determinados. Por exemplo:

$$P(R) = P(\bar{L}R) + P(LR) \text{ ou } P(\bar{R}) = P(\bar{L}\bar{R}) + P(L\bar{R}) \quad (4.5)$$

A matriz de contingência é uma generalização da matriz de confusão, que é a base padrão para computar-se medidas de avaliação de hipóteses em problemas de classificação binária.

Várias medidas podem ser usadas para avaliar o desempenho de um classificador, sendo a precisão a mais comum. Entretanto, novas medidas considerando

novidade, simplicidade e facilidade de compreensão humana são necessárias (TODOROVSKI et al., 2000).

Utilizando como base a matriz de contingência, é possível definir a maioria das medidas sobre regras, por exemplo, a confiabilidade positiva (prel), confiabilidade negativa (nrel), suporte (sup), sensibilidade (sens), especificidade (spec), precisão total (tacc), cobertura (cov), novidade (nov) e satisfação (sat), definidas através das equações a seguir:

$$\text{prel}(L \rightarrow R) = P\left(\frac{R}{L}\right) = \frac{lr}{l} \quad (4.6)$$

$$\text{nrel}(L \rightarrow R) = P\left(\frac{\bar{R}}{\bar{L}}\right) = \frac{\bar{l}\bar{r}}{\bar{l}} \quad (4.7)$$

$$\text{sup}(L \rightarrow R) = P(LR) = \frac{lr}{n} \quad (4.8)$$

$$\text{sens}(L \rightarrow R) = P\left(\frac{L}{R}\right) = \frac{lr}{r} \quad (4.9)$$

$$\text{spec}(L \rightarrow R) = P\left(\frac{\bar{L}}{\bar{R}}\right) = \frac{\bar{l}\bar{r}}{\bar{r}} \quad (4.10)$$

$$\text{tacc}(L \rightarrow R) = P(LR) + P(\bar{L}\bar{R}) = \frac{lr + \bar{l}\bar{r}}{n} \quad (4.11)$$

$$\text{cov}(L \rightarrow R) = P(L) = \frac{l}{n} \quad (4.12)$$

$$\text{nov}(L \rightarrow R) = P(LR) - P(L)P(R) = \frac{lr}{n} - \frac{l \times r}{n^2} \quad (4.13)$$

$$\text{sat}(L \rightarrow R) = \frac{P(\bar{R}) - P(\bar{R}/L)}{P(\bar{R})} = 1 - \frac{n \times \bar{l}\bar{r}}{l \times \bar{r}} \quad (4.14)$$

- **cobertura** – considerando uma regra $L \rightarrow R$, os exemplos que satisfazem a parte L da regra constituem o seu conjunto de cobertura, ou seja, os exemplos são cobertos pela regra ou a regra dispara para esses exemplos. Exemplos que satisfazem tanto a condição L como a conclusão R são cobertos incorretamente pela regra. Por outro lado, os exemplos que não satisfazem a condição L não são cobertos pela regra. Baranauskas (2001) apresentou um resumo dessas quatro situações através da tabela 4.6.

Tabela 4.6 – Definições de cobertura da regra *if L then R*

Exemplos satisfazendo:	são:
$\bar{L}$	não cobertos pela regra
L	cobertos pela regra
$L \wedge R$	cobertos corretamente pela regra
$L \wedge \bar{R}$	cobertos incorretamente pela regra

4.4 Indutores

Kohavi et al. (1996) apresentaram uma biblioteca denominada MLC++ que continha vários indutores. Dentre eles, o *C4.5*, o *C4.5RULES*, o *C5.0*, o *CN2*, o *IB* (*Instance-Based*) e o *NB* (*Naive Bayes*).

Esses indutores são bem difundidos na comunidade de AM. Representam três abordagens de aprendizado. *NB* é um indutor estatístico muito simples, *IB* é um indutor preguiçoso (*lazy*) e *C4.5*, *C4.5RULES* e *CN2* são indutores ávidos (*eager*).

Algoritmos puramente preguiçosos armazenam todos os dados de treinamento e respondem a uma requisição de informação através da combinação de seus dados armazenados, descartando a resposta obtida e qualquer resultado intermediário. Em contraste, algoritmos ávidos compilam os dados de treinamento de maneira “gulosa” por uma descrição de conceito intencional, tais como um conjunto de regras ou uma árvore de decisão, descartando os dados após este processo.

Mitchell (1998) ressaltou que a principal diferença entre os métodos preguiçosos e ávidos é que os primeiros permitem a opção da seleção de uma hipótese diferente a cada novo exemplo a ser classificado enquanto os segundos estão comprometidos com uma única hipótese que cobre todo o espaço de exemplos.

Neste trabalho, foi utilizado um sistema comercial para plataforma Windows denominado *See5*¹ que inclui melhorias sobre algoritmos *C4.5* e *C4.5RULES* (Quinlan, 1993), os quais têm sido usados, freqüentemente, para comparar seu bom desempenho com outros algoritmos de AM. O sistema *See5* foi projetado para trabalhar com conjuntos de dados relativamente grandes. Como seus precursores, manipula atributos com valores discretos ou contínuos, induzindo conceitos expressos como árvores de decisão ou conjuntos de regras *if-then* não ordenadas (BARANAUSKAS, 2001). Os arquivos de entrada para o *See5* são dois, um contendo os dados no formato atributo-

¹ <http://www.rulequest.com/see5-info.html>

valor e o outro contendo a descrição do tipo dos atributos, ambos com o mesmo nome, porém com extensões **.data** e **.names**, respectivamente. A saída do algoritmo pode ser uma árvore ou regras de decisão, dependendo dos parâmetros especificados pelo usuário.

4.5 Finalização

Baranauskas (2001) ressaltou que na mídia atual é comum o anúncio de produtos nos quais o fabricante afirma “você aperta o botão e nós fazemos o resto”. Para o autor seria interessante, em extração de conhecimento, poder fazer o mesmo: coletar dados e deixar que o algoritmo de indução faça o resto. Porém, uma grande quantidade de esforço é necessária ao selecionar exemplos, escolher atributos, integrar o conhecimento prévio sobre o domínio, selecionar o método que fornece a melhor representação, avaliar os resultados e repetir todo esse processo.

Aprendizado de máquina pode não substituir a engenharia do conhecimento como técnica para construir sistemas baseados em conhecimento, entretanto, representa um progresso significativo em direção à automação. Ao que tudo indica, a indução de regras e outros métodos de aprendizado tornar-se-ão progressivamente dominantes, à medida que seus benefícios se tornarem mais compreensíveis.

MODELO COMPUTACIONAL

Discovery consists of seeing what everybody has seen and thinking what nobody has thought.

- Albert von Szent-Gyorgyi

5.1 Generalidades

Quando a demanda total de um sistema de distribuição de água para abastecimento é menor que a capacidade total de distribuição há uma potencial redução nos custos de operação de bombas e válvulas. A busca por uma operação ótima do sistema, como sendo um problema de controle ótimo, que visa encontrar as decisões para operação de válvulas e bombas, dentro de um intervalo de tempo pré-definido, deve possibilitar que as demandas sejam atendidas, as condições de pressões satisfeitas e o custo total de operação minimizado (ZESSLER e SHAMIR, 1989).

Na resolução de problemas de engenharia como esse, procura-se, inicialmente, representar o sistema real de maneira mais fidedigna possível, para que possa reproduzir o comportamento do mesmo. Os recursos computacionais atualmente disponíveis mostram que as simplificações, que há algumas décadas eram necessárias para resolução de problemas de engenharia em geral, vem se tornando cada vez mais dispensáveis. Entretanto, para a resolução do modelo matemático entendido como sendo o conjunto completo e consistente de equações que representam o sistema físico real, é necessária a inserção de parâmetros complementares obtidos, em geral, em campo, visando a obtenção de resultados numéricos significativos (KOELLE, 1992).

Técnicas de otimização vêm sendo tradicionalmente usadas no dimensionamento de unidades de um sistema de distribuição de água, objetivando menores custos. A interface entre o modelo de simulação hidráulica e o modelo de otimização deve ser

cuidadosamente elaborada no sentido de dar transparência ao modelo, de forma a facilitar o seu uso e, ao mesmo tempo, permitir a análise de problemas complexos, envolvendo múltiplos objetivos. Atualmente, algoritmos baseados em processo estocástico de busca vêm sendo empregados com sucesso, devido à facilidade de aplicação e praticamente à ilimitada condição de uso como os algoritmos genéticos (AG).

Além da redução dos custos operacionais, um objetivo importante, quando se pretende otimizar operacionalmente um sistema de abastecimento de água potável, é a maximização de sua confiabilidade. A confiabilidade é a medida da performance (desempenho) do sistema. A performance hidráulica depende das condições ideais que viabilizam: a interação entre o sistema de transporte, a conservação da capacidade de armazenamento, a conservação da capacidade de distribuição e o funcionamento do sistema de segurança, através do funcionamento adequado de válvulas reguladoras e bombas e maximização da confiabilidade dos componentes individuais do sistema (BAO e MAYS, 1990).

A confiabilidade pode ser descrita através da minimização de dois tipos de falhas: mecânicas e hidráulicas. Falhas mecânicas são do tipo: rompimento de tubulações, falhas no bombeamento, falhas no controle de válvulas e perda de potência, dentre outras. São consideradas como falhas hidráulicas, as ocorrências de vazões e pressões inadequadas, para um ou mais pontos de demanda do sistema.

Usualmente, a política global empregada na operação de um sistema de abastecimento de água não é definida de maneira clara. A principal razão é a dificuldade em obter as informações, devido ao fato de que estas, geralmente, são baseadas apenas no conhecimento e experiência do operador. Outras razões são a variedade de mecanismos de controle e as frequentes mudanças na topologia do sistema de abastecimento de água (LEÓN et al., 2000).

O objetivo da operação de sistemas de abastecimento de água é o atendimento, com riscos aceitáveis, das necessidades de consumo e minimização dos custos operacionais e, de maneira implícita, um melhor aproveitamento do sistema de transporte e reservação, de forma a retardar investimentos com ampliações. A operação é a seqüência de manobras exercidas sobre os elementos ativos do sistema, como válvulas e bombas, de forma a atender os objetivos.

Como o sistema de macro distribuição de água é usualmente composto de estações de bombeamento e reservatórios, o procedimento básico de operação deve

considerar o controle das vazões e cargas ou o número de bombas em operação e o nível de água disponível nos reservatórios, de forma a garantir o atendimento independente da variação diária das demandas. De acordo com Ko et al. (1997), o objetivo de uma efetiva operação de sistemas de abastecimento de água é minimizar os custos operacionais, mantendo a operação das bombas de forma estável, além de garantir uma alta confiabilidade do sistema, através da manutenção dos níveis de água nos reservatórios próximos dos máximos. A redução nos custos de energia elétrica pode ser obtida com uma maior utilização dos conjuntos elevatórios no período fora de pico de consumo de água, mas isto vai depender de uma complicada correlação entre demanda horária e tarifa de energia elétrica.

A otimização proposta neste trabalho, leva em consideração dois objetivos: a minimização de custos operacionais decorrentes do consumo de energia elétrica nas estações de bombeamento e a maximização dos benefícios hidráulicos (considerando o índice de atendimento da demanda, os níveis adequados de água nos reservatórios e as pressões mínimas e máximas nos pontos de demanda, para um período de análise de 24 horas).

Com o objetivo de reduzir os custos de operação relativos à utilização de estações de bombeamento, além de incrementar os benefícios hidráulicos do sistema, este trabalho propõe o desenvolvimento de um modelo a ser utilizado para obtenção de regras de operação de unidades de um sistema real, com operação não automática, que satisfaça os objetivos pré-fixados. O modelo proposto é composto de três módulos distintos: um módulo de simulação hidráulica que utiliza os códigos do *EPANET2*, um módulo otimizador que utiliza algoritmos genéticos multiobjetivo e um módulo indutor de regras através do algoritmo de aprendizado de máquina.

Com o objetivo de avaliar os resultados obtidos pelo modelo proposto, complementarmente é implementada uma rotina denominada ordem de preferência - OP (*Preference Ordering Routine-POR*) utilizada para a escolha e extração de soluções (estratégias operacionais) do conjunto Pareto ótimo, em substituição ao módulo indutor de regras.

O desenvolvimento do modelo proposto visa produzir uma ferramenta útil e prática que minimize as intervenções subjetivas por parte dos operadores, os quais, normalmente, executam a tarefa de acordo com o grau de experiência pessoal adquirida no exercício de suas atividades diárias.

A Figura 5.1 apresenta um fluxograma representativo do modelo proposto.

5.2 Simulação hidráulica

Os modelos de simulação de redes de distribuição de água consistem basicamente do equacionamento das vazões nos trechos de uma rede e das cargas hidráulicas nos nós entre trechos consecutivos. Podem ser aplicados em estudos de dimensionamento de redes, isoladamente, onde a solução é obtida por análises sucessivas, com alterações impostas pelos usuários, ou em combinação com modelos de otimização, onde a solução é obtida automaticamente.

A simulação hidráulica avalia as respostas do sistema em termos das suas variáveis de estado: pressão, vazão e níveis nos reservatórios, face às diversas conjunturas operacionais. É, portanto, uma ferramenta essencial à avaliação dos objetivos estabelecidos. Neste trabalho, o simulador utilizado se baseia no código original do *EPANET2* (ROSSMAN, 2001), acoplado ao C++ (*Borland Builder C++ 5.0*).

O *EPANET2* é um programa de domínio público desenvolvido pela *Environmental Protection Agency (EPA)* para analisar redes de distribuição de água em regime permanente. O pacote computacional é distribuído gratuitamente sob duas formas: executável e código fonte (*Toolkit Library*). A forma executável é aplicável a estudos que envolvem apenas as simulações hidráulicas. A forma de código fonte é recomendável para estudos que envolvem o acoplamento de outros modelos (calibração, vazamentos, demandas variáveis, entre outros) com o *EPANET2*.

A *Toolkit* é uma biblioteca de vínculo dinâmico (*DLL*) a qual contém funções que permitem aos programadores o desenvolvimento de aplicações específicas. Essas funções podem ser incorporadas em aplicações Windows escritas em: C/C++, Pascal (Delphi), Visual Basic ou outras linguagens de programação que possibilitem chamar as funções através de *DLLs*. Além disso, essa biblioteca permite abrir e fechar um arquivo que descreve a rede (*.INP*), bem como ler e modificar os vários parâmetros de projeto e operação, simular períodos extensivos acessando resultados a cada passo e escrever os resultados das simulações nos arquivos de saída (*.RTP*).

Figura 5.1 - Fluxograma do modelo proposto

Para acoplar o *EPANET2* com o compilador C++ (*Borland Builder C++ 5.0*), primeiramente, necessrio copiar os arquivos fonte (*Toolkit*) atravs de *download* da pgina na *web*² da *Environmental Protection Agency (EPA)*. A *Toolkit* inclui diferentes

² <http://www.epa.gov/epanet.html#Downloads>

arquivos de cabeçalho(.h), diferentes arquivos de funções e arquivos .lib que simplificam a tarefa de interface da biblioteca com o compilador (C/C++, Pascal(Delphi), Visual Basic). Essa biblioteca contém também um arquivo de ajuda para orientar a utilização adequada das várias funções existentes através de alguns programas exemplo.

5.3 Otimização operacional de sistemas de abastecimento de água potável

As técnicas de otimização vêm sendo utilizadas na busca de soluções ótimas para problemas operacionais específicos. Se o objetivo for, por exemplo, o mínimo custo operacional, a função custo estará associada à tarifa de energia elétrica, ao rendimento das bombas, às perdas de carga nas instalações, etc. Por outro lado, há restrições impostas pelo próprio sistema, tais como, níveis máximos e mínimos dos reservatórios, limites de pressão e de potência e quantidade de água disponível.

Algumas dificuldades decorrem da utilização das técnicas de otimização para a solução do problema do controle. Estas dificuldades estão associadas ao grande número de equações a serem resolvidas, à dificuldade de formulação do problema geral, ao tratamento de restrições operacionais não explícitas e à aquisição e manutenção de dados.

A otimização da operação de um sistema de abastecimento público de água é um complexo processo de objetivos múltiplos envolvendo uma negociação (*trade-off*) entre os custos e benefícios, associado às decisões operacionais.

Como citado no capítulo 3, a literatura fornece uma grande quantidade de algoritmos evolucionários apropriados para a aplicação a problemas de otimização multiobjetivo.

Cheung et al (2003), trabalhando com a otimização de reabilitação de sistemas de distribuição de água, apresentou um estudo comparativo entre dois métodos evolucionários multiobjetivo, o *MOGA* (não elitista) e o *SPEA* (elitista). As análises foram conduzidas para um sistema de distribuição hipotético simples, para o qual a minimização de custos e de pressões requeridas foi tratada como um problema de dois objetivos. Os autores concluíram que, para aquele tipo de problema, o *SPEA* apresentou uma performance melhor que o *MOGA* em termos de frentes Pareto produzidas e do tempo de processamento computacional requerido.

Neste trabalho são implementados e testados três algoritmos evolucionários multiobjetivos que foram descritos no capítulo 3. Um não elitista, o *Nondominated*

Sorting Genetic Algorithm (NSGA) e dois elitistas, o *Nondominated Genetic Algorithm II (NSGA II)* e o *Strength Pareto Evolutionary Algorithm (SPEA)*. Estes algoritmos foram avaliados através da biblioteca de metas-heurísticas multiobjetivo *MOMHLib++*³ (JASKIEWICZ, 1999), de domínio público, escrita em C++ por Andrzej Jaskiewicz em 1999. A *MOMHLib++* não implementa em seu código os operadores genéticos de recombinação e mutação, os quais devem ser introduzidos pelo próprio usuário, de acordo com o problema investigado.

Os operadores genéticos implementados e investigados durante o desenvolvimento deste trabalho foram: recombinação BLX, uniforme e linear e mutação aleatória, limite e não uniforme, cujas descrições foram apresentadas nos itens 3.5.6.2 e 3.5.6.3.

5.4 Extração de regras utilizando o algoritmo *See5*

A principal função de um algoritmo de aprendizado de máquina é construir os modelos de classificação na forma de árvores de decisão ou conjunto de regras que possam determinar corretamente a classe de novos exemplos ainda não rotulados. Neste trabalho, este não é o principal objetivo. O classificador denominado conjunto de regras (*rulesets*) é utilizado para a seleção da(s) melhor(es) estratégia(s) e posterior extração de regras operacionais, a partir de um conjunto de exemplos (casos) fornecidos pelo modelo de otimização através das soluções pertencentes à frente Pareto. Para tanto é utilizado o sistema *See5*, que é a mais recente versão do indutor *C4.5* descrito por Quinlan (1993), conforme descrito no item 4.4 deste trabalho.

5.5 Método de ordem de preferência (OP)

Neste trabalho, o método de ordem de preferência (OP) é utilizado para a escolha de um número menor de soluções da frente Pareto com o objetivo de definir uma estratégia operacional ótima dentre as várias soluções possíveis constantes da frente. O principal objetivo é avaliar os resultados obtidos com a aplicação desse processo e compará-los com os obtidos através da utilização do algoritmo de aprendizado de máquina, que é a metodologia aqui proposta.

Este novo método, ainda pouco utilizado em sistemas de distribuição de água, é destinado à classificação dos conjuntos de parâmetros ótimos Pareto, de acordo com

³ <http://www-idss.cs.put.poznan.pl/~jaskiewicz/MOMHLib>

suas habilidades de dominação num sub-espço (N) Pareto, onde N é a dimensão das funções objetivo.

Para melhor entendimento do processo, apresenta-se a seguir um exemplo simples de um problema de minimização de três funções objetivo (N=3). A tabela 5.1 apresenta os conjuntos de pontos ótimos Pareto para N=3. Há um total de 12 pontos ótimos Pareto em que nem todos esses pontos são totalmente dominados pelos outros. As figuras 5.2, 5.3 e 5.4 apresentam os pontos da tabela 5.1 plotados no espaço (N-1)-dimensional ou bidimensional.

Tabela 5.1 – Valores numéricos das três funções objetivo

Conjunto	objetivo 1	objetivo 2	objetivo 3
A	7,82	10,89	4,08
B	15,79	12,99	2,65
C	2,74	14,90	4,43
D	7,84	15,72	3,16
E	15,57	17,46	1,76
F	3,45	18,98	1,52
G	13,09	19,61	1,27
H	0,32	41,24	1,25
I	9,96	52,47	0,44
J	3,71	70,77	0,25
K	1,17	72,78	0,98
L	2,12	96,08	0,45

Figura 5.2 – Soluções - espaço bidimensional objetivo1 x objetivo 2

Figura 5.3 – Soluções – espaço bidimensional objetivo 1 x objetivo 3

Figura 5.4 – Soluções – espaço bidimensional objetivo 3 x objetivo 2

Nas figuras 5.2 a 5.4, os pontos I(círculo) e J(triângulo) são ótimos Pareto em seus respectivos espaços bidimensionais. Analisando as figuras, várias observações podem ser listadas:

- nem todos os pontos ótimos Pareto no espaço tridimensional são ótimos no espaço bidimensional;
- alguns dos pontos ótimos Pareto do espaço bidimensional [objetivo1, objetivo2] não aparecem como pontos ótimos Pareto nos outros dois espaços, [objetivo1, objetivo3] e [objetivo2, objetivo3]; e
- pode haver um ou mais pontos que estão no espaço Pareto ótimo bidimensional para todas as combinações das funções objetivo.

Se existem alguns pontos ou soluções que satisfazem este último requisito, então elas serão as soluções preferidas. No caso do exemplo, o ponto I é a solução desejada e corresponde, na tabela 5.1, ao ponto H (0,32; 41,24; 1,25).

O conceito de ordem de preferência descrito anteriormente foi formulado primeiramente como dois teoremas e provados matematicamente por Das (1999). Estes dois teoremas, de acordo com o autor, tem um critério de dominação tão confiável quanto o de dominação Pareto e são apresentados da seguinte forma:

- **Eficiência de ordem k (ou k-pontos ótimos Pareto)**

Considerando todos os possíveis sub-espacos k -dimensionais de um espaco de funcao objetivo N -dimensional ($1 \leq k \leq N$), um ponto e definido como sendo eficiente de ordem k , se este ponto no e dominado por nenhum dos outros pontos em nenhum dos sub-espacos k -dimensionais.

- Eficincia de ordem k com grau m (ou $[k,m]$ – pontos timos Pareto)

Dependendo dos valores de N e k , pode haver um nmero de $(k+1)$ -solucoes timas Pareto porm, nenhuma dessas, so k -solucoes timas Pareto. Se o nmero de $(k+1)$ -solucoes tima Pareto e muito grande, e necessrio subjugar estas solucoes para promover um minucioso exame. Sendo k o sub-espaco dimensional das funcoes objetivo e m o nmero de relacoes entre essas funcoes ou o nmero de sub-espacos plotados, um ponto(soluco) e definido como sendo **eficiente de ordem k com grau m** se este no e dominado por nenhum outro ponto no possvel sub-espaco k -dimensional para todos os m sub-espacos plotados, e e denominado de $[k,m]$.

Para melhor esclarecimento do teorema descrito, aplicar-se- o conceito de ordem de preferncia no exemplo da tabela 5.1. Desta aplicacao, pode-se encontrar pontos que so de ordem 3 (N)- timo Pareto, $[2,1]$ -timo Pareto, $[2,2]$ - timo Pareto e $[2,3]$ - timo Pareto, onde os valores entre colchetes so $[k,m]$ definidos no teorema anterior, e podem ser vistos na tabela 5.2. Para o exemplo, a solucao que apareceu nas quatro condicoes descritas, no caso a solucao representada pelo ponto **H**, foi a escolhida.

Na tabela 5.2, os espacos sombreados representam os pontos cujas solucoes pertencem ao grau de eficincia, previamente definidos.

5.6 Descricao do problema em estudo

O propsito deste trabalho e desenvolver uma ferramenta flexvel e facilmente utilizvel pelos supervisores da operacao de sistemas de abastecimento de gua, que forneça um conjunto eficiente de regras operacionais, atendendo os objetivos pr-definidos e as condicoes de funcionamento de cada unidade do sistema. Para anlise do modelo proposto e avaliacao dos resultados, foi utilizada uma parte do sistema de macro distribuicao de gua potvel da cidade de Goinia, no estado de Gois.

Tabela 5.2 – Aplicação do método OP nos dados da tabela 5.1

<i>Conjunto</i>	<i>3-ótimo Pareto</i>	<i>[2,1]-ótimo Pareto</i>	<i>[2,2]-ótimo Pareto</i>	<i>[2,3]-ótimo Pareto</i>
A				
B				
C				
D				
E				
F				
G				
H				
I				
J				
K				
L				

5.6.1 Sistema de abastecimento de água de Goiânia

O sistema de abastecimento de água da cidade de Goiânia é suprido por dois mananciais superficiais, o ribeirão João Leite e o rio Meia Ponte. Estes mananciais abastecem o sistema de forma independente, porém, a rede de distribuição de ambos é interligada em alguns pontos. A população atualmente atendida pelos subsistemas é de aproximadamente 1.2 milhão de habitantes, com uma vazão total de aproximadamente $4.5 \text{ m}^3/\text{s}$. Cada subsistema abastece 50% da população da cidade. Cada um conta com unidades de bombeamento de água bruta, estação de tratamento, estações elevatórias de água tratada, adutoras, reservatórios de distribuição e redes de distribuição de água principais e secundárias.

Optou-se pela utilização de apenas parte do sistema de abastecimento de água de Goiânia denominado macro distribuição João Leite e composto por unidades de bombeamento, adução, reservação e nós de demandas, que são interligados aos reservatórios e a pontos de abastecimento da rede de distribuição de água. Na realidade, os nós de demanda correspondem à derivação para setores de redes representativos da

área de influência de cada reservatório. Assim, as demandas atribuídas a estes nós representam as vazões de abastecimento de tais setores. Observa-se que a simulação hidráulica realizada neste trabalho não considerou as possíveis variações de vazões produzidas a partir da estação de tratamento de água do João Leite, considerando que os níveis de água no reservatório pulmão SENAC são compatíveis à sucção pela estação elevatória. Da mesma forma, não foram consideradas características específicas dos setores de redes de distribuição como vazões máximas e mínimas e perdas de carga nas tubulações e cargas mínimas e máximas nos nós, assim como perdas e transientes. Como todas as unidades do sistema são interligadas, considerou-se o atendimento da demanda máxima horária em cada área de influência dos reservatórios e as pressões mínimas nos pontos mais desfavoráveis da rede de distribuição. Os gráficos da distribuição horária das demandas para as áreas de atendimento de cada reservatório são apresentados na figura 5.6. Estas curvas de demandas foram fornecidas pela SANEAGO e representam, segundo os responsáveis pela operação do sistema, a efetiva quantidade de água medida na saída dos reservatórios, incluindo, portanto, os consumos e as perdas na rede de distribuição. Para a manutenção das pressões mínimas na rede de distribuição de água, sugerida pela SANEAGO, no valor de 10 m, o mínimo valor nos nós de demanda, de acordo com informações dos responsáveis pela operação e manutenção do sistema de Goiânia, deve ser da ordem de 15m. A pressão máxima considerada para os nós de demanda foi de 60m, seguindo, da mesma forma, orientações dos operadores do sistema em estudo.

Para melhor entendimento da aplicação proposta, apresenta-se um esquema do sistema estudado (figura 5.5), com suas características principais (figura 5.6 e tabelas 5.3 a 5.8), cuja operação pode ser descrita sucintamente da seguinte forma:

- a estação elevatória de água tratada EE-Senac succiona água do reservatório Senac (1) e recalca (bomba) para quatro grandes reservatórios de distribuição de jusante, Pedro Ludovico (14), Amazônia (19), Itapoã (29) e Atlântico (24).

- os quatro reservatórios apresentam variações de níveis de água e características operacionais diferenciadas, dependentes das demandas em cada área específica. Em função destas características, foram previstas estruturas de controle (ECs) na entrada de cada reservatório: ECs Pedro Ludovico (29), Amazônia (30), Itapoã (32) e Atlântico (31). Estas ECs são compostas de válvulas de controle automático de vazão (VCVs), que têm como objetivo o controle das vazões nas entradas dos reservatórios em função das variações dos níveis de água de acordo com as demandas ao longo do dia.

5.6.2 Definição das funções objetivo

Vários objetivos podem ser listados quando se avalia um sistema real de abastecimento de água, visando a sua operação ótima. Um deles é a minimização do custo operacional, incluindo a manutenção e a operação do sistema. Para isto, necessita-se de dados sobre todos os parâmetros mecânicos e hidráulicos relevantes para esta avaliação. Vários trabalhos já desenvolvidos demonstraram que, de todos os parâmetros relacionados a esta questão, o mais relevante é o custo relativo ao consumo de energia elétrica nas estações elevatórias de água. Um outro objetivo que deve ser focado é a

Figura 5.5 – Esquema de macro-adição de Goiânia

Figura 5.6 – Goiânia – curvas das demandas horárias por áreas de reservação – Saneago S.A.(2004)

confiabilidade do sistema em atender os consumidores de forma adequada. Neste caso, vários parâmetros podem ser enumerados.

A confiabilidade dos sistemas de abastecimento de água pode ser considerada sob o aspecto hidráulico ou mecânico. O primeiro está relacionado aos parâmetros físicos, mutáveis em função de mudanças operacionais do sistema, enquanto o segundo considera as possíveis intervenções sobre os equipamentos. Confiabilidade mecânica é a capacidade que os componentes do sistema têm de atender continuamente as necessidades de operação sem a necessidade de freqüentes reparos, modificações ou substituições de componentes ou sub-componentes. Quando quantificada, a confiabilidade mecânica é meramente uma expressão da probabilidade de um equipamento operar de maneira adequada num determinado período de tempo. Confiabilidade hidráulica é definida como a capacidade que o sistema tem de atender as demandas dos usuários com um nível aceitável de interrupção, independente de condições anormais que possam ocorrer. A avaliação da confiabilidade hidráulica está diretamente relacionada com a função básica do sistema de distribuição de água, que é distribuir uma quantidade específica de água, num local determinado, com uma pressão adequada ao atendimento do usuário, num período de tempo pré-fixado. É importante que a confiabilidade hidráulica seja considerada na determinação da confiabilidade mecânica.

Este trabalho propõe avaliar dois objetivos básicos. O objetivo econômico e o objetivo de confiabilidade hidráulica, aqui denominado benefícios, relativos aos sistemas de macro distribuição de água potável, sendo que este último considera três parâmetros distintos: suprimento das vazões nos nós de demanda com pressões adequadas, atendimento de níveis mínimos de água nos reservatórios e pressões mínimas nos referidos nós para garantia do abastecimento de água nos pontos de consumo.

Tabela 5.3 – Goiânia – dados de demanda

Nós de Demanda	Cota(m)	Demandas Máx.(l/s)
15	843,50	616
20	846,70	496
25	820,65	474
30	846,80	542

Fonte: Saneago S.A., 2004.

Tabela 5.4 – Goiânia – dados dos reservatórios

Reservatórios	Volume(m^3)	Nív.Mín(m)	Nív.Máx(m)	Elevação(m)
P.Ludovico(14)	10.000	1,5	6,0	858,0
Amazônia(19)	5.000	1,5	5,5	861,5
Atlântico(24)	10.000	1,5	7,0	836,5
Itapoã(29)	3.000	1,5	5,0	863,0

Fonte: Saneago S.A., 2004.

Tabela 5.5 – Goiânia - dados das bombas da estação Senac

Bombas	Vazão(l/s)	Carga(m)
26	895	85
27	895	85
28	895	85

Fonte: Saneago S.A., 2004.

Tabela 5.6 – Goiânia – dados das válvulas de controle

Válvulas	Nó 1	Nó 2	Diâm.(mm)	Tipo(*)	Vazão(l/s)
29	12	13	600	VCV	616
30	17	18	600	VCV	496
31	27	28	800	VCV	542
32	22	23	500	VCV	474

Fonte – Saneago S.A., 2004.

* - Válvula Controladora de Vazão(VCV)

Tabela 5.7 – Goiânia – dados dos nós

NÓS				
Número	Cota(m)	Demandas Máximas(l/s)	Hora	Multipl. Deman.
2	788,10	0	1	0,3
3	788,15	0	2	0,15
4	788,12	0	3	0,2
5	788,10	0	4	0,45
6	788,12	0	5	0,43
7	788,13	0	6	0,55
8	801,20	0	7	0,6
9	823,10	0	8	0,8
10	845,15	0	9	0,9
11	840,05	0	10	1
12	845,12	0	11	0,9
13	845,12	0	12	0,8
15	843,50	616	13	0,7
16	848,20	0	14	0,65
17	850,15	0	15	0,65
18	850,15	0	16	0,6
20	846,70	496	17	0,6
21	852,20	0	18	0,63
22	832,30	0	19	0,68
23	832,25	0	20	0,65
25	820,65	474	21	0,6
26	850,10	0	22	0,3
27	850,15	0	23	0,3
28	850,13	0	24	0,3
30	846,80	542		
1	Reservatório Senac			
14	Reservatório P. Ludovico			
19	Reservatório Amazônia			
24	Reservatório Atlântico			
29	Reservatório Itapoã			

Fonte: Saneago S.A., 2004.

Tabela 5.8 – Goiânia – dados dos trechos

TUBULAÇÕES					
Número	Nó Inicial	Nó Final	Comp(m)	Diam(mm)	Coef.(HW)
1	1	2	50	1500	100
2	2	3	5	1200	100
3	2	6	5	1200	100
4	4	5	5	1200	100
5	5	7	5	1200	100
6	5	8	2050	1500	100
7	8	9	2840	600	100
8	9	10	3990	800	100
9	10	12	200	800	100
10	8	11	4725	1372	100
11	11	10	120	800	100
12	13	14	50	800	100
13	14	15	10	800	100
14	11	16	1050	1372	100
15	16	21	5368	1200	100
16	21	22	441	800	100
17	23	24	50	800	100
18	24	25	10	800	100
19	21	26	2070	800	100
20	26	27	50	1000	100
21	28	29	50	1000	100
22	29	30	10	1000	100
23	16	17	473	800	100
24	18	19	50	800	100
25	19	20	10	800	100
26	Estação Elevatória Senac-Bomba 1				
27	Estação Elevatória Senac-Bomba 2				
28	Estação Elevatória Senac-Bomba 3				
29	Válvula de Controle 1 – Pedro Ludovico				
30	Válvula de Controle 2 - Amazônia				
31	Válvula de Controle 4 - Itapoã				
32	Válvula de Controle 3 - Atlântico				

Fonte: Saneago S.A., 2004.

No caso do objetivo econômico, pretende-se minimizar os custos relativos ao consumo de energia elétrica na estação elevatória de água Senac.

Para composição do custo de energia elétrica, foram consideradas duas parcelas: a do consumo em kWh e a de demanda máxima mensal em kW.

A *demanda máxima mensal* é definida como sendo o maior valor registrado para a potência elétrica exigida pelos motores do sistema elétrico, php (rendimento bomba x motor). Este valor é integralizado a cada 15 (quinze) minutos pelo medidor de energia e registrado o maior no mês (quando a avaliação for mensal). Para unidades com demandas maiores que 300 kW a SANEAGO estabelece contratos no regime de tarifas horo-sazonais (tarifas diferenciadas para o período seco e para o úmido) que consideram tarifas no período de ponta (18 às 21h) e fora de ponta. Neste caso há dois tipos: a azul e a verde. A verde tem tarifa única. A azul tem tarifas diferenciadas nos dois horários. No caso do sistema de abastecimento de água de Goiânia, a Saneamento de Goiás S.A. - SANEAGO adota, para unidades maiores que 300kW, que não podem parar no horário de ponta, a tarifa azul, considerada neste trabalho.

O consumo de energia é o valor efetivamente medido, totalizado durante o período de faturamento, expresso em kWh.

O custo diário atribuído a cada bomba de uma estação elevatória é dado pela soma do custo relativo ao fator de demanda máxima mais o custo do consumo medido.

De acordo com dados fornecidos pela SANEAGO, para a estação elevatória Senac, os valores(Jun-2003) utilizados são os seguintes:

- para o consumo no horário de pico (18 – 21hs), R\$ 0,17076/kWh e fora deste horário, R\$ 0,0816/kWh;

- para a demanda máxima no período entre 18 e 21hs o valor total médio é de R\$ 26,38/kW e, fora deste intervalo, R\$ 8,66/kW.

Adotando como objetivo econômico a minimização dos custos de energia elétrica, relativos à operação de estações elevatórias, a função objetivo (FO_1) pode ser expressa utilizando a equação 5.1.

$$FO_1 = \sum_{t=1}^{24} \sum_{k=1}^{nb} Cu(t) * \frac{Q(k, t) * H(k, t) * ?}{?(k, t)} + D * Tarifa \quad (5.1)$$

onde :

k = número de bombas na estação elevatória ;

t = hora do dia;

nb = número total de bombas na elevatória ;

Cu(t) = custo unitário da tarifa(R\$/ kWh);

Q(i, t) = vazão bombeada(m³/s);

H(i, t) = carga hidráulica (m);

?(i, t) = rendimento do conjunto(%);

D = fator de demanda máxima(kW) ;

Tarifa = tarifa correspondente ao fator de demanda máxima(R\$/ kW).

? = coeficiente de conversão de unidades

A quantidade de água que um sistema de abastecimento pode efetivamente distribuir aos usuários, a uma pressão adequada, é um dos principais fatores na determinação da performance e da confiabilidade deste sistema. Para tanto, a relação entre as demandas e as cargas nodais deve ser incorporada na medida da confiabilidade.

Como citado anteriormente, a função objetivo benefício hidráulico será formulada para quantificar os benefícios resultantes da adequação de pressões nos nós de demanda, dos níveis de água nos reservatórios de distribuição e garantia de atendimento às demandas horárias.

Baseado em estudos desenvolvidos por Gargano e Pianese (2000), Tanyimboh et al. (2001) e Righetto (2002), este trabalho propõe a avaliação dos benefícios hidráulicos através da adoção de coeficientes ou índices de performance. Para consideração das pressões nos nós de demanda utilizou-se o coeficiente denominado de benefício de atendimento às pressões mínimas nodais (ψ_{bp}), de acordo com o seguinte equacionamento:

$$\psi_{bp}(i, t) = \left(\frac{P_{at}(i, t) - P_{mín}}{P_{req}(t) - P_{mín}} \right)^{1/2} \text{ se } P_{mín} \leq P_{at}(i, t) \leq P_{req}(t) \quad (5.2)$$

$$\psi_{bp}(i, t) = 0 \text{ se } P_{at}(i, t) < P_{mín} \text{ ou } P_{at}(i, t) > P_{req}(t) \quad (5.3)$$

sendo :

$P_{at(i, t)}$ = pressão no nó de demanda i no tempo t (fornecida pelo simulador hidráulico);

$P_{mín}$ = pressão mínima adotada no caso 15m; e

$P_{req(t)}$ = pressão máxima admitida no tempo t.

O benefício hidráulico relativo ao atendimento adequado da pressão nodal (BH_{PN}) será dado pela seguinte equação:

$$BH_{PN} = \sum_{t=1}^{24} \sum_{i=1}^{nn} ?_{bp(i, t)} \quad (5.4)$$

Para avaliação dos benefícios relativos aos níveis de água nos reservatórios, considerou-se o coeficiente denominado de benefício dos níveis adequados de água nos reservatórios (ψ_{bn}), dado pela seguinte equação:

$$?_{bn(j, t)} = \left(\frac{N_{at(j, t)} - N_{mín(j)}}{N_{req(j, t)} - N_{mín(j)}} \right)^{1/2} \quad \text{se } N_{mín} \leq N_{at(j, t)} \leq N_{req(j, t)} \quad (5.5)$$

$$?_{bn(j, t)} = 0 \quad \text{se } N_{at(j, t)} < N_{mín} \quad \text{ou } N_{at(j, t)} > N_{req(j, t)} \quad (5.6)$$

sendo :

$N_{at(j, t)}$ = nível de água no reservatório j no tempo t(fornecido pelo simulador) ;

$N_{mín(j)}$ = nível mínimo de água no reservatório j; e

$N_{req(j, t)}$ = nível de água requerido no tempo t.

O benefício hidráulico relativo à adequação dos níveis de água nos reservatórios (BH_{NR}) é dado pela seguinte equação:

$$BH_{NR} = \sum_{t=1}^{24} \sum_{j=1}^{nr} ?_{bn(j, t)} \quad (5.7)$$

Com relação à garantia de atendimento das demandas considerou-se o benefício (BH_{AD}) de acordo com o seguinte equacionamento:

$$BH_{AD} = \sum_{t=1}^{24} \sum_{i=1}^{nn} \left(\frac{P_{at(i,t)} - P_{mín}}{P_{req(t)} - P_{mín}} \right)^{1/2} * \left(\frac{Q_{dem(i,t)}}{\sum_{i=1}^{nn} Q_{dem(i,t)}} \right) \quad (5.8)$$

onde :

$Q_{dem(i,t)}$ = demanda horária nó de demanda i no tempo t; e

$\sum_{i=1}^{nn} Q_{dem(i,t)}$ = demanda horária total no tempo t.

nn = número total de nós de demanda.

A função objetivo (FO_2) relativa aos benefícios hidráulicos será expressa como o somatório desses três benefícios, ou seja:

$$FO_2 = BH_{PN} + BH_{NR} + BH_{AD} \quad (5.9)$$

O problema de minimização de FO_1 e de maximização de FO_2, conforme equações 5.1 e 5.9 formulado anteriormente é de otimização multiobjetivo. As funções objetivo custo e benefício hidráulico são conflitantes entre si. Assim, não há uma solução ótima simples que possa satisfazer o ótimo global de ambos os objetivos, mas um conjunto de soluções ótimas designadas soluções não dominadas ou não inferiores que não podem ser melhoradas simultaneamente, sem sacrificar pelo menos um objetivo. O conjunto destas soluções é chamado frente ótima Pareto. Os algoritmos genéticos têm se mostrado efetivos na identificação destas soluções.

5.7 Implementação dos métodos evolucionários multiobjetivo para o problema de operação ótima de sistemas de distribuição de água

5.7.1 Representação das soluções

Para o caso específico do sistema em estudo, cada vetor representativo de uma possível solução S1 para a estratégia operacional do sistema tem a seguinte característica:

$$S_1 = \left[\underbrace{B_{(0,1)}, B_{(0,2)}, B_{(0,3)}, V_{(0,1)}, V_{(0,2)}, V_{(0,3)}, V_{(0,4)}}_{\text{Hora 0}} \dots \underbrace{B_{(23,1)}, B_{(23,2)}, B_{(23,3)}, V_{(23,1)}, V_{(23,2)}, V_{(23,3)}, V_{(23,4)}}_{\text{Hora 23}} \right]$$

Sendo :

S_1 = vetor solução ;

B e V = variáveis de decisão relativas à operação de bombas e válvulas, respectivamente;

$B_{(0,1)}$ = condição de funcionamento da bomba número 1 na hora zero (0/1 – desligada/ligada) ;

$V_{(0,1)}$ = condição de funcionamento da válvula número 1 na hora zero (0/1 – fechada/aberta).

Assim, cada possível solução define valores para o funcionamento de três bombas e quatro válvulas de controle, durante o período de 24 horas, totalizando 168 variáveis de decisão.

5.7.2 Definição do método evolucionário multiobjetivo e dos operadores genéticos e critério de parada

6.7.2.1 Definição do método evolucionário multiobjetivo

Existe uma tendência em utilizar algoritmos evolucionários para problemas de otimização envolvendo objetivos múltiplos. A literatura apresenta diversos métodos capazes de buscar em uma única simulação as soluções ótimas Pareto. Entretanto, a escolha do método mais eficiente não é uma tarefa fácil (CHEUNG, 2004).

Visando à avaliação dos algoritmos, várias simulações utilizando 5 diferentes dimensões de populações iniciais (100, 200, 300, 400 e 500) foram desenvolvidas, empregando *NSGA*, *NSGA II* e o *SPEA*, para o problema de otimização operacional em estudo. A figura 5.7 apresenta as melhores frentes Pareto para cada algoritmo, do conjunto analisado. Estas frentes Pareto foram obtidas utilizando população inicial de 300.

Figura 5.7 – Frentes Pareto - algoritmos multiobjetivo

Da análise visual dos resultados relativos ao conjunto de soluções de cada algoritmo, percebe-se que o *SPEA* apresenta uma frente Pareto com soluções melhores distribuídas do que o *NSGA* e *NSGA II*. Apesar do número menor de soluções identificado pelo *SPEA*, quando comparado ao *NSGA II*, este fornece, para os mesmos custos (minimização) mostrados pelos outros algoritmos, maiores (melhores) valores para os benefícios hidráulicos (maximização), ou seja, as soluções obtidas através *SPEA* representam as soluções não-dominadas quando comparadas com as dos outros algoritmos. Em outras palavras, as soluções do *SPEA* dominam as demais. Além disso, o tempo de processamento computacional para o algoritmo *SPEA* foi 30% menor que o gasto para os outros dois algoritmos testados. Dessa forma, decidiu-se adotar, para este trabalho, o *Strength Evolutionary Pareto Algorithm-SPEA* como método para a otimização multobjetivo em estudo.

5.7.2.2 Definição dos operadores genéticos e critério de parada

Após a identificação do método mais eficiente (*SPEA*) para o problema de otimização da operação de sistemas de distribuição de água considerando objetivos múltiplos, foram realizados estudos em relação aos operadores genéticos, de recombinação e mutação, com o objetivo de selecionar os mais adequados para o

problema. Para tanto, três operadores de recombinação e três de mutação foram avaliados conforme descrito no item 5.3.

Para uma melhor visualização das frentes Pareto assim obtidas, foram elaborados alguns gráficos apresentando as soluções para um mesmo operador recombinação e diferentes operadores de mutação. As figuras de 5.8 a 5.10 apresentam estas frentes.

A figura 5.11 apresenta as melhores frentes Pareto dentre aquelas avaliadas. Percebe-se que, apesar de apresentar um número maior de soluções, a frente utilizando recombinação uniforme e mutação limite fornece menores valores para a função objetivo de maximização (FO_2-benefício hidráulico), para valores similares da função objetivo de minimização (FO_1-custos), quando comparada com a frente obtida utilizando operadores recombinação uniforme e mutação não uniforme. Desta forma, para este trabalho, foram adotados os operadores recombinação uniforme e mutação não uniforme.

Figura 5.8 – Frentes Pareto – operador mutação – recombinação BLX-a

Depois da identificação do algoritmo e dos operadores genéticos mais adequados para o problema de otimização da operação do sistema de macro distribuição da cidade de Goiânia, o modelo foi testado utilizando quatro diferentes números de gerações, 500, 1000, 4000 e 5000. Vale ressaltar que o número de gerações é o parâmetro adotado como critério de parada para o algoritmo adotado (*SPEA*). O objetivo era identificar

aquele que apresentava a frente Pareto mais adequada para o problema em estudo. A figura 5.12 apresenta as frentes Pareto para cada geração testada.

Figura 5.9 – Frentes Pareto – operador mutação – recombinação uniforme

Figura 5.10 – Frentes Pareto – operador mutação – recombinação linear

Figura 5.11 – Frentes Pareto – recombinação uniforme – mutações uniforme e não uniforme

Figura 5.12 – Frentes Pareto para diversos números de gerações-algoritmo SPEA

Analisando a figura 5.12, percebe-se que as frentes para 1000, 3000 e 5000 praticamente são coincidentes nos pontos relativos às melhores soluções (faixa média das frentes), com melhores valores para a frente utilizando 5000 gerações quando

comparados os valores de FO_2 com os mesmos valores de FO_1 para as duas outras frentes mencionadas. Além disso, o tempo de processamento computacional despendido quando utilizadas 5000 gerações foi muito próximo dos tempos de processamento quando utilizados números menores de gerações. Dessa forma, para o modelo proposto, utilizou-se, como critério de parada, 5000 gerações.

5.8 Arquivo de dados para o sistema *See5*

O emprego do modelo otimizador através dos algoritmos genéticos multiobjetivo produz uma frente Pareto com as soluções operacionais de melhor compatibilização entre os objetivos pré-definidos. A utilização do sistema *See5*, sobre este conjunto de soluções, aqui chamadas de exemplos ou casos, visa extrair as melhores regras (estratégias) operacionais do conjunto apresentado após a aplicação do método *SPEA*.

O processo de classificação neste trabalho visa a identificação das melhores estratégias operacionais de um conjunto pré-definido através da frente Pareto. Para isto, é necessário verificar no arquivo de saída do modelo de otimização, os registros contendo os atributos que caracterizam a estratégia operacional. Cada registro deve ser rotulado por um especialista da área operacional. Este especialista tem várias opções em termos da melhor condição operacional para o sistema a ser operado, podendo ser destacadas a flexibilidade e a confiabilidade do sistema. Para este trabalho, considerou-se a confiabilidade mecânica representada pela condição operacional de bombas e válvulas como rótulo para definição das melhores regras. Consideraram-se as seguintes classes: condição operacional boa e condição operacional ótima.

A escolha dos parâmetros a serem utilizados no modelo visou o atendimento dos objetivos pré-definidos, quais sejam, menor custo de energia elétrica e maior benefício hidráulico. Porém, a classe relativa à **condição operacional**, definida para o aprendizado de máquina, priorizou, através do conhecimento e de informações do operador (especialista), a maximização da confiabilidade mecânica em termos de uma operação mais eficaz das bombas na estação de bombeamento e das válvulas de controle na entrada dos reservatórios. Desta forma, a classe **ótima** representa as soluções, fornecidas pela frente Pareto, que fornecem os melhores valores em termos do atendimento das demandas, pressões nos nós e os melhores níveis de água no reservatório, com os menores custos de operação em termos de energia elétrica e, também, a melhor estratégia operacional em termos de preservação e manutenção mecânica das bombas e válvulas.

Righetto (2002) propôs cinco parâmetros distintos para avaliar a performance de uma determinada regra operacional num período de 24 horas. Um desses parâmetros era o índice de condição operacional de bombas. Neste trabalho foi adotada a mesma sistemática de definição de índices operacionais de Righetto (2002). Porém, o julgamento do especialista para definição da classe condição operacional, foi implementado através de um índice operacional que avalia a estratégia de funcionamento de bombas e válvulas.

O índice foi definido com base no número de bombas e válvulas em funcionamento num período de 24 horas de operação do sistema. Foi chamado de índice de condição operacional (K_{pV}) e definido pela seguinte expressão:

$$K_{BV}(i) = \sum_{k=1}^{nb} \sum_{t=1}^{24} ?_b(k, t) + \sum_{j=1}^{nv} \sum_{t=1}^{24} ?_v(j, t) \quad (5.10)$$

Onde:

k, j = número de bombas e válvulas , respectivamente;

t = intervalo de tempo considerado;

nb = número de bombas na estação elevatória;

nv = número de válvulas no sistema;

i = solução(estratégia) operacional da frente Pareto; e

λ_b e λ_v = parâmetros de funcionamento das bombas e válvulas , respectivamente ;

Estes parâmetros têm as seguintes definições:

$$?(^{(k,t)}_b \text{ ou } ?^{(j,t)}_v = 1 \text{ se } ?_b(k, t) = ?_b(k, t-1) \text{ ou } ?_v(j, t) = ?_v(j, t-1) \quad (5.11)$$

$?^{(k,t)}_b \text{ ou } ?^{(j,t)}_v = 0$ em qualquer outra situação

A classe condição operacional será ótima (**o**) se $K_{bv(i)} > \frac{1}{ns} * C$ e boa (**b**) se

$K_{bv(i)} \leq \frac{1}{ns} * C$. Onde $C = \sum_{i=1}^{ns} K_{bv(i)}$ e ns é o número total de soluções na frente

Pareto.

Para utilização do sistema *See5* são necessários, como mencionado no item 4.4, dois arquivos de entrada distintos. O arquivo *names* com a definição dos atributos e classes e o arquivo *data*, contendo os exemplos (definidos através dos valores de seus

atributos) a serem analisados. Os formatos de parte dos arquivos, utilizando os dados obtidos da frente Pareto são apresentados a seguir.

Para o nosso caso, os atributos são representados pelas variáveis de decisão (**B**=bombas e **V**=válvulas) cujos valores podem ser **0** (desligadas/fechadas) ou **1** (ligadas/abertas) e as classes, definidas para rotulação de cada exemplo, foram condição operacional ótima (**o**) ou boa (**b**), como apresentadas no arquivo **.names**. Cada exemplo representa uma solução ou um ponto da frente Pareto resultante do modelo otimizador, como pode ser visto no arquivo **.data**.

Formato do **arquivo.names**:

/ Dados do sistema de abastecimento de água de Goiânia.

/ Por Ivaltemir(Tinil) Barros Carrijo usando Algoritmos Genéticos Multiobjetivo

/ Variáveis 0,1 para declarar operação das unidade. b=boa; o=ótima.

/ B=bomba; V=válvula; B01(Bomba 1 , hora 0).

condição operacional.

condição: b,o.

B01: 0,1.

B02: 0,1.

B03: 0,1.

V01: 0,1.

V02: 0,1.

V03: 0,1.

V04: 0,1.

B11: 0,1.

B12: 0,1.

B13: 0,1.

V11: 0,1.

V12: 0,1.

V13: 0,1.

V14: 0,1.

.

.

.

B231: 0,1.

B232: 0,1.

B233: 0,1.

V231: 0.1.

V232: 0,1.

V233: 0,1.

V234: 0.1.

Formato do **arquivo.data**:

/Saída MOMHLib para aplicação do aprendizado de máquina

/Crossover-Uniforme; Mutação-Não-Uniforme; Pcross-0.90; Pm-0.006

/Pop.-300; Ger.-5000

/Saída das Melhores Soluções

$b, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, \dots$ – Exemplo 1

$b, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, \dots$ —Exemplo 2

$b, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, \dots$ – Exemplo 3

$b, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, \dots$ —Exemplo 4

[illegible][illegible]

$b, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, \dots$ —Exemplo 7

[illegible]

$b, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, \dots$ —Exemplo 9

$0,1,0,0,1,0,0,1,1,0,0,1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,0,0,0,0,0,0,1,1,1,1,\dots$ —Exemplo 10

$b, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, \dots$ – Exemplo 11

•

.

.

[illegible]

$o, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, \dots$ – *Exemplo 49*

[illegible][illegible]

$b, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, \dots$ – Exemplo 52

O modelo fornece como saída de dados, um conjunto típico de regras com as seguintes características:

Rule 4 : (cover 15)

B62 = 0

? class o [confidence 93%]

onde:

- *cover* especifica o número de casos em que a regra é aplicada;e
- *confidence* representa o percentual de número de vezes que a regra se aplica corretamente;

Assim, a regra acima sugere que todos os exemplos (estratégias operacionais) nos quais a bomba 2 esteja desligada às 6 horas da manhã, sejam classificados como ótimos.

A tabela 5.9 apresenta a descrição dos atributos dos conjuntos de dados da frente Pareto.

5.9 Modelo computacional proposto

Um modelo computacional foi implementado compreendendo basicamente três módulos: primeiro o de avaliação hidráulica, utilizando códigos da biblioteca *Toolkit-Epanet2*), um de implementação de AGs multiobjetivos, utilizando códigos fornecidos pela biblioteca , e um último para criação do classificador de conjunto de regras a partir do sistema *See5* .

As principais características do sistema macro-adutor de Goiânia, utilizado como aplicação neste trabalho, foram apresentadas através das figuras 5.5 e 5.6 e das tabelas 5.3 a 5.8.

Ressalta-se que as análises aqui efetuadas consideram um período de simulação de 24 horas, iniciando a avaliação a zero hora. Como essa avaliação considera um plano para obtenção da melhor regra operacional para as 24 horas subseqüentes, optou-se por impor que os níveis de água nos reservatórios, no final do período de análise, sejam idênticos aos considerados no início da simulação.

Os parâmetros finais utilizados nos módulos de simulação hidráulica e otimização foram:

- níveis iniciais máximos nos reservatórios de distribuição (zero hora);

- aproveitamento total dos volumes de reservação até o nível de água de 1,5m;
- recombinação uniforme ;
- mutação não-uniforme;
- tamanho da população – 300;
- número máximo de gerações – 5000;
- probabilidade de recombinação – 0,90;
- probabilidade de mutação – 0,006.

Para o módulo indutor de regras foram utilizados os arquivos (**.data e .names**) com os formatos similares aos apresentados no item 5.8.

Tabela 5.9 – Goiânia – frente Pareto – descrição dos atributos

Atributo	tipo	descrição	valor
1 - condição	discreto	condição operacional	ótima (o) , boa (b)
2 - B01	contínuo	bomba 1, hora 0	0 (desligada), 1 (ligada)
3 - B02	contínuo	bomba 2, hora 0	0 (desligada), 1 (ligada)
4 - B03	contínuo	bomba 3, hora 0	0 (desligada), 1 (ligada)
5 - V01	contínuo	válvula 1, hora 0	0 (fechada), 1 (aberta)
6 - V02	contínuo	válvula 2, hora 0	0 (fechada), 1 (aberta)
7 - V03	contínuo	válvula 3, hora 0	0 (fechada), 1 (aberta)
8 - V04	contínuo	válvula 4, hora 0	0 (fechada), 1 (aberta)
.	.	.	.
.	.	.	.
.	.	.	.
163 - B231	contínuo	bomba 1, hora 23	0 (desligada), 1 (ligada)
164 - B232	contínuo	bomba 2, hora 23	0 (desligada), 1 (ligada)
165 - B233	contínuo	bomba 3, hora 23	0 (desligada), 1 (ligada)
166 - V231	contínuo	válvula 1, hora 23	0 (fechada), 1 (aberta)
167 - V232	contínuo	válvula 2, hora 23	0 (fechada), 1 (aberta)
168 - V233	contínuo	válvula 3, hora 23	0 (fechada), 1 (aberta)
169 - V234	contínuo	válvula 4, hora 23	0 (fechada), 1 (aberta)

5.10 Verificação complementar com a utilização da rotina de ordem de preferência-OP

Como citado no item 5.5, a literatura apresenta várias técnicas que podem ser utilizadas com o objetivo de extrair um número menor de soluções do conjunto Pareto ótimo. São os chamados métodos multicriteriais.

Com o objetivo de avaliar uma metodologia pouco implementada em problemas envolvendo otimização de sistemas de abastecimento de água, este trabalho propõe a utilização do método denominado ordem de preferência (*Preference Ordering Routine-POR*) para analisar os resultados obtidos com a aplicação do algoritmo de aprendizado de máquina na escolha das melhores estratégias da frente Pareto.

O método OP é aplicável a problemas com, no mínimo, três funções objetivo. Como descrito na metodologia, o problema aqui estudado considerou apenas dois objetivos: minimização dos custos e maximização dos benefícios hidráulicos. Para avaliação dos resultados com a aplicação do OP, definiu-se uma terceira função objetivo (FO_3) denominada confiabilidade mecânica. Na realidade, considerou-se que o índice de condição operacional (K_{PV}), definido para aplicação do algoritmo de aprendizado de máquina, é esta FO_3 utilizada na aplicação do OP ao problema em estudo.

As funções objetivo FO_1 e FO_2 são as mesmas definidas pelas equações 5.1 e 5.9. A função objetivo FO_3 é dada pela equação 5.12:

$$K_{BV}(i) = \sum_{k=1}^{nb} \sum_{t=1}^{24} ?_b(k, t) + \sum_{j=1}^{nv} \sum_{t=1}^{24} ?_v(j, t) \quad (5.12)$$

Onde:

k, j = número de bombas e válvulas, respectivamente;

t = intervalo de tempo considerado;

nb = número de bombas na estação elevatória;

nv = número de válvulas no sistema;

i = solução(estratégia) operacional da frente Pareto; e

$?_b, ?_v$ = parâmetros de funcionamento de bombas e válvulas, respectivamente.

Estes parâmetros têm as seguintes definições:

$$?(k,t)_b \text{ ou } ?(j,t)_v = 1 \text{ se } ?_b(k, t) = ?_b(k, t-1) \text{ ou } ?_v(j, t) = ?_v(j, t-1) \quad (5.13)$$

$$?(k,t)_b \text{ ou } ?(j,t)_v = 0 \text{ em qualquer outra situação} \quad (5.14)$$

RESULTADOS E ANÁLISES

Suppression of truth is a false representation.

- Latin proverb

Conforme indicado no item 5.6, alguns parâmetros utilizados na avaliação do modelo proposto neste trabalho foram fornecidos pela Saneamento de Goiás S.A.- SANEAGO, empresa responsável pelo gerenciamento do sistema de Goiânia. Um aspecto comum, em quase todas as concessionárias dos serviços de água no Brasil, é a falta de confiabilidade dos dados operacionais das unidades componentes dos sistemas. Informações relacionadas às demandas totais nas áreas de influência e ao comportamento dos níveis diários de água de determinado reservatório, dados relativos ao funcionamento de bombas nas estações elevatórias, dentre outros, demonstram incoerências, quando avaliadas.

Consciente das deficiências provenientes dos dados fornecidos pela SANEAGO buscou-se, para avaliação do modelo desenvolvido, utilizar dados e curvas típicos dos parâmetros do sistema e compará-los com os fornecidos pelo modelo. Tais resultados são apresentados no item 6.1.

Métodos de análise multicriterial vem sendo aplicados com o objetivo de classificar as soluções obtidas na frente Pareto. Neste trabalho, o algoritmo de aprendizado de máquina é utilizado como uma alternativa aos métodos de análise multicriterial. Para comparação e avaliação dos resultados apresentados, utilizando o modelo proposto (item 6.1), utilizou-se um método denominado ordem de preferência – OP (*Preference Ordering Routine-POR*). Os resultados obtidos aplicando o OP ao problema em estudo, são apresentados no item 6.2.

Com o objetivo de avaliar a aplicabilidade da metodologia proposta frente às deficiências dos dados da operação, foram analisadas mais três situações distintas. Na

primeira, apresentada no item 6.3, foram considerados diferentes cenários para as demandas com o objetivo de avaliar a possibilidade de adoção de uma regra operacional geral, fornecida pelo algoritmo de aprendizado de máquina, para diferentes consumos. Esta é uma situação normal em um sistema de abastecimento de água, pois a variação na demanda depende de vários aspectos, muitas vezes não considerados na rotina operacional, como por exemplo, a expansão do sistema, os vazamentos e o crescimento populacional fora dos índices considerados no projeto, entre outros.

Para obtenção dos resultados apresentados no item 6.1, consideraram-se, no início do período de avaliação (zero hora), níveis máximos de água nos quatro reservatórios do sistema. Esta condição operacional é adequada, mas nem sempre possível, para um sistema real. No item 6.4, são apresentados e avaliados os resultados fornecidos pelo modelo quando são considerados níveis médios de água nos reservatórios no início do período de simulação.

Um dos maiores problemas em sistemas de abastecimento de água, tanto em nível de projeto como de operação, é a aquisição de dados necessários para a avaliação da variação das demandas ao longo do dia. É comum, realizarem-se avaliações com base na adoção da demanda máxima diária considerada no projeto do sistema. Este procedimento acarreta um super dimensionamento das unidades, orientando para o aumento dos custos de implantação e operação. No item 6.5, são apresentados os resultados provenientes das avaliações com base na demanda máxima diária constante, na área de influência de cada reservatório do sistema de Goiânia.

6.1 Resultados e análises considerando os parâmetros típicos fornecidos pela SANEAGO

Aplicando os módulos relativos ao simulador hidráulico e ao otimizador operacional e utilizando os parâmetros descritos no item 5.9, no sistema de macro distribuição de água João Leite, em Goiânia, foi gerada a frente Pareto apresentada na figura 6.1.

As 52 soluções, mostradas na figura, representam as melhores estratégias operacionais para o sistema de macro distribuição de água de Goiânia. Porém, o objetivo do trabalho é fornecer um conjunto menor de regras operacionais ao operador do sistema, evitando que ele opere as unidades, baseado apenas em sua experiência. É importante, por outro lado, fornecer mais de uma estratégia para que, na complexa tarefa da operação, haja uma relativa flexibilidade para que fatores, não considerados no

modelo, possam, eventualmente, ser incorporados sem comprometimento do atendimento dos usuários do sistema.

Figura 6.1 – Goiânia – algoritmo multiobjetivo SPEA – frente Pareto

Com o objetivo de obter um conjunto menor de soluções para a operação do sistema, a partir das estratégias componentes da frente Pareto, foi aplicado o algoritmo de AM *See5*, conforme descrito no capítulo 4, obtendo-se duas estratégias finais. Estas estratégias foram produzidas após a utilização do índice de condição operacional (K_{PV}) definido no item 5.8 e implementação do algoritmo de aprendizado de máquina sobre o arquivo de saída do otimizador.

O conjunto de dados para aplicação do *See5* consistiu dos 52 exemplos (soluções de melhor compromisso fornecidas pela frente Pareto), representados por 169 atributos (168 contínuos e 1 discreto), conforme descritos na tabela 5.9 e duas classes predizendo a condição operacional **ótima** (44,2% dos exemplos) e **boa** (55,8% dos exemplos), conforme descrito na tabela 6.1.

Tabela 6.1 – Goiânia – frente Pareto – descrição do conjunto de dados

# Ex	atr (discretos, contínuos)	classes (ótima, boa)
52	169 (1, 168)	2 (44,2%, 55,8%)

Utilizando os arquivos **.data** e **.names**, conforme descritos no item 5.8, foi executado o algoritmo de AM *See5* para indução de regras considerando *10-fold cross-validation*. As regras selecionadas são mostradas na tabela 6.2.

Tabela 6.2 – Regras selecionadas e suas matrizes de contingência

Regras selecionadas	matriz de contingência
Regra 1 (R ₁)	18 5 23
IF B81 =1	6 23 29
AND V171 = 1	24 28 52
THEN CLASS = ótima (o)	
Regra 2 (R ₂)	
IF V03 = 0	20 3 23
AND B81 = 1	9 20 29
AND B133 = 0	
THEN CLASS = ótima (o)	29 23 52

No capítulo 4 foram apresentadas algumas definições de parâmetros normalmente utilizados para avaliar a qualidade das regras produzidas pelo *See5*. São apresentados na sequência, os cálculos e valores desses principais parâmetros para avaliação das regras escolhidas.

- precisão das regras :

$$\text{Acc} (R_1) = \frac{lr}{n} = \frac{18}{23} = 0,782$$

$$\text{Acc} (R_2) = \frac{lr}{n} = \frac{20}{23} = 0,87$$

- o erro de cada regra será:

$$\text{Err} (R_1) = 1 - \text{Acc}(R_1) = 0,217$$

$$\text{Err} (R_2) = 1 - \text{Acc}(R_2) = 0,13$$

Outras medidas genéricas de avaliação, descritas anteriormente, e que servem para avaliação das regras R₁ e R₂ são apresentadas na tabela 6.3.

Tabela 6.3 – Medidas genéricas para regras R1 e R2

Medida	regra 1(R ₁)	regra 2 (R ₂)
Precisão total (tacc)	0,789	0,769
Cobertura (cov)	0,462	0,558
Satisfação (sat)	0,552	0,443

Com base na avaliação das tabelas anteriores, é possível avaliar vários dados interessantes, tais como: o número de regras induzidas, os atributos mais relevantes utilizados na descrição das regras e os erros associados às regras induzidas pelo *See5*. Como visto no capítulo 4, a precisão de uma regra destina-se a avaliar regras individuais e, por essa razão, tende a favorecer a precisão de exemplos positivos. A precisão do conjunto de regras que constituem a hipótese (classificador) e que é definida em função da matriz de confusão, é uma medida mais apropriada para o nosso caso. A tabela 6.4 apresenta a matriz de confusão para o caso estudado e a precisão total para a hipótese, considerando o erro aparente.

Outros parâmetros úteis e que podem ser utilizados para avaliação da hipótese, são apresentadas nas tabela 6.5 e têm os seguintes significados:

- # Ex – número de exemplos;
- ótima e boa – número de exemplos com classes “ótima” e “boa”;
- ErrA % - porcentagem do erro aparente do *See5*, isto é, o erro obtido quando se utiliza todo o conjunto de dados como treinamento e teste;
- Err10cv % - porcentagem da estimativa do erro verdadeiro de *See5* obtido usando *10-fold cross validation* e o erro padrão;
- # R – número de regras induzidas por *See5* utilizando todos os exemplos para treinar e testar; e
- # MR – número médio de regras induzidas por *See5* obtido usando *10-fold cross validation* e o erro padrão;

Ressalta-se que uma medida apropriada para avaliação da hipótese é o percentual da estimativa do erro verdadeiro dado por ***Err10cv***.

Tabela 6.4 – Matriz de confusão

Classes	preditas ótimas	preditas boas	precisão da classe	precisão total
ótima	17	0	1,00	0,885
boa	6	29	0,83	

Tabela 6.5 – Goiânia – resultados do See5 – condição operacional

# Ex	ótima	boa	ErrA %	Err10cv %	#R	# MR
52	23	29	12	17 ± 0,1	4	3,9 ± 0,1

Os resultados apresentados indicam que as regras R1 e R2 selecionadas após a execução do algoritmo *See5* e utilizadas com o objetivo de escolher a(s) melhor(es) solução(ões) da frente Pareto, apresenta uma boa precisão total (88,5%) e um erro verdadeiro também aceitável (17,3%). Percebe-se que mesmo com seis exemplos mal classificados como classe ótima, mas que pertencem à classe boa, através do índice de condição operacional (K_{pv}), a capacidade preditiva do modelo induzido por *See5* é boa.

A figura 6.2 mostra a frente Pareto com as soluções definidas pelo especialista como classe de condição operacional boa e ótima, e as duas soluções finais, resultantes da aplicação do algoritmo *See5*, representando as melhores regras (estratégias) operacionais para o sistema em estudo, quando considerados os objetivos previamente definidos e o índice operacional (K_{pv}).

Após a aplicação do algoritmo *See5* e definição das duas melhores estratégias operacionais, é necessária a verificação do comportamento hidráulico do sistema segundo tais soluções. Os parâmetros hidráulicos considerados nos objetivos e no índice de avaliação do especialista, quais sejam: níveis de água nos reservatórios, pressões nos nós de demanda e frequência de funcionamento de bombas e válvulas devem ser avaliados e criticados para a verificação da eficácia do modelo proposto.

Para facilidade de compreensão, as soluções fornecidas pelo AM serão chamadas de AM1 (2494,3; 130,1) e AM2 (2650,4; 130,8).

As figuras 6.3 e 6.4 apresentam, respectivamente, os níveis de água nos quatro reservatórios, para as soluções AM1 e AM2, assim como, as figuras 6.5 e 6.6 mostram as pressões nos quatro nós de demanda, respectivamente para as mesmas soluções.

Analisando primeiramente os resultados relativos aos níveis de água nos quatro reservatórios, pode-se observar que há uma variação de comportamento dos níveis dos diversos reservatórios. Porém, as duas alternativas, indicam o mesmo comportamento. Antes de analisar especificamente os resultados fornecidos pelo modelo, é pertinente uma explicação do atual funcionamento do sistema em estudo, o que será apresentado a seguir.

Figura 6.2 – Goiânia – frente Pareto - soluções ótimas e boas – melhores soluções AM

Figura 6.3 – Goiânia – Níveis de água nos reservatórios – solução AM1

Figura 6.4 – Goiânia – Níveis de água nos reservatórios – solução AM2

O sistema João Leite, como visto anteriormente, é composto por quatro reservatórios que estão interligados às redes de distribuição principal (anéis) e secundárias e abastecem zonas de influência independentes. Há características de planejamento e projeto, que interferem no bom funcionamento e aproveitamento global do sistema.

Por exemplo, o reservatório Pedro Ludovico, com volume de reservação de 10.000 m^3 e cota de 858,00 m, conta com uma área de influência limitada em função de estar localizado em local elevado. Com uma demanda máxima diária de 616 l/s, em sua zona de atendimento, o seu volume não é totalmente aproveitado. O mesmo ocorre com o reservatório Atlântico, com volume de 10.000 m^3 e a cota de 836,50 m, cujo sub-aproveitamento ocorre pelo fato da unidade ter sido construída num ponto de baixa elevação, restringindo assim a sua área de atendimento. A demanda máxima diária para esta unidade é de 474 l/s.

No caso do reservatório Itapoã, é evidente o sub-dimensionamento da unidade. A área de atendimento, com uma demanda máxima diária de 542 l/s, é muito densa para um volume de apenas 3.000 m^3 . O reservatório Amazônia, diferentemente dos demais,

apresenta uma relação compatível entre volume de reservação, cota e área de atendimento.

Em termos de avaliação global do sistema, ou seja, interligação entre os quatro reservatórios para um possível aproveitamento do excedente de volume de reservação, percebe-se que, pela localização das unidades, esta estratégia não é representativa para a melhoria da qualidade da operação. O reservatório que apresenta dificuldades no atendimento diário da demanda é o Itapoã. Esta unidade está localizada numa cota, 863,00 m, superior às cotas dos reservatórios com excesso de volume. Portanto, este aproveitamento somente seria possível através da implantação de mais uma estação elevatória intermediária.

Analisando-se os gráficos das figuras 6.3 e 6.4 percebe-se que estes aspectos relatados anteriormente são evidenciados pelos resultados. Os reservatórios Atlântico e Pedro Ludovico apresentam níveis de água próximos aos máximos durante quase todo o período de análise. O Amazônia apresenta uma variação de nível adequada em termos de uma operação eficiente do sistema, enquanto no Itapoã o nível de água se mantém próximo do mínimo durante o período diurno de pico, porém garantindo o atendimento das demandas na área de abastecimento.

Comparando as curvas de variação do nível de água nos reservatórios de Goiânia apresentadas através das figuras 6.3 e 6.4 e comparando-as com a curva típica de variação do nível no reservatório Amazônia, pode-se perceber que, mesmo tendo trajetórias diferenciadas, com exceção do reservatório Itapoã, os níveis mínimos estão acima do mínimo admitido para cada reservatório (1,5m), o que garante o suprimento dos nós de demanda com pressões adequadas.

Se forem comparadas apenas as curvas relativas ao reservatório Amazônia (figuras 6.3 e 6.4), fornecidas pelo modelo desenvolvido, com a curva típica de variação diária do nível de água no reservatório Amazônia (figura 6.8) fornecida pela área operacional da SANEAGO, há apenas um aspecto a ressaltar. No caso das soluções fornecidas pelo modelo, há um maior aproveitamento do volume do reservatório durante o período de 24 horas; o que, segundo os operadores do sistema, favorece, entre outros, a qualidade da água a ser distribuída por dificultar uma maior deposição e, conseqüente, concentração de cloro no reservatório.

Figura 6.5 – Goiânia – Pressões nos nós de demanda – solução AM1

Figura 6.6 – Goiânia – Pressões nos nós de demanda – solução AM2

As pressões nos nós de demanda, apresentadas através dos gráficos das figuras 6.5 e 6.6, mostram valores compatíveis com aquele definido como mínimo (15m).

Em termos do funcionamento das bombas, os resultados apresentados através da figura 6.7 mostraram que é possível manter um número razoável de bombas desligadas ao longo do dia, principalmente no horário de pico (18 h às 21 h), em que o valor da tarifa é duas vezes maior que fora deste horário. Neste período, foi possível reduzir em aproximadamente 60% o funcionamento das unidades.

Quando comparadas as regras do modelo, em termos de número de bombas em funcionamento ao longo do período de 24 horas (figura 6.7), com o funcionamento típico das bombas na estação de bombeamento SENAC (figura 6.9), percebe-se que o modelo fornece uma solução mais econômica, atendendo ao primeiro objetivo, que era uma economia em termos de custos de energia elétrica. Vale ressaltar que os dados relativos à figura 6.9 foram fornecidos pela SANEAGO e tiveram como base, relatórios de funcionamento diário da estação elevatória SENAC por um período de um ano (Junho-2002 a Julho-2003).

Figura 6.7 – Número de bombas em funcionamento – soluções AM1 e AM2

Para a solução AM1, apenas em dois períodos (12 hs e 16 hs) havia três bombas em funcionamento, enquanto que, para a solução AM2, isso ocorreu em apenas um período (16 hs). Segundo o funcionamento típico da elevatória SENAC, as três bombas funcionam durante um longo período, das 8 às 19 hs, ininterruptamente.

Como pode ser visto na tabela 6.6, as soluções AM1 e AM2 apresentaram uma economia de aproximadamente 40% nos custos de energia elétrica, quando comparadas com a solução típica de funcionamento das bombas.

Figura 6.8 – Curva típica de nível de água – reservatório Amazônia

Figura 6.9 – Estratégia de funcionamento típico das bombas na estação elevatória SENAC

Tabela 6.6 – Comparação de custos das soluções

Regra operacional	custos de energia elétrica(\$R/dia)
AM1	2494,3
AM2	2650,3
típica	4327,6

As figuras 6.10, 6.11, 6.12 e 6.13 apresentam o cronograma de funcionamento para cada bomba e válvula ao longo das 24 horas, para as regras operacionais AM1 e AM2. Os resultados demonstram que as duas soluções ou estratégias operacionais, fornecidas através da aplicação do método de aprendizado de máquina, podem ser caracterizadas como condição operacional “ótima” pois, conforme pressuposto no modelo, evitam o liga/desliga e abre/fecha de bombas e válvulas, respectivamente. Percebe-se, através das figuras, que é priorizada a continuidade de funcionamento de cada equipamento, como pressuposto pelo modelo através do índice de condição operacional (K_{pv}).

Figura 6.10 – Cronograma de funcionamento diário das bombas – solução AM1

Figura 6.11 – Cronograma de funcionamento diário das válvulas – solução AM1

Figura 6.12 – Cronograma de funcionamento das bombas – solução AM2

Figura 6.13 – Cronograma de funcionamento das válvulas - Solução AM2

Os resultados demonstraram a eficácia do modelo, tanto em termos de otimização, ou seja, do atendimento dos objetivos de menor custo de energia elétrica e maior benefício hidráulico, quanto em termos do algoritmo de aprendizado de máquina, por demonstrar que as estratégias operacionais selecionadas através do aprendizado apresentaram as melhores estratégias operacionais para o sistema em estudo.

Um importante e interessante aspecto da performance do sistema segundo as regras operacionais selecionadas, é que a capacidade total de bombeamento não foi utilizada durante o período de maiores demandas, porque nestes períodos estas foram atendidas, mesmo que parcialmente, pelos reservatórios de distribuição. Conseqüentemente, houve uma redução dos custos de energia elétrica durante os períodos em que as tarifas apresentam os maiores valores.

6.2 Resultados e análises utilizando processo da ordem de preferência (OP)

Implementando o otimizador *SPEA*, utilizando três funções objetivo conforme descrito no item 5.10, e utilizando os mesmos parâmetros definidos no item 5.9, foram obtidas as frentes Pareto, representadas nos espaços bi-dimensionais das funções objetivo, mostradas nas figuras 6.14, 6.15 e 6.16.

Com o objetivo de reduzir o número de soluções fornecidas pela frente Pareto, foi implementado o método de ordem de preferência. Implementando a rotina do OP de acordo com o teorema descrito no item 5.5 deste trabalho, sobre o conjunto de soluções

ótimas Pareto das figuras 6.14 a 6.16, obteve-se uma solução final no espaço bi-dimensional FO_1 x FO_2 e FO_1 x FO_3 . As figuras 6.17 e 6.18 mostram a solução selecionada pelo OP. É importante ressaltar que, segundo teorema, a melhor solução é dada por aquela cuja eficiência é de ordem 2 com grau 3 ou seja [2,3] – pontos ótimos Pareto porém, em nosso caso, esta solução não foi encontrada. Adotou-se, dessa forma, a solução com eficiência de ordem 2 e grau 2 , ou seja, [2,2]- pontos ótimos Pareto.

Figura 6.14 – Soluções FO_1 x FO_2 e frente Pareto -utilização do OP

Figura 6.15 – Soluções FO_1 x FO_3 e frente Pareto - utilização do OP

Figura 6.16 – Soluções FO_2 x FO_3 e frente Pareto - utilização do OP

Figura 6.17 – Estratégia operacional selecionada – espaço FO_1 x FO_2 - OP

Figura 6.18 – Estratégia operacional selecionada – espaço FO_1 x FO_3 - OP

Após a aplicação do método OP, e definição da estratégia operacional com melhor índice de eficiência, é necessária a verificação do comportamento hidráulico do sistema para esta estratégia. Os parâmetros hidráulicos considerados nos objetivos, quais sejam: níveis de água nos reservatórios, pressões nos nós de demanda e níveis de

funcionamento de bombas e válvulas devem ser avaliados e criticados para a verificação da eficácia do modelo proposto.

As figuras 6.19 e 6.20 apresentam, respectivamente, os níveis de água nos quatro reservatórios e as pressões nos quatro nós de demanda, para a solução OP. A figura 6.21 apresenta o funcionamento diário das bombas.

Figura 6.19 – Goiânia – Níveis de água nos reservatórios – solução OP

Avaliando os resultados da figura 6.19, percebe-se que mesmo com um custo total baixo, quando comparado com os das soluções do AM, os níveis de água nos reservatórios apresentam uma variação adequada à flutuação das demandas. Porém, no reservatório Itapoã, durante maior parte do período de análise, o nível tende a permanecer próximo do mínimo, o que não é conveniente operacionalmente. Nestas condições, o sistema fica sobrecarregado e não há garantia de suprimento dos consumidores atendidos pelo reservatório Itapoã, caso haja a necessidade de utilização da reserva de incêndio para outras eventualidades.

Figura 6.20 – Goiânia – Pressões nos nós de demanda – solução OP

Figura 6.21 – Número de bombas em funcionamento – Solução OP

As figuras 6.22 e 6.23 apresentam, respectivamente, as estratégias operacionais de bombas e válvulas da solução OP. Em termos de funcionamento das bombas, a solução apresenta uma disposição adequada ao índice de confiabilidade mecânica considerado, pois além de evitar o liga-desliga, como pode ser visto através da figura

6.22, dá a condição de utilização de uma das três bombas como reserva “quente”. Por apresentar um custo menor de energia elétrica, a solução OP tem menos bombas funcionando durante o período. Porém, isto acarreta uma incapacidade de abastecimento do reservatório Itapoã durante a maior parte do dia.

Avaliando os resultados da solução selecionada pelo método OP percebe-se que o mesmo prioriza as soluções com os melhores valores em termos dos objetivos custo (mínimo) e confiabilidade (máxima), em detrimento dos benefícios hidráulicos.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Tempo(hs)
																								Bomba 1
																								Bomba 2
																								Bomba 3

Legenda :

Figura 6.22 – Cronograma de funcionamento das bombas – solução OP

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Tempo(hs)
																								Valv 1
																								Valv 2
																								Valv 3
																								Valv 4

Legenda:

Figura 6.23 – Cronograma de funcionamento das válvulas – solução OP

Apesar da estratégia operacional selecionada pelo método de ordem de preferência apresentar um custo menor, de aproximadamente 25%, além de uma estratégia de funcionamento de bombas e válvulas bem mais adequadas do que aquelas fornecidas pelo método do aprendizado de máquina, considera-se que no conjunto de atendimento dos objetivos, as estratégias propostas pelo AM são melhores por garantirem o atendimento adequado dos consumidores durante as 24 horas do dia.

6.3 Resultados e análises considerando diferentes cenários de demandas

Como elucidado no início deste capítulo, os dados relativos às demandas, fornecidos pela SANEAGO, não são consistentes. Para avaliar a validade do modelo quando implementado a um sistema com comportamento de consumo diversificado, previram-se quatro cenários distintos para a demanda diária, imaginando-se atingir um intervalo de valores que cobrisse diversas possibilidades em termos de operação real do

sistema e, também, de planejamento em médio prazo em termos de expansão deste. O objetivo principal é o de avaliar a possibilidade de obtenção de uma regra operacional comum, fornecida pelo método otimizador e de aprendizado de máquina, que possa ser aplicada ao sistema sob diferentes cenários de demandas, sem prejuízo de atendimento dos objetivos previamente estabelecidos.

As figuras 6.24 a 6.27 apresentam as curvas de demanda consideradas para cada reservatório, de acordo com cada qual desses cenários. As curvas de demanda, utilizadas como referência, foram aquelas apresentadas na figura 5.6. Ressalta-se que, para aplicação do modelo desenvolvido e obtenção de uma regra operacional comum, foram utilizados cinco cenários, pois incluiu-se, no estudo, o cenário referência da figura 6.2, aqui chamado de cenário 2, cujas curvas não são mostradas novamente.

Figura 6.24 – Cenário de demanda 1 ($0.75 \times$ demandas de referência)

Figura 6.25 – Cenário de demanda 3 (1.2*demandas de referência)

Figura 6.26 – Cenário de demanda 4 (1.4*demandas referência)

Figura 6.27 – Cenário de demanda 5 (1.5*demandas de referência)

Na tabela 6.7, são descritos os conjuntos de dados utilizados na aplicação do algoritmo *See5* para cada cenário. Nas tabelas 6.8 a 6.11 são mostradas as matrizes de confusão para cada caso, para os quais são apresentadas as precisões totais a partir dos erros aparentes, demonstrando a validade das hipóteses. Na tabela 6.12 são apresentados alguns parâmetros relevantes para avaliação da hipótese. Não é apresentada a matriz de confusão para o cenário 2, pois esta já foi mostrada no item 6.1.

Tabela 6.7 – Descrição dos conjuntos de dados-AM

Cenário	# Ex	atr(discretos,contínuos)	classes (ótima, boa)
1	14	169 (1,168)	2 (42,9; 57,1)
3	73	169 (1,168)	2 (50,7; 49,3)
4	53	169 (1,168)	2 (35,8; 64,2)
5	101	169 (1,168)	2 (40,6;59,6)

Tabela 6.8 – Matriz de confusão – cenário 1

Classes	preditas ótimas	preditas boas	precisão da classe	precisão total
ótima	5	0	1,00	0,93
boa	1	8	0,89	

Tabela 6.9 – Matriz de confusão – cenário 3

Classes	preditas ótimas	preditas boas	precisão da classe	precisão total
ótima	33	2	0,94	0,92
boa	4	34	0,89	

Tabela 6.10 – Matriz de confusão – cenário 4

Classes	preditas ótimas	preditas boas	precisão da classe	precisão total
ótima	14	0	1,00	0,91
boa	5	34	0,87	

Tabela 6.11 – Matriz de confusão – cenário 5

Classes	preditas ótimas	preditas boas	precisão da classe	precisão total
ótima	37	3	0,92	0,93
boa	4	57	0,93	

Tabela 6.12 – Goiânia – resultados do See5 – condição operacional

Cenário	# Ex	ótima	boa	ErrA %	Err10cv %	#R	# MR
1	14	6	8	7	19± 0,1	2	1,8± 0,1
2	52	23	29	12	17± 0,1	4	3,9 ± 0,1
3	73	37	36	8	14± 0,1	5	4,7± 0,1
4	53	19	34	9	11± 0,1	4	3,8± 0,1
5	101	41	60	7	9± 0,1	6	5,7± 0,1

Após a implementação do algoritmo de otimização e *See5*, foram obtidas as frentes Pareto juntamente com a definição das regras operacionais boas e ótimas, além das regras selecionadas para os cenários 1, 3, 4 e 5. Tais resultados são apresentados nas figuras 6.28 a 6.31. A figura 6.32 apresenta as estratégias operacionais selecionadas para cada cenário.

Partindo das melhores estratégias selecionadas para cada cenário, aplicou-se novamente o algoritmo *See5* considerando, portanto, apenas as dez soluções

apresentadas na figura 6.32. As tabelas 6.13, 6.14 e 6.15 apresentam, respectivamente, a descrição dos conjuntos de dados, a matriz de confusão e parâmetros gerais para o caso. A estratégia final (EF) selecionada pelo *See5* e que será utilizada como padrão para avaliação dos parâmetros hidráulicos de cada cenário de demanda, coincidiu com uma das estratégias obtida para o cenário 2 e tem a seguinte característica: EF (2494,34; 130,1), conforme apresentada no item 6.1.

Figura 6.28 – Frente Pareto e melhores soluções AM – cenário de demanda 1

Figura 6.29 – Frente Pareto e melhores soluções AM – cenário de demanda 3

Figura 6.30 – Frente Pareto e melhores soluções AM – cenário de demanda 4

Figura 6.31 – Frente Pareto e melhores soluções AM – cenário de demanda 5

Figura 6.32 – Cenários de demanda – estratégias selecionadas pelo AM

Tabela 6.13 – Descrição do conjunto de dados – cenários de demandas-geral

# Ex	atr (discretos, contínuos)	classes (ótima, boa)
10	169 (1, 168)	2 (60%, 40%)

Tabela 6.14 – Matriz de confusão – cenários de demandas - geral

Classes	preditas ótimas	preditas boas	precisão da classe	precisão total
ótima	6	0	1,00	1,00
boa	0	4	1,00	

Tabela 6.15 – Goiânia – resultados do See5 – condição operacional

# Ex	ótima	boa	ErrA %	Err10cv %	#R	# MR
10	6	4	0	$2 \pm 0,1$	2	$1,9 \pm 0,1$

A estratégia operacional final (EF) foi utilizada para avaliação da variação dos níveis de água nos reservatórios. As figuras 6.33 a 6.36 apresentam estas curvas para os quatro cenários de demanda avaliados. Ressalta-se que a figura relativa ao cenário 2 não foi apresentada pois é a mesma mostrada na figura 6.3, item 6.1.

Figura 6.33 – Níveis de água nos reservatórios – estratégia geral – cenário 1

Figura 6.34 – Níveis de água nos reservatórios – estratégia geral – cenário 3

Figura 6.35 – Níveis de água nos reservatórios – estratégia geral – cenário 4

Figura 6.36 – Níveis de água nos reservatórios – estratégia geral – cenário 5

Analisando a figura 6.33 percebe-se que os níveis de água nos reservatórios são praticamente os mesmos apresentados pela figura 6.3. Da mesma forma que ocorre nesta última, o reservatório Itapoã apresenta, num pequeno intervalo de tempo, níveis de água próximos do mínimo, o que não compromete o atendimento dos consumidores pertencentes à sua área de influência. A variação dos níveis de água dos reservatórios para o cenário 3, apresentadas na figura 6.34, mostra que mesmo para o reservatório Itapoã, onde durante o intervalo de tempo entre 14 e 17 horas os níveis estão próximos do mínimo, o atendimento aos consumidores é apropriado. Para os demais reservatórios as variações apresentam-se dentro da normalidade.

Considerando os cenários 4 e 5 (figuras 6.35 e 6.36), os níveis de água nos reservatórios Amazônia e Itapoã apresentam valores próximos dos mínimos para um grande intervalo de tempo no período de análise. Esta situação se agrava no cenário 5, principalmente para o reservatório Itapoã, que mantém, praticamente durante todo o período de análise, níveis próximos do mínimo.

Desta avaliação, pode-se concluir que a estratégia operacional selecionada somente poderia ser aplicada ao sistema de distribuição de água em estudo, se houvesse um aumento de, no máximo, 20% nas demandas atualmente consideradas. Para demandas superiores a esta, outras estratégias operacionais deveriam ser avaliadas e selecionadas com o objetivo de atender aos objetivos previamente definidos.

6.4 Resultados e análises considerando níveis iniciais médios nos reservatórios

Como dito anteriormente, uma situação desejável, em termos de operação de sistemas de distribuição de água, é manter níveis de água próximos dos máximos nos reservatórios no final da noite (zero hora). Esta foi a consideração feita na análise do modelo e mostrada no item 6.1. Como esta não é uma situação comum no sistema de Goiânia, como pode ser constatado através da figura 6.8, com o objetivo de avaliar o modelo para esta nova situação, neste item são apresentados os resultados da implementação considerando níveis de água médios iniciais nos quatro reservatórios do sistema.

As tabelas 6.16, 6.17 e 6.18 apresentam, respectivamente, a descrição dos conjuntos de dados, a matriz de confusão e os parâmetros gerais para o caso. A figura 6.37 apresenta a frente Pareto com as regras operacionais boas e ótimas definidas pelo especialista, além das soluções selecionadas através do AM.

Tabela 6.16 – Descrição do conjunto de dados – níveis iniciais médios

# Ex	atr (discretos, contínuos)	classes (ótima, boa)
46	169 (1, 168)	2 (47,8%, 52,2%)

Tabela 6.17 – Matriz de confusão – níveis iniciais médios

Classes	preditas ótimas	preditas boas	precisão da classe	precisão total
ótima	20	2	0,91	0,93
boa	1	23	0,96	

Tabela 6.18 – Goiânia – resultados do See5 – condição operacional

# Ex	ótima	boa	ErrA %	Err10cv %	#R	# MR
46	22	25	7	$9 \pm 0,1$	4	$3,8 \pm 0,1$

Figura 6.37 – Frente Pareto – níveis iniciais médios nos reservatórios - soluções AM

Após a definição das estratégias selecionadas pelo AM, implementou-se o simulador para avaliação do comportamento do sistema para a regra operacional obtida. A figura 6.38 apresenta as curvas de variação dos níveis de água nos reservatórios. Na figura 6.39 podem ser vistas as variações horárias das pressões nos nós de demanda e na figura 6.40 é mostrada a estratégia de funcionamento das bombas para a situação estudada. Como os parâmetros para as duas soluções selecionadas são similares optou-se por apresentar apenas as características da solução S1 (2425,6; 123,2).

Figura 6.38 – Níveis de água nos reservatórios – níveis iniciais médios de água

Figura 6.39 – Pressões nos nós de demanda – níveis iniciais médios de água

Figura 6.40 – Número de bombas em funcionamento – níveis iniciais médios de água

Os resultados demonstram a eficácia do modelo proposto em termos dos dois objetivos previamente definidos. Os níveis de água nos reservatórios apresentam valores adequados assim como as pressões nos nós de demanda. A estratégia de funcionamento das bombas apresenta uma pequena economia nos custos com energia elétrica quando comparada com as estratégias selecionadas no item 6.1 e que são usadas como referência para comparações da qualidade dos resultados.

Comparando esta solução com as apresentadas no item 6.1 (figuras 6.3 e 6.4) percebe-se que, em termos de curvas de variação diária dos níveis de água nos reservatórios, ela apresenta uma característica bastante diferenciada. Nas figuras do item 6.1, é possível observar que para os reservatórios Amazônia e Itapoã houve um aproveitamento quase que total do volume disponível no período estudado. O mesmo não aconteceu com relação aos reservatórios Atlântico e Pedro Ludovico. No caso da solução considerando níveis iniciais médios iniciais esta situação não ocorre. Através da figura 6.38 é possível observar este aspecto, com exceção do reservatório Itapoã que tem um aproveitamento inferior a 50% do volume útil. Percebe-se que, mesmo com as soluções selecionadas no item 6.1 apresentando baixos níveis de água no reservatório Itapoã, num pequeno período de tempo, a estratégia se mostra mais apropriada por aproveitar todo o volume do reservatório no período.

Com relação à estratégia de funcionamento das bombas, mesmo sendo bastante similares, em termos de custos de energia elétrica, a estratégia selecionada utilizando os níveis iniciais médios tem como desvantagem o fato de contar com um maior número de bombas em funcionamento no período de pico.

Além dos aspectos mencionados anteriormente, comparando as frentes Pareto das soluções (figuras 6.2 e 6.37), é possível observar que as soluções que consideram níveis iniciais máximos apresentam maiores valores do benefício hidráulico (FO_2) com, praticamente, os mesmos valores dos custos de energia elétrica (FO_1).

6.5 Resultados e análises considerando demandas máximas constantes

É comum no gerenciamento de sistemas de abastecimento de água a falta de dados relativos a alguns parâmetros fundamentais para a análise da performance do mesmo. Um desses parâmetros é o comportamento da variação diária da demanda de água. Por falta de informação deste elemento, é comum, tanto no projeto quanto na operação, a consideração de uma demanda constante. Por temor de um subdimensionamento, o que pode acarretar a falta de abastecimento de determinados

usuários do sistema, é comum realizarem-se avaliações considerando uma demanda com valor superestimado.

O comportamento do modelo proposto para este tipo de situação, foi avaliado considerando demandas máximas constantes para a área de influência de cada reservatório, de acordo com a tabela 6.19.

Tabela 6.19 – demandas máximas constantes

Reservatórios	demandas máximas(l/s)
Pedro Ludovico	616
Amazônia	496
Atlântico	474
Itapoã	542

A descrição dos conjuntos de dados, a matriz de confusão e os parâmetros gerais para o caso, são apresentadas, respectivamente, nas tabelas 6.20, 6.21 e 6.22. A figura 6.41 apresenta a frente Pareto com as regras operacionais boas e ótimas definidas pelo especialista, além das soluções selecionadas através do AM.

Tabela 6.20 – Descrição do conjunto de dados – demandas constantes máximas

# Ex	atr (discretos, contínuos)	classes (ótima, boa)
28	169 (1, 168)	2 (46,4%, 53,6%)

Tabela 6.21 – Matriz de confusão – demandas constantes máximas

Classes	preditas ótimas	preditas boas	precisão da classe	precisão total
ótima	13	3	0,81	0,89
boa	0	12	1,00	

Tabela 6.22 – Goiânia – resultados do See5 – condição operacional

# Ex	ótima	boa	ErrA %	Err10cv %	#R	# MR
28	13	15	11	15 ± 0,1	3	2,8 ± 0,1

Figura 6.41 – Frente Pareto – demandas máximas constantes - soluções AM

Após a seleção das estratégias pelo AM, simulou-se o comportamento do sistema segundo a regra operacional obtida. A figura 6.42 apresenta as curvas de variação dos níveis de água nos reservatórios. Na figura 6.43 podem ser vistas as variações horárias das pressões nos nós de demanda e na figura 6.44 é mostrada a estratégia de funcionamento das bombas para a situação estudada. Como os parâmetros para as duas soluções selecionadas são praticamente similares, escolheu-se, para apresentação, a solução S1 (3642,3; 112,5).

Figura 6.42 – Níveis de água nos reservatórios – demandas máximas constantes

Figura 6.43 – Pressões nos nós de demanda – demandas máximas constantes

Como previsto, as estratégias selecionadas (figura 6.41) apresentam altos valores de custos de energia elétrica com baixos valores dos benefícios hidráulicos, quando comparadas com as selecionadas no item 6.1 (figura 6.2). Este é um importante aspecto, pois, mesmo com custos aproximadamente 40% maiores, o atendimento aos consumidores pode ser comprometido na área de influência do reservatório Itapoã, conforme pode ser constatado através da figura 6.42.

Figura 6.44 – Número de bombas em funcionamento – demandas máximas constantes

Observando a figura 6.44 é possível constatar como a adoção de uma demanda constante super estimada, na área de influência de cada reservatório de distribuição, pode causar prejuízos na operação do sistema. Mesmo com o funcionamento quase que ininterrupto de pelo menos duas bombas, e, em grande parte do período de análise, de três bombas, não foi possível atender convenientemente os benefícios hidráulicos.

Ressalta-se que é possível realizar várias simulações, adotando diferentes condições operacionais do sistema de distribuição de água, porém, entende-se que as cinco situações apresentadas neste capítulo cobrem uma boa parte dessas condições. Além disso, os resultados apresentados demonstraram a validade do modelo desenvolvido para várias situações que ocorrem com frequência, tanto em termos de operação quanto de projeto, no gerenciamento de sistemas de abastecimento de água potável. Para cobrir uma fatia maior de situações reais, além destas apresentadas, é imprescindível a obtenção de dados operacionais a partir da gerência do sistema ou através de uma pesquisa de campo específica.

CONCLUSÕES E ESTUDOS FUTUROS

One never notices what has been done. One can only see what remains to be done.

- Marie Curie

Neste trabalho foi desenvolvido um modelo computacional para otimização da operação de sistemas de macro distribuição de água potável considerando objetivos múltiplos. O modelo foi concebido considerando três módulos: o módulo de simulação hidráulica, o módulo de otimização multiobjetivo e o módulo de extração das melhores regras operacionais para o sistema. Com o objetivo de avaliar o algoritmo adotado para escolha das melhores estratégias, foi implementada ainda uma rotina que utiliza critérios de pontuação para definição de melhores soluções.

Os estudos iniciais para implementação do módulo de simulação hidráulica objetivaram o acoplamento do simulador hidráulico *EPANET2* ao compilador *Borland Builder C++ 5.0*. Esse acoplamento foi desenvolvido utilizando a biblioteca de funções *Toolkit*. Esse procedimento é necessário para que as rotinas do *EPANET2* sejam avaliadas dentro do compilador *C++*.

Implementado e testado o simulador hidráulico, numa segunda fase dos estudos foram avaliados os algoritmos de otimização multiobjetivo. Três algoritmos evolucionários foram investigados: um não elitista, *NSGA*, e dois elitistas, *SPEA* e *NSGA II*. Acoplados ao módulo simulador hidráulico, estes métodos de otimização foram testados com intuito de avaliar qual deles apresentava o melhor conjunto de soluções, quando comparados os objetivos previamente definidos.

Após a identificação do método de otimização multiobjetivo a utilizar, foi necessária a definição de um algoritmo para a escolha e seleção das melhores estratégias

de um conjunto fornecido pelo módulo otimizador. Várias técnicas de análise multicriterial são apresentadas pela literatura, em geral, considerando pesos para definição das melhores soluções. Neste trabalho, optou-se por utilizar um algoritmo de AM, ainda pouco usado na avaliação de problemas de sistemas de abastecimento de água.

Estabelecido o algoritmo de aprendizado de máquina como módulo para seleção das melhores estratégias operacionais a partir do conjunto de soluções fornecidas pelo otimizador, o modelo computacional para otimização operacional de sistemas de distribuição de água foi concluído. Após testá-lo, utilizando dados do sistema de macro distribuição de água de Goiânia, foi implementado, também, um método de análise multicriterial chamado ordem de preferência - OP (*preference ordering routine*). Para tal implementação, foi criada uma terceira função objetivo para o problema.

A investigação do modelo computacional desenvolvido neste trabalho, assim como a do estudo complementar com o método **OP**, apresentou vários resultados que possibilitaram conclusões interessantes sobre as ferramentas adotadas e sobre a aplicação do modelo proposto para o sistema de distribuição de água estudado. Algumas dessas conclusões são relatadas a seguir.

A literatura apresenta diversos trabalhos de aplicação de alguns algoritmos evolucionários multiobjetivo em problemas envolvendo sistemas de abastecimento de água. Porém, a aplicação dos métodos *SPEA*, *NSGA II*, *MOMGA*, dentre outros, só começou a ser difundida a partir de 1999 e, mesmo assim, não se tem notícia de trabalhos utilizando tais algoritmos em problemas de operação de sistemas de abastecimento de água.

A definição do melhor algoritmo, assim como de parâmetros como dimensão da população, número de gerações, definição e probabilidade dos operadores recombinação e mutação, é motivo de questionamento por parte dos pesquisadores. Com o objetivo de identificar o algoritmo mais eficiente para o problema em estudo, algumas verificações foram realizadas. Para tanto, foram realizadas análises comparativas entre três métodos, através da análise visual das frentes Pareto produzidas, baseada na uniformidade de distribuição das soluções. Os três algoritmos investigados, para determinação das 168 variáveis de decisão do problema de otimização correspondente, foram o *NSGA*, o *SPEA* e o *NSGA II*. Para a verificação, foram adotadas diferentes dimensões para população e números de gerações. 5000 gerações mostraram-se suficientes para estabelecer as frentes Pareto, para todos os três algoritmos utilizados, adotando-se a

população de soluções com dimensão de 300. Os resultados foram obtidos utilizando a biblioteca *MOMHLib* ++. As análises demonstraram um melhor desempenho do *SPEA*, mas com um bom desempenho, também do *NSGA II*. O *NSGA* não apresentou bons resultados para o caso estudado.

Definido o algoritmo *SPEA* para o problema em estudo, foram desenvolvidas análises comparativas para definição dos operadores genéticos de recombinação e de mutação. Foram investigados os operadores de recombinação BLX-a, linear e uniforme e três operadores mutação: aleatória, limite e não-uniforme. O operador recombinação com melhor desempenho foi o uniforme. Para o operador mutação, a aleatória não apresentou bons resultados e o desempenho entre a não-uniforme e limite foi praticamente similar. Neste trabalho adotou-se a mutação não-uniforme.

Depois de identificado o algoritmo de otimização multiobjetivo mais adequado e os operadores genéticos de recombinação e mutação, foi implementado o módulo do algoritmo de aprendizado de máquina, adotando-se o indutor *See5*. O algoritmo de aprendizado de máquina é uma ferramenta nova para este tipo de aplicação. Este algoritmo, normalmente é utilizado em arquivos com grande quantidade de dados analisados e classificados por especialistas, com o objetivo de rotular dados futuros ou dados ainda não rotulados do arquivo. No estudo proposto, utilizou-se o classificador para, a partir de um conjunto de soluções fornecidas pela frente Pareto, rotuladas pelo especialista através do índice de condição operacional (Kpv), encontrar as melhores estratégias operacionais dentro do conjunto definido como **operação ótima**. A adoção do índice Kpv para rotulação das soluções, quanto às condições operacionais, é uma das contribuições desse trabalho.

Definidos todos os módulos pertencentes ao modelo computacional proposto, foi possível aplicá-lo ao sistema de macro distribuição de Goiânia. Antes de comentar os resultados obtidos através dessa aplicação, alguns detalhes precisam ser relatados. O objetivo de trabalhar com o sistema de Goiânia era o de aplicar a um sistema existente e que pudesse, dessa forma, contar com informações atualizadas da operação das unidades consideradas no modelo, facilitando a comparação e verificação dos resultados obtidos. Entretanto, as informações relativas aos níveis diários de água nos reservatórios mostraram-se inconsistentes, com exceção do reservatório Amazônia. Os macro-medidores instalados nas saídas dos reservatórios Itapoã e Atlântico também apresentaram problemas e as curvas de demandas para estes reservatórios foram obtidas através de informações do boletim operacional de Goiânia, baseadas em demandas

máximas dentro da área de cada reservatório e coeficientes horários de demandas. Os cenários de demanda adotados foram sugeridos pela área de projeto da SANEAGO. Os dados de funcionamento das bombas da estação elevatória Senac apresentaram consistência e são típicos de dias em que não houve interrupção do funcionamento das bombas, devido a problemas de manutenção dos equipamentos. Devido a esta falta de informação, não foi possível calibrar o modelo que consideramos uma etapa fundamental para sistemas existentes.

Ressalta-se que, mesmo com as dificuldades decorrentes da deficiência dos dados, foi possível fazer uma comparação entre os resultados obtidos pelo modelo proposto e algumas variáveis operacionais do sistema. Essas comparações demonstraram a validade do modelo, tanto quando comparados os custos operacionais (objetivo de minimização), quanto os benefícios hidráulicos (objetivo de maximização).

Todas as simulações propostas, apresentadas no capítulo 6, mostraram resultados que comprovam a validade do modelo proposto para a otimização operacional de sistemas de distribuição de água utilizando dois objetivos. Por outro lado, os resultados obtidos através da aplicação do método de ordem de preferência apresentaram dois aspectos interessantes e que devem ser ressaltados. Considerando-se três objetivos, observou-se que houve uma prevalência dos objetivos 1(custos) e 3(confiabilidade mecânica) sobre o objetivo 2 (benefícios hidráulicos). Com relação à utilização do **OP** para seleção de melhores soluções da frente Pareto, para o caso em estudo, os resultados demonstraram inconsistência, pois as soluções selecionadas, apesar de apresentarem custos operacionais 40% menores que as soluções selecionadas pelo modelo proposto, não apresentaram valores adequados para os parâmetros adotados na avaliação do objetivo de benefício hidráulico. Dessa forma, o método da **OP** deve ser melhor avaliado, antes de ser adotado em substituição ao algoritmo de aprendizado de máquina, para problemas envolvendo operação de sistemas de abastecimento de água.

Consideram-se como maiores contribuições desse trabalho, a utilização do algoritmo evolucionário multiobjetivo *SPEA* ao problema de otimização operacional de um sistema de distribuição de água composto de estação de bombeamento, estruturas com válvulas de controle de água e grandes reservatórios; a utilização do algoritmo de aprendizado de máquina, um método ainda pouco explorado e usado para este fim, com o objetivo de selecionar as melhores soluções (estratégias) de um conjunto fornecido pelo algoritmo de otimização, que, pode-se considerar um uso pioneiro para problemas com tais características; e, por fim, a aplicação de um novo método denominado ordem

de preferência, ainda não utilizado por pesquisadores da área, para escolha das melhores soluções da frente Pareto em problemas com três ou mais objetivos. Os resultados apresentados com a aplicação do modelo ao sistema de Goiânia demonstram a consistência dessas contribuições.

Ressalta-se que, o primeiro passo para a implantação do modelo proposto a um sistema de distribuição de água existente, com o objetivo de auxiliar o operador do sistema, é muni-lo de dados operacionais das unidades componentes desse sistema. Um segundo passo seria a calibração do modelo, a partir de informações de campo, relativas às vazões abastecidas por cada reservatório e pressões simultâneas em diversos pontos de monitoramento da rede. Um terceiro passo seria a automatização da operação do sistema, além de um cadastro técnico geo-referenciado e a implantação de um sistema SCADA envolvendo todas as unidades do sistema. O desenvolvimento de um modelo para operação ótima envolvendo um sistema automático que, através do SCADA, forneça informações operacionais e, através dos CLPs, possa comandar o controle da operação à distância, é o objetivo futuro do autor desse trabalho.

O autor não poderia deixar de citar a importância, durante o desenvolvimento desse trabalho, do estágio feito, pelo período de 1 ano, no *Centre for Water Systems* da Universidade de Exeter, na Inglaterra. Neste centro, foi consultada grande parte do material referenciado na revisão bibliográfica desse trabalho, assim como desenvolvidos os estudos e a implementação dos módulos relativos ao aprendizado de máquina e ordem de preferência. Além disso, as constantes discussões sobre o modelo desenvolvido, com os Professores Godfrey Walters, Dragan Savic e outros pesquisadores do centro, serviram para a implementação de melhorias, principalmente com relação ao tempo de processamento computacional.

Considera-se que os objetivos desse trabalho foram cumpridos. Cabe ao autor, quando retornar a operar o sistema de abastecimento de água de Goiânia, viabilizar a coleta de informações consistentes sobre o sistema em estudo e ampliar a automatização do sistema para contemplar as unidades componentes do sub sistema João Leite. Com o modelo de otimização proposto e informações consistentes, é possível selecionar regras operacionais diárias para as unidades do sistema que, com uma operação automática, poderão ser aplicadas e avaliadas diariamente e, eventualmente, alteradas em função de variáveis não consideradas neste trabalho.

Com relação à qualidade do modelo proposto, um importante aspecto é a reavaliação das funções objetivo propostas. Neste caso, alguns critérios importantes

podem, no futuro ser incorporados, como por exemplo, a qualidade da água e a confiabilidade mecânica. É importante, também, uma melhor avaliação do índice de condição operacional (Kpv). Considerou-se, neste trabalho, os *status* de funcionamento de bombas e válvulas como única condição para definição desse índice. Acredita-se que este índice possa ser aprimorado, considerando a flexibilidade operacional do sistema. De qualquer forma, considerando os objetivos propostos pelo trabalho, acredita-se ter atingido a meta previamente proposta, que era a de desenvolver e testar um modelo de otimização e de extração de regras operacionais para um sistema de macro distribuição de água potável.

BIBLIOGRAFIA

It is only the limitations of the human mind that make the possible, impossible.

- Marc Drake

8.1 – Artigos apresentados pelo autor

CARRIJO, I.B.; CHEUNG, P.B.; SILVA, F.G; SOARES, A.K.; REIS, L.F.R. (2003). *Otimização da Operação de Sistemas de Distribuição de Água Utilizando Algoritmos Genéticos Multiobjetivo*. In: XV Simpósio Brasileiro de Recursos Hídricos. Curitiba, Brasil, Novembro. CD-Rom. 12p.

CARRIJO, I.B.; REIS, L.F.R. (2004). *Multiobjective Optimization in the Operation of a Water Distribution System using Elitist Evolutionary Algorithm*. In: EPSRC-Research Network Seminar-ACTUI2004- Decision Support in the Water Industry under Conditions of Uncertainty. Exeter, UK, Março. 12p.

CARRIJO, I.B.; REIS, L.F.R. (2004). *Operational Optimization of WDS using GA with Multiobjective Function and Operating Rules Extracted through Data Mining*. In: Simposio Internazionale di Ingegneria Sanitaria Ambientale-SIDISA2004. Taormina, Itália, Junho. CD-Rom. 9p.

CARRIJO, I.B.; REIS, L.F.R.; WALTERS, G.A.; SAVIC, D.A. (2004). *Operational Optimization of Water Distribution System using Multi-objective Evolutionary Algorithm and Operating Rules Extracted through Machine Learning*. In: Critical Transitions in Water and Environmental Resources Management-ASCE2004. Salt Lake City, USA, Julho. CD-Rom. 8p.

CARRIJO, I.B.; Reis, L.F.R. (2004). *Extração de Regras Operacionais Ótimas de Sistemas de Distribuição de Água através de Algoritmos Genéticos e Aprendizado de*

Máquina. Simpósio Hispano-Brasileiro de Redes de Distribuição de Água-SEREA2004. João Pessoa, Brasil, Novembro. CD-Rom. 12p.

8.2 Referências bibliográficas

ALMEIDA, R.; BARBOSA, P.S.F. (2002). *Operação Ótima de Sistemas de Distribuição de Água via Modelo de Programação Não Linear Inteira Mista*. In:Seminário-Planejamento, Projeto e Operação de Redes de Abastecimento de Água.O Estado da Arte e Questões Avançadas. João Pessoa. CD-Rom. 19 p.

ALMEIDA, R.; BARBOSA, P.S.F. (2002). *Efeito da Reservação Inicial sobre o Desempenho Operacional de Sistemas de Abastecimento de Água*. In:Seminário - Planejamento, Projeto e Operação de Redes de Abastecimento de Água. O Estado da Arte e Questões Avançadas. João Pessoa. CD-Rom. 15 p.

ALPEROVITS, E.; SHAMIR, U. (1977). *Design of Optimal Water Distribution System*. Water Resouces Research. Vol.13, n.6, p.885-900.

ANDERSON, E.J.; AL-JAMAL, K.H. (1995). *Hydraulic-Network Simplification*. Journal of Water Resources Planning and Management, vol. 121, n. 3, May/June, p. 235-240.

ANGEL, R.P.L.; HERNANDÉZ, R.J.A.; AGUDELO, R.J.J. (1999). *Fuzzy Expert System Model for The Operation of an Urban Water Supply System*. Water Industry Systems:modeling and optimization applications. Volume 1. Research Studies Press Ltd.Exeter, England, p. 449-457.

BABOVIC, V.; DRÉCOURT,J.P.; KEIJZER,M.; HANSEN, P.F. (2002). *A Data Mining Approach to Modelling of Water Supply Assets*. Urban Water, vol.4, p. 401-414.

BACK, T. (1996). *Evolutionary Algorithms in Theory and Practice*. New York. Oxford university Press.

BANZHAF, W.; NORDIN, P.; KELLER, R.; FRANCONI, F.D. (1997). *Genetic Programming – An Introduction*. Morgan Kaufmann Publishers, Inc.San Francisco, CA, USA, 451 p.

BAO, Y.; MAYS, L.W. (1990). *Model for Water Distribution System Reliability*. Journal of Hydraulic Engineering, v.116, n.9, p.1119-1137.

BARANAUSKAS, J.A. (2001). *Extração Automática de Conhecimento por Múltiplos Indutores*. São Carlos, SP. 178p. Tese(Doutorado)-Instituto de Ciências Matemáticas e de Computação-ICMC-USP.

- BARCELLOS, J.C.H. (2000). *Algoritmos Genéticos Adaptativos:Um Estudo Comparativo*. São Paulo. 184p. Tese(Mestrado)-Escola Politécnica da Universidade de São Paulo.
- BESSLER, F.T. (1998). *Linear Network Optimisation and Identification of Control Rules-Develop a General Operating Policy for a Water Supply System*. Exeter, U.K.. 134p. Dissertation-Master of Engineering. Centre for Water Systems, School of Engineering. University of Exeter, UK.
- BESSLER, F.T.;WALTERS, G.A.; SAVIC, D.A. (2001). *Water System Reliability Predictions Using Data Mining*. Meeting The World's Water and Environmental Resources Challenges-State of The Pactice. Orlando, Florida,US. EWRI-ASCE. 9p.
- BESSLER, F.T.; WALTERS, G.A.; SAVIC, D.A. (2003). *Water Reservoir Control with Data Mining*. Journal of Water Resources Planning and Management, v.129, n.1, January, p.26-34.
- BHATTACHARYA, B.; LOBBRECHT, A.H.; SOLOMATINE, D.P. (2003). *Neural Networks and Reinforcement Learning in Control of Water Systems*. Journal of Water Resources Planning and Management, v.129, n.6, november, p.458-465.
- BHAVE, P.R. (1988). *Extended Period Simulation of Water Systems-Direct Solution*. Journal of Environmental Engineering, v.114, n.5, october, p.1146-1159.
- BLAKE, C.L.; MERZ, C.J. (1998). *UCI Repository of Machine Learning Databases*. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- BRION, L.M.; MAYS, L.W. (1991). *Methodology for Optimal Operation of Pumping Stations in Water Distribution Systems*. Journal of Water Resources Planning and Management, v.117, n.11, november, p.1551-1569.
- BRUNONE, B.; MORELLI, L. (1999). *Automatic Control Valve-Induced Transients in Operative Pipe System*. Journal of Hydraulic Engineering, vol. 125, n. 5, May, p. 534-542.
- BURROWS, R.; CROWDER, G.S.; ZHANG, J. (1999). *Application of Network Modelling for Operational Management of Water Distribution Systems*. Water Industry Systems:modeling and optimization applications. Volume 1. Research Studies Press Ltd. Exeter, England, p. 229-241.
- BUSH, C.A.; UBER, J.G. (1998). *Sampling Methods for Water Distribution Model Calibration*. Journal of Water Resources Planning and Management, v.124, n.6, november, p.334-344.

- CALIMAN, R.O. ; REIS, L.F.R. ; SILVA, F.G.B.; PORTO, R.M.; CHAUDHRY, F.H. (2001). *Determinação de Parâmetros do Modelo Pressão x Vazamento para um Sub-Sector da Cidade de São Carlos-S.P., com a Utilização de Algoritmos Genéticos*. In:XIV Simpósio Brasileiro de Recursos Hídricos. Aracaju, Se, Brasil, 15p.
- CARRIJO, I.B. (1989). *Otimização de Redes de Distribuição de Água com Expansão: Aplicação à Cidade de Goiânia*. São Carlos. 238p. Dissertação (Mestrado). Escola de Engenharia de São Carlos. Universidade de São Paulo, Brasil.
- CARUANA, R.A.; SCHAFFER, J.D. (1988). *Representation and Hidden Bias: Gray vs. Binary Coding for Genetic Algorithms*. Fifth International Conference on Machine Learning. Proceedings. University of Michigan. Ann Arbor.
- CHEESEMAN, P; STUTZ, J. (1990). *Bayesian Classification (Autoclass): Theory and Results Advances in Knowledge Discovery and Data Mining*. <http://ic.arc.nasa.gov/ic/projects/bayes-group/autoclass-c-program.html>.
- CHEUNG, P.B; SOUZA, R.S (2001). *Influência do Zoneamento de uma Rede Hidráulica na Acurácia de um Modelo de Calibração*. In:21º Congresso Brasileiro de Engenharia Sanitária e Ambiental. João Pessoa-PB, Brasil, 15p.
- CHEUNG, P.B.; REIS, L.F.R.; FORMIGA, K.T.M.; CHAUDHRY, F.H.; TICONA, W.G.C. (2003). *Multiobjective Evolutionary Algorithms Applied to the Rehabilitation of a Water Distribution System: A Comparative Study*. Evolutionary Multi-Criterion Optimization. Second International Conference, EMO-2003. Faro, Portugal, April. Proceedings. Springer. New York, p. 662-676.
- CHEUNG, P.B.; REIS, L.F.R.; CARRIJO, I.B. (2003). *Multiobjective Optimization to the Rehabilitation of a Water Distribution Network*. In Proceedings of the International Conference on Computing and Control for the Water Industry. Londres, UK, Setembro, p. 318-328.
- CHEUNG, P.B.(2004). *Análise de Reabilitação de Redes de Distribuição de Água de Abastecimento via Algoritmos Genéticos Multiobjetivo*. São Carlos - S.P. 268p. Tese(Doutorado)-Escola de Engenharia de São Carlos da Universidade de São Paulo.
- COELLO, C.A.C. (1996). *An Empirical Study of Evolutionary Techniques for Multiobjective Optimization in Engineering Design*. 420p. Tese de Doutorado. Tulane University, USA.
- COLOMBO, A.F.; KARNEY, B.W. (2002). *Energy and Costs of Leaky Pipes: Toward Comprehensive Picture*. Journal of Water Resources Planning and Management, vol. 128, n. 6, November, p. 441-450.

- COOMBS, S.; DAVIS, L. (1987). *Genetic Algorithms and Communication Link Speed Design: Constraints and Operators*. Genetic Algorithms and their Applications: Proceedings. Second International Conference on Genetic Algorithms. J.J. Grefenstett. Lawrence Erlbaum, Hillsdale, New Jersey, p. 257-260.
- CORNE, D.; KNOWLES, J.; OATES, M. (2000). *The Pareto Envelope-Based Selection Algorithm for Multi-Objective Optimization*. In Proceedings of The Sixth International Conference on Parallel Problem Solving from Nature VI (PPSN-VI), p. 839-848.
- CUNHA, M.C; SOUSA, J. (1999). *Water Distribution Network Design Optimization: Simulated Annealing Approach*. Journal of Water Resources Planning and Management, v.125, n.4, july, p. 215-221.
- DANDY, G.C.; SIMPSON, A.R.; MURPHY, L.J. (1996). *An Improved Genetic Algorithm for Pipe Network Optimization*. Water Resources Research, v.32, No 2, February, p. 449-458.
- DANDY, G.C.; SIMPSON, A.R.; MURPHY, L.J. (1997). *Comparison of Methods for Yield Assessment of Multiple Reservoir Systems*. Journal of Water Resources Planning and Management, v.123, n.6, November, p. 350-358.
- DAS, I. (1999). *A Preference Ordering Among Various Pareto Optimal Alternatives*. Structural Optimization, vol. 18 (1), p. 30-35.
- DAVIS, A.L; JEPPSON, R.W. (1979). *Developing a Computer Program for Distribution System Analysis*. Journal of The American Water Works Association, may, p. 236-241.
- DAVIS, L. (1991). *Handbook of Genetic Algorithms*. New York . Van Nostrand Reinhold.
- DEB, K.; GOYAL, M (1996). *A Combined Genetic Adaptive Search(GeneAS) for Engineering Design*. Computer Science and Informatics, 26(4), p. 30-45.
- DEB, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley&Sons, Ltd.Chichester, England, 491p.
- DEB, K. et al. (2002). *A Fast and Elitist Multiobjective Genetic Algorithm: NSGA II*. IEEE Transactions on Evolutionary Computation. 6(2), April, p. 182-197.
- DEJONG, K.A. (1975). *An Analysis of The Behavior of a Class of Genetic Adaptive Systems*. PhD Thesis, Ann Arbor, MI: University of Michigan.
- ERICKSON, M.; MAYER, A.; HORN, J. (2001). *The Niched Pareto Genetic Algorithm 2 Applied to The Design of Groundwater Remediation Systems*. In Proceedings of The

First International Conference on Evolutionary Multi-Criterion Optimization (EMO-2001), p. 265-283.

ERICSSON, A.A. (1989). *Controle Automático de Sistemas Urbanos de Distribuição de Água*. In: 15º Congresso Brasileiro de Engenharia Sanitária e Ambiental, Belém. Anais, v.2, Tomo II, p. 506-521.

ESHELMAN, L.J.; SCHAFFER, J.D. (1993). *Real-Coded Genetic Algorithms and Interval-Schemata*. In Foundations of Genetic Algorithms 2 (FOGA 2), p. 187-202.

EUSUFF, M. M.; LANSEY, K. E. (2003). *Optimization of Water Distribution Network Design Using the Shuffled Frog Leaping Algorithm*. Journal of Water Resources Planning and Management, v.129, n.3, May, p. 210-225.

FOGEL, D.B; FOGEL, L.J.; ATMAR, W.; FOGEL, G.B. (1992). *Hierarchic Methods in Evolutionary Programming*. In Proceedings of The First Annual Conference on Evolutionary Programming, p. 175-182.

FONSECA, C.M.; FLEMING, P.J. (1993). *Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization*. In: Fifth Internacional Conference, San Mateo, CA, 8p.

FONSECA, C.M.; FLEMING, P.J. (1995). *An Overview of Evolutionary Algorithms in Multi-Objective Optimization*. Evolutionary Computation Journal, 3(1), p. 1-16.

FORMIGA, K.T.M; CHAUDHRY, F.H. (2002). *Utilização de Algoritmos Evolucionários Multiobjetivos no Dimensionamento de Redes de Abastecimento de Água*. In: Seminário-Planejamento, Projeto e Operação de Redes de Abastecimento de Água. O Estado da Arte e Questões Avançadas. João Pessoa. CD-Rom. 15p.

FORMIGA, K.T.M.; CHAUDHRY, F.H. ; CHEUNG, P.B. ; REIS, L.F.R. (2003). *Optimal Design of Water Distribution System by Multiobjective Evolutionary Methods. Evolutionary Multi-Criterion Optimization*. Second International Conference, EMO-2003. Faro, Portugal, April. Proceedings. Springer. New York, p. 677-691.

FRANCATO, A.L. (2002). *Operação Multiobjetivo para a Operação de Sistemas Urbanos de Abastecimento de Água*. Tese de Doutorado. Universidade de Campinas(UNICAMP), Brasil.

FREITAS, M.S.H. (1989). *Gerenciamento da Distribuição de Água- Organização da Estrutura Operacional por Objetivos*. In: 15º Congresso Brasileiro de Engenharia Sanitária e Ambiental, Belém. Anais, v.2, Tomo II, p. 443-458.

- GALVÃO et. al (1999). *Sistemas Inteligentes-Aplicações a Recursos Hídricos e Ciências Ambientais*. Editora da Universidade. Universidade Federal do Rio Grande do Sul, Porto Alegre, 246p.
- GARGANO, R.; PIANESE, D. (2000). *Reliability as Tool for Hydraulic Network Planning*. Journal of Hydraulic Engineering, v.126, n.5, May, p.354-364.
- GEN, M; CHENG, R. (1997). *Genetic Algorithms and Engineering Design*. Canada. John Wiley e Son.
- GOLDBERG, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Reading, Massachusetts, 403p.
- GOLDBERG, D.E.; KUO, C. (1987). *Genetic Algorithms in Pipeline Optimization*. Journal of Computing in Civil Engineering, 1(20), p. 128-141.
- HALHAL, D.; WALTERS, G.A.; OUAZAR, D.; SAVIC, D.A. (1997). *Water Network Rehabilitation with Structured Messy Genetic Algorithm*. Journal of Water Resources Planning and Management, v.123, n.3, May/June, p. 137-146.
- HAMBERG, D.; SHAMIR, U. (1988). *Schematic Models for Distribution System Design*. Journal of Water Resources Planning and Management, v.114, n.2, March, p. 129-140.
- HANSEN, N.; OSTERMEIER, A. (1996). *Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation*. In Proceedings of The IEEE International Conference on Evolutionary Computation, p. 312-317.
- HAUPT, R.L.; HAUPT, S.E. (1998). *Practical Genetic Algorithms*. Wiley-Interscience. New York.
- HERNANDÉZ, V.; MARTÍNEZ, F.; VIDAL, A.M.; ALONSO, J.M.; ALVARRUIZ, F.; GUERRERO, D.; RUIZ, P.A.; VERCHER, J. (1999). *HIPERWATER: A High Performance Computing EPANET-Based Demonstrator for Water Network Simulation and Leakage Minimisation*. Water Industry Systems: modeling and optimization applications. Volume 1. Research Studies Press Ltd. Exeter, England, p. 141-153.
- HOLLAND, J.H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MIT Press.
- HORN, J.; NAFPLOITIS, N.; GOLDBERG, D. (1994). *A Niche Pareto Genetic Algorithm for Multi-Objective Optimization*. In: Proceedings of The First IEEE Conference on Evolutionary Computation, p. 82-87.

- HORST, P.S.; MONARD, M.C. (2000). *Um Sistema Computacional para Avaliação de Regras Induzidas por Algoritmos de Aprendizado de Máquina*. Proceedings of the IBERAMIA/SBIA, Atibaia, S.P, p. 167-176.
- JASZKIEWICZ, A. (1999). *Multiobjective Methods Metaheuristic Library for C++*. <http://www-idss.cs.put.poznan.pl/~jaszkiewicz/>.
- JOWITT, P.W; XU, C. (1990). *Optimal Valve Control in Water-Distribution Networks*. Journal of Water Resources Planning and Management, v.116, n.4, July, p. 455-472.
- JOWITT, P.W; GERMANOPOULOS, G. (1992). *Optimal Pump Scheduling in Water-Supply Networks*. Journal of Water Resources Planning and Management, v.118, n.4, July, p. 416-422.
- KADAR, Y; BARIUDIN V. (2001). *Optimization Model for the Design of Water Supply Network*. Meeting The World's Water and Environmental Resources Challenges-State of The Pactice-Orlando, Florida, US.EWRI-ASCE, 10p.
- KAPELAN, Z.S.; SAVIC, D.A.; WALTERS, G.A. (2002). *Hybrid GA for Calibration of Water Distribution Hydraulic Models*. 2002 Conference on Water Resources Planning and Management. Roanoke, VA, USA. May. Proceedings, 11p.
- KAPELAN, Z.S.; SAVIC, D.A.; WALTERS, G.A. (2003). *Multiobjective Sampling Design for Water Distribution Model Calibration*. Journal of Water Resources Planning and Management, v.129, n.6, November, p. 466-479.
- KAZANTZIS, M.D. et al. (2002). *A New Methodology for Optimizing the Daily Operations of a Pumping Plant*. 2002 Conference on Water Resources Planning and Management. Roanoke, VA, USA. May. Proceedings, 10p.
- KLEMOUS, R.; KOTOWSKI, J.; NIKODEM, J.; ULASIEWICZ, J. (1997). *Optimization Algorithms of Operative Control in Water Distribution Systems*. Journal of Computational and Applied Mathematics. N. 84, p. 81-99.
- KNOWLES, J.D.; CORNE, D.W. (2000). *Approximating The Non-Dominated Front Using The Pareto Archived Evolution strategy*. Evolutionary Computation Journal, 8(2), p. 149-172.
- KO, S.K.; OH, M.H.; FONTANE, D.G. (1997). *Multiobjective Analysis of Service-Water-Transmission Systems*. Journal of Water Resources Planning and Management, v.123, n.2, March, p.79-83.
- KOELE, E. (1992). *Control Valves Inducing Oscillatory Flow in Hydraulic Networks*. In: The International Conference on Unsteady Flow and Fluid Transients, Durham, United Kingdom. Proceedings, v.1, 15p.

- KOHAVI, R.; SOMMERFIELD, D.; DOUGHERTY, J. (1996). *Data Mining Using MLC++: A Machine Learning Library in C++*. In Tools with Artificial Intelligence. IEEE Compute Society Press, p. 234-245
- KONONENKO, I. (1993). *Inductive and Bayesian Learning in Medical Diagnosis*. Applied Artificial Intelligence, 7, p. 317-337.
- KRAVOSAC, A.C.B. et al. (1989). *Automação de Sistemas : Aplicação de CLP no Controle de Estação Elevatória de Água*. In: 15º Congresso Brasileiro de Engenharia Sanitária e Ambiental, Belém. Anais, v.2, Tomo II, p. 183-196.
- KRITPIPHAT, W. et al. (1998). *Pipeline Network Modelling and Simulation for Intelligent Monitoring and Control: a Case Study of a Municipal Water Supply System*. Ind. Eng. Chem. Res, v.37, p. 1033-1044.
- KUBAT, M; HOLTE, R.C.; MATWIN, S. (1998). *Machine Learning for Detection of Oil Spills in Satellite Radar Images*. Machine Learning 30 (2/3), p. 195-215.
- LANSEY, K.E.; AWUMAH, K. (1994). *Optimal Pump Operations Considering Pump Switches*. Journal of Water Resources Planning and Management, v.120, n.1, January, p. 17-34.
- LÉON, C.; MARTÍN, S.; ELENA, J.M. ; LUQUE, J. (2000). *Explore-Hybrid Expert System for Water Networks Management*. Journal of Water Resources Planning and Management, v.126, n.2, March, p. 65-74.
- LINGIREDDY, S.; WOOD, J.D. (1998). *Improved Operation of Water Distribution Systems Using Variable-Speed Pumps*. Journal of Energy Engineering. Vol. 124, n.3, December, p. 91-103.
- LUVIZOTTO JÚNIOR, E. (1995). *Controle Operacional de Redes de Abastecimento de Água Auxiliado por Computador*. São Paulo. 143p. Tese(Doutorado)-Escola Politécnica da Universidade de São Paulo.
- MAKSIMOVIC, C.; CARMI, N. (1999). *Gis Supported Analysis of Pressure Dependant Vulnerability of Distribution Networks to Leakage*. Water Industry Systems: modeling and optimization applications. Volume 2. Research Studies Press Ltd. Exeter, England, p. 85-96.
- MAIER, H.R.; SIMPSON, A.R.; ZECCHIN, A.C.; FOONG, W.K.; PHANG, K.Y.; SEAH, H.Y.; TAN, C.L. (2003). *Ant Colony Optimization for Design of Water Distribution Systems*. Journal of Water Resources Planning and Management, v.129, n.3, May, p. 200-209.

- MARTÍNEZ, F. et al. (2003). *Calibration of Valve-Driven Water Distribution System using jointly GIS and Genetic Algorithms. Application to the WDS of Valencia*. The International Conference on Computing and Control for The Water Industry. 2003. London, UK, September, 17p.
- MARTINS, C. A. (2003). *Uma Abordagem para Pré-processamento de Dados Textuais em Algoritmos de Aprendizado*. São Carlos-S.P. 184p. Tese (Doutorado)- Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, ICMC-USP.
- MAYS, L.W. (1989). *Reliability Analysis of Water Distribution Systems*. American Society of Civil Engineers. New York . 510p.
- MAYS, L.W.; YEOU-KOUNG, T. (1992). *Hydrosystems Enginnering and Management* . MC Graw Hill. USA, 544p.
- McCORMICK, G.; POWELL, R.S. (2003). *Optimal Pump Scheduling in Water Supply Systems with Maximum Demand Charges*. Journal of Water Resources Planning and Management, v.129, n.5, September, p. 372-379.
- MEDAWAR, P; MEDAWAR, J. (1978). *Ciência da Vida*. Rio de Janeiro. Zahar Editores.
- MICHALSKI, R.S.; BRATKO, I.; KUBAT, M. (1998). *Machine Learning and Data Mining Methods and Applications*. West Sussex, England : John Wiley & Sons Ltd.
- MICHALEWICZ, Z. (1992). *Genetic Algorithms + Data Structures=Evolution Programs*. Spring-Verlag. New York.
- MICHALEWICZ, Z.; ATTIA, N. (1994). *Evolutionary Optimization of Constrained Problems*. In Proceedings of The Third Annual Conference on Evolutionary Programming, p. 98-108.
- MITCHELL, T.M. (1998). *Machine Learning*. McGraw Hill.
- MORGAN, D.R.; GOULTER, I.C. (1985). *Optimal Urban Water Distribution Design*. Water Resouces Research, v.21, n.5, may, p. 642-652.
- MURPHY, L.J.; SIMPSON, A.R. (1992). *Pipe Optimisation using Genetic Algorithms*. Research Report No R 93. Dept.of Civil and Environmental Engineering. The University of Adelaide, Australia, 43 p.
- NITIVATTANANON, V.; SADOWSKI, E.C.; QUIMPO, R.G. (1996). *Optimization of Water Supply System Operation*. Journal of Water Resources Planning and Management, v.122, n.5, September/October, p. 374-384.
- OBRADOVIC, D.; LONSDALE, P. (1998). *Public Water Supply – Models, Data and Operational Management*. St Edmundsbury Press, Suffolk, UK. 453p.

- OBRADOVIC, D. (2000). *Modelling of Demand and Losses in Real-Life Water Distribution Systems*. UrbanWater, vol. 2, p. 131-139.
- OLIVEIRA, R.; LOUCKS, D.P. (1997). *Operating Rules for Multireservoir Systems*. Water Research Resources. V.33, n.4, April, p. 839-852.
- ONIZUKA, K. (1986). *Systems Dynamics Approach to Pipe Network Analysis*. Journal of Hydraulic Engineering. ASCE. Vol 112, n.8, August, p. 728-749.
- ORMSBEE, L.E.; WOOD, D.J. (1986). *Explicit Pipe Network Calibration*. Journal of Water Resources Planning and Management, v.112, n.2, april, p. 166-182.
- ORMSBEE, L.E.; WALSKI, T.M.; CHASE, D.V.; SHARP, W.W. (1989). *Methodology for Improving Pump Operation Efficiency*. Journal of Water Resources Planning and Management, v.115, n.2, March, p. 148-164.
- ORMSBEE, L.; KESSLER, A. (1990). *Optimal Upgrading of Hydraulic-Network Reliability*. Journal of Water Resources Planning and Management, v.116, n.6, November, p. 784-802.
- ORMSBEE, L.E.; LANSEY, K.E. (1994). *Optimal Control of Water Supply Pumping Systems*. Journal of Water Resources Planning and Management, v.120, n.2, March, p. 237-252.
- ORMSBEE, L.E.; REDDY, S.L. (1995). *Nonlinear Heuristic for Pump Operations*. Journal of Water Resources Planning and Management, v.121, n.4, July, p. 302-309.
- OSTFELD, A.; KOGAN, D.; SHAMIR, U. (2002). *Reliability Simulation of Water Distribution Systems – Single and Multiquality*. Urban Water, 4(2002, p. 53-61.
- PESSIS-PASTERNAK, G. (1993). *Do Caos à Inteligência Artificial*. São Paulo. Editora da Universidade Estadual Paulista.
- PEZESHK, S.; HELWEG, O.J. ; OLIVER, K.E. (1994). *Optimal Operation of Ground-Water Supply Distribution Systems*. Journal of Water Resources Planning and Management, v.120, n.5, Sep./Oct., p. 573-586.
- PEZESHK, S.; HELWEG, O.J (1996). *Adaptive Search Optimization in Reducing Pump Operating Costs*. Journal of Water Resources Planning and Management, v.122, n.1, Jan./Feb., p. 57-63.
- POLL, H.G. (1999). *Válvulas de Controle Acopladas a Redes Hidráulicas*. São Paulo. Tese de Doutorado. Escola Politécnica da Universidade de São Paulo.
- PRASAD, T.D.; PARK, N. (2004). *Multiobjective Genetic Algorithms for Design of Water Distribution Networks*. Journal of Water Resources Planning and Management, v.130, n.1, January, p. 73-82.

- PRESCOTT, S.L.;ULANICKI, B. (2003). *Dynamic Model of Pressure Reducing Valves*. Journal of Hydraulic Engineering, v. 129, n. 10, October, p. 804-812.
- QUINDRY, G.E.; BRILL, E.D.; LIEBMAN, J.C. (1981). *Optimization of Looped Water Distribution Systems*. Journal of The Environmental Engineering Division, v.107, n.EE4, August, p. 665-679.
- QUINLAN, J.R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers. San Francisco, CA.
- RAHAL, H. (1994). *Water Distribution Network Steady State Simulation AQUANET Model*. In:International Conference on Hydraulic Engineering Software, 5, Greece. Proceedings. V.1, p. 395-402.
- RAO, H.S.; BREE, W.D. (1977). *Extended Period Simulation of Water Systems-Part A*. Journal of The Hydraulics Division, v.103, n.HY2, February, p. 97-108.
- REIS, L.F.R.; PORTO, R.M.; CHAUDHRY, F.H. (1997). *Optimal Location of Control Valves In Pipe Networks By Genetic Algorithm*. Journal of Water Resources Planning and Management, v.123, n.6, November, p. 317-326.
- REIS, L.F.R.; AKUTSU, J. (2002). *Estratégias Operacionais para Sistemas de Reservatórios via Algoritmo Genético(AGs)*. Revista Brasileira de Recursos Hídricos.Volume 7, número 3 , Jul./Set., p. 5-17.
- REZENDE, S.O.; PUGLIESI, J.B. (1998). *Aquisição de Conhecimento Explícito ou Manual*. Technical Report 37, ICMC-USP.
- RIBEIRO, C.R. (1985). *Controle de Escoamentos em Instalações de Conduitos Forçados*. São Paulo. 182p.Tese de Doutorado. Escola Politécnica da Universidade de São Paulo.
- RIBEIRO, F.J.L et al. (1994). *Genetic Algorithm Programming Environments*. Computer Science.
- RIGHETTO, A.M. (2001). *Calibração de Modelo Hidráulico de Rede de Distribuição*. Revista ABRH, v.6, n.3, p.33-44, Setembro.
- RIGHETTO, A.M. (2002). *Operação Ótima de Sistema Urbano de Distribuição de Água*. In:Seminário-Planejamento, Projeto e Operação de Redes de Abastecimento de Água.O Estado da Arte e Questões Avançadas.João Pessoa.CD-Rom. 16p.
- RITZEL, B.J.; EHEART, J.W. (1994). *Using Genetic Algorithms to Solve a Multiple Objective Groundwater Pollution Containment Problem*. Water Resources Research. Vol.30, n.5, May, p. 1589-1603.

- ROSSMAN, L.A. (2001). *EPANET2-Users Manual*. U.S. Environmental Protection Agency. Cincinnati, Ohio.
- RUDOLPH, G. (1998). *On a Multi-Objective Evolutionary Algorithm and its Convergence to The Pareto Set*. In: Proceedings of The 5th IEEE Conference on Evolutionary Computation, p. 511-516.
- RUSSEL, S.; NORVIG, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall.
- SAKARYA, A.B.A.; MAYS, L.W. (2000). *Optimal Operation of Water Distribution Pumps considering Water Quality*. Journal of Water Resources Planning and Management, v.126, n.4, July/August, p. 210-220.
- SANEAMENTO DE GOIÁS S.A.-SANEAGO (2004). Boletim de Informações Operacionais (BIO). Março-Maio-2004. 203p..
- SAVIC, D.A.; WALTERS, G.A. (1997a). *Genetic Algorithms for Least-Cost Design of Water Distribution Networks*. Journal of Water Resources Planning and Management, v.123, n.2, March, p. 67-77.
- SAVIC, D.A.; WALTERS, G.A. (1997b). *Evolving Sustainable Water Networks*. Hydrological Sciences Journal, v.42, No 4, August, p. 549-564.
- SCHINDLER D.F.; GARRARD, T.P. (1999). *GIS as an Active Management Tool in Water Distribution*. Water Industry Systems: modeling and optimization applications. Volume 2. Research Studies Press Ltd. Exeter, England, p. 95-111.
- SCHWAB, M.; SAVIC, D.A.; WALTERS, G.A. (1996). *Multi-objective Genetic Algorithm for Pump Scheduling in Water Supply Systems*. Report No 96/02, University of Exeter, UK.
- SHAMIR, U. (1974). *Optimal Design and Operation of Water Distribution Systems*. Water Resources Research, v.10, n.1, p. 27-36.
- SHAWCROSS, J.F. (1985). *Modelling Complex Water Distribution Systems*. Computer Applications in Water Resources. H.C. Torno. New York, p. 438-447.
- SHEPHERD, A.; ORTOLANO, L. (1996). *Water-Supply System Operations: Critiquing Expert-System Approach*. Journal of Water Resources Planning and Management, v.122, n.5, September, p. 348-355.
- SHINSTINE, D.S.; AHMED, I.; LANSEY, K.E. (2002). *Reliability/Availability Analysis of Municipal Water Distribution Network: Case Studies*. Journal of Water Resources Planning and Management, v.128, n.2, March, p. 140-151.

- SILVA, F.G.B.; REIS, L.F.R.; SOARES, A.K. (2002). *Calibração de Rede de Distribuição de Água considerando as Perdas por Vazamentos Explicitamente no Modelo com o uso de Algoritmos Genéticos-Aplicação para a Rede de São Carlos-SP*. In:Seminário-Planejamento, Projeto e Operação de Redes de Abastecimento de Água. O Estado da Arte e Questões Avançadas. João Pessoa. CD-Rom. 15p.
- SIMONOVIC, S.P. (1992). *Reservoir Systems Analysis:Closing Gap between Theory and Practice*. Journal of Water Resources Planning and Management, v.118, n.3, May, p. 121-130.
- SIMPSON, A.R.; DANDY, G.C.; MURPHY, L.J. (1994). *Genetic Algorithms Compared to Other Techniques for Pipe Optimization*. Journal of Water Resources Planning and Management, v.120, n.4, July, p. 423-441.
- SIMPSON, A.R. (1999). *Modeling of Pressure Regulating Devices: A Major Problem Yet to be Satisfactorily Solved in Hydraulic Simulation*.Water Distribution System Conference. Tempe, Arizona, June, 11p.
- SIWON, Z. (1998). *Hydraulic Analysis of Water Distribution Systems*. Environmental Protection Engineering, v.24, n. 3, p. 121-130.
- SOARES, A.K.; REIS, L.F.R.; SILVA, F.G.B. (2002). *Modelos de Simulação Hidráulica Dirigidos pela Pressão(MSHDP) com Suporte do EPANET*. In:Seminário-Planejamento, Projeto e Operação de Redes de Abastecimento de Água. O Estado da Arte e Questões Avançadas. João Pessoa.CD-Rom. 15p.
- SOARES, A.K. (2003). *Calibração de Modelos de Redes de Distribuição de Água para Abastecimento Considerando Vazamentos e Demandas Dirigidas pela Pressão*. São Carlos-S.P. 153p. Dissertação(Mestrado)-Escola de Engenharia de São Carlos da Universidade de São Paulo.
- SOARES, A.K.; REIS, L.F.R.; CARRIJO, I.B. (2003). *Head-Driven Simulation Model(HDSM) for Water Distribution System Calibration*. In: Proceedings of the International Conference on Computing and Control for the Water Industry, September, London, UK, p. 197-208.
- SOUSA, E.R.; SILVA, U.L. (1987). *Planejamento e Análise de Sistemas de Abastecimento de Água*.In: Simpósio Luso-Brasileiro de Hidráulica e Recursos Hídricos, III, Salvador, 1987. Anais, v.1, p. 149-161.
- SOUZA, R.S. (1994). *Aspectos Computacionais da Análise de Redes de Distribuição de Água com Componentes Hidráulicos em Regime Permanente*. São Carlos. 236p.

Dissertação (Mestrado). Escola de Engenharia de São Carlos. Universidade de São Paulo, Brasil.

SOUZA, R.S. (1998). *Controle Operacional Otimizado de Redes de Distribuição de Água usando a Teoria do Regulador Quadrático Linear*. São Carlos. 247p. Tese (Doutorado). Escola de Engenharia de São Carlos. Universidade de São Paulo, Brasil.

SRINIVAS, N; DEB, K (1994). *Multi-Objective Function Optimization Using Non-Dominated Sorting Genetic Algorithms*. Evolutionary Computation Journal, 2(3), p. 221-248.

SU, Y.; MAYS, L.W.; DUAN, N.; LANSEY, K.E. (1987). *Reliability-Based Optimization Model for Water Distribution Systems*. Journal of Hydraulic Engineering, v.114, n.12, December, p. 1539-1556.

SURRY, P.D. ; RADCLIFFE, N.J.; BOYD, I.D. (2001). *A Multi-Objective Approach to Constrained Optimisation of Gas Supply Networks: The COMOGA Method*. Meeting The World's Water and Environmental Resources Challenges-State of The Practice-Orlando, Florida, US.EWRI-ASCE. 15p.

TAHER, S.A.; LABADIE, J.W. (1996). *Optimal Design of Water-Distribution Networks with GIS*. Journal of Water Resources Planning and Management, v.122, n.4, July, p. 301-311.

TANNER, T.L. (1985). *Adapting a Pipe Network Analysis Computer Program*. Computer Application in Water Resources. H.C. Torno. New York, p. 430-437.

TANYIMBOH, T.T.; TABESH, M.; BURROWS, R. (2001). *Appraisal of Source Head Methods for Calculating Reliability of Water Distribution Networks*. Journal of Water Resources Planning and Management, v.127, n.4, July-August, p. 206-213.

TARDELLI, J.F. (1987). *Aspectos Gerais Relativos a um Centro de Controle Operacional de um Sistema de Abastecimento de Água*. Relatório apresentado à OPS e ao Serviço Nacional de Abastecimento de Água Potável do Peru. 38p.

TODOROVSKI, L.; FLACH, P.; LAVRAC, N. (2000). *A Report on Experiments with Weighted Relative Accuracy in CN2*. Technical Report CSTR-00-003, Department of Computer Science, University of Bristol.

TOLSON, B.A.; MAIER, H.R.; SIMPSON, A.R.; LENCE, B.J. (2004). *Genetic Algorithms for Reliability-Based Optimization of Water Distribution Systems*. Journal of Water Resources Planning and Management, v.130, n.1, January, p. 63-72.

- VELDHUIZEN, V.D.; LAMONT, G.B. (2000). *Multi-Objective Evolutionary Algorithms. Analyzing The State-of-the-Art*. Evolutionary Computation Journal, 8(2), p. 125-148.
- WALSKI, T.M. (1983). *Technique for Calibration Network Models*. Journal of Water Resources Planning and Management, v.109, n.4, October, p. 360-372.
- WALSKI, T.M. et. al. (1987). *Battle of The Nework Models: Epilogue*. Journal of Water Resources Planning and Management, v.113, n.2, March, p. 191-203.
- WALSKI, T.M. (1993). *Tips for Saving Energy in Pumping Operations*. Journal AWWA - Management & Operations. July, p. 49-53.
- WALSKI, T.M. (1995). *Optimization and Pipe-Sizing Decisions*. Journal of Water Resources Planning and Management, v.121, n.4, July, p. 340-343.
- WALSKI, T.M. (2001). *The Wrong Paradigm-Why Water Distribution Optimization Doesn't Work*. Journal of Water Resources Planning and Management, v.122, n.4, July, p. 203-205.
- WALTERS, G.A.; LOHBECK, T. (1993). *Optimal Layout of tree Networks using Genetic Algorithms*. Engineering Optimization, v.22, p. 27-48.
- WALTERS, G.A.; HALHAL, D.; SAVIC, D.A.; OUAZAR, D. (1999). *Improved Design of "Anytown" Distribution Network Using Structured Messy Genetic Algorithms*. Urban Water, v.1, p. 23-38.
- WARDLAW, R.; SHARIF, M. (1999). *Evaluation of Genetic Algorithms for Optimal Reservoir System Operation*. Journal of Water Resources Planning and Management, v.125, n.1, Jan, p. 25-33.
- WEISS, S.M.; KULIKOWSKI, C.A. (1991). *Computer Systems that Learn*. San Mateo, California, USA. Morgan Kaufmann.
- WEISS, S.M.; INDURKHIA, N. (1998). *Predictive Data Mining: A Pratical Guide*. San Francisco, Califórnia: Morgan Kaufmann.
- WESTPHAL, K.S.; VOGEL, R.M.; KIRSHEN, P.; CHAPRA, S.C. (2003). *Decision Support System for Adaptive Water Supply Management*. Journal of Water Resources Planning and Management, v.129, n.3, May, p. 165-177.
- WOOD, D.J. (1980). *User's Manual – Computer Analysis of flow in Pipe Networks Including Extended Period Simulations*. University of Kentucky, Lexington, Ky.
- WRIGHT, T.M. (1991). *Genetic Algorithms for Real Parameter Optimization*. In Foundations of Genetic Algorithms 2 (FOGA-1), p. 205-218.

- WU, Z.; SIMPSON, A.R. (1996). *Messy Genetic Algorithms for Optimisation of Water Distribution Systems*. Department of Civil and Environmental Engineering. Research Report R 140, November. 62p.
- WU, Z.; BOULOS, P.; SCHAETZEN, W.; ORR, C. ; MOORE, M. (2001). *Using Genetic Algorithms for Water Distribution System Optimization*. Meeting The World's Water and Environmental Resources Challenges-State of The Pactice-Orlando, Florida, US.EWRI-ASCE. 8p.
- WU, Z. et al. (2002). *Optimal Capacity Design of Water Distribution Systems*. 2002 Conference on Water Resources Planning and Management. Roanoke, VA, USA. May. Proceedings, 10p.
- XU, C.; GOULTER, I.C. (1998). *Probabilistic Model for Water Distribution Reliability*. Journal of Water Resources Planning and Management, v.124, n.4, Jul./Aug., p. 218-228.
- XU, C.; GOULTER, I.C. (1999). *Reliability-Based Optimal Design of Water Distribution Networks*. Journal of Water Resources Planning and Management, v.125, n.6, November, p. 352-362.
- YEH, W.G. (1985). *Reservoir Management and Operations Model: A State-of-art Review*. Water Resources Research. 21(2), p. 1797-1818.
- YU, G.; POWELL, R.S.; STERLING, M.J.H. (1994). *Optimized Pump Scheduling in Water Distribution Systems*. Journal of Optimization Theory and Application. V.83, n.3, p. 463-488.
- ZAHED FILHO, K. (1990). *Previsão de Demanda de Consumo em Tempo Real no Desenvolvimento Operacional de Sistemas de Distribuição de Água*. São Paulo. 135p. Tese(Doutorado). Escola Politécnica da Universidade de São Paulo.
- ZESSLER, U.; SHAMIR, U. (1989). *Optimal Operation of Water Distribution Systems*. Journal of Water Resources Planning and Management, v.115, n.6, November, p. 735-752.
- ZITZLER, E.; THIELE, L. (1998). *An Evolutionary Algorithm for Multi-Objective Optimization: The Strength Pareto Approach*. Technical Report 43, Zurich, Switzerland: Computer Engineering and Networks Laboratory(TIK), Swiss Federal Institute of Technology(ETH).
- ZITLER, E.; LAUMANN, M.; THIELE, L. (2001). *SPEA2: Improving The Strength Pareto Evolutionary Algorithm*. Technical Report 103. Computer Engineering and

Networks Laboratory (TIK). Swiss Federal Institute of Technology (ETH). Zurich, Switzerland, May, 19p.

ZONGXUE, X.; JINNO, K. ; KAWAMURA, A. ; TAKESAKI, S. ; ITO, K. (1998). *Performance Risk Analysis for Fukuoka Water Supply System*. Water Resources Management, 12, p. 13-30.

ZYL, J.E.; SAVIC, D.A. ; WALTERS, G.A. (2001). *An Explicit Integration Technique for Dinamic Modeling of Water Distribution Systems*. Meeting The World's Water and Environmental Resources Challenges-State of The Pactice-Orlando, Florida, US.EWRI-ASCE. 9p.

ZYL, J.E. (2001). *A Methodology for Improved Operational Optimization of Water Distribution System*. Exeter, U.K. 187p. Tese – PhD of Engineering. School of Engineering and Computer Science. University of Exeter.

ZYL, J.E.; SAVIC, D.A.; WALTERS, G.A. (2004). *Operational Optimization of Water Distribution Systems Using a Hybrid Genetic Algorithm*. Journal of Water Resources Planning and Management, v.130, n.2, March, p. 160-170.

ANEXO A – Saída de dados do See5 para a simulação descrita no item 6.1.

See5 [Release 1.20a] Wed Sep 15 20:18:36 2004

Options:
Rule-based classifiers

Class specified by attribute `condiçãõ`

Read 52 cases (169 attributes) from Test1_24.08.04.data

Rules:

Rule 1: (35/7, lift 1.4)
 P143 = 1
 -> class b [0.784]

Rule 2: (17/1, lift 2.0)
 P143 = 0
 -> class o [0.895]

Default class: b

Evaluation on training data (52 cases):

Rules		
No	Errors	
2	8(15.4%)	<<
(a)	(b)	<-classified as
28	1	(a): class b
7	16	(b): class o

Time: 0.0 secs

See5 [Release 1.20a] Wed Sep 15 20:19:57 2004

Options:
Rule-based classifiers
Cross-validate using 10 folds

Class specified by attribute `condiçãõ`

Read 52 cases (169 attributes) from Test1_24.08.04.data

[Fold 0]

Rules:

Rule 1: (21/1, lift 1.7)
 P62 = 1
 P143 = 1
 -> class b [0.913]

Rule 2: (32/7, lift 1.4)
 P143 = 1
 -> class b [0.765]

Rule 3: (15/1, lift 2.0)
 P143 = 0
 -> class o [0.882]

```

Rule 4: (21/4, lift 1.8)
        V171 = 1
        V234 = 1
        -> class o [0.783]

```

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
4	0(0.0%)	<<

[Fold 1]

Rules:

```

Rule 1: (21/1, lift 1.7)
        P62 = 1
        P143 = 1
        -> class b [0.913]

```

```

Rule 2: (20/1, lift 1.6)
        V31 = 0
        P143 = 1
        -> class b [0.909]

```

```

Rule 3: (16/1, lift 2.0)
        P143 = 0
        -> class o [0.889]

```

```

Rule 4: (23/6, lift 1.6)
        V31 = 1
        -> class o [0.720]

```

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
4	0(0.0%)	<<

[Fold 2]

Rules:

```

Rule 1: (15, lift 1.7)
        P51 = 0
        V104 = 0
        -> class b [0.941]

```

```

Rule 2: (20/1, lift 1.6)
        V104 = 0
        V204 = 0
        -> class b [0.909]

```

```

Rule 3: (19/2, lift 1.9)
        V104 = 1
        -> class o [0.857]

```

```

Rule 4: (8/1, lift 1.8)
        P51 = 1
        V204 = 1
        -> class o [0.800]

```

Default class: b

Evaluation on hold-out data (5 cases):

Rules

```

-----
      No      Errors
      4      2(40.0%)  <<

[ Fold 3 ]

Rules:

Rule 1: (19, lift 1.7)
      V31 = 0
      P143 = 1
      -> class b [0.952]

Rule 2: (22/2, lift 1.6)
      P143 = 1
      V171 = 0
      -> class b [0.875]

Rule 3: (16, lift 2.1)
      P143 = 0
      -> class o [0.944]

Rule 4: (15, lift 2.1)
      V31 = 1
      V171 = 1
      -> class o [0.941]

Default class: b

Evaluation on hold-out data (5 cases):

      Rules
-----
      No      Errors
      4      2(40.0%)  <<

[ Fold 4 ]

Rules:

Rule 1: (31/6, lift 1.4)
      P143 = 1
      -> class b [0.788]

Rule 2: (16/1, lift 2.0)
      P143 = 0
      -> class o [0.889]

Default class: b

Evaluation on hold-out data (5 cases):

      Rules
-----
      No      Errors
      2      1(20.0%)  <<

[ Fold 5 ]

Rules:

Rule 1: (31/6, lift 1.4)
      P143 = 1
      -> class b [0.788]

Rule 2: (16/1, lift 2.0)
      P143 = 0
      -> class o [0.889]

Default class: b

```

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
2	1 (20.0%)	<<

[Fold 6]

Rules:

Rule 1: (22/1, lift 1.7)
 P62 = 1
 P143 = 1
 -> class b [0.917]

Rule 2: (10/1, lift 1.5)
 V41 = 1
 P143 = 1
 -> class b [0.833]

Rule 3: (15/1, lift 2.0)
 P143 = 0
 -> class o [0.882]

Rule 4: (6, lift 2.0)
 V41 = 0
 P62 = 0
 -> class o [0.875]

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
4	1 (20.0%)	<<

[Fold 7]

Rules:

Rule 1: (32/6, lift 1.4)
 V104 = 0
 -> class b [0.794]

Rule 2: (15/1, lift 2.1)
 V104 = 1
 -> class o [0.882]

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
2	1 (20.0%)	<<

[Fold 8]

Rules:

Rule 1: (29/5, lift 1.4)
 V104 = 0
 -> class b [0.806]

Rule 2: (17/2, lift 1.9)
 V104 = 1
 -> class o [0.842]

Default class: b

Evaluation on hold-out data (6 cases):

Rules		
No	Errors	
2	1(16.7%)	<<

[Fold 9]

Rules:

Rule 1: (22/1, lift 1.6)
 P62 = 1
 P143 = 1
 -> class b [0.917]

Rule 2: (19/1, lift 1.6)
 V31 = 0
 P143 = 1
 -> class b [0.905]

Rule 3: (14/1, lift 2.0)
 P143 = 0
 -> class o [0.875]

Rule 4: (5, lift 2.0)
 V31 = 1
 P62 = 0
 -> class o [0.857]

Default class: b

Evaluation on hold-out data (6 cases):

Rules		
No	Errors	
4	1(16.7%)	<<

[Summary]

Fold	Rules	
----	No	Errors
0	4	0.0%
1	4	0.0%
2	4	40.0%
3	4	40.0%
4	2	20.0%
5	2	20.0%
6	4	20.0%
7	2	20.0%
8	2	16.7%
9	4	16.7%
Mean	3.2	19.3%
SE	0.3	4.2%

(a)	(b)	<-classified as
----	----	
25	4	(a): class b
6	17	(b): class o

Time: 0.0 secs

See5 [Release 1.20a] Wed Sep 15 20:21:17 2004

Options:
Rule-based classifiers

Class specified by attribute `condiçãõ`

Read 52 cases (169 attributes) from Test1.1_24.08.04.data

Rules:

Rule 1: (21/1, lift 1.6)
V31 = 0
V104 = 0
-> class b [0.913]

Rule 2: (16/1, lift 1.6)
V22 = 0
V104 = 0
V204 = 0
-> class b [0.889]

Rule 3: (15/1, lift 2.0)
V22 = 1
V31 = 1
-> class o [0.882]

Rule 4: (12/1, lift 1.9)
V31 = 1
V204 = 1
-> class o [0.857]

Rule 5: (19/2, lift 1.9)
V104 = 1
-> class o [0.857]

Default class: b

Evaluation on training data (52 cases):

Rules		
No	Errors	
5	4(7.7%)	<<
(a)	(b)	<-classified as
27	2	(a): class b
2	21	(b): class o

Time: 0.0 secs

See5 [Release 1.20a] Wed Sep 15 20:22:06 2004

Options:
Rule-based classifiers
Cross-validate using 10 folds

Class specified by attribute `condiçãõ`

Read 52 cases (169 attributes) from Test1.1_24.08.04.data

[Fold 0]

Rules:

Rule 1: (19/1, lift 1.6)
 V31 = 0
 V104 = 0
 -> class b [0.905]

Rule 2: (21/3, lift 1.5)
 V04 = 1
 V104 = 0
 -> class b [0.826]

Rule 3: (14, lift 2.1)
 V04 = 0
 V31 = 1
 -> class o [0.938]

Rule 4: (17/2, lift 1.9)
 V104 = 1
 -> class o [0.842]

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
4	1(20.0%)	<<

[Fold 1]

Rules:

Rule 1: (17, lift 1.7)
 V31 = 0
 V104 = 0
 -> class b [0.947]

Rule 2: (16/1, lift 1.6)
 V22 = 0
 V104 = 0
 V204 = 0
 -> class b [0.889]

Rule 3: (14/1, lift 2.0)
 V22 = 1
 V31 = 1
 -> class o [0.875]

Rule 4: (12/1, lift 1.9)
 V31 = 1
 V204 = 1
 -> class o [0.857]

Rule 5: (18/2, lift 1.9)
 V104 = 1
 -> class o [0.850]

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
5	1(20.0%)	<<

[Fold 2]

Rules:

```
Rule 1: (30/6, lift 1.4)
        V104 = 0
        -> class b [0.781]
```

```
Rule 2: (5, lift 1.9)
        V41 = 0
        P62 = 0
        -> class o [0.857]
```

```
Rule 3: (17/2, lift 1.9)
        V104 = 1
        -> class o [0.842]
```

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
3	1(20.0%)	<<

[Fold 3]

Rules:

```
Rule 1: (13, lift 1.7)
        V22 = 0
        V104 = 0
        V204 = 0
        -> class b [0.933]
```

```
Rule 2: (20/1, lift 1.6)
        V31 = 0
        V104 = 0
        -> class b [0.909]
```

```
Rule 3: (13, lift 2.1)
        V22 = 1
        V31 = 1
        -> class o [0.933]
```

```
Rule 4: (10, lift 2.1)
        V31 = 1
        V204 = 1
        -> class o [0.917]
```

```
Rule 5: (17/1, lift 2.0)
        V104 = 1
        -> class o [0.895]
```

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
5	2(40.0%)	<<

[Fold 4]

Rules:

```
Rule 1: (19/1, lift 1.6)
        V31 = 0
        V104 = 0
        -> class b [0.905]
```

```
Rule 2: (15/1, lift 1.6)
        V22 = 0
        V104 = 0
```



```

V204 = 0
-> class b [0.882]

Rule 3: (14/1, lift 2.0)
V22 = 1
V31 = 1
-> class o [0.875]

Rule 4: (11/1, lift 1.9)
V31 = 1
V204 = 1
-> class o [0.846]

Rule 5: (17/2, lift 1.9)
V104 = 1
-> class o [0.842]

Default class: b

Evaluation on hold-out data (5 cases):

```

Rules		
No	Errors	
5	0(0.0%)	<<

[Fold 5]

Rules:

```

Rule 1: (19/1, lift 1.6)
V31 = 0
V104 = 0
-> class b [0.905]

Rule 2: (14/1, lift 1.6)
V22 = 0
V104 = 0
V204 = 0
-> class b [0.875]

Rule 3: (13/1, lift 1.9)
V22 = 1
V31 = 1
-> class o [0.867]

Rule 4: (11/1, lift 1.9)
V31 = 1
V204 = 1
-> class o [0.846]

Rule 5: (17/2, lift 1.9)
V104 = 1
-> class o [0.842]

```

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
5	0(0.0%)	<<

[Fold 6]

Rules:

```

Rule 1: (30/5, lift 1.5)
V104 = 0
-> class b [0.813]

Rule 2: (17/1, lift 2.0)

```

```

V104 = 1
-> class o [0.895]

Default class: b

Evaluation on hold-out data (5 cases):

```

```

      Rules
-----
      No      Errors
      2      2(40.0%)  <<

```

```
[ Fold 7 ]
```

```
Rules:
```

```

Rule 1: (16, lift 1.6)
        P51 = 0
        V104 = 0
        -> class b [0.944]

```

```

Rule 2: (19/2, lift 1.5)
        V52 = 0
        V104 = 0
        -> class b [0.857]

```

```

Rule 3: (18/2, lift 2.0)
        V104 = 1
        -> class o [0.850]

```

```

Rule 4: (8/1, lift 1.9)
        P51 = 1
        V52 = 1
        -> class o [0.800]

```

```
Default class: b
```

```
Evaluation on hold-out data (5 cases):
```

```

      Rules
-----
      No      Errors
      4      3(60.0%)  <<

```

```
[ Fold 8 ]
```

```
Rules:
```

```

Rule 1: (20/1, lift 1.6)
        V31 = 0
        V104 = 0
        -> class b [0.909]

```

```

Rule 2: (24/2, lift 1.6)
        V104 = 0
        P151 = 0
        -> class b [0.885]

```

```

Rule 3: (5, lift 2.0)
        V31 = 1
        P151 = 1
        -> class o [0.857]

```

```

Rule 4: (17/2, lift 1.9)
        V104 = 1
        -> class o [0.842]

```

```
Default class: b
```

```
Evaluation on hold-out data (6 cases):
```

```

      Rules
-----

```

No	Errors	
4	3 (50.0%)	<<

[Fold 9]

Rules:

Rule 1: (18/1, lift 1.6)
 V31 = 0
 V104 = 0
 -> class b [0.900]

Rule 2: (14/1, lift 1.5)
 V22 = 0
 V104 = 0
 V204 = 0
 -> class b [0.875]

Rule 3: (13/1, lift 2.0)
 V22 = 1
 V31 = 1
 -> class o [0.867]

Rule 4: (16/2, lift 1.9)
 V104 = 1
 -> class o [0.833]

Rule 5: (9/1, lift 1.9)
 V31 = 1
 V204 = 1
 -> class o [0.818]

Default class: b

Evaluation on hold-out data (6 cases):

Rules		
No	Errors	
5	0 (0.0%)	<<

[Summary]

Fold	Rules	
	No	Errors
0	4	20.0%
1	5	20.0%
2	3	20.0%
3	5	40.0%
4	5	0.0%
5	5	0.0%
6	2	40.0%
7	4	60.0%
8	4	50.0%
9	5	0.0%
Mean	4.2	25.0%
SE	0.3	6.9%

(a)	(b)	<-classified as
22	7	(a): class b
6	17	(b): class o

Time: 0.0 secs

--

See5 [Release 1.18] Wed Sep 15 20:43:10 2004

Options:
Rule-based classifiers

Class specified by attribute `condição`

Read 52 cases (169 attributes) from Test1.2_24.08.04.data

Rules:

Rule 1: (20/2, lift 1.5)
 P82 = 0
 -> class b [0.864]

Rule 2: (26/4, lift 1.5)
 V171 = 0
 -> class b [0.821]

Rule 3: (13/3, lift 1.3)
 V41 = 1
 -> class b [0.733]

Rule 4: (17, lift 2.1)
 V41 = 0
 P82 = 1
 V171 = 1
 -> class o [0.947]

Default class: b

Evaluation on training data (52 cases):

Rules		
No	Errors	
4	6 (11.5%)	<<
(a)	(b)	<-classified as
29	17	(a): class b
6		(b): class o

Time: 0.0 secs

See5 [Release 1.18] Wed Sep 15 20:43:41 2004

Options:
Rule-based classifiers
Cross-validate using 10 folds

Class specified by attribute `condição`

Read 52 cases (169 attributes) from Test1.2_24.08.04.data

[Fold 0]

Rules:

Rule 1: (16, lift 1.6)
 P62 = 1
 P81 = 0
 -> class b [0.944]

```
Rule 2: (8, lift 1.6)
        V41 = 1
        P81 = 0
        -> class b [0.900]
```

```
Rule 3: (24/4, lift 1.4)
        V171 = 0
        -> class b [0.808]
```

```
Rule 4: (14/1, lift 2.1)
        P81 = 1
        V171 = 1
        -> class o [0.875]
```

```
Rule 5: (5, lift 2.0)
        V03 = 0
        P81 = 1
        P133 = 0
        -> class o [0.857]
```

Default class: o

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
5	0(0.0%)	<<

[Fold 1]

Rules:

```
Rule 1: (16, lift 1.7)
        P62 = 1
        P81 = 0
        -> class b [0.944]
```

```
Rule 2: (6, lift 1.6)
        V41 = 1
        P81 = 0
        -> class b [0.875]
```

```
Rule 3: (24/4, lift 1.5)
        V171 = 0
        -> class b [0.808]
```

```
Rule 4: (15/1, lift 2.0)
        P81 = 1
        V171 = 1
        -> class o [0.882]
```

```
Rule 5: (5, lift 1.9)
        V03 = 0
        P81 = 1
        P133 = 0
        -> class o [0.857]
```

Default class: o

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
5	0(0.0%)	<<

[Fold 2]

Rules:

```
Rule 1: (26/4, lift 1.5)
        P81 = 0
```

```

-> class b [0.821]

Rule 2: (10/2, lift 1.4)
V03 = 1
-> class b [0.750]

Rule 3: (19/2, lift 1.9)
V03 = 0
P81 = 1
-> class o [0.857]

Default class: b

Evaluation on hold-out data (5 cases):

```

Rules		
No	Errors	
3	2(40.0%)	<<

```

[ Fold 3 ]

Rules:

Rule 1: (17/1, lift 1.6)
P82 = 0
-> class b [0.895]

Rule 2: (24/3, lift 1.5)
V171 = 0
-> class b [0.846]

Rule 3: (12/2, lift 1.4)
V41 = 1
-> class b [0.786]

Rule 4: (16, lift 2.1)
V41 = 0
P82 = 1
V171 = 1
-> class o [0.944]

Default class: b

Evaluation on hold-out data (5 cases):

```

Rules		
No	Errors	
4	1(20.0%)	<<

```

[ Fold 4 ]

Rules:

Rule 1: (17, lift 1.7)
P62 = 1
P81 = 0
-> class b [0.947]

Rule 2: (7, lift 1.6)
V41 = 1
P81 = 0
-> class b [0.889]

Rule 3: (23/4, lift 1.4)
V171 = 0
-> class b [0.800]

Rule 4: (14/1, lift 2.0)
P81 = 1
V171 = 1
-> class o [0.875]

```

```

Rule 5: (5, lift 1.9)
      V03 = 0
      P81 = 1
      P133 = 0
      -> class o [0.857]

```

```

Rule 6: (7/1, lift 1.7)
      V41 = 0
      P62 = 0
      -> class o [0.778]

```

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
6	0(0.0%)	<<

[Fold 5]

Rules:

```

Rule 1: (15, lift 1.7)
      P62 = 1
      P82 = 0
      -> class b [0.941]

```

```

Rule 2: (23/4, lift 1.4)
      V171 = 0
      -> class b [0.800]

```

```

Rule 3: (29/10, lift 1.4)
      P82 = 1
      -> class o [0.645]

```

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
3	0(0.0%)	<<

[Fold 6]

Rules:

```

Rule 1: (18, lift 1.7)
      P62 = 1
      P81 = 0
      -> class b [0.950]

```

```

Rule 2: (8, lift 1.6)
      V41 = 1
      P81 = 0
      -> class b [0.900]

```

```

Rule 3: (18/1, lift 2.0)
      V03 = 0
      P81 = 1
      -> class o [0.900]

```

```

Rule 4: (5, lift 1.9)
      V41 = 0
      P62 = 0
      -> class o [0.857]

```

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
4	3 (60.0%)	<<

[Fold 7]

Rules:

Rule 1: (19/2, lift 1.5)
P82 = 0
-> class b [0.857]

Rule 2: (25/4, lift 1.5)
V171 = 0
-> class b [0.815]

Rule 3: (19/2, lift 1.9)
P82 = 1
V171 = 1
-> class o [0.857]

Default class: b

Evaluation on hold-out data (5 cases):

Rules		
No	Errors	
3	1 (20.0%)	<<

[Fold 8]

Rules:

Rule 1: (18/2, lift 1.5)
P82 = 0
-> class b [0.850]

Rule 2: (22/3, lift 1.5)
V171 = 0
-> class b [0.833]

Rule 3: (24/7, lift 1.6)
V171 = 1
-> class o [0.692]

Default class: b

Evaluation on hold-out data (6 cases):

Rules		
No	Errors	
3	1 (16.7%)	<<

[Fold 9]

Rules:

Rule 1: (17/2, lift 1.5)
P82 = 0
-> class b [0.842]

Rule 2: (22/3, lift 1.5)
V171 = 0
-> class b [0.833]

Rule 3: (24/7, lift 1.6)


```
V171 = 1
-> class o [0.692]
```

Default class: b

Evaluation on hold-out data (6 cases):

Rules		
No	Errors	
3	1 (16.7%)	<<

[Summary]

Fold	Rules	
----	No	Errors
0	5	0.0%
1	5	0.0%
2	3	40.0%
3	4	20.0%
4	6	0.0%
5	3	0.0%
6	4	60.0%
7	3	20.0%
8	3	16.7%
9	3	16.7%
Mean	3.9	17.3%
SE	0.3	6.3%

(a)	(b)	<-classified as
----	----	
24	5	(a): class b
4	19	(b): class o

Time: 0.0 secs