OPERATIONAL OPTIMIZATION OF WDS USING A GENETIC ALGORITHM WITH MULTIOBJECTIVE FUNCTION AND OPERATING RULES EXTRACTED THROUGH DATA MINING

Ivaltemir(Tinil) Barros Carrijo* and Luisa Fernanda Ribeiro Reis**

** Centre forWater Systems, School of Engineering, Computer Science and Mathematics, Exeter, EX4 4QF, UK, I.B.Carrijo@ex.ac.uk.*
*** São Carlos School of Engineering, University of São Paulo, Av. Trabalhador São Carlense,400, São Carlos, S.P.,13566-590, Brazil, fernanda@sc.usp.br .*

ABSTRACT

The rapid growth of cities, associated with the lack of investments in basic sanitation, has rendered water supply systems highly complex and difficult to operate. 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. This paper aims to develop a computational model for the optimal operation control of macro water distribution systems using the EPANET2 hydraulic simulator, multiobjective genetic algorithms as an optimization tool, and data mining to extract operational rules for the system. These studies are to be based on the macro system of the city of Goiânia, in Brazil, owing to the complexity of its topology and the possible optimal operating rules. The results show that solutions for satisfactory operation can be quickly produced as a substitute to the personal judgment of the operator.

KEYWORDS

Operational optimization; multiobjective evolutionary algorithms; data mining process.

INTRODUCTION

The operation of an urban water supply system in real time is a highly complex task which has received special attention from researchers because of the need to ensure the reliability of services, the economic use of equipment (electrical energy and maintenance), to ensure that demands are met with the desired pressures, and to avoid or postpone the need for investments in the expansion of units. This reliability should also encompass service to the consumer under abnormal conditions occasioned by damaged units in the system.

The concept of systems operation, understood by laypersons as a mere sequence of equipment commands whose objective is to meet the demand (Zahed Filho, 1990), is actually far more complex, involving aspects of planning, control and supervision, and infrastructural consumer support and services, considered simultaneous and interdependent.

The operation plan requires that at least four basic conditions be met: a) a clear definition of the objectives to be achieved; b) the availability of mathematical analysis models; c) equipment to process these models; and d) knowledge of the system (Luvizotto Jr, 1995).

Righetto (2002) emphasizes that the interface between models for hydraulic simulation, optimization and definition of operating rules must be built carefully to make the model transparent, facilitate the introduction of restrictive inequalities and obtain objective function values in the successive steps required by the optimizer.

The purpose of this work is to present a methodology to achieve the optimal operation of water distribution systems, essentially macro systems (skeleton), concerning the costs of the operation and the hydraulic benefits. It represents an attempt to provide adequate operation rules in order to minimize costs and maximize the hydraulic benefits. Based on the knowledge of the system, provided by technical and commercial geo-referenced records, the purpose is to optimize its operation through multiobjective genetic algorithms (MOGAs), supported by a realistic hydraulic simulation model of the system behavior, and to produce operational rules through the data mining process.

METHODOLOGY

The purpose of operating water supply systems is, at acceptable risks, to meet the needs of consumption and minimize operational costs and, implicitly, to take the best possible advantage of the transport and reservoir system so as to retard expansion-related investments. The system's operation is a sequence of actions taken on the active elements of the system, such as valves and pumps in order to meet the objectives.

The optimization model implemented here takes into account two objectives: the minimization of the operational costs and the maximization of the hydraulic benefits, considering the index of demand met, adequate levels of water in the tanks, and minimum and maximum pressures at the demand points for a 24-hour period of analysis.

Hydraulic Simulation of the System

The hydraulic simulation evaluates the system's response to operational decisions in terms of the state variables, i.e., pressure, flow rate and level in the tanks. It is therefore an essential tool for the computational routine, which evaluates the established objectives. EPANET2, via Toolkit Library, Rossman (2001), is used for this purpose.

Operational Optimization using Multiobjective Genetic Algorithms

Optimization techniques are used to search for optimal trade-off for a specific operational problem. If the objective is, for instance, to minimize the operational cost, the cost function will be associated with the price of electrical energy, pump discharge, load losses at the facilities, etc. On the other hand, to meet the objective of minimum cost, the system itself imposes restrictions, such as maximum and minimum reservoir levels, limits of pressure and of power and quantity of available water.

According to Deb(2001) and Deb et al.(2002), since 1993, different evolutionary algorithms have been proposed for the solution of multiobjective optimization problems. The Multi-Objective Genetic Algorithm *(MOGA),* the Niched-Pareto Genetic Algorithm *(NPGA)*, and the Non-dominated Sorting Genetic Algorithm *(NSGA)*, were the precursors of this technique, whose basic characteristics are: evaluation of the members of a population based on the Pareto dominance concept and on preservation of the diversity of solutions. Although these algorithms have proven efficient to obtain multiple non-dominant solutions to various engineering problems, researchers have suggested the introduction of elitism to improve their convergence properties. Several algorithms stand out among the multiobjective evolutionary algorithms that consider elitism, i.e., the Strength Pareto Evolutionary Algorithm *(SPEA* and *SPEA II)*, the Pareto Archived Evolution Strategy (*PAES)*, the elitist GA of Rudolph, Pareto Envelope-based Selection Algorithm *(PESA* and *PESA II*) and Non-dominated Sorting Genetic Algorithm *(NSGA II*).

In Zitzler and Thiele's (1998) work, elitism was introduced through the maintenance of an external (secondary) population. This population stores a fixed number of non-dominant solutions found at the beginning of the simulation. In each generation, new non-dominant solutions are compared with the existing external population and the resulting non-dominant solutions are preserved. This algorithm not only preserves the best solutions but also uses them to participate in the genetic operations in order to reach better regions within the search space. This work uses the elitism-based SPEA method. Uniform crossover and non-uniform mutation were adopted, following an analysis of the results of several tests using various different operators (Cheung et al., 2003).

Extraction of Rules using Data Mining

There is a set of methods known as expert systems or knowledge-based systems whose classification models can be developed according to two main routes. The first obtains rules for the model through interviews based on experts and the inclusion of previous knowledge in the system. The second creates an inductive model through the generalization of a large record of collected and classified data (Monard, 1999).

According to Bessler et al. (2003), the method called data mining used in this work belongs to the second aforementioned route, which creates a classification model through the discovery and analysis of patterns that can be found in the data records. To apply the algorithms, several specific characteristics of the data must be analyzed. All the information about the cases (or examples) has to be presented in the form of attributes and each case is allocated to belong to a discrete predefined class.

The main function of a data mining program generally is to construct classification models as decision trees for later application. That, however, is not the main objective in this work. The classifier called *rulesets* is used to extract operational rules from a set of examples (cases) supplied by the optimization model (Pareto front) and labeled for an expert. The decision tree tool SEE5, which is the most recent version of the C4.5 inducer described by Quinlan (1993), is used for this purpose.

DESCRIPTION OF THE PROBLEM

Focusing on the development of a flexible tool that is easily handled by water supply systems operators, clearly providing a set of operational rules according to the working conditions of each unit of the system, part of the macro piping system of Goiânia, Brazil, was considered for analysis and evaluation of the results. For a clearer picture of the proposed application a diagram is shown of the system under study (Figure 1), with its main characteristics (Tables 1, 2, and 3).

DEFINITION OF THE OBJECTIVE FUNCTIONS

Several studies developed in the past showed that, of all the parameters relating to the operational costs, the most relevant one is the cost of the electrical energy consumption at the water pumping stations. Another possibility is the system's reliability in meeting consumer needs coherently. In this case, several parameters can be listed. According to Bao and Mays (1990), the reliability of the water supply systems can be considered from a hydraulic or mechanical standpoint. The former involves physical parameters that vary according to the operational changes in the system, while the latter involves the possible interventions on equipment. As in Walters et al. (1999), this work evaluates two basic objectives, the economic objective and the objective of hydraulic benefits of the water distribution systems.

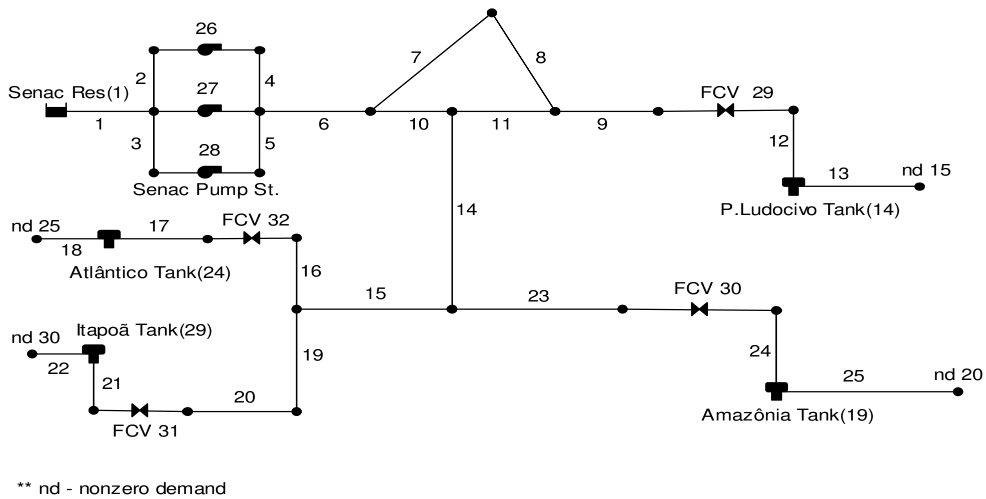

Figure 1. Diagram of the macro pipeline of Goiânia

Table 1. Nonzero demand node data – Goiânia

Nonzero Demand	Elevation(m)	Average Demand(l/s)
15	849	308
20	851	248
25	825	322
30	850	271

Table 2. Tank data – Goiânia

Tanks	Volume(m^3)	Min.lev.(m)	Max.lev.(m)	Elevation(m)
P.Ludovico	10,000	1.5	6.0	858.0
Amazônia	5,000	1.5	5.5	861.5
Atlântico	10,000	1.5	7.0	836.5
Itapoã	3,000	1.5	5.0	863.0

Table 3. Pump data – Goiânia

Pumps	Flow(l/s)	Head(m)
26	895	85
27	895	85
28	895	85

In the case of the economic objective, the intention is to minimize the costs of electrical energy consumption at the pumping stations. The daily cost for each pump at a pumping station is given by the sum of the cost of the maximum demand factor and the measured cost of consumption.

Considering the electrical energy costs to operate the pumping stations as an objective function, one has the following expression:

$$FO\_1 = [\ \sum_{t=1}^{24}\sum_{k=1}^{np} Cu(t) * \frac{Q(k,t) * H(k,t) * \gamma}{\eta(k,t)}\ ] + D * Rate$$

where k is the number of pumps at the pump station, t is time(h), $Cu(t)$ is the unit cost of the rate($R\$ / kWh$), $Q(k,t)$ is the pumped outflow(m^3 / s), $H(k,t)$ is the hydraulic head(m), $\eta(k,t)$ is the efficiency of the set(%), D is the maximum demand factor(kW), and $Rate$ is the rate for maximum demand factor(*R$/kW*).

Three hydraulic parameters of the system were evaluated as hydraulic benefit attributes: pressures at the demand nodes; minimum levels in the tanks, and the degree to which the demands are met. Based on Gargano and Pianese (2000), it was decided to adopt an optimal performance index to evaluate the hydraulic benefits. In this work, the quality of the water and the mechanical reliability were not considered, although their importance in the operational optimization of the system is recognized.

To consider the benefit of meeting the pressures at the demand nodes, an index called the pressure adequacy benefit (ψ_{pb}) was adopted using the following equation:

$$\psi_{pb(i,t)} = \left(\frac{P_{at(i,t)} - P_{mín}}{P_{req(t)} - P_{mín}} \right)^{1/2} \quad if \quad P_{mín} \le P_{at(i,t)} \le P_{req(t)}$$

$$\psi_{pb(i,t)} = 0 \quad if \quad P_{at(i,t)} < P_{mín} \quad or \quad P_{at(i,t)} > P_{req(t)}$$

where $P_{at(i,t)}$ is the actual nodal pressure (by EPANET), $P_{\min}$ is the minimum pressure(15m), and $P_{req(t)}$ is the maximum permissible pressure(time t).

The hydraulic benefit in terms of adequate nodal pressure (HB_{NP}), was evaluated according to the following equation:

$$HB_{NP} = \sum_{t=1}^{24}\sum_{i=1}^{nn} \psi_{pb(i,t)}$$

For the benefit in terms of water levels in the tanks, an index called the tank water level benefit (ψ_{lb}), was used as follows:

$$\psi_{lb(j,t)} = \left(\frac{N_{at(j,t)} - N_{mín(j)}}{N_{req(j,t)} - N_{mín(j)}} \right)^{1/2} \quad if \quad N_{mín} \le N_{at(j,t)} \le N_{req(j,t)}$$

$$\psi_{lb(j,t)} = 0 \quad if \quad N_{at(j,t)} < N_{mín} \quad or \quad N_{at(j,t)} > N_{req(j,t)}$$

where $N_{at(j,t)}$ is the tank water level at time *t* (by EPANET), $N_{\min(j)}$ is the minimum water level in the tank *j*, and $N_{req(j,t)}$ is the required water level in the tank *j* at the time *t*.

The hydraulic benefit in terms of satisfactory water levels in the tanks (BH_{WL}), will be:

$$HB_{WL} = \sum_{t=1}^{24}\sum_{j=1}^{nr} \psi_{lb(j,t)}$$

For the demand, the hydraulic benefit (HB_{SD}), can be defined as:

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

where $Q_{dem(i,t)}$ is the hourly demand at the node *i* at time *t*, and $\sum_{i=1}^{nn} Q_{dem(i,t)}$ is the total hourly demand at time *t*.

Thus, the objective function of hydraulic benefits to be maximized was expressed as:

$$FO\_2 = HB_{NP} + HB_{WL} + HB_{SD}$$

The objectives (minimum cost and maximum benefits) defined by FO_1 and FO_2 are conflicting. So, trade-off solutions that define the set of non-dominated alternatives called the Pareto front, have to be identified, in this case using multiobjective genetic algorithms.

REPRESENTATION OF THE SOLUTIONS

Each vector representative of a possible solution S_1 for the system's operational strategy has the following characteristic:

$$S_1 = \left[\underbrace{P_{(0,1)}, P_{(0,2)}, P_{(0,3)}, V_{(0,4)}, V_{(0,5)}, V_{(0,6)}, V_{(0,7)}}_{Hour\ 0} ..... \underbrace{P_{(23,1)}, P_{(23,2)}, P_{(23,3)}, P_{(23,4)}, P_{(23,5)}, P_{(23,6)}, P_{(23,7)}}_{Hour\ 23} \right]$$

where S_1 represents the solution vector, $P\ and\ V$ are the decision variables, $P_{(0,1)}$ represents the status of pump number 1 at time zero(0/1=off/on), and $V_{(0,1)}$ is the status of the valve number 1 at time zero(0/1=closed/opened).

DATA FILE FOR THE SEE5

As result of multiobjective genetic algorithm SPEA application, a Pareto front is obtained with the operational solutions that comprise the trade-off surface between both predefined objectives. The data mining process is then used on the full set of Pareto optimal solution to identify rules that lead to solutions which an expert judges are excellent in terms of overall operational practice. The set of solutions, here called examples or cases, are therefore classified according to expert judgment.

Two different input files are required to use the SEE5 inductor: the data.file, containing the data on the variables (attributes) to be analyzed and the names.file, containing the definition of the attributes and classes. A typical rule from the rule set looks like this:

Rule 4: (cover 15)
P213 = 0
$\rightarrow class\ e[confidence\ 80\%]$

where cover specifies the number of cases that the rule applies to and confidence is the percentage of the number of times the rule was correct. Thus, the above rule suggests that every example (solution) for which pump 3 is off at 9pm is classified as excellent.

In this specific study, the attributes can assume the values 0 or 1. The data.file supplies the values of each attribute for each case or example. Each example represents a solution or a point of the Pareto front resulting from the optimizer model. The class, which is defined in the names.file, can be either *g* (good) or *e* (excellent), according to an expert. For the purposes of this study, as a substitute for expert judgment the classification was derived automatically on the basis of the number of pump and valve switches, in the following way. An index called operating switch $K_{PV}(i)$, is defined by the following expression:

$$K_{PV}(i) = \sum_{k=1}^{np}\sum_{t=1}^{24}\lambda_p(k,t) + \sum_{j=1}^{nv}\sum_{t=1}^{24}\lambda_v(j,t)$$

where i is the operational rule of the Pareto front, λ_p and λ_v are the pump switch parameter and the valve switch parameter, respectively and have the following definitions:

$$\lambda_p \ or \ \lambda_v = 1 \ \ if \ \ \lambda_p(k,t) = \lambda_p(k,t-1) \ \ or \ \lambda_v(j,t) = \lambda_v(j,t-1)$$

$$\lambda_p \ \ or \ \lambda_v = 0 \ \ otherwise \ .$$

The class condition is e (excellent) if $K_{pv(i)} > \frac{1}{ns} * C$ and g (good) if $K_{pv} <= \frac{1}{ns} * C$. $C = \sum_{i=1}^{ns} K_{pv(i)}$ and ns is the total number of solutions in the Pareto front.

RESULTS

Figure 2 presents a Pareto front identified by the multiobjective genetic algorithm SPEA in the objective space. The figure also indicates two particular solutions, DM Sol.1 and DM Sol.2, classified as *excellent,* and which represent the strategy covered for rule indicated by the classifier. The strategy of operation represented by DM Sol. 1 in terms of the pump status and water tank levels along 24-hours period of a typical day is presented in figure 3 and figure 4.

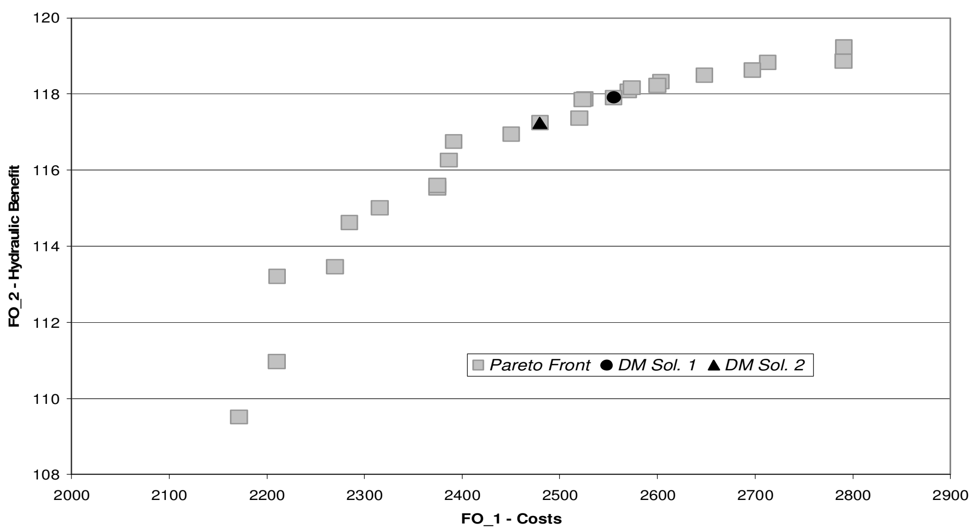

Figure 2. Goiânia – Pareto front – optimal solutions

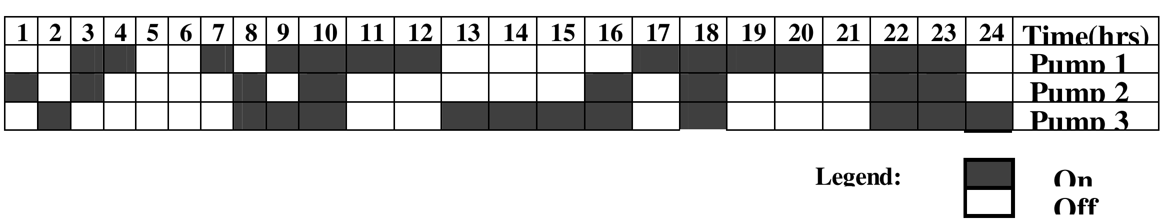

Figure 3. Goiânia – Pump operation – DM Sol. 1 solution

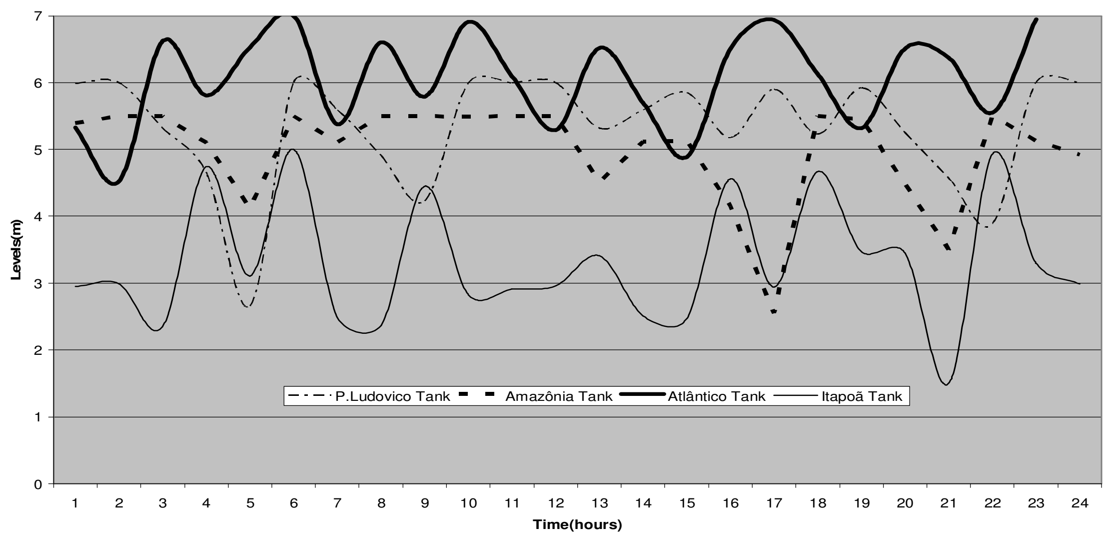

Figure 4. Goiânia – Water tank level – DM Sol. 1 solution

CONCLUSIONS

In order to give a efficient tool to the operator of WDS was developed a simulation-optimization model for determining optimal operational rules in a 24-hour horizon, which has demonstrated to be an efficient and practical tool.

The concept of Pareto optimality has been used to choose among multiple alternatives based on multiple criteria of choice. The aim this work was to apply the data mining process to narrow down the number of operational solutions supplied by Pareto front. With a small number of operational solutions, the developed model provides the operator proper conditions for rationally operating the system's units without depending exclusively on personal experience.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Saneamento de Goiás S.A. (Goiás State Sanitation Company)-SANEAGO S.A. and Department of Hydraulic and Sanitary Engineering – EESC-USP for support to the first author of this work.

REFERENCES

Bao, Y. and Mays, L.W. (1990). Model for water distribution system reliability. *Journal of Hydraulic Engineering*, v.116, n.9, 1119-1137.

Deb, K. (2001). *Multiobjective optimization using evolutionary algorithms*. John Wiley & Sons, Ltd. Chichester, England.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm:NSGA II, *IEEE Transactions on Evolutionary Computation*, 6(2), April, 182-197.

Bessler, F.T., Savic, D.A. and Walters, G.A. (2003). Water reservoir control with data mining. *Journal of Water Resources Planning and Management*, v.129, n.1, January, 26-34.

Cheung, P.B., Reis, L.F.R. and Carrijo, I.B. (2003). Multiobjective optimization to the rehabilitation of a water distribution network. *Proc. of International Conference on Computing and Control for the Water Industry*, London, UK, Maksimovic, C., Butler, D.and Memon, F., 315-325.

Gargano, R. and Pianese, D. (2000). Reliability as tool for hydraulic network planning. *Journal of Water Resources Planning and Management,* v.126, n.5, May, 354-364.

Luvizotto Júnior, E. (1995).Controle operacional de redes de abastecimento de água auxiliado por computador. PhD thesis, Escola Politécnica da Universidade de São Paulo, Brazil.

Monard, M.C. (1999). Data preparation, reduction and prediction in the context of data mining : a case study with insurance policies. *Technical Report 81,* ICMC-USP, Brazil.

Quinlan, J.R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers, San Francisco, CA.

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, Brazil, CD-Rom, 16 pages.

Rossman, L.A. (2001). *EPANET2 - Users manual.* U.S. Environmental Protection Agency, Cincinnati, Ohio.

Walters, G.A., Hallhal, D., Savic, D and Quazar, D. (1999). Improved design of ´Anytown ´ distribution network using structured messy genetic algorithms. *Urban Water*, v.1 , 23-38.

Zahed, Filho K. (1990). Previsão de demanda de consumo em tempo real no desenvolvimento operacional de sistemas de distribuição de água. PhD thesis, Escola Politécnica da Universidade de São Paulo, Brazil.

Zitzler, E. and 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).