

Operational Optimization of WDS based on Multiobjective Genetic Algorithms and Operational Extraction Rules using Data Mining

Ivaltemir B. Carrijo¹, Luisa Fernanda R.Reis², Godfrey A. Walters³, Dragan Savic⁴

Abstract

The growth of cities, associated with the lack of investment in basic infrastructure, has rendered water supply systems complex and difficult to operate. The efficient operation of such systems is a fundamental tool for extending the system's service life as much as possible, thus ensuring a reliable service to the consumers while keeping electrical energy and maintenance costs at acceptable levels. Efficient operation requires knowledge of the system, supported by tools such as models for hydraulic simulation, optimization, and definition of rules, provides the operator with proper conditions for the rational operation of the system's units. This paper aims to develop a computational model for the optimal operational control of macro water distribution systems using the EPANET2 hydraulic simulator, *SPEA (Strength Pareto Evolutionary Algorithm)* multiobjective genetic algorithms, and data mining to extract operational rules for the system. The studies were conducted on the macro system of the city of Goiânia, Brazil, chosen due to its complex characteristics, showing that solutions for its satisfactory operation can be quickly produced as a substitute to the personal judgment of the operator.

Introduction and Rationale

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 simultaneously and interdependently.

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).

Seeking greater reliability in the establishment of the system operating rules, new hydraulic techniques associated with optimization algorithms have been developed. Righetto (2002) emphasizes that the interface between models for

¹ Designer Manager, SANEAGO S.A., Goiânia, Brazil and Honorary University Fellow, Department of Engineering, School of Engineering and Computer Science, University of Exeter, North Park Road, Exeter, EX4 4QF, UK; phone 44 1392 263646; tinil@sc.usp.br and I.B.Carrijo@ex.ac.uk

² Associate Professor – São Carlos School of Engineering, University of São Paulo, Av. Trabalhador São Carlense, 400, São Carlos, São Paulo, Brazil; phone 55 16 2739545; fernanda@sc.usp.br

³ Professor, Department of Engineering, School of Engineering and Computer Science, University of Exeter, North Park Road, Exeter, EX4 4QF, UK; phone 44 1392 263633; G.A.Walters@ex.ac.uk

⁴ Professor, Department of Engineering, School of Engineering and Computer Science, University of Exeter, North Park Road, Exeter, EX4 4QF, UK; phone 44 1392 263637; D.Savic@ex.ac.uk

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 an 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 appropriate operation rules in order to minimize costs and maximize hydraulic benefits. Based on the knowledge of the system, provided by technical and commercial georeferenced records, the purpose is to optimize its operation through multiobjective genetic algorithms (MOGAs). This is supported by a realistic hydraulic simulation model of the system behavior, and the production of operational rules through the data mining process.

Methodology

The optimization model implemented here takes into account two objectives: the minimization of the operational costs and the maximization of the hydraulic benefits. Where the hydraulic benefits are considered as 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 tank level. 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 (MOGAs). 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 Niche-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 in obtaining 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 SPEAII)*, *the Pareto Archived Evolution Strategy (PAES)*, *the elitist GA of Rudolph*, *Pareto Envelope-based Selection Algorithm (PESA and PESAI)* and *Non-dominated Sorting Genetic Algorithm (NSGAI)*.

This work uses the elitism-based *SPEA* method. Uniform crossover and nonuniform 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.

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 a discrete predefined class.

The main function of a data mining program is generally 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). 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).

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 electric energy consumption at water pumping stations. Another possibility is the system's reliability in meeting consumer needs coherently. In this case, several parameters can be listed. The reliability of 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.

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.

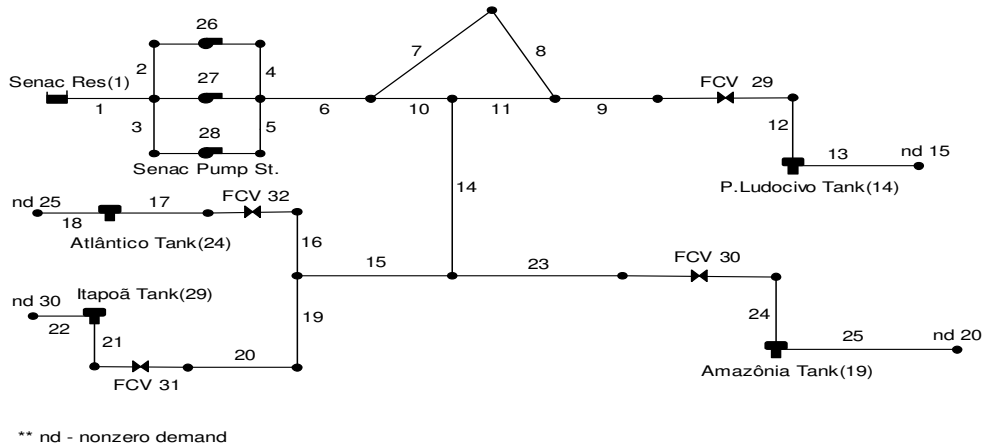


Figure 1 – Diagram of the macro pipeline of Goiânia

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

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

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/h), $H(k,t)$ 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 pressure adequacy benefit (ψ_{pb}) was adopted using the following equation :

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

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

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

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

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

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

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

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:

$$FO_2 = HB_{NP} + HB_{WL} + HB_{SD}$$

Both the objectives (minimum cost and maximum benefits) defined by FO_1 and FO_2 are self-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 S1 for the system's operational strategy has the following characteristic:

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

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 a result of the 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 judgement.

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)
P62 = 0
→ class e[confidence 53%]

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 2 is off at 6am 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:

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

$$\lambda_p \text{ or } \lambda_v = 0 \text{ otherwise .}$$

The class condition is e (excellent) if $K_{pv(i)} > \frac{1}{ns} * C$ and g (good) if $K_{pv} \leq \frac{1}{ns} * C$.

$$C = \sum_{i=1}^{ns} K_{pv(i)} \text{ and } ns \text{ is the total number of solutions in the Pareto front.}$$

Results

Figure 2 presents a Pareto front identified by the multiobjective genetic algorithm SPEA in the objective space with solutions classified as *excellent or good* for an expert. The figure also indicates two particular solutions, DM Sol.1 and DM Sol.2, classified as *excellent*, whose result is summarized in Figure 3. The result in figure 3 show that *Rule 4* classifies all solution in which pump 1 is *on* at 4am as *excellent*. Sol.1 and Sol.2 represent the strategy covered for *Rule 4*. The strategy of operation represented by DM Sol.1 in terms of the pump status along 24-hours period of a typical day is presented in figure 4.

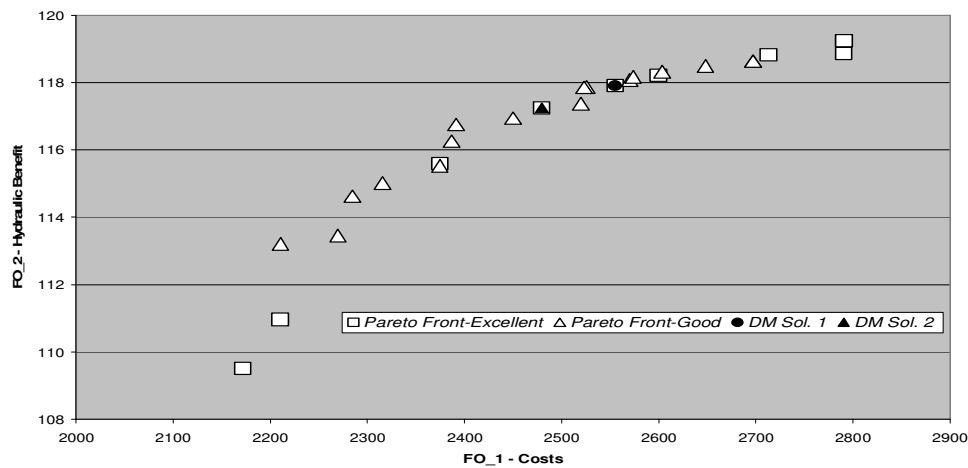


Figure 2 – Pareto front and excellent operational solutions from data mining

Rule 4 : (cover 7)
P41 = 1
→ class e [confidence 89%]

Figure 3 – Output file produced by the data mining process

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | Time(hrs) |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----------|
| | | | | | | | | | | | | | | | | | | | | | | | | Pump 1 |
| | | | | | | | | | | | | | | | | | | | | | | | | Pump 2 |
| | | | | | | | | | | | | | | | | | | | | | | | | Pump 3 |

Legend: On
 Off

Figure 4 – Pump operational – data mining solution- DM Sol. 1

Conclusions

In order to provide an efficient tool to the operator of Water Distribution Systems, a simulation-optimization model was developed for determining optimal operational rules for a 24-hour horizon. This has been demonstrated to be an efficient and practical tool.

The preliminary results obtained show that the concept of Pareto optimality has been successfully used to identify multiple alternatives taking into account two objectives: the minimization of operational costs and the maximization of hydraulic benefits. The data mining process used to narrow down the number of operational solutions supplied by the Pareto front, also represents a valuable tool in providing rules for rationally operating the system's units without depending exclusively on personal experience. As a result, one can say that the simulation-optimization model developed can be useful as a fast tool for identifying operational strategies and the rules that support them.

Acknowledgments

The authors wish to express their gratitude to Saneamento de Goiás S.A. (Goiás State Sanitation Company)-SANEAGO S.A. for support to the first author of this work, including the one-year visit to Exeter University, UK, for which he is grateful also to the Centre for Water Systems. Support of Brazilian Research Council (CNPq) under the CT-HIDRO Program (process 550070/2002-8) is also acknowledged.

References

- Bessler, F.T., Savic, D.A. and Walter, G.A. (2003), "Water Reservoir Control with Data Mining", *Journal of Water Resources Planning and Management*, v.129, n.1, January, 26-34.
- Deb, K. (2001). "Multi-Objective Optimization using Evolutionary Algorithms", *John Wiley & Sons, Ltd*, Chichester, England, p.491.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.A. (2002), "Fast and Elitist Multiobjective Genetic Algorithm: NSGA II", *IEEE Transactions on Evolutionary Computation*, 6(2), April, 182-197.
- Cheung, P.B. Reis, L.F.R. and Carrijo, I.B. (2003), "Multiobjective Optimization to the Rehabilitation of a Water Distribution Network", *Proc. International Conference on Computing and Control for the Water Industry*, London, UK, Maksimovic, C., Butler, D., Memon, A. 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", São Paulo, p.143, *Tese (Doutorado)*, Escola Politécnica da Universidade de São Paulo.
- 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, Brasil, CD-Rom, p.16.
- Rossman, L.A. (2001), "EPANET2-Users Manual", *U.S. Environmental Protection Agency*, Cincinnati, Ohio.
- Walters, G.A., Halhal, 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", São Paulo, p.135, *Tese (Doutorado)*, Poli-USP.