Water Resources Management manuscript No.
(will be inserted by the editor)

An Improved Stochastic Optimization Model for Water Supply Pumping Systems in Urban Networks

Jonathan De La Vega · Douglas Alem

Received: date / Accepted: date

Abstract This study investigates a pump scheduling problem for the collection, transfer and storage of water in water supply systems in urban networks. The objective of this study is to determine a method to minimize the electricity costs associated with pumping operations. To address the dynamic and random nature of water demand, we propose a two-stage stochastic programming model with recourse, where the random variables are represented by a finite, discrete set of realizations or scenarios. The developed mathematical model is an extension of a previously developed deterministic model in the literature and reflects the basic assumption that a fixed cost could be incurred by the activating/deactivating activities of the hydraulic pumps. To control the possible violations of the water demand constraints for different scenarios, we also analyze a robustness technique in an attempt to obtain "almost feasible" solutions. In addition, we adopt the so-called Mean Absolute Deviation criterion, which is a risk-aversion criterion, to obtain second-stage costs that are less dependent on the realizations of the scenarios. The scenarios were generated using a Monte Carlo simulation procedure that may use any probability distribution to produce empirical probabilities of the random variables. Because the proposed pump scheduling problem with fixed cost is a two-stage stochastic mixed $0-1$ program, we developed an efficient hybrid heuristic to obtain acceptable solutions for practical scenarios with a reasonable computational time. The overall results evidence the stability of the scenario generation method, the sensitivity of the solution based on the key parameters of the mathematical model, and the efficiency of the heuristic in solving large scenarios. Finally, we show that it is possible to conserve resources by solving the stochastic programming model rather than adopting simpler approaches based on the expected value.

Keywords Pump Scheduling · Water Supply Systems in Urban Networks · Stochastic Water-Demand · Stochastic Programming · Robustness Analysis · Mean Absolute Deviation

1 Introduction

The conservation of energy and water resources so that they are available for future generations is a topic that has created considerable concern worldwide. This concern motivates the proposal of new techniques that lead to efficient

J. De La Vega
Federal University of São Carlos, Rodovia João Leme dos Santos (SP-264), Km 110 Bairro do Itinga, Sorocaba, São Paulo, Brasil, CEP 18052-780
E-mail: fauthor@example.com

D. Alem
Federal University of São Carlos, Rodovia João Leme dos Santos (SP-264), Km 110 Bairro do Itinga, Sorocaba, São Paulo, Brasil, CEP 18052-780

water management and the rational use of energy. In Brazil, the sector responsible for the water and wastewater treatment supply accounts for 3% of the country's electricity consumption, and over 90% of this consumption comes from water pumping operations. Furthermore, it is estimated that between 35% and 42% of treated water is lost as a result of leakage (PROSAB 2009), which compromises the supply and contributes to the increased cost of pumping operations and water distribution in urban networks. Therefore, scheduling pumps for water collection and planning activities related to water storage and distribution are essential measures to minimize the costs of electricity, minimize water loss and guarantee the efficiency of urban water supply systems.

Guaranteeing that the demand for water at specified pressure levels is met is the responsibility of the Water Supply Systems (abbreviated SAA, in Portuguese). A water supply system consists of a set of pumps, pipes, valves, reservoirs and similar hydraulic components. To analyze the interactions between these components, it is common for SAAs to be represented by a network known as a *water supply network* or a *distribution network* (Amit and Ramachandran 2009). Basically, the operation of an SAA entails activation of hydraulic pumps to transfer the water from the water wells to the reservoirs. Subsequently, water is transferred to different consumption centers via the distribution system. An example of a water supply network topology is shown in Figure 1. In such a network, water is collected from water wells with an unlimited capacity and, after passing through the treatment plant, is transferred to the reservoirs through the supply network. Such reservoirs, in turn, can only distribute water to low areas of consumption or to towers directly connected to them. In addition, each tower can only distribute water to associated areas of high consumption. The presence of pumps between two plants indicates that pumping operations must be performed for both water collection and distribution. In contrast, the absence of pumps between two plants in the network indicates that water transport is performed by the gravitational force, which does not require the use of electricity.

[Insert Figure 1 here.]

To function efficiently, the sector responsible for the planning of the SAAs must make a number of decisions involving hydraulic pump activation policies, the volume of water to be moved in the network, and the volume of water that will remain in reservoirs over the periods of the planning horizon. Thus, it is necessary to consider the following: (1) the availability of hydraulic pumps in the system, (2) the tariff structure of energy consumption, (3) water consumption patterns for the reservoirs (demand), (4) water loss caused by leaks in the network, and (5) the capacity of the reservoirs.

The Brazilian Electricity Regulatory Agency (abbreviated ANEEL, in Portuguese) is responsible for defining the criteria for the classification of electricity consumers and, depending on this classification, applying different consumption tariffs. One of the criteria adopted by ANEEL to classify consumers relates to their daily consumption patterns, and, at the time of billing, consumers are classified as either high-voltage or low-voltage consumers. The SAAs are classified as high-voltage consumers because of their high energy consumption rate and the electric power demands of their pumping operations. For this classification, the tariff structure of electricity consumption and power demand involves applying different rates throughout the day, where this rate is higher during peak hours, which consists of the period between 6 pm and 9 pm.

Contrary to what common sense would suggest, activating the hydraulic pumps outside the peak periods to collect as much water as possible would be a bad strategy for SAAs because, in general, network leakage is proportional to the volume of water maintained in the reservoirs. Thus, appropriate levels of water in the reservoirs should be established to minimize this loss. Therefore, it is important to schedule the activation of pumps and plan water distribution in urban supply networks to minimize the costs of energy consumption associated with pumping activities and, simultaneously, minimize water loss, while also meeting the water demand in reservoirs and respecting the relevant constraints of the system.

The scheduling of pumps determines which pumps will be used and at what time of day they will be activated. The planning of water distribution through the supply networks consists of defining the length of time the pumps will be operated to collect water and meet an expected demand. The problem of scheduling the activation of hydraulic pumps and planning water distribution in supply networks to optimize the costs associated with energy consumption is known in the literature as an Optimal Control Problem (OCP), (Zessler and Shamir 1989; Ormsbee and Lansey 1994). To address the OCP, the planning horizon is usually divided into discrete intervals of 1 hour. Optimal decisions regarding pump scheduling and water transport are made during each interval. Typically, OCP involves multiple hydraulic constraints such as conservation of mass, conservation of energy, performance criteria of the system (limiting pressures and speeds), and operational constraints, such as demand requirements, reservoir volumes, and the maximum activation frequency of the pumps.

Depending on the decision variables of the optimization model, the OCP associated with the SAAs can be formulated using a direct or indirect approach. The direct formulation of the optimal control problem is obtained when, for each discrete interval of the planning horizon, one decision variable is assigned to each pump of the network, which indicates the length of time that it will remain active. In contrast, the indirect formulation of the OCP occurs when, in each discrete interval of the planning horizon, the volume of water in the reservoirs corresponds to the decision variable of the problem. To differentiate these two formulations, the optimal control problem associated with the direct formulation is commonly cited in the literature as the Pump Scheduling Problem (PSP). In this article, the terms OCP and PSP will be used to refer to the problem of collecting, storing and transferring water in urban water supply networks.

However, to determine an optimal policy for the collection, storage and transfer of water, it is necessary to know the demand for water in the supply network. In practice, it is common for such demand to be specified in hours because this clearly presents an important daily cycle, and the tariffs vary throughout the day. Nevertheless, even short-term demand possesses a dynamic nature that is difficult to predict because it depends on a number of inherently random factors such as temperature, rainfall and humidity. Moreover, the demand also varies based on the day of the week, month, season, socioeconomic and cultural profile of the population and on the cost of the water itself (Shvartser et al 1993; Odan and Reis 2012).

Typically, the random nature of water demand is dealt with using forecasting methods, which can be long- or short-term forecasts (Zhou et al 2000; Jain et al 2001; Froukh 2001; Sjobom and Assfalk 2005; Zhang et al 2006; Alvisi et al 2007; Adamowski 2008; Herrera et al 2010; Mohamed and Al-Mualla 2010; Nasseri et al 2011; Odan and Reis 2012). Thus, optimal control policies can be implemented based on the demand profiles generated by the forecasting methods. However, if at any given time a significant difference is found between the estimated and actual demand during the daylight hours, it will be necessary to re-optimize the model of optimal control by using the updated data while still considering the estimates on future demands (Alvisi et al 2007). A possible disadvantage of implementing the optimal control policies using the estimated demands is the inefficiency presented during water collection, storage and transfer when the estimates are inaccurate. If the estimated demand ignores the pessimistic historical realizations, extra pumping operations will be necessary during critical periods when the cost of electricity is higher. In contrast, if the demand estimate is too pessimistic, the hydraulic pumps will most likely be used for longer periods of time than necessary, resulting in a significant and unnecessary increase in energy consumption.

One way of making optimal control policies less dependent on a single estimated demand is to consider a set of possible realizations for this demand that would be weighted based on their probability of occurrence. Thus, part of these policies will be "forced" to simultaneously meet all the demand profiles, generating *reasonably good* decisions or even suboptimal decisions to the problem. Among the optimization techniques that are based on this concept is the two-stage stochastic programming technique, which is one of the most widespread optimization techniques in the literature that explicitly takes into account the uncertainties involved in a problem. In this case, the control policies that must simultaneously meet all the demand realizations are called *first-stage decisions*. Moreover, each possible demand realization is associated with a set of decision variables called second-stage decisions, whose purpose is to correct the infeasibility that arises once the random variables are determined given the first-stage choice.

In this context, this paper proposes a two-stage stochastic programming model for the problem of collection, storage and transfer of water in urban networks, considering the demand for water as a random variable that can be reasonably approximated by a set of realizations. The mathematical model developed is an extension of the deterministic model proposed in the work of Toledo et al (2008) and reflects the basic assumption that there may be a fixed cost incurred by the activation/deactivation of the hydraulic pumps. In scenario-based stochastic programming a set of realizations of random variable are simultaneously considered in the optimization model. For this reason, measures of risk aversion and measures to quantify “deficit” and “surplus” situations are incorporated into the optimization problem to make the solution and the stochastic programming model less sensitive to the different realizations of the random variables. The possible realizations of the random variables were determined using Monte Carlo sampling. Because the proposed mathematical model is a large-scale mixed-integer $0-1$ stochastic programming problem, we propose an efficient heuristic to obtain acceptable solutions within a reasonable computational time.

The remainder of the paper is organized as follows. Section 2 presents a literature review of important papers that address the problem of pump scheduling. Section 3 develops a mixed-integer $0-1$ stochastic programming model for the problem of pump scheduling with stochastic demand. Section 4 proposes a method for generating scenarios based on Monte Carlo sampling. Section 5 describes the heuristic algorithm. Section 6 discusses the numerical results. Finally, Section 8 provides the concluding remarks and a discussion on future research.

2 Literature Review

Water supply systems are controlled to meet several objectives, including hydraulic performance of the network and economic efficiency. Hydraulic performance measures involve pressure levels at the nodes and reservoirs. Economic efficiency is influenced by the costs of energy consumption and pump maintenance. In an SAA, the costs related to energy consumption associated with pumping activities account for the largest share of the total operating costs. For this reason, most studies that address OCP in the literature attempt to minimize such costs (Ormsbee et al 2009). An optimal control system in SAAs consists of defining the pump scheduling over a predefined time period so that the demand for water is met without delay, the hydraulic network conditions are met, and the total cost of the electricity consumed is minimized (Zessler and Shamir 1989). Ormsbee and Lansey (1994) notes that the optimal control system in an SAA should contain three main components: (1) a hydraulic network model, which ensures that the hydraulic operating conditions of the network are met; (2) a demand forecasting model that provides an estimate of the water consumption patterns in reservoirs over the periods of the planning horizon; and (3) an optimal control model of the system where the decisions regarding the operating policies of the system that minimize a certain cost are determined. In the following, we present how the literature has addressed these main features and summarize the information associated with urban network configuration and the types of hydraulic models, demand models, control algorithms, and control policies given in Table 1.

2.1 Hydraulic Network Models

Hydraulic network models can be represented by mass-balance models (MB), regression models (RM), simplified hydraulic models (SH) and full hydraulic models (FH) (Ormsbee and Lansey 1994).

Mass-Balance Models: In mass-balance models, the flow rate within the system is determined by the demand for water plus the rate of variation in water levels in the reservoirs. The pressure requirements for obtaining the flow of water in the reservoirs are neglected and it is assumed that, to obtain the desired water volume in the reservoirs, a combination of pumps are available. Furthermore, if the water level in the reservoirs remains within a previously specified range, the pressure requirements on the nodes of the network will also be omitted. Multidimensional MB models have also been developed (Sterling and Coulbeck 1975b; Fallside and Perry 1975; Toledo et al 2008; Soler 2008; Ikonen et al 2012). Such models incorporate important functional relations between the level of water in the reservoirs and the pumped flow. Because hydraulic characteristics of the network in MB models are only determined using water balance restrictions in the reservoirs, the multidimensional MB models are generally converted into linear optimization programs. Thus, the main advantage of mass-balance models is the relative tractability of the corresponding optimization problem.

Regression Models: Unlike mass-balance models, regression models are used as a more accurate representation of the hydraulic characteristics of the network. In such models, the pressure requirements in the nodes and reservoirs are considered explicitly and modeled using a set of nonlinear equations. Several strategies can be used to model the non-linear equations that represent the hydraulic conditions of the network. For example, for different water volumes in the reservoirs and different demand estimates, it is possible to perform simulations for each possible pump combination in the network in an attempt to determine the hydraulic conditions under which the pumps must operate. The results of these simulations can be plotted and are termed pump operating curves because they approximate the hydraulic requirements associated with a particular pump operation (Ormsbee and Lansey 1994). Ormsbee et al (1989); Awumah and Lansey (1992); Beckwith and Wong (1995) and Eker et al (2003) used regression curves to approximate the hydraulic restrictions of supply network operations. The curves illustrate the hydraulic operational requirements for each possible pump combination in the supply system for different water levels in the reservoirs and for a series of demand criteria. Once these curves are defined, the information contained in them is used as an input in the optimization model. Another approach to generatinge operating curves is to use hydraulic simulators. In Magalhães et al (2010); Fang et al (2010); Bagirov et al (2013), the authors developed algorithms that combine optimization techniques with hydraulic simulation models (EPANET) to determine optimal control policies that minimize electricity consumption costs in SAAs. An EPANET-based simulator allows one to determine the hydraulic conditions under which the network should operate. Once the operating hydraulic conditions of the network are defined, an optimization model is used to determine the optimal control policies. Readers interested in the EPANET simulator can consult Rossman (2002). The main advantage of regression models is the possibility of incorporating the non-linearity of the hydraulic network. However, the regression curves only provide information for the configuration of the network under control and for a given series of demand criteria. Any change in the network configuration or change in the demand series can result in inefficient control policies (Ormsbee et al 1989; Ormsbee and Lansey 1994).

Full Hydraulic Models: In the full hydraulic models, the hydraulic requirements for the operation of the network are considered explicitly in the optimization model through a set of nonlinear equations and therefore nonlinear programming models are obtained. The equations that represent the hydraulic conditions of the network include mass-balance and energy conservation (Almeida and Barboza 2002). These models are more adaptable to changes in the supply system and to demand variations, which is in contrast to mass-balance and regression models, which require more data in their formulation and thus require extra effort for calibration. Brion and Mays (1991); Almeida and Barboza (2002) used a complete hydraulic model and developed a mixed-integer $0-1$ nonlinear programming model to determine the optimal control policy of a water supply network that minimizes energy consumption costs. The configuration of this network consists of multiple reservoirs and a single water treatment plant that supplies them.

Simplified Hydraulic Models: The simplified hydraulic model is defined as an intermediate model between the regression model and the full hydraulic model. In this model, the network may be approximated or analyzed using a macroscopic network model that uses a linear hydraulic equations system. Macroscopic models represent the system by using a simplified network. In general, by using simplified models, it may be possible to represent the hydraulic operating conditions of the network using linear equations and, in this way, linear programming models can be constructed. Jowitt et al (1988); Little and McCrodden (1989); Jowitt and Germanopoulos (1992); Pasha and Lansey (2009) and Price and Ostfeld (2013) developed linear programming models to address the optimal control problem of water supply systems.

2.2 Demand Forecasting Models

The demand for water is a key factor in determining optimal control policies for water supply systems. Because, in most cases, the demand is not known when the control policies of the SAAs are implemented, it is commonly estimated using forecasting methods. Such forecasting methods can be short-term, mid-term and long-term (Nasseri et al 2011). Short-term prediction is concerned with low scale planning and management. The resolution of this type of approach varies from 1 h to several days (Zhou et al 2000; Jain et al 2001; Sjobom and Assfalk 2005; Zhang et al 2006; Alvisi et al 2007; Adamowski 2008; Herrera et al 2010; Odan and Reis 2012). Mid-term prediction is applied in mid-time management and its resolution is between one month and one year (Nasseri et al 2011). Long-term prediction is concerned with large scale planning and management. Most of the long-term development programs in urban management are based on this type of prediction. The prediction resolution is equal to or greater than one year (Froukh 2001; Mohamed and Al-Mualla 2010).

To determine the water demand profile in upcoming periods, different methodologies have been used, such as linear and/or multiple regression (Mohamed and Al-Mualla 2010) and time series analyses (Zhou et al 2000; Froukh 2001). Recently, artificial neural networks have attracted special interest from researchers as a methodology to estimate the water demand profile of water supply systems (Jain et al 2001; Sjobom and Assfalk 2005; Zhang et al 2006; Alvisi et al 2007; Adamowski 2008; Herrera et al 2010; Odan and Reis 2012). Genetic algorithms have also been used to forecast the water demand in water supply systems (Nasseri et al 2011). To develop the above methodologies, data inputs are required, among them, historical water consumption data, rainfall, temperature, and humidity. According to Ormsbee and Lansey (1994), the water demand forecast can be incorporated into optimal control models using the following approaches: (1) lumped, where the system demand is typically represented by a single aggregate value; (2) proportional, where the regression relations are derived from a single demand pattern that can vary in proportion to the total demand of the system and (3) distributed, where the aggregate demand is distributed both spatially and temporally among the different points of demand in the network.

McCormick and Powell (2004) proposed the use of linear regression analyzes to estimate the demand for water and revealed that variations of future demands depend only on the demands of the immediately preceding periods. Thus, the authors used a Markov process to model it. To determine the control policies that minimize the expected operating costs, the authors used stochastic dynamic programming. A discrete set of possible states of demand was assumed from the results of the regression analysis, which also made it possible to estimate the probabilities of transitions of the states.

2.3 Optimal Control Model

Finally, the third component of an optimal control system is the optimization model. In this model, the optimal control policies are defined to optimize an operational objective of the SAAs, respecting the relevant constraints of the system. In general, the objective of the optimal control problem is to minimize the energy consumption incurred during the operation of a set of pumps over the planning horizon. The energy consumption associated with pumping activities varies mainly with the pumping duration and on the electricity unit cost (Ormsbee et al 2009). Most studies on optimal control problems in SAAs found in the literature attempt to minimize such costs.

In addition to electricity costs, several studies also consider pump maintenance costs (Lansey and Awumah 1994; Barán et al 2005). An operating program where pumps are frequently switched on and off can reduce power consumption. However, such a program might accelerate the wear and tear of the pumps and result in increased pump maintenance costs. Lansey and Awumah (1994) emphasized that pump maintenance costs are difficult to measure, but it can be assumed that the cost is proportional to the pump activation frequency. Thus, the number of times the pumps are activated during the planning period may be used as an alternative measure to the intangible costs associated with the wear and tear of the pumps.

Generally, system restrictions include limitations on the volume of water stored in reservoirs, quantities that can be supplied from a treatment plant, pump and valve configurations, and the maximum activation frequency of the pumps, among others. Depending on the decision variables used in the optimization model, direct or indirect formulations of the optimal control problem may be developed. The direct formulation is used, for each period of the specified horizon, when the decision variable corresponds to the fraction of time that the pumps will be in operation. In this approach, the optimal control policy can be classified as explicit (Ormsbee and Lansey 1994). The indirect formulation is used when the water level in the reservoir is used as the control variable. Therefore, the objective is to determine the optimal path that defines the optimal levels of water storage in the reservoirs (Ormsbee et al 1989). The pump operation policy generated from these formulations is classified as implicit ((Zessler and Shamir 1989; Ormsbee and Lansey 1994).

Implicit formulations typically require the solution of two subproblems. The first subproblem involves determining the optimal decision path. The optimal decision path can be defined as the volumes of water to be stored in reservoirs over the planning horizon and is determined in such a way as to minimize the cost of electricity. The second problem consists of determining pump operating policies that lead to the optimal decision path. The difficulty associated with implicit formulations relates to the fact that a considerable number of pump combinations produce the desired path. Moreover, among all the pump combinations that produce the optimal decision path, one should choose a combination that incurs the minimal cost (Ormsbee et al 2009).

The optimal control problem for water supply systems with multiple reservoirs, treatment plants and pumps is a large-scale nonlinear optimization problem with discrete and continuous variables, which makes it one of the most difficult mathematical problems to solve (Ormsbee et al 2009; Fang et al 2010; Bagirov et al 2013). Thus, efficient optimal control algorithms must be proposed to solve such problems. Dynamic programming techniques have successfully been used to solve the optimal control problem of SAAs with a single reservoir (Ormsbee and Lansey 1994). For example, Sterling and Coulbeck (1975b); Coulbeck (1984); Solanos and Montoliu (1988); Ormsbee et al (1989); Lansey and Awumah (1994); Zhuan and Xia (2013) developed control algorithms based on dynamic programming to determine optimal operating policies for water supply systems in real time. However, extensions of these approaches for systems with multiple reservoirs are limited because of the combinatorial nature of the problem.

Other operational research techniques that address the optimal control problem of supply water systems have also been used, such as linear programming (Jowitt et al 1988; Jowitt and Germanopoulos 1992; Toledo et al 2008; Pasha and Lansey 2009; Price and Ostfeld 2013), mixed-integer linear programming (Little and McCrodden 1989; Awumah and Lansey 1992; Toledo et al 2008) and nonlinear programming (Sterling and Coulbeck 1975a; Whaley and Hume 1986; Brion and Mays 1991; Chase and Ormsbee 1993; Yu et al 1994; Almeida and Barboza 2002; Cembrano et al 2000; Eker et al 2003). The differences between these studies primarily involve the physical dimension of the network, in terms of the number of treatment plants and reservoirs; type of hydraulic models used; manner in which demand was incorporated into the optimization model; control algorithm used and the objective to be minimized. These studies are similar in the fact that the control policies used can be considered explicit. Another common aspect between these studies is the disregard of inherent demand randomness in the problem.

Furthermore, methods based on both mono-objective and multi-objective evolutionary algorithms have also been developed to address the optimal control problem in SAAs. Beckwith and Wong (1995); Mackle et al (1995); Savic et al (1997); Barán et al (2005); López-Ibañez et al (2005); Magalhães et al (2010); Fang et al (2010) and Kougias and Theodossiou (2012) have developed evolutionary algorithms (EA) to determine pump activation policies that minimize the costs of electricity consumption and pump maintenance. Most of these studies use Genetic Algorithms (GA) as control algorithms. Constructive heuristics based on a simple analysis of the operational policies of the SAAs to determine the operating conditions and identify strategies to reduce costs have also been used to solve the problem under consideration (Soler 2008; Gergely and János 2012; Ikonen et al 2012).

In addition, the optimal control problem can be addressed solely to define optimal reservoir operation policies. In such cases, the main decision variable corresponds to the amount of water to be delivered from the reservoirs to their related consumption centers during the day. In general, the hydraulic network model used corresponds to the mass-balance. The objective function, however, is not aimed at minimizing the energy consumption costs associated with pumping activities, but at meeting the demand. For example, Ahmed and Sarma (2005); Reddy and Kumar (2006); Yun et al (2010); Jothiprakash et al (2011) proposed, as their main objective, minimizing the sum of the squared deviations of the unmet demand for water in reservoirs. These studies used both mono-objective and multi-objective evolutionary algorithms to determine the optimal pump activation policies.

Table 1 summarizes the most relevant characteristics of several of the aforementioned papers. This table is an extension of the table presented in Lansey and Ormsbee (1994). The studies were classified based on the following criteria: (1) the physical dimension of the network, i.e., the number of pumping stations and reservoirs (columns 2 and 3); (2) type of hydraulic model (column 4); (3) type of demand model (column 5); (4) demand representation (column 6); (5) risk measure (column 7); (6) control algorithm (column 8); (7) control policy (column 9); (8) transfer of water (column 10) and (8) objectives under optimization (column 11). From Table 1, it is possible to infer that most of the cited papers address the problem of optimal control under an explicit control policy. Moreover, there are few studies that consider the full hydraulic model. This reflects the fact that the resulting optimization models are nonlinear, which limits the use of efficient control algorithms. Most studies cited also ignore the dynamic and random nature of demand in an attempt to obtain simpler and/or implementable mathematical models. In several studies, the random nature of demand is treated using forecasting methods. However, erroneous water demand forecasts may generate inefficient water collection, transfer and storage policies. One method of explicitly considering the randomness of the demand for water in the pump scheduling problem is to describe it considering a finite set of realizations or scenarios. Thus, under the assumption that all the possible occurrences of demand (or the most likely ones) are considered in the mathematical model, it is possible to establish feasible control policies independent of the random variable. Moreover, it is also possible to incorporate measures of risk aversion in this model to generate feasible policies that are less sensitive to the numerous demand realizations in an attempt to reduce the reprogramming requirements for the pumps for each new demand realization. To date, no studies have been found in the literature that consider risk aversion measures and/or the robustness in the pump scheduling problem and related problems.

[Insert Table 1 here]

3 Problem Definition and Mathematical Model

Here, we closely follow the deterministic mathematical model presented in Toledo et al (2008). To facilitate the description of the mathematical model, consider the urban supply network shown in Figure 2. In this network, each well can only fill the reservoir with which it has a direct link. Similarly, each reservoir may supply water to the related consumption center. However, because the reservoirs are located in regions of the city at higher elevations, it is not necessary to designate pumps between them and their consumption centers because water distribution takes place using the gravitational force. In addition, there are pumps between different reservoirs to indicate the existence of water distribution between them. However, because there may be pairs of reservoirs that have no direct connection with a water well, the water flow between this pair of reservoirs must pass through the reservoir allocated between them. Notice that both reservoir j and pump j sometimes have an interchangeable meaning because each well is associated with a reservoir and a pump.

[Insert Figure 2 here.]

Basically, the proposed pump scheduling problem under stochastic demand involves determining which pumps should be operated and how long the pumps must be in operation to meet an uncertain demand at minimal electricity costs. Each hydraulic pump can be either off or on. In addition, activating a pump incurs a fixed charge. To explicitly consider the cost of electricity consumption associated with the activating/deactivation of the water pumps, let Y_{jt} be the binary decision variable that takes the value 1 if pump j performs water collection in period t (or, similarly, if pump j is used in period t), and let X_{jt} be the fraction of the time period t in which pump j remains in operation (with water being pumped). Both variables X_{jt} and Y_{jt} must be mathematically related by the expression $X_{jt} \leq Y_{jt}$. Notice that, for $X_{jt} > 0$, the variable Y_{jt} is activated; i.e., if there is water collection occurring (transfer between wells and reservoirs), the pump must necessarily be in operation. However, the mathematical expression $X_{jt} \leq Y_{jt}$ only informs whether pump j is in operation at time t but not if the pump is turned on.

To determine whether pump j is turned on or off during period t, one should know the state of the pump in the immediately preceding period, i.e., the assumed value of the decision variable $X_{j(t-1)}$. Let Z_{jt} be a binary variable that indicates whether pump j is activated at the beginning of period t. Therefore, the mathematical relation that determines whether pump j is activated or not during period t is $Z_{jt} \geq Y_{jt} - X_{j(t-1)}$. From this mathematical relation, it is possible to see that if the binary variable Y_{jt} takes the value of 1 and pump j is not turned on during the entire immediately preceding period, the variable Z_{jt} is activated and this incurs a pump activation cost of β_{jt}.

In this paper, the uncertain water demand in each consumption center k during period t (d_{kt}) is modeled as a random variable in a probability space $(\mathscr{W}, \mathcal{F}, \Pi)$, where $\mathscr{W}$ is a set of scenarios (with a specific realization $\xi(\omega)$ for scenario $\omega \in \mathscr{W}$) equipped with a $\sigma-$algebra of events $\mathcal{F}$ with a probability measure denoted by Π. We also assume that the random demand for scenario ω is represented by the random vector $\boldsymbol{\xi}^T = \{\xi(\omega = 1), \xi(\omega = 2), \cdots, \xi(\omega = |\mathscr{W}|)\}$ with a probability of occurrence $\pi(\omega)$ such that $\pi(\omega) > 0$ and $\sum_{\omega \in \mathscr{W}} \pi(\omega) = 1$. Each component of the random vector $\boldsymbol{\xi}$ can be written as the following random matrix:

$$\xi(\omega) = \begin{pmatrix} d_{11}(\omega) & d_{12}(\omega) & \cdots & d_{1|\mathscr{T}|}(\omega) \\ d_{21}(\omega) & d_{22}(\omega) & \cdots & d_{2|\mathscr{T}|}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ d_{|\mathscr{K}|1}(\omega) & d_{|\mathscr{K}|2}(\omega) & \cdots & d_{|\mathscr{K}||\mathscr{T}|}(\omega) \end{pmatrix},$$

where $|\mathscr{K}|$ and $|\mathscr{T}|$ represent the number of consumption centers and the number of time periods, respectively.

All the scenarios $\omega \in \mathscr{W}$ are considered simultaneously in the optimization model by using the two-stage stochastic programming with recourse paradigm. Using this paradigm, we partition the variables of the problem into *first* and *second-stage* decisions. The first-stage or scenario-independent variables are decisions that must be taken without complete information on the random variables. The second-stage or scenario-dependent variables are corrective decisions that serve to adjust the first-stage decisions for the demand in each scenario. Initially, because the main purpose of water supply systems is to meet the water demand of the consumption centers, it is mandatory to collect water from wells, which means activating a set of hydraulic pumps even though water demands are only partially known. Waiting for the realization of the water demands to then make decisions regarding water collection can cause a high repressed demand or result in the hydraulic pumps having to operate at peak times, which significantly increases the electricity consumption costs. Therefore, Y_{jt}, X_{jt} and Z_{jt} are defined as first-stage decision variables.

The impact of first-stage decisions determined using partial knowledge of the water demands can be minimized by recourse or second-stage decisions. For example, suppose the water demand realization is high and that, in addition, the water taken from the wells in the current period fails to meet such demand. To meet this repressed demand, the following recourse decisions may be taken: maintain an adequate store of water in the reservoirs during the previous period and/or transfer water from reservoirs with lower demands (or greater volumes of water) to those with repressed demand. Thus, it is sufficient to define the following second-stage (recourse) decisions: $I_{jt}(\omega)$ is the water store in reservoir j at the end of period t in scenario ω and $V_{j\ell t}(\omega)$ as the fraction of period t in which water is transferred from reservoir j to ℓ in scenario ω. By multiplying $V_{j\ell t}(\omega)$ by the corresponding pump flow rate to transfer water from the reservoirs, we obtain the total amount of water transferred.

Next, we will display a complete list of symbols for the two-stage Stochastic Pump Scheduling Problem (abbreviated SPSP).

Indices and sets:

$(j, \ell) \in \mathscr{R}$ Set of reservoirs or pumps.
$k \in \mathscr{K}$ Set of consumption centers.
$t \in \mathscr{T}$ Set of periods.
$\omega \in \mathscr{W}$ Set of scenarios.
$\mathscr{K}_j$ $= \{k,$ such that the center consumer k is supplied by reservoir $j\}$.
$\mathscr{R}_j$ $= \{\ell,$ such that ℓ is a reservoir that can receive water from reservoir $j\}$.
$\mathscr{R}_j^\star$ $= \{\ell,$ such that ℓ is a reservoir that can send water to reservoir $j\}$.

Deterministic Parameters:

α_{jt} Cost of maintaining the pump j in operation during period t.
β_{jt} Fixed cost of switching on and off pump j during period t.
$\delta_{j\ell t}$ Cost of transferring water from reservoir j to ℓ during period t.
q_{jt} Flow rate of pump j during period t.
$\nu_{j\ell t}$ Pump flow rate to transfer water from j to ℓ during period t.
$h_j^{\min}$ Minimum storage level of water allowed in reservoir j.

$h_j^{\max}$ Maximum storage level of water allowed in reservoir j.
h_j^0 Water stored in reservoir j at the beginning of the planning horizon.
θ_{jt} Fraction of water lost by leaks in reservoir j during period t.
γ Goal programming weight to control the *tradeoff* between feasibility and expected total cost.

Stochastic Parameters:

$d_{kt}(\omega)$ Water demand of the consumption center k during period t in scenario ω.
$\pi(\omega)$ Probability of scenario ω.

First-Stage Decision Variables:

X_{jt} Fraction of period t in which pump j remains in operation.
Y_{jt} $(= 1)$ If pump j performs water collection during period t (i.e., $X_{jt} > 0$).
Z_{jt} $(= 1)$ If pump j is switched on at the beginning of period t.

Second-Stage Decision Variables:

$I_{jt}(\omega)$ Water stored in reservoir j at the end of period t in scenario ω.
$V_{j\ell t}(\omega)$ Fraction of period t in which water is transferred from reservoir j to ℓ in scenario ω.

Let $\mathcal{X}$ be the set of all the first-stage decision variables, and let $\boldsymbol{\xi}$ be the random vector formed by all the water demands scenarios. The two-stage stochastic programming model with recourse in its equivalent deterministic form for the pump scheduling problem with stochastic demand can be posed as

$$\min \sum_{j\in\mathscr{R}}\sum_{t\in\mathscr{T}} \alpha_{jt}X_{jt} + \sum_{j\in\mathscr{R}}\sum_{t\in\mathscr{T}} \beta_{jt}Z_{jt} + \sum_{\omega\in\mathscr{W}} \mathcal{Q}[\mathcal{X},\boldsymbol{\xi}] \tag{1}$$

$$\text{s.t.: } X_{jt} \le Y_{jt},\ j\in\mathscr{R},\ t\in\mathscr{T} \tag{2}$$

$$Z_{jt} \ge Y_{jt} - X_{j(t-1)},\ j\in\mathscr{R},\ t\in\mathscr{T} \tag{3}$$

$$0 \le X_{jt} \le 1,\ j\in\mathscr{R},\ t\in\mathscr{T} \tag{4}$$

$$Y_{jt} \in \{0,1\},\ j\in\mathscr{R},\ t\in\mathscr{T} \tag{5}$$

$$Z_{jt} \in \{0,1\},\ j\in\mathscr{R},\ t\in\mathscr{T}, \tag{6}$$

where $\mathcal{Q}[\mathcal{X},\boldsymbol{\xi}]$ is the optimal value of the second-stage problem (7)−(11):

$$\min \sum_{j\in\mathscr{R}}\sum_{\ell\in\mathscr{R}_j}\sum_{t\in\mathscr{T}}\sum_{\omega\in\mathscr{W}} \pi(\omega)\delta_{j\ell t}V_{j\ell t}(\omega) \tag{7}$$

$$\text{s.t.: } I_{jt}(\omega) = (1-\theta_{j(t-1)})I_{j(t-1)}(\omega) + q_{jt}X_{jt} + \sum_{\ell\in\mathscr{R}_j^\star} \nu_{\ell jt}V_{\ell jt}(\omega) - $$
$$- \sum_{\ell\in\mathscr{R}_j} \nu_{j\ell t}V_{j\ell t}(\omega) - \sum_{k\in\mathscr{K}_j} d_{kt}(\omega), j\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W} \tag{8}$$

$$I_{j0}(\omega) = h_j^0,\ j\in\mathscr{R},\ \omega\in\mathscr{W} \tag{9}$$

$$h_j^{\min} \le I_{jt}(\omega) \le h_j^{\max},\ j\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W} \tag{10}$$

$$0 \le V_{j\ell t}(\omega) \le 1,\ j\in R,\ \ell\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W}. \tag{11}$$

The problems (1)−(6) involves all the first-stage decisions that must be optimally defined to minimize the costs associated with water collection (part 1), the activation/deactivation of the pumps (part 2), and with the expected value of the costs incurred by the second-stage decisions (part 3). Whereas constraints (2) express whether pump j is used during period t, constraints (3) inform whether the pump is activated during this period. Constraints (4) indicate that the hydraulic pump for collecting water from wells can be operated throughout the whole period ($X_{jt} = 1$) or simply for a part of it ($X_{jt} < 1$). Finally, constraints (4) and (5) indicate the domains of the first-stage variables. In the second-stage problem (7)−(11), the decision variables and the constraints depend on the scenario ω, and, therefore, the total cost depends on the expectancy $Q[\cdot]$ of the second-stage costs, i.e., the costs associated with the transfer of water between reservoirs (part 1). Constraints (8) express the water storage balance in the reservoirs. Constraints (10) ensure that the minimum and maximum water store in each reservoir is satisfied. Once the decision variables $V_{j\ell t}(\omega)$ represent the fraction of the time period t that the pumps between reservoirs will be in operation, constraints (11) will indicate that these variables can transfer water throughout the period ($V_{j\ell t}(\omega) = 1$) or a part of it ($V_{j\ell t}(\omega) < 1$).

3.1 An Improved Full Recourse-Based Formulation

In our current application, it is desirable that any solution $\mathcal{X}^\star$ satisfying the first-stage constraints (2)−(6) also have a feasible completion for all scenarios $\omega \in \mathscr{W}$ in the second-stage problem (7)−(11). In other words, we would like to ensure that any implementable operational policy regarding water collection can be adjusted independently of the behavior of the water demands by using water storage and/or water transfer among reservoirs. We say that the stochastic program with this property has a *relatively complete recourse*. However, notice that model (1)−(11) does *not* have a relatively complete recourse unless any amount of water can be stored in reservoirs and/or transferred among them. Thus, to build a relatively complete recourse model, it suffices to allow that the water storage decision variable $I_{jt}(\omega)$ can violate constraints (10) as long as these violations are compensated via penalization in the objective function. Indeed, consider the following equivalent formulation for the constraints (8):

$$I_{jt}(\omega) = \prod_{\tau=1}^{t}(1-\theta_{j\tau})h_j^0+\sum_{\tau=1}^{t}\left(q_{j\tau}X_{j\tau} + \sum_{\ell\in\mathscr{R}_j^\star}\nu_{\ell j\tau}V_{\ell j\tau}(\omega) - \sum_{\ell\in\mathscr{R}_j}\nu_{j\ell\tau}V_{j\ell\tau}(\omega) - \sum_{k\in\mathscr{K}_j}d_{k\tau}(\omega)\right),\ j\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W}.$$

As constraints (10) must be true for all $\omega \in \mathscr{W}$, then

$$h_j^{\min} \le \prod_{\tau=1}^{t}(1-\theta_{j\tau})h_j^0+\sum_{\tau=1}^{t}\left(q_{j\tau}X_{j\tau} + \sum_{\ell\in\mathscr{R}_j^\star}\nu_{\ell j\tau}V_{\ell j\tau}(\omega) - \sum_{\ell\in\mathscr{R}_j}\nu_{j\ell\tau}V_{j\ell\tau}(\omega) - \sum_{k\in\mathscr{K}_j}d_{k\tau}(\omega)\right) \le h_j^{\max},\ j\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W}.$$

Now, define $\mathcal{E}_{jt}^+(\omega) \ge 0$ as the water excess in reservoir j during period t in scenario ω and $\mathcal{E}_{jt}^-(\omega) \ge 0$ as the water shortage in reservoir j during period t in scenario ω. It is clear from the previous inequalities that, as long as $\mathcal{E}_{jt}^-(\omega) = h_j^{\min} - I_{jt}(\omega)$ and $\mathcal{E}_{jt}^+(\omega) = I_{jt}(\omega) - h_j^{\max}$, for all $j \in \mathscr{R},\ t \in \mathscr{T},\ \omega \in \mathscr{W}$, then the minimum and maximum water storage level always holds true. This is equivalent to rewriting the second-stage problem as follows:

$$\min \sum_{\omega\in\mathscr{W}} \pi(\omega) \left(\sum_{j\in\mathscr{R}} \sum_{\ell\in\mathscr{R}_j} \sum_{t\in\mathscr{T}} \delta_{j\ell t} V_{j\ell t}(\omega) + \gamma \sum_{j\in\mathscr{R}} \sum_{t\in\mathscr{T}} [\rho^+_{jt}\mathcal{E}^+_{jt}(\omega) + \rho^-_{jt}\mathcal{E}^-_{jt}(\omega)] \right)$$

$$\begin{aligned} \text{s.t.: } & I_{jt}(\omega) - (1-\theta_{j(t-1)})I_{j(t-1)}(\omega) - q_{jt}X_{jt} - \sum_{\ell\in\mathscr{R}^\star_j} \nu_{\ell jt} V_{\ell jt}(\omega) + \\ & + \sum_{\ell\in\mathscr{R}_j} \nu_{j\ell t} V_{j\ell t}(\omega) + \sum_{k\in\mathscr{K}_j} d_{kt}(\omega) = \mathcal{E}^+_{jt}(\omega) - \mathcal{E}^-_{jt}(\omega), j\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W} \\ & I_{j0}(\omega) = h^0_j,\ j\in\mathscr{R},\ \omega\in\mathscr{W} \\ & h^{\min}_j \le I_{jt}(\omega) \le h^{\max}_j,\ j\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W} \\ & 0 \le V_{j\ell t}(\omega) \le 1,\ j\in R,\ \ell\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W} \\ & \mathcal{E}^+_{jt}(\omega),\ \mathcal{E}^-_{jt}(\omega) \ge 0,\ j\in\mathscr{R},\ t\in\mathscr{T},\ \omega\in\mathscr{W}. \end{aligned} \tag{12}$$

The parameters ρ^+_{jt} and ρ^-_{jt} represent the penalty incurred for violating the maximum and the minimum storage level allowed in reservoir j during period t, respectively, and the parameter γ is a goal programming weight that can be used to reflect robustness preferences. For instance, if $\gamma = 0$, the violations occur frequently. In contrast, if γ is assigned to be sufficiently large, the number of violations in the water storage balancing constraint is expected to be very low at the expense of a higher expected total cost. Several authors refer to this solution as *robust* in the sense that it is "almost" feasible for any realization of the scenarios or, at least, feasible for most scenarios. The idea of "almost feasible" originates from the concept of *robustness* advocated by Mulvey et al (1995) and particularly from the concept of *model robustness.* The precise definition of "almost" can be devised by a proper choice of the penalty function. According to Sniedovich (2012), model (12) can also be called the *robust satisficing model.* It is still possible to show that model (12) is a special case of a relatively recourse model called a *complete* (or *full*) *recourse model*, where the second-stage problem is always feasible for all ω and any feasible or *not* $\mathcal{X}$.

Another important aspect of the pump scheduling problem in urban supply systems relates to the risk of failing to meet the demand for water. As mentioned previously, to reduce this risk, it is possible to resort to the transfer of water between reservoirs. However, the decision to do this may vary between different scenarios, making it necessary to reprogram the connection between pumps to transfer water between reservoirs for every new demand realization. Because this can be impractical and/or too costly, it is possible to reduce the variability of such a decision by enforcing the cost of this decision to vary little or vary according to the degree of tolerance of the decision maker. Thus, it is expected that the water transfer decisions have a greater stability among the scenarios. In this study, we used the mean absolute deviation (MAD) to reduce the variability of the costs of transferring water between reservoirs. For the pump scheduling problem investigated here, MAD can be formulated as follows:

$$\sum_{t\in\mathscr{T}} \sum_{j\in\mathscr{R}} \sum_{\ell\in\mathscr{R}_j} \delta_{j\ell t} V_{j\ell t}(\omega) - \sum_{t\in\mathscr{T}} \sum_{j\in\mathscr{R}} \sum_{l\in\mathscr{R}_j} \sum_{\underline{\omega}\in\mathscr{W}} \pi(\underline{\omega}) \delta_{j\ell t} V_{j\ell t}(\underline{\omega}) = \Delta^+(\omega) - \Delta^-(\omega),\ \omega\in\mathscr{W}. \tag{13}$$

The positive and negative deviations $\Delta^+(\omega)$ and $\Delta^-(\omega)$ are minimized in the objective function and weighted by the risk parameter λ, which controls the conflict between the total expected cost and the risk. The more risk-averse the decision maker is, the greater the value that should be assigned to λ. Finally, the second-stage problem with robust satisficing constraints and risk aversion, which we call hereafter ISPSP (**I**mproved **S**tochastic **P**ump **S**cheduling **P**roblem), can be posed as follows:

$$
\begin{aligned}
\min \sum_{\omega \in \mathscr{W}} \pi(\omega) \left(\sum_{j \in \mathscr{R}} \sum_{\ell \in \mathscr{R}_j} \sum_{t \in \mathscr{T}} \delta_{j\ell t} V_{j\ell t}(\omega) + \gamma \sum_{j \in \mathscr{R}} \sum_{t \in \mathscr{T}} [\rho^+_{jt} \mathcal{E}^+_{jt}(\omega) + \rho^-_{jt} \mathcal{E}^-_{jt}(\omega)] + \lambda [\Delta^+(\omega) + \Delta^-(\omega)] \right) \\
\text{s.t.: } & I_{jt}(\omega) - (1 - \theta_{j(t-1)}) I_{j(t-1)}(\omega) - q_{jt} X_{jt} - \sum_{\ell \in \mathscr{R}^\star_j} \nu_{\ell jt} V_{\ell jt}(\omega) + \\
& + \sum_{\ell \in \mathscr{R}_j} \nu_{j\ell t} V_{j\ell t}(\omega) + \sum_{k \in \mathscr{K}_j} d_{kt}(\omega) = \mathcal{E}^+_{jt}(\omega) - \mathcal{E}^-_{jt}(\omega), j \in \mathscr{R},\ t \in \mathscr{T},\ \omega \in \mathscr{W} \\
& \sum_{t \in \mathscr{T}} \sum_{j \in \mathscr{R}} \sum_{l \in \mathscr{R}_j} \delta_{j\ell t} V_{j\ell t}(\omega) - \sum_{t \in \mathscr{T}} \sum_{j \in \mathscr{R}} \sum_{l \in \mathscr{R}_j} \sum_{\underline{\omega} \in \mathscr{W}} \pi(\underline{\omega}) \delta_{j\ell t} V_{j\ell t}(\underline{\omega}) = \Delta^+(\omega) - \Delta^-(\omega),\ \omega \in \mathscr{W} \\
& I_{j0}(\omega) = h^0_j,\ j \in \mathscr{R},\ \omega \in \mathscr{W} \\
& h^{\min}_j \leq I_{jt}(\omega) \leq h^{\max}_j,\ j \in \mathscr{R},\ t \in \mathscr{T},\ \omega \in \mathscr{W} \\
& 0 \leq V_{j\ell t}(\omega) \leq 1,\ j \in R,\ \ell \in \mathscr{R},\ t \in \mathscr{T},\ \omega \in \mathscr{W} \\
& \mathcal{E}^+_{jt}(\omega),\ \mathcal{E}^-_{jt}(\omega) \geq 0,\ j \in \mathscr{R},\ t \in \mathscr{T},\ \omega \in \mathscr{W} \\
& \Delta^+(\omega),\ \Delta^-(\omega) \geq 0,\ \omega \in \mathscr{W}.
\end{aligned}
\tag{14}
$$

For each scenario ω, the mean absolute deviation (MAD) risk is defined as the difference between the cost associated with the transfer of water between reservoirs and the mathematical expectation of this cost based on all the scenarios. The reformulated objective function minimizes the expected cost of transferring water between reservoirs (part 1), penalties associated with water excess and shortage in the reservoirs (part 2), and the risk involved (part 3). Note that variables that quantify the deviation are non-negative. For additional information on risk/variability measures, the reader is referred to, e.g., Ogryczak and Ruszczyński (1999); Takriti and Ahmed (2004); Dupačová (2008); Alem and Morabito (2013).

4 Generation of Scenarios via Monte-Carlo Simulation

In this paper, we propose a method for generating scenarios based on Monte Carlos sampling (MC). The main objective of this methodology is to determine the realizations of the random variables involved in the optimization problem, i.e., to construct the discrete probability distribution of the uncertain parameters. This distribution represents an approximation of the marginal distribution of the random variables. Subsequently, it becomes possible to combine the generated realizations in a scenario tree structure to, for example, analyze the correlation between the events. The MC sampling allows the assignment of the probabilities of the scenarios based on the relative frequency of the realizations of the random variables, and, thus, they are not subjectively attributed. It is also important to mention that, based on previous studies, the minimum and maximum estimates of water demand during each hour of the day in the reservoirs are known, and such information motivates the development of this methodology. The steps of the scenario generation procedure are described as follows:

1. Determine $|\mathscr{W}|$ class intervals from the minimum and maximum water consumption estimates.
2. For each class defined in the previous step, generate a random number uniformly distributed between the limits of this class to attribute a numerical value to the associated achievement.

3. Because the realizations are arranged in increasing order, once additional $|\mathscr{W}|$ class intervals are constructed, with the difference that, in this case, the upper and lower limits of the interval ω will correspond to the realization of classes $\omega - 1$ and ω, respectively.
4. Generate N random numbers ϕ following a probability distribution D between the value of the minimum estimate and the nominal value of the maximum realization. This step corresponds to the implementation of the Monte Carlo simulation.
5. Quantify the relative frequency related to each class ω. This frequency determines the probability of a given realization between the limits of this class.
6. Construct the scenario tree using a combination of the realizations and the product of the probabilities by assuming statistical independence among the random variables.

In this study, we tested four probability distributions D: uniform, empirical, triangular and lognormal.

5 A Hybrid Heuristic Method for the ISPSP

The proposed hybrid heuristic (HH) solves our problem in two phases. In the first phase, a constructive heuristic determines a pump activation policy, i.e., the first-stage solution. Given this solution, the second phase solves optimally the resulting stochastic problem via the barrier method to determine the second-stage decision variables. In the following, we describe these two phases in detail.

5.1 A Constructive Heuristic (CH) for the First-Stage Decision Variables

The CH that served as the basis for this study was first developed in Soler (2006) in an attempt to obtain solutions that were calculated quickly from practical instances. Before explaining the development of the CH, it is necessary to classify the periods of the planning horizon as critical, pre-critical or non-critical.

Definition 1 *Critical periods* are those periods where the demand for water is high, i.e., peak periods, and, in addition, the cost of electricity consumption during this period is higher. Such periods are known *a priori* and correspond to the time intervals between 6 pm and 9 pm.

Definition 2 *Pre-critical periods* precede critical periods. In general, given a level of water leakage in the reservoirs, the water storage requirements necessary to meet the demand during critical periods is carried out in the pre-critical periods in an attempt to avoid pump operations during critical periods when the cost of electricity is higher.

Definition 3 Periods that are not classified as critical or pre-critical are defined as *non-critical.*

For each type of period, a strategy exists for the operation of the pumps. For non-critical periods, if the water storage levels in the reservoirs at the end of period t are greater than or equal to the minimum level permitted, the pumps are kept off. Otherwise, two strategies are defined, and the one that incurs a lower cost is chosen. For the case of meeting the demand by activating the pump during this period, if the pump is not turned on during the entirety of the immediately preceding period, pump activation costs are incurred. For the case of meeting the demand of the period during the last period in which the pump was turned on, pump activation costs are not incurred but electricity costs associated with water losses are. The operational strategy for the pre-critical periods consists of keeping the pumps turned on until the capacity of the reservoirs is reached to store water and meet the demand of critical periods. Finally, if the water accumulated in the reservoirs is not sufficient to meet the demand of critical periods, then it is possible to use two strategies during critical periods: transferring water between reservoirs and/or repeatedly turning on pumps to meet the demand. If transferring water between reservoirs is required, the cost of electricity consumption between each pair of reservoirs must be known so that the reservoir that incurs lower electricity costs can be chosen.

To determine the cost of electricity between each pair of reservoirs, we used the Floyd's algorithm. This algorithm determines the shortest path between all pairs of nodes in a graph.

The CH proposed in Soler (2008) can be explained as follows. After the planning horizon's periods have been classified as critical, pre-critical and non-critical, the heuristic algorithm routes all the planning horizon's periods, and, based on the type of period, it determines a lower-cost pump operating strategy that meets the demand of this period. For further details about this heuristic algorithm, the reader is referred to Soler (2008). In this paper, we also analyze the alternative of meeting the demand for water during pre-critical periods through non-critical periods. When this alternative is evaluated, there is a possibility of larger volumes of water being stored in the reservoirs during pre-critical periods. As a result, extra pumping activities to meet the water demand of the critical periods are avoided. This alternative was incorporated into the constructive heuristics developed in this paper; pre-critical periods are covered and analyzed as if they were not critical. Thus, the operational strategies of such periods are assessed during the pre-critical periods. The algorithm then routes the periods until the strategy of meeting the demand through pump activation in the period under analysis (strategy 1) is more cost effective than the strategy of meeting the demand during the last preceding period in which the pump was turned on (strategy 2). Once strategy 1 is selected, it is necessary to return to the first pre-critical period and perform the corresponding strategy during this period. Using this alternative, preliminary tests indicated an improvement of approximately 5% in the optimality gap of the heuristic compared to the original proposal by Soler (2008).

Because the CH was designed to define the best pump activation policy for a single realization of the demand (deterministic problem), it cannot be used directly on a problem having multiple demand realizations (stochastic problem). For this reason, we develop a strategy to approximate the uncertainties of the second-stage problem to replace the set of scenarios by one *reference scenario*, e.g., best-case scenario, expected-value scenario and/or worst-case scenario. Given the random vector $\boldsymbol{\xi}$, the best-case scenario represents the realization whose value $\xi(\omega)$ is the smallest among all the realizations; the expected value scenario corresponds to the expectation, i.e., $\bar{\xi} = \sum_{\omega \in \mathscr{W}} \pi(\omega)\xi(\omega)$; and the realization $\xi(\omega)$ whose value is greater than the value of the remaining realizations corresponds to the worst-case scenario. In this paper, these three strategies were used in the computational tests to determine which performs best in solving the first-stage problem. These tests were performed as follows: 5 instances of each class were resolved with the solution method proposed in this paper and the three mentioned strategies. The average optimality gap's obtained by solving the 5 instances for each class of the ISPSP were $\approx 3\%$, $\approx 17\%$ and $\approx 25\%$ for the strategies worst-case scenario, expected-value scenario and the best-case scenario, respectively. Therefore, the solution methodology proposed in this paper was based on the worst-case scenario.

5.2 Determining the Second-Stage Decision Variables

In stochastic programming, the number of rows and columns of the technological matrix increases significantly with the number of scenarios, frequently turning the equivalent deterministic problem into a large-scale optimization problem and making the technological matrix very sparse. Table 2 gives the model dimensions with respect to the number of constraints (m), number of continuous variables (nc), number of $0-1$ variables ($n01$), number of non-zero elements in the constraint matrix (nnz) and the matrix density (Den in %) for each combination of number of periods and scenarios. Note that the number of variables and constraints of the problem increases based on the size of the horizon and the number of scenarios. As expected, regardless of the combination of horizon and number of scenarios, the associated coefficient matrix is very sparse because its density is rather low (mostly zero entries). To exploit the sparsity of the matrix in an attempt to improve the efficiency of the solution method, we use the barrier method as the solution method for the second-stage problem. This method generates a sequence of strictly positive primal and dual solutions and ends when the gap of these solutions is within a predetermined tolerance limit. If the solution determined by the barrier method is not optimal, a procedure called a crossover can be executed. This procedure consists of applying the simplex method from the basic feasible solution determined using the barrier method so that the optimal solution of the problem can then be determined. To avoid the execution of the simplex method at the end of the barrier algorithm, thus reducing the execution times, the crossover was disabled.

[Insert Table 2 here]

The solution method procedure developed in this paper is summarized in Figure 3. The heuristic is called a hybrid procedure because it combines constructive heuristic rules and exact linear programming methods. Notice that, whereas we approximate the scenarios in determining a pump activation policy in the first phase, we restart all the scenarios in the second phase to solve the stochastic problem with fixed first-stage variables.

[Insert Figure 3 here.]

6 Model Analysis and Evaluation

The computational experiments were intended to (1) identify whether the proposed scenario generation method provides scenario trees that describe good approximations of the random variables distribution; (2) investigate if the form in which the probabilities are attributed to the scenarios influence the optimal value of the expected costs and the first- and second-stage decisions; (3) analyze the first- and second-stage decisions to obtain insights that are useful in the decision making of this problem; (4) determine the sensitivity of the expected costs and of the decision variables with respect to variations of the parameters λ and γ; (5) determine the impact of uncertainties on the problem by calculating the EVPI and the VSS; and (6) evaluate the performance of the hybrid heuristic (HB) both in regards to the quality of the solution and computational efficiency compared to the CPLEX 11.2 solver. The mathematical model and the hybrid heuristic were coded using the General Algebraic Modeling System (GAMS) and solved by the CPLEX 11.2 software package using the barrier method without crossover, which is used to solve linear problems. The scenario generation algorithm and the Floyd's algorithm were implemented using the MATLAB version 2012a software package. All the computational experiments were performed on a 2.8 GHz Core i7 PC with 8.0 GB of RAM running on the Windows 7 operating system. For the analysis, a section of the topology of the urban water supply network of the city of São Carlos (Brazil) was considered. The network consists of three reservoirs, as shown in Figure 2. However, other topologies could have been used because the mathematical model is flexible.

6.1 Data Description

Model ISPSP was tested for: (1) planning horizons of 2, 7 and 30 days; (2) scenario-trees with 27, 64 and 125 scenarios; and (3) leakage levels in the network of 0%, 10% and 20%. Therefore, there are 27 classes of generated examples from the combination of the above parameters. For each class, 10 instances were randomly generated. Each class was described by the triplet $|\mathscr{T}|/|\mathscr{W}|/\theta$, where $|\mathscr{T}|$ and $|\mathscr{W}|$ represent, in the same order, the cardinality of the sets that represent the number of days in the planning horizon and the possible realizations of the random variable, and θ describes the percentage of leakage in the network. Because the estimated demand for water in each reservoir is given in periods of one hour, each day in the planning horizon was divided into 24 sub-periods corresponding to one hour, which resulted in sub-periods of 48, 168 and 720 hours for the planning horizons of 2, 7 and 30 days, respectively. The water demand in the three reservoirs vary according to the intervals shown in Table 3. Such intervals describe the minimum and maximum estimates of water consumption for every hour of the day.

[Insert Table 3 here]

The realizations of the water demands in each reservoir were determined using the proposed method for generating scenarios. Once these realizations had been determined, they were combined with the three reservoirs to generate the scenario tree. To illustrate the construction of the scenario tree, each random variable associated with the reservoirs was qualitatively described by three, four and five possible realizations for the trees with 27, 64 and 125 scenarios, respectively. Such realizations were (1) low, medium and high for the tree with 27 scenarios (3^3); (2) relatively low, low, medium and high for the tree with 64 scenarios (4^3); and (3) relatively low, low, medium, high and relatively high for the tree with 125 scenarios (5^3). Numerical values obtained using the scenario generation method were associated to these qualitative realizations. Figure 4 illustrates a scenario tree for three realizations and three reservoirs, totaling 27 scenarios.

[Insert Figure 4 here.]

For each of the classes considered in this paper, the minimum and maximum volume of water stored in each reservoir is 270 m^3 and 2000 m^3 for the first reservoir and 270 m^3 and 1000 m^3 for the remaining reservoirs. It was considered that, at the beginning of the planning horizon, reservoir 1 contains 1000 m^3 of water and reservoirs 2 and 3, 500 m^3. The flow rate of the hydraulic pumps located between the wells and the reservoirs is 300 m^3 per hour. For pumps located between the reservoirs, the flow rate is 60 m^3 per hour. The capacity of the wells is considered unlimited. The cost of maintaining the pumps in operation during each sub-period within the planning horizon is 30 monetary units (m.u.), except for peak periods at 19 h, 20 h and 21 h, when this cost is doubled. The cost of electricity consumption associated with the transfer of water between reservoirs is 3 m.u. It is worth mentioning that the values of the data are those used in Toledo et al (2008). The variation intervals for the parameters λ, γ and ρ were assumed to be $[0, 0.05]$, $[0, 0.3]$, and $[0, 50]$, respectively.

6.2 Analysis of the Scenario Generation Method

To assess if the proposed scenario generation method generates reliable solutions, i.e., if it does not cause an instability of the solutions, we evaluate the quality indicators' so-called in-sample and out-of-sample stability according to (Kaut and Wallace 2003; Kaut et al 2007). Basically, in-sample stability requires that the solutions must not vary across scenario trees of the same size. This is evaluated by generating $\mathcal{S}$ scenario trees of identical size and subsequently solving the $\mathcal{S}$ corresponding stochastic programming problems. If the $\mathcal{S}$ values of the objective function are approximately equal, the method is said to have in-sample stability. Out-of-sample stability assesses if the "true" objective value is approximately equal to the objective values from the different scenario trees. To assess the out-of-sample stability, it is sufficient to obtain the objective value of the stochastic programming problem with the "true" probability distribution and to compare it to the objective value of the $\mathcal{S}$ approximate stochastic programming problems. The differences between those objective values are commonly denoted as the *error*.

We verify the in-sample variation of the objective values (expected total cost) across 100 scenario-trees having 27 scenarios each by considering the class 2/27/0% and the proposed four probability distributions. Because the "true" probability distribution is not available, we tested the out-of-sample and error properties by analyzing the most suitable scenario tree size. For this purpose, we solved the stochastic programming model for scenario trees composed of 8, 27, 64, 125, 216 and 343 scenarios. Figure 5 shows the box-plot results over 20 simulations for each scenario tree size, where the squares indicate the average results and the vertical lines link both the maximum and minimum values among all the instances. Even in the largest variation (from 8 to 27 scenarios), the average objective value only increased by approximately 1%. From 125 scenarios on, the variation of the average objective value is negligible for practical purposes and rather small from a theoretical standpoint. Therefore, we consider that $|\mathscr{W}| = 125$ defines a good approximation to the true probability distribution.

From Figure 6, we infer that, independently from the probability distribution, the stochastic programming problems are approximately equal among themselves because the dispersion of the objective values is minimal ($\approx 0.7\%$). The dispersion of errors is relatively low as well ($\approx 0.7\%$), suggesting that the objective values of the approximate problems are approximately equal to the objective value of the "original" problem (i.e., the 125-scenario tree). We consider these values very satisfactory and affirm that the proposed scenario generation method satisfies the properties of in-sample stability and out-of-sample stability and does not subsequently cause instability in the solutions of the SPSP.

[Insert Figures 5 and 6 here.]

6.3 Effect of the Probabilities in the ISPSP Model

We performed an analysis of variance (ANOVA) to determine whether statistically significant differences existed in solving the stochastic programming model using each of the four probability distributions considered in the scenario generation method. The response variables of the ANOVA correspond to the following attributes: the value of the objective function; the first-stage variables X_{jt}, Y_{jt} and Z_{jt}; the second-stage variables $I_{jt}(\omega)$ and $V_{j\ell t}(\omega)$; and the auxiliary error variables $\mathcal{E}^{-}_{jt}(\omega)$ and $\mathcal{E}^{+}_{jt}(\omega)$. Table 4 shows no statistically significant effect of the probability distribution of the first-stage decision variables at a confidence level of 95% ($p-$values range from 0.1683 to 0.2919). The practical advantage of this result essentially lays in the fact that the imprecise information about the demands does not cause serious disruptions in the implementable pump scheduling activities, which in most cases must be taken in advance of the realization of the water demands. In contrast, as expected, the manner in which the scenario probabilities are attributed results in a significant change in the optimal value of the objective function and in the second-stage decision variables ($p-$values range from 0.0001 to 0.0036). Therefore, it is necessary to be cautious when choosing the distribution that will be used in the scenario generation method. For the sake of brevity, we focus the next results on the uniform distribution of the scenario generation method. However, the results for the remaining distributions is similar.

[Insert Table 4 here]

6.4 Analysis of the First- and Second-Stage Policies

This section summarizes the following behavior of the policies related to the collection, transfer and storage of water: (1) the water collection from wells (X_{jt}); (2) pump activation (Z_{jt}); (3) water transfer between the reservoirs ($V_{j\ell t}(\omega)$); and (4) the storage of water volumes in the reservoirs ($I_{jt}(\omega)$):

1. *Water Collection Policies:* Regardless of the number of periods and scenarios, during the peak periods, no water collection is performed or small volumes of water are collected from wells. Consequently, the pumps are not used or are operated for only a small duration thus reducing electricity consumption. In contrast, when there is water leakage, i.e., for classes $|\mathscr{T}|/|\mathscr{W}|/10\%$ and $|\mathscr{T}|/|\mathscr{W}|/20\%$, the results indicate that it is cheaper to meet water demands during critical periods using the water collected in wells during current periods than with water stored in the reservoirs during previous periods.
2. *Pump activation policies:* Although water collection is performed during critical periods, hydraulic pumps do not necessarily have to be switched on to carry out such a collection because this depends on the state of the pumps in the immediately preceding period. The results reveal that, independent of the class, during the peak periods, the pumps are in operation most of the time thus avoiding the fixed costs of switching on the pumps (see Figures 7). In addition, the daily pump activation frequency remains constant as the number of scenarios and periods increase. In contrast, the pump activation frequency increases substantially as the level of water leakage grows, e.g., the activation frequency of pump 1 increases from 7 ($\theta = 0$) to 21 ($\theta = 10\%$), reaching 23 activations for $\theta = 20\%$. From a practical perspective, the advantage of using binary variables to define such decisions lays in the relative ease of implementing the pump scheduling scheme by any system operator by simply analyzing the pump activation frequency and the fraction of the time interval in which pumps remain in operation. Of course, software having a user-friendly interface would be required for this task.

3. *Water transfer between reservoirs:* Based on the solution methodology proposed in this paper, such decisions can only be made during critical periods. Regardless of the class, the volumes of water transferred between reservoirs exhibit a similar behavior. The reservoir of greater capacity (reservoir 1) transfers water to reservoir 2. Some of the water transferred to reservoir 2 is used to meet the demand for this reservoir, and the remaining amount is transferred to reservoir 3. Notice in Figure 2 that because reservoir 1 has no direct connection to reservoir 3, the water must necessarily pass through reservoir 2. Finally, it is also important to mention that the amount of water to be transferred between the reservoirs decreases as the level of water leakage increases (see Table 5) so as to mitigate water losses.
4. *Water volumes in the reservoirs:* As shown in Figure 9, water storage in each reservoir at the end of each hour during the scheduling period exhibits similar behavior. In general, the reservoirs store an adequate volume of water in the periods prior to periods of high demand. Thus, the hydraulic pumps are not being operated or are switched on for a short period of time during the latter periods. However, the volume of water stored in the reservoirs varies according to the level of water leakage in the network. For example, when there is no water leakage in the network, i.e., for classes $|\mathscr{T}|/|\mathscr{W}|/0\%$ large volumes of water are stored in the reservoirs. In contrast, low volumes of water are stored in the reservoirs during the periods prior to the critical periods, when a level of water loss in the network is attributed. In such cases, the solution is to maintain water levels in the reservoirs close to their minimum capacity so that water losses can be minimized.

[Insert Figures 7 here.]

[Insert Table 5 here.]

[Insert Figure 9 here.]

6.5 Model and Solution Sensitivity with Parameters λ and γ

This subsection investigates the sensitivity of the objective value and the second-stage decision variables as λ and γ vary. For the sake of brevity, we present an instance of the class 7/64/10%. Preliminary computational tests suggest that λ and γ vary in the range $[0, 0.05]$ and $[0, 0.075]$, respectively. Tables 6 and 7 summarize the main results, considering the objective value; robustness price, defined as the relative difference between the objective value for $\gamma = 0$ ($\lambda = 0$) and the objective value for the robust (risk-averse) problems; amount of water transferred $V = \sum_{j\ell t\omega} \pi(\omega) V_{j\ell t}(\omega)$; amount of water stored $I = \sum_{jt\omega} \pi(\omega) I_{jt}(\omega)$; total positive deviation $\Delta^+ = \sum_{\omega} \pi(\omega)\Delta(\omega)^+$; total negative deviation $\Delta^- = \sum_{\omega} \pi(\omega)\Delta(\omega)^-$; total water surplus $\mathcal{E}^+ = \sum_{j\omega t} \pi(\omega)\mathcal{E}_{jt}(\omega)^+$; total water shortage $\mathcal{E}^- = \sum_{j\omega t} \pi(\omega)\mathcal{E}_{jt}(\omega)^-$; and the correspondent relative reductions in violations (deviations) as γ (λ) increases.

The observations based on Table 6 and Figure 10 are summarized as follows. As expected, as γ increases at intervals of 0.00375 in the range $[0, 0.0750]$, the expected total cost increases, but the expected shortage/surplus of water in the reservoirs decreases. For $\gamma = 0$, because of the minimization of the expected total cost, no water collection is performed, and, consequently, the total shortage and surplus of water storage are maximized. Notice that this solution is useless in practice. When γ increases from 0 to 0.00375, there is a drastic reduction in the total violation, $\approx 98\%$, at the expense of a minor increase in the expected total cost of 2%. Subsequently, the price of robustness increases substantially, reaching 20% for $\gamma = 0.0675$, although without leading to a significant improvement in the surplus or shortage of water. As robustness is enforced, there is a clear increasing trend in both the level of water in the reservoirs and the amount of water transferred in an attempt to reduce the number of violations. From $\gamma = 0.03375$ on, the volume of water transferred increases by more than 27%, and the water storage level increases almost 29% compared to the first useful solution for $\gamma = 0.00375$. From Table 7 and Figure 11, we note that there is a clear, three-phase transition phenomenon as risk-aversion is enforced. In the first phase, from 0 to 0.0075, positive and negative deviations are insensitive to λ, but the objective value increases accordingly. In the second phase, from 0.01 to 0.045, deviations are greatly reduced - approximately 81% at a negligible price of less than 0.5% in the objective value. In addition, water transfer activities increase by approximately 88%. Finally, in the third phase, deviations decrease by nearly 100% without resulting in a significantly increased cost , but the amount of water transferred continues to increase at a rate three times higher than in the second phase.

[Insert Tables 6 and 7 here.]

[Insert Figures 10 and 11 here.]

7 Expected Value of Perfect Information and Value of the Stochastic Solution

Although many real-world problems involve uncertain parameters, it is common to use strategies to eliminate them to simplify the decision making process. For example, replacing random variables by their expected values is a common strategy to approximate these uncertainties. Another strategy involves solving various deterministic problems corresponding to possible realizations of the random variables and subsequently combininge these solutions using heuristic rules. One question commonly arises: How good is the solution provided by such approximation schemese? From a theoretical perspective, this issue can be addressed using the concepts of the expected value of perfect information (EVPI) and the value of the stochastic solution (VSS). EVPI reflects the maximum value a decision maker would be willing to pay for complete and accurate information about the future. This concept also expresses how much can be saved by having perfect information. The EVPI can be determined as follows: (1) Solve the deterministic problems (wait-and-see problems) associated with each scenario ω and define $\mathrm{WS}^\star(\omega)$ to be the corresponding optimal solution. Next, (2) determine the expected value of the *wait-and-see* solutions using $\mathrm{WS} = \sum_\omega \pi(\omega)\mathrm{WS}^\star(\omega)$. Finally (3), the expected value of perfect information can be determined as the difference between the wait-and-see and the here-and-now solutions, i.e., $\mathrm{EVPI} = \mathrm{RP} - \mathrm{WS}$. The higher the EVPI value, the greater the impact of uncertainty in the problem.

Because the WS strategy provides a set of solutions that correspond to the various scenarios, this strategy is not implementable in practice. However, heuristic rules for combining WS solutions can be used as a strategy to obtain a single feasible policy. A possible strategy involves considering the expected value of the random variable, or the average scenario $\bar{\xi} = \sum_\omega \pi(\omega)\xi(\omega)$, and solving the corresponding deterministic problem. This problem is known in the literature as the expected value (EV). Let $x^\star(\bar{\xi})$, commonly known as the expected value solution, be the optimal solution of this problem. A particularly interesting question arises from the possibility of using the solution $x^\star(\bar{\xi})$ and indicates whether a decision is acceptable when compared to the decision of the stochastic problem. To determine this, one must define the *expected result obtained by using the EV solution*, or simply, the EEV. The EEV problem is the stochastic problem RP itself, where the first-stage variables are fixed in advance based on the solution of the EV problem $x^\star(\bar{\xi})$. The resulting problem is only solved in the second-stage variables, and, for this reason, the EEV problem is said to be *decomposable in scenarios*. The solution of the EEV problem expresses how the second-stage decisions are optimally defined as a function of $x^\star(\bar{\xi})$ and ξ. Finally, the value of the stochastic solution is defined as $\mathrm{VSS} = \mathrm{EEV} - \mathrm{RP}$, and this value shows the gain obtained by explicitly considering the uncertainties rather than by using the approximate expected value strategy. In addition, the VSS also indicates the cost incurred from ignoring the randomness. The EV problem can be defined using the average scenario, the most likely scenario, and/or the worst-case scenario. In all the cases, however, the VSS calculation is performed in a similar manner.

To present the EVPI and the VSS values, 10 instances for each problem class under study were solved. Table 8 presents the average values of the RP solution, WS solution, absolute value of EVPI, relative value $\mathrm{EVPI\%} = (\mathrm{EVPI}/\mathrm{RP}) \times 100\%$, EV solution, EEV solution, and the VSS absolute value and relative $\mathrm{VSS\%} = (\mathrm{VSS}/\mathrm{RP}) \times 100\%$ value for each class. The EVPI results indicate that it would be possible to save a considerable amount of money $-$ on average, $6,631$, or 33.33% $-$ if perfect information about the stochastic parameters are known when the problem is solved. Moreover, the results suggest that, as the number of periods and scenarios increases, randomness plays a more important role in the ISPSP. The analysis of the wait-and-see problems indicates that, in general, the most significant savings obtained by solving the problems with perfect information are given by the water collection costs

($\approx$ 65%) of the total savings, followed by the fixed costs of activating the pumps ($\approx$ 20%), and finally, by the cost of transferring water between the reservoirs ($\approx$ 5%). The remaining cost savings are a result of penalties related to water shortage/surplus.

Although the solution of the first-stage problem of the EEV problem generates lower water collection and pump activation costs compared to the RP problem, such a solution of the first-stage problem of the EEV problem can not meet all the possible realizations of water demand. This results in the second-stage problem of the EEV problem being highly infeasible and, hence, the cost penalty grows at an intractable rate ($\approx$ 75%). The remainder ($\approx$ 25%) is the cost of water transfer between the reservoirs. Regarding the VSS, it is possible to have an indication of the additional gain that would be obtained on average, $3,195$, or 14.13% - by solving the stochastic problem rather than by adopting a strategy to approximate the uncertainties. In this case, it can be said that significant gains are achieved because high values of the VSS and the VSS% were obtained. Therefore, it is more convenient to solve the problem by considering its inherent uncertainties rather than by adopting the approximate solution of the EV problem.

[Insert Table 8 here.]

7.1 Performance of the Proposed Hybrid Heuristic

We use the performance profile technique proposed by Dolan and Moré (2002) to compare solutions provided by the proposed Hybrid Heuristic (HH) to those obtained by the commercial solver CPLEX 12.3 (MIP). Roughly speaking, a performance profile of a solver can be defined as the cumulative distribution function for a given performance metric. The performance of the HH and MIP strategies were evaluated and compared based on the optimality gap metric. The stopping criterion for both strategies was an execution time of 3,600 seconds. If neither strategy finds a solution to a particular problem during this time, it is said that the strategy failed to solve this problem. Figure 12 shows the performance profile of the two solution strategies in relation to the gap by considering 270 instances, i.e., 10 instances of the problem for each of the 27 classes discriminated by $|\mathscr{T}|/|\mathscr{W}|/\theta$.

[Insert Figure 12 here.]

The performance profile curve of the MIP strategy was superior in relation to the HH strategy for approximately 66% of the problems ($\tau = 0$). However, this strategy failed to solve all the problems because its performance curve did not reach the value of 1 for $\tau \in [0, r_M]$. Notice that the MIP strategy solved only $\approx$ 82% of the problems, which means that the probability that the MIP strategy will determine a solution is approximately 82%. The HH strategy, in contrast, managed to solve all the problems because its performance profile curve achieved the value of 1, as can be noticed for τ approximately equal to 0.15. Thus, it can be inferred that the HH strategy was able to solve every problem within a factor of approximately $2^{0.15}$ of the best performance of the MIP strategy. The cumulative distribution function curve relating to the computation-time metric could not be drawn because the MIP strategy required a much greater computation time than the HH strategy to find a feasible solution, and the differences between both were very high. Table 9 presents, for each of the classes considered in this paper, the average computation time required to find a feasible solution using the MIP and the HH strategies. The symbol "$-$" in the tables indicates that no solution was found after 3,600 seconds of computation time. When the MIP strategy managed to find a feasible solution, the corresponding computation time was excessively high in relation to the HH strategy, which took, on average, 15 seconds. Notice that not only does increasing the number of periods and scenarios degrade the performance of both strategies (as expected), but the level of water leakage of the network does as well. However, the increase in the computation times caused by leakage decreases as the number of periods and scenarios increases. Taken altogether, the proposed HH strategy was proven to be efficient in solving practical instances.

[Insert Table 9 here]

8 Summary and Concluding Remarks

Water supply pumping systems in urban networks involve numerous decisions related to when the hydraulic pumps must be activated and deactivated, how long the pumps should operate, the level of water stored in the reservoirs and the transferring policies between reservoirs. These activities are even more complex because water demands are inherently dynamic and involve uncertain values. In this paper, we proposed an improved stochastic programming model for the pump scheduling problem in urban network supply systems, considering the fixed cost of turning the pumps on and off and modeling the water demand as a random variable. We also added to the initial formulation a complete recourse to ensure that no outcome can result in an infeasible solution. Moreover, a risk-averse variability measure was adopted in an attempt to explicitly incorporate preferences of the decision maker into the problem. Because of the combinatorial nature of our stochastic $0-1$ model, we proposed a hybrid heuristic algorithm to determine acceptable solutions using reasonable computation times. The solution method determines the first-stage decision variables by using a constructive heuristic and subsequently fixes the first-stage variables in the mathematical formulation of the ISPSP. To solve the second-stage problem, we used the barrier method without crossover in the commercial solver CPLEX 11.2. The overall results showed that the hybrid heuristic was able to solve all the instances in 15 seconds, on average, whereas the MIP strategy of the CPLEX solver solved only 82% of the instances with a much higher computation time. In addition, the analysis of the decision variables suggested that, in general, water collection is avoided during critical periods, which reduces the pump activation frequency and the overall electricity costs. In contrast, the water leakage level causes extra pumps activations, which may substantially increase electricity consumption costs. For this reason, from a practical viewpoint, it is very important to investigate methods of reducing water leakage in urban networks. The solutions presented in this paper can also reflect different robustness and risk-aversion preferences. Based on the data set analyzed in this application, even the most conservative choices for γ and λ do not lead to unsatisfactory solutions in the sense that both violations and deviations can be substantially reduced at a fair price in the objective value. The high values of EVPI and VSS indicate that the problem of pump scheduling with stochastic water demands is heavily influenced by uncertainties.

For future research, we plan to extend the proposed methodology to consider a multi-stage stochastic version of the proposed model. This would likely require significantly more computational effort; however, more efficient solution methods may be developed. Another promising line of research involves developing alternative interval-based robust optimization models for the pump scheduling problem studied here. Robust optimization includes several appealing features. It is capable of addressing several uncertain parameters while remaining computationally tractable, and it does not require scenario discretization, which can avoid the dimensionality curse, especially in large-scale problems. Although the comparison between robust stochastic and robust optimization models is not trivial, it would be interesting to provide some guidance for their use in the context of water planning and management.

Acknowledgements This work was supported by academic research grants from FAPESP 2013/08303-2 and CNPq 470154/2013-6. These support are gratefully acknowledged. The first author also thank CAPES for his scholarship.

References

Adamowski JF (2008) Peak daily water demand forecast modeling using artificial neural networks. Journal of Water Resources Planning and Management 134(2):119–128

Ahmed JA, Sarma AK (2005) Genetic algorithm for optimal operating policy of a multipurpose reservoir. Water Resources Management 19(2):145–161

Alem D, Morabito R (2013) Risk-averse two-stage stochastic programs in furniture plants. OR Spectrum 35(4):773–806

Almeida R, Barboza PSF (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 - PB, Brasil

Alvisi S, Franchini M, Marinelli A (2007) A short-term, pattern-based model for water-demand forecasting. Journal of Hydroinformatics 9(1):39

Amit RK, Ramachandran P (2009) Optimal design of water distribution networks a review. The Selected Works of R K Amit

Awumah K, Lansey K (1992) Energy efficient pump station operation with a pump switching constraint. In: Water Resources Planning and Management: Saving a Threatened Resource. In Search of Solutions, pp 604–609

Bagirov A, Barton A, Mala-Jetmarova H, Al Nuaimat A, Ahmed S, Sultanova N, Yearwood J (2013) An algorithm for minimization of pumping costs in water distribution systems using a novel approach to pump scheduling. Mathematical and Computer Modelling 57(3):873–886

Barán B, von Lücken C, Sotelo A (2005) Multi-objective pump scheduling optimisation using evolutionary strategies. Advances in Engineering Software 36(1):39–47

Beckwith S, Wong K (1995) A genetic algorithm approach for electric pump scheduling in water supply systems. In: Evolutionary Computation, 1995., IEEE International Conference on, vol 1, pp 21–26

Brion LM, Mays LW (1991) Methodology for optimal operation of pumping stations in water distribuition systems. Journal of Hydraulic Engineering 117(11):1551–1569

Cembrano G, Wells G, Quevedo J, Perez R, Arguelaguet R (2000) Optimal control of a water distribution network in a supervisory control system. Control Engineering Practice 8:1177–1188

Chase D, Ormsbee L (1993) Computer-generated pumping schedules for satisfying operational objectives. Journal / American Water Works Association 85(7):54–61

Coulbeck B (1984) Optimisation of water networks. Institute of Measurent and Control 6(5):271–279

Dolan ED, Moré JJ (2002) Benchmarking optimization software with performance profiles. Mathematical Programming Series A 91:201–213

Dupačová J (2008) Risk objectives in two-stage stochastic programming models. Kybernetika 44(2):227–242

Eker, Grimble M, Kara T (2003) Operation and simulation of city of Gaziantep water supply system in Turkey. Renewable Energy 28(6):901–916

Fallside F, Perry PF (1975) Hierarchical optimisation of a water-supply network. Institute of Eletrical Engineers 122(2):202–208

Fang H, Zhang J, Jin-liang G (2010) Optimal operation of multi-storage tank multi-source system based on storage policy. Applied Physics & Engineering 11(8):571–579

Froukh ML (2001) Decision-support system for domestic water demand forecasting and management. Water Resources Management 15(6):363–382

Gergely J, János C (2012) Finding least-cost pump schedules for reservoir filling with a variable speed pump. Journal of Water Resources Planning and Management 138(6):682–686

Herrera M, Torgo L, Izquierdo J, Pérez-García R (2010) Predictive models for forecasting hourly urban water demand. Journal of Hydrology 387(1-2):141–150

Ikonen E, Selek I, Bene J (2012) Optimization of pumping schedules using the genealogical decision tree approach. Chemical Product and Process Modeling 7(1):1934–2659,

Jain A, Varshney A, Joshi U (2001) Short-term water demand forecast modelling at iit kanpur using artificial neural networks. Water Resources Management 15(5):299–321

Jothiprakash V, Shanthi G, Arunkumar R (2011) Development of operational policy for a multi-reservoir system in India using genetic algorithm. Water Resources Management 25(10):2405–2423

Jowitt P, Germanopoulos G (1992) Optimal pump scheduling in water-supply networks. Journal of Water Resources Planning and Management 118(4):406–422

Jowitt P, Garrett R, Cook S, Germanopulos G (1988) Real-time forecasting and control for water distribuition. Computer Applications in Water Supply 2:329–355

Kaut M, Wallace SW (2003) Evaluation of scenario-generation methods for stochastic programming. In: World Wide Web, Stochastic Programming E-Print Series, pp 14–2003

Kaut M, Wallace SW, Vladimirou H, Zenios S (2007) Stability analysis of portfolio management with conditional value-at-risk. Quantitative Finance 7(4):397–409

Kougias IP, Theodossiou NP (2012) Multiobjective pump scheduling optimization using harmony search algorithm (HSA) and polyphonic HSA. Water Resources Management 27(5):1249–1261

Lansey K, Awumah K (1994) Optimal pump operations considering pump switches. Journal of Water Resources Planning and Management 120(1):17–35

Little KW, McCrodden BJ (1989) Minimization of raw water pumping costs using MILP. Journal of Water Resources Planning and Management 115(4):511–522

López-Ibañez M, Devi Prasad T, Ben P (2005) Multi-objective optimisation of the pump scheduling problem using SPEA2. Evolutionary Computation 1:435–442

Mackle G, Savic D, Walters G (1995) Application of genetic algorithms to pump scheduling for water supply. In: Genetic Algorithms in Engineering Systems: Innovations and Applications, 1995. GALESIA. First International Conference on (Conf. Publ. No. 414), pp 400–405

Magalhães L, Holanda M, Ramos H (2010) Utilização de um algoritmo genético híbrido para operação ótima de sistemas de abastecimento de água. Engenharia Sanitária e Ambiental 15(2):187–196

McCormick G, Powell RS (2004) Optimal pump scheduling in water supply systems with maximum demand charges. Journal of Water Resources Planning and Management 129(5):372–379

Mohamed MM, Al-Mualla Aa (2010) Water demand forecasting in umm Al-quwain (UAE) using the IWR-MAIN specify forecasting model. Water Resources Management 24(14):4093–4120
Mulvey J, Vanderbei R, Zenios S (1995) Robust optimization of large-scale systems. Operations Research 43:264–281
Nasseri M, Moeini A, Tabesh M (2011) Forecasting monthly urban water demand using extended kalman filter and genetic programming. Expert Systems with Applications 38(6):7387 – 7395
Odan F, Reis L (2012) Hybrid water demand forecasting model associating artificial neural network with fourier series, Water Distribution Systems Analysis 2010, chap 115, pp 1287–1305
Ogryczak W, Ruszczyński A (1999) From stochastic dominance to mean-risk models: semideviations as risk measures. European Journal of Operational Research 116:33–50
Ormsbee L, Lansey K (1994) Optimal control of water supply pumping systems. Journal of Water Resources Planning and Management 120(2):237–252
Ormsbee L, Walski TM, Chase DV, Sharp WW (1989) Methodology for improving pump operation efficiency. Journal of Water Resources Planning and Management 115(2):148–164
Ormsbee L, Lingireddy S, Chase D (2009) Optimal pump scheduling for water distribution systems. In: Multidisciplinary International Conference on Scheduling : Theory and Applications, pp 341–356
Pasha M, Lansey K (2009) Optimal pump scheduling by linear programming. In: World Environmental and Water Resources Congress 2009, pp 1–10
Price E, Ostfeld A (2013) Iterative linearization scheme for convex nonlinear equations: Application to optimal operation of water distribuition systems. Journal of Water Resources Planning and Management 139(3):299–312
PROSAB (2009) Uso racional de água e energia: Conservação de água e energia em sistemas prediais e públicos de abastecimento de água. Universidade Federal do Espíritu Santo, programa de pesquisa em saneamento básico.
Reddy MJ, Kumar DN (2006) Optimal reservoir operation using multi-objective evolutionary algorithm. Water Resources Management 20(6):861–878
Rossman L (2002) Epanet 2 users manual. Water supply and Water Resources Division, National Risk Management Research Laboratory, Cincinnati.
SABESP (2004) Programe de controle e redução de perda. Tech. rep., Diretoria Metropolitana de Distribução - M, cia de Saneamento Básico do Estado de São Paulo
Savic D, Walters G, Schwab M (1997) Multiobjective genetic algorithms for pump scheduling in water supply. In: Corne D, Shapiro J (eds) Evolutionary Computing, Lecture Notes in Computer Science, vol 1305, Springer Berlin Heidelberg, pp 227–235
Shvartser L, Shamir U, Feldman M (1993) Forecasting hourly water demands by pattern recognition approach. Journal of Water Resources Planning and Management 119(6):611–627
Sjobom A, Assfalk M (2005) XII-008 - Feramenta para a Previsão de Nível de Rerservatórios em Sistemas de Abastecimento de Água. In: 23 Congresso Brasileiro de Engenharia Sanitária e Ambiental, 1, pp 1–8
Sniedovich M (2012) Robust-optimization: the elephant in the robust-satisficing room. Working paper sm-1-12, Department of Mathematics and Statistics, The University of Melbourne, Melbourne, VIC, Australia.
Solanos J, Montoliu J (1988) The optimal operation of water system. Computer applications in water supply 2:356–375
Soler E (2008) Otimização dos custos de energia elétrica na programação do armazenamento e distribuição de água em redes urbanas. Master's thesis, Instituto de ciências matemáticas e de computação. Universidade de São Paulo.
Sterling M, Coulbeck B (1975a) Optimization of water pumping costs by hierarchical methods. ICE proceedings 59(4):789–797
Sterling M, Coulbeck B (1975b) Technical note. A dymamic programming solution to optimization of pumping costs. ICE Proceedings 59(4):813–818
Takriti S, Ahmed S (2004) On robust optimization of two-stage systems. Mathematical Programming Series A 99:109–126
Toledo FMB, dos Santos MO, Arenales MN, Júnior PS (2008) Logística de distribuição de água em redes urbanas - Racionalização energética. Pesquisa Operacional 28(1):75–91
Whaley RS, Hume R (1986) An optimization algorithm for looped water networks. Pipeline Simulation Interest Group pp 1–19
Yu G, Powell S, Sterling M (1994) Optimized pump scheduling in water distribuition systems. Journal of Optimization theory and Applications 83(3):463–488
Yun R, Singh VP, Dong Z (2010) Long-term stochastic reservoir operation using a noisy genetic algorithm. Water Resources Management 24(12):3159–3172
Zessler U, Shamir U (1989) Optimal operation of water distribution systems. Journal of Water Resources Planning and Management 115(6):735–752
Zhang J, Song R, Bhaskar N, French M (2006) Short-term water demand forecasting: A case study, Water Distribution Systems Analysis Symposium 2006, chap 48, pp 1–14
Zhou S, McMahon T, Walton A, Lewis J (2000) Forecasting daily urban water demand: a case study of melbourne. Journal of Hydrology 236(3):153 – 164
Zhuan X, Xia X (2013) Optimal operation scheduling of a pumping station with multiple pumps. Applied Energy 104:250–257

Table 1: Literature review of pump scheduling problems featuring urban network configuration and the types of hydraulic models, demand models, control algorithm, and control policies realized.

Reference	Number of Reservoirs	Number of Sources	Hydraulic Model	Demand Model	Demand Representation	Risk Measure	Control Algorithm	Control Policy	Transfer Water	Objective
Sterling and Coulbeck (1975b)	Single	Multiple	MB	Lumped	Deterministic	None	PD	Implicit	none	Electrical Energy
Sterling and Coulbeck (1975a)	Multiple	Multiple	MB	Lumped	Deterministic	None	PNL	Implicit	none	Electrical Energy
Coulbeck (1984)	Single	Single	SM	Lumped	Deterministic	None	PD	Implicit	none	Electrical Energy
Whaley and Hume (1986)	Multiple	Multiple	FH	Proportional	Deterministic	None	PNL	Explicit	none	Electrical Energy
Solanos and Montoliu (1988)	Multiple	Multiple	MB	Lumped	Deterministic	None	PD	Implicit	none	Electrical Energy
Jowitt et al (1988)	Multiple	Multiple	MB	Lumped	Deterministic	None	PL	Explicit	none	Electrical Energy
Ormsbee et al (1989)	Single	Multiple	RM	Proportional	Deterministic	None	PD	Explicit	none	Electrical Energy
Little and McCrodden (1989)	Single	Single	SH	None	Deterministic	None	PLIM	Explicit	none	Electrical Energy
Brion and Mays (1991)	Multiple	Multiple	FH	Distributed	Deterministic	None	PNL	Explicit	none	Electrical Energy
Chase and Ormsbee (1993)	Multiple	Multiple	FH	Distributed	Deterministic	None	PNL	Explicit	none	Electrical Energy
Jowitt and Germanopoulos (1992)	Multiple	Multiple	SH	Distributed	Deterministic	None	PL	Explicit	yes	Electrical Energy
Awumah and Lansey (1992)	Single	Multiple	RM	Proportional	Deterministic	None	PLIM	Explicit	none	Electrical Energy
Lansey and Awumah (1994)	Multiple	Single	RM	Proportional	Deterministic	None	PD	Implicit	none	Electrical Energy
									none	and Maintenance
Yu et al (1994)	Multiple	Multiple	FH	Distributed	Deterministic	None	PNL	Explicit	none	Electrical Energy
Beckwith and Wong (1995)	Multiple	Multiple	RM	Proportional	Deterministic	None	AE	Explicit	none	Electrical Energy
Mackle et al (1995)	Single	Single	MB	Proportional	Deterministic	None	AE	Explicit	none	Electrical Energy
Savic et al (1997)	Single	Single	MB	Proportional	Deterministic	None	AE	Explicit	none	Electrical Energy
									none	and Maintenance
Cembrano et al (2000)	Multiple	multiple	MB	Distributed	Deterministic	None	PNL	Implicit	none	Electrical Energy
Almeida and Barboza (2002)	Multiple	Single	FH	Distributed	Deterministic	None	PNL	Explicit	none	Electrical Energy
Eker et al (2003)	Multiple	Multiple	RM	Proportional	Deterministic	None	PNL	Explicit	none	Electrical Energy
McCormick and Powell (2004)	Single	Single	RM	Proportional	Stochastic	None	PDE	Implicit	none	Electrical Energy
Ahmed and Sarma (2005)	Single		MB	Proportional	Deterministic	None	AE-PD	Explicit	none	Meet Demand
Barán et al (2005)	Single	Single	SH	Proportional	Deterministic	None	AE	Explicit	none	Electrical Energy
									none	and Maintenance
López-Ibañez et al (2005)	Single	Single	FH	Distributed	Deterministic	None	AE		none	Electrical Energy
									none	and Maintenance
Reddy and Kumar (2006)	Single		MB	Proportional	Deterministic	None	AE	Explicit	none	Meet Demand
Toledo et al (2008)	Multiple	Multiple	MB	Distributed	Deterministic	None	PLIM	Explicit	yes	Electrical Energy
										and Maintenance
Soler (2008)	Multiple	Multiple	MB	Distributed	Deterministic	None	AH	Explicit	yes	Electrical Energy
									none	and Maintenance
Pasha and Lansey (2009)	Single	Single	SH	Proportional	Deterministic	None	PL	Explicit	none	Electrical Energy
Magalhães et al (2010)	Multiple	Multiple	RM	Proportional	Deterministic	None	AE	Explicit		Electrical Energy
Fang et al (2010)	Multiple	Multiple	RM	Distributed	Deterministic	None	AE	Implicit	none	Electrical Energy
Yun et al (2010)	Single		MB	Proportional	Deterministic	None	AE-AGN	Explicit	none	Satisfy Demand
Jothiprakash et al (2011)	Multiple		MB	Distributed	Deterministic	None	AE-PDE	Explicit	none	Satisfy Demand
Gergely and János (2012)	Single	Single	SH	Proportional	Deterministic	None	AH	Explicit	none	Electrical Energy
Kougias and Theodossiou (2012)	Single	Single	MB	Distributed	Deterministic	None	AE	Explicit	none	Electrical Energy
									none	and Maintenance
Ikonen et al (2012)	Multiple	Multiple	MB	Distributed	Deterministic	None	AH	Implicit	none	Electrical Energy
Bagirov et al (2013)	Multiple	Multiple	RM	Distributed	Deterministic	None	PNL	Explicit	none	Electrical Energy
Zhuan and Xia (2013)	Single	Single	MB	Distributed	Deterministic	None	PD	Implicit	none	Electrical Energy
									none	and Maintenance
Price and Ostfeld (2013)	Multiple	Multiple	MB	Distributed	Deterministic	None	PL	Explicit		Electrical Energy
This paper	Multiple	Multiple	MB	Distributed	Stochastic	Mean	AH	Explicit	none	Electrical Energy
						Absolute			yes	and
						Deviation				Maintenance

MB: Mass Balance; RM: Regression Models; SH: Simplified Hydraulic; FH: Full Hydraulic; LP: Linear Programming; MILP: Mixed Integer Linear Programming.
NLP: Nonlinear Programming; DP: Dynamic Programming; SDP: Stochastic Dynamic Programming; HA: Heuristic Algorithms; EA: Evolutionary Algorithms;
EA-DP: Evolutionary Algorithms-Dynamic Programming; EA-SDP: Evolutionary Algorithms-Stochastic Dynamic Programming.

Table 2: Model dimensions with respect to the number of constraints (m), number of continuous variables (nc), number of $0-1$ variables ($n01$), number of non-zero elements in the constraint matrix (nnz) and matrix density (Den in %).

Horizon	Scenarios	m	nc	$n01$	nnz	$Den(\%)$
	27	4,204	17,689	288	184,123	0.248
2	64	9,569	41,332	288	889,713	0.225
	125	18,414	80,311	288	3,200,759	0.216
	27	14,644	60,889	1,008	643,483	0.072
7	64	33,329	14,252	1,008	3,111,753	0.067
	125	64,134	276,391	1,008	11,198,279	0.063
	27	62,668	259,609	4,320	2,756,539	0.019
30	64	142,625	606,484	4,320	13,333,140	0.015
	125	274,446	1,178,359	4,320	47,986,870	0.014

Table 3: Minimum and maximum water demands (in m^3) in each reservoir.

Hour	$0-6$	$6-7$	$7-8$	$8-9$	$9-10$	$10-11$	$11-12$	$12-13$	$13-14$
Min	10	90	105	121	190	152	167	183	198
Max	30	110	125	141	210	172	187	203	218
Hour	$14-15$	$15-16$	$16-17$	$17-19$	$19-20$	$20-21$	$21-22$	$22-23$	$23-24$
Min	214	229	245	260	198	135	73	10	5
Max	234	249	265	270	218	155	93	30	20

Table 4: Average ANOVA results over 10 instances of the class 2/27/0%. .

	Objective Value	X_{jt}	Y_{jt}	Z_{jt}	$I_{jt}(\omega)$	$V_{j\ell t}(\omega)$	$\mathcal{E}^+_{jt}(\omega)$	$\mathcal{E}^-_{jt}(\omega)$
$p-$value	0.00021	0.1683	0.2919	0.2919	0.00055	0.00360	0.00010	0.00015

Table 5: Amount of transferred water (in m^3) between the reservoir in the critical periods for one instance of each class 2/27/0%, 2/27/10% and 2/27/20%.

	2/27/0%			2/27/10%			2/27/20%		
	Res. 1	Res. 2	Res. 3	Res. 1	Res. 2	Res. 3	Res. 1	Res. 2	Res. 3
Res. 1	0.0	108	0.0	0.0	72	0.0	0.0	48	0.0
Res. 2	0.0	0.0	54	0.0	0.0	42	0.0	0.0	30
Res. 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 6: Results of an instance of the class 7/64/10% for γ varying in the range of $[0, 0.075]$ and for a fixed value of λ equals to 0.05.

γ	Obj. value	Price (%)	V	I	Δ^+	Δ^-	$\mathcal{E}^+$	$\mathcal{E}^+$ Red. (%)	$\mathcal{E}^-$	$\mathcal{E}^-$ Red. (%)
0.00000	9,489	–	0.00	141,270	0.0000	0.0000	5,796	–	21,630	–
0.00375	9,645	1.644	16.35	182,191	0.0000	0.0000	150.2	97.41	422.1	98.05
0.00750	9,751	2.761	17.10	182,103	0.0000	0.0000	147.9	97.45	415.2	98.08
0.01125	9,857	3.878	17.31	182,091	0.0000	0.0000	147.6	97.45	414.3	98.08
0.01500	9,962	4.985	17.75	182,071	0.0000	0.0000	145.5	97.49	414.2	98.09
0.01875	10,067	6.091	17.86	182,067	0.0000	0.0000	145.2	97.49	414.1	98.09
0.02250	10,172	7.198	17.99	182,064	0.0000	0.0000	145.0	97.50	414.0	98.09
0.02625	10,276	8.294	18.26	182,057	0.0000	0.0000	144.5	97.51	413.8	98.09
0.03000	10,381	9.400	18.62	182,050	0.0000	0.0000	144.3	97.51	413.3	98.09
0.03375	10,484	10.49	27.46	181,890	0.0000	0.0000	144.3	97.51	396.6	98.17
0.03750	10,585	11.55	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.04125	10,686	12.61	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.04500	10,788	13.69	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.04875	10,889	14.75	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.05250	10,990	15.82	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.05625	11,091	16.88	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.06000	11,192	17.95	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.06375	11,293	19.01	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.06750	11,395	20.09	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.07125	11,496	21.15	27.46	181,878	0.1800	0.0800	144.2	97.51	395.4	98.17
0.07500	11,597	22.22	27.46	181,877	0.1800	0.0800	144.2	97.51	395.4	98.17

Table 7: Results of an instance of the class 7/64/10% for λ varying in the range of $[0, 0.05]$ and for a fixed value of γ equals to 0.075.

λ	Obj. value	Price (%)	V	I	Δ^+	Δ^+ Red. (%)	Δ^-	Δ^- Red. (%)	$\mathcal{E}^+$	$\mathcal{E}^-$
0.0000	11,541	–	9.640	181,877	15.26	–	5.380	–	144.1	395.5
0.0025	11,550	0.0780	9.640	181,877	15.26	0.000	5.380	0.000	144.1	395.5
0.0050	11,559	0.1560	9.640	181,877	15.26	0.000	5.380	0.000	144.1	395.5
0.0075	11,567	0.2253	9.640	181,877	15.26	0.000	5.380	0.000	144.1	395.5
0.0100	11,572	0.2686	16.90	181,877	3.850	74.77	1.450	73.05	144.1	395.5
0.0125	11,574	0.2859	17.10	181,877	3.600	76.41	1.360	74.72	144.1	395.5
0.0150	11,576	0.3033	17.31	181,877	3.410	77.65	1.290	76.02	144.1	395.5
0.0175	11,578	0.3206	17.75	181,877	3.100	79.69	1.150	78.62	144.1	395.5
0.0200	11,580	0.3379	17.86	181,877	3.020	80.21	1.120	79.18	144.1	395.5
0.0225	11,582	0.3553	18.24	181,878	2.700	82.31	1.070	80.11	144.1	395.5
0.0250	11,583	0.3639	18.24	181,878	2.700	82.31	1.070	80.11	144.2	395.5
0.0275	11,585	0.3812	18.26	181,878	2.700	82.31	1.060	80.30	144.2	395.5
0.0300	11,586	0.3899	18.31	181,878	2.680	82.44	1.040	80.67	144.2	395.5
0.0325	11,588	0.4072	18.34	181,878	2.670	82.50	1.030	80.86	144.2	395.5
0.0350	11,590	0.4246	18.34	181,878	2.670	82.50	1.030	80.86	144.2	395.5
0.0375	11,591	0.4332	18.53	181,878	2.650	82.63	0.960	82.16	144.2	395.5
0.0400	11,593	0.4506	18.62	181,878	2.630	82.77	0.950	82.34	144.2	395.5
0.0425	11,594	0.4592	19.11	181,877	2.500	83.62	0.890	83.46	144.2	395.4
0.0450	11,596	0.4766	19.11	181,877	2.500	83.62	0.890	83.46	144.2	395.4
0.0475	11,597	0.4852	27.46	181,877	0.180	98.82	0.080	98.51	144.2	395.4
0.0500	11,597	0.4852	27.46	181,877	0.180	98.82	0.080	98.51	144.2	395.4

Table 8: Average values of EVPI and VSS.

Class	$\lvert\mathscr{T}\rvert/\lvert\mathscr{W}\rvert/\theta$	RP	WS	EVPI	EVPI%	EV	EEV	VSS	VSS%
1	2/27/0	1,918	1,392	526.5	27.45	1,250	2,262	343.8	15.20
2	2/27/10	2,756	1,824	932.2	33.82	1,530	3,289	532.1	16.18
3	2/27/20	2,767	1,825	942.1	34.05	1,856	3,320	553.1	16.66
4	2/64/0	2,658	2,171	486.9	18.32	1,256	3,170	511.9	16.15
5	2/64/10	2,776	1,826	949.2	34.20	1,540	3,311	535.4	16.17
6	2/64/20	3,310	2,053	1,257	37.97	1,861	3,748	438.1	11.69
7	2/125/0	3,325	2,244	1,081	32.50	1,259	3,919	594.1	15.16
8	2/125/10	3,336	2,052	1,284	38.48	1,545	3,788	451.5	11.92
9	2/125/20	3,341	2,052	1,289	38.57	1,858	3,788	447.0	11.80
10	7/27/0	9,845	7,138	2,707	27.50	5,560	11,205	1,360	12.14
11	7/27/10	9,909	6,696	3,214	32.43	5,568	11,901	1,992	16.74
12	7/27/20	9,942	6,696	3,246	32.65	5,570	11,926	1,984	16.64
13	7/64/0	9,752	6,290	3,462	35.50	5,562	11,294	1,542	13.65
14	7/64/10	9,868	6,696	3,172	32.14	5,569	11,856	1,988	16.77
15	7/64/20	11,800	7,418	4,383	37.14	5,572	13,477	1,677	12.44
16	7/125/0	10,568	7,863	2,705	25.60	5,558	11,801	1,233	10.45
17	7/125/10	11,844	7,415	4,428	37.39	5,564	13,534	1,690	12.49
18	7/125/20	11,897	7,421	4,476	37.62	5,571	13,525	1,628	12.04
19	30/27/0	36,890	24,274	12,616	34.20	20,550	41,712	4,822	11.56
20	30/27/10	42,580	29,099	13,481	31.66	20,567	51,080	8,500	16.64
21	30/27/20	42,785	29,102	13,683	31.98	20,630	51,560	8,776	17.02
22	30/64/0	38,980	26,000	12,980	33.30	20,558	44,472	5,492	12.35
23	30/64/10	42,905	29,102	13,803	32.17	20,570	51,537	8,633	16.75
24	30/64/20	50,764	32,098	18,666	36.77	20,680	58,082	7,318	12.60
25	30/125/0	47,515	32,263	15,252	32.10	20,561	56,337	8,822	15.66
26	30/125/10	51,025	32,100	18,925	37.09	20,572	58,300	7,276	12.48
27	30/125/20	51,191	32,107	19,084	37.28	20,578	58,324	7,133	12.23

Table 9: Average computation time (in seconds) of the MIP and HH strategies to find a feasible solution for the ISPSP instances.

Class	$\lvert\mathscr{T}\rvert/\lvert\mathscr{W}\rvert/\theta$	MIP	HH
1	2/27/0%	7.550	0.100
2	2/27/10%	13.03	0.220
3	2/27/20%	100.8	0.270
4	2/64/0%	120.7	0.650
5	2/64/10%	252.7	0.700
6	2/64/20%	458.9	0.780
7	2/125/0%	1,817	1.170
8	2/125/10%	2,017	1.330
9	2/125/20%	2,396	1.560
10	7/27/0%	2,634	1.890
11	7/27/10%	2,849	2.480
12	7/27/20%	3,048	2.750
13	7/64/0%	3,234	2.800
14	7/64/10%	3,568	3.000
15	7/64/20%	–	3.330
16	7/125/0%	–	3.370
17	7/125/10%	–	4.000
18	7/125/20%	–	4.250
19	30/27/0%	–	5.220
20	30/27/10%	–	5.760
21	30/27/20%	–	6.980
22	30/64/0%	–	7.250
23	30/64/10%	–	10.15
24	30/64/20%	–	15.32
25	30/125/0%	–	74.56
26	30/125/10%	–	77.84
27	30/125/20%	–	86.02

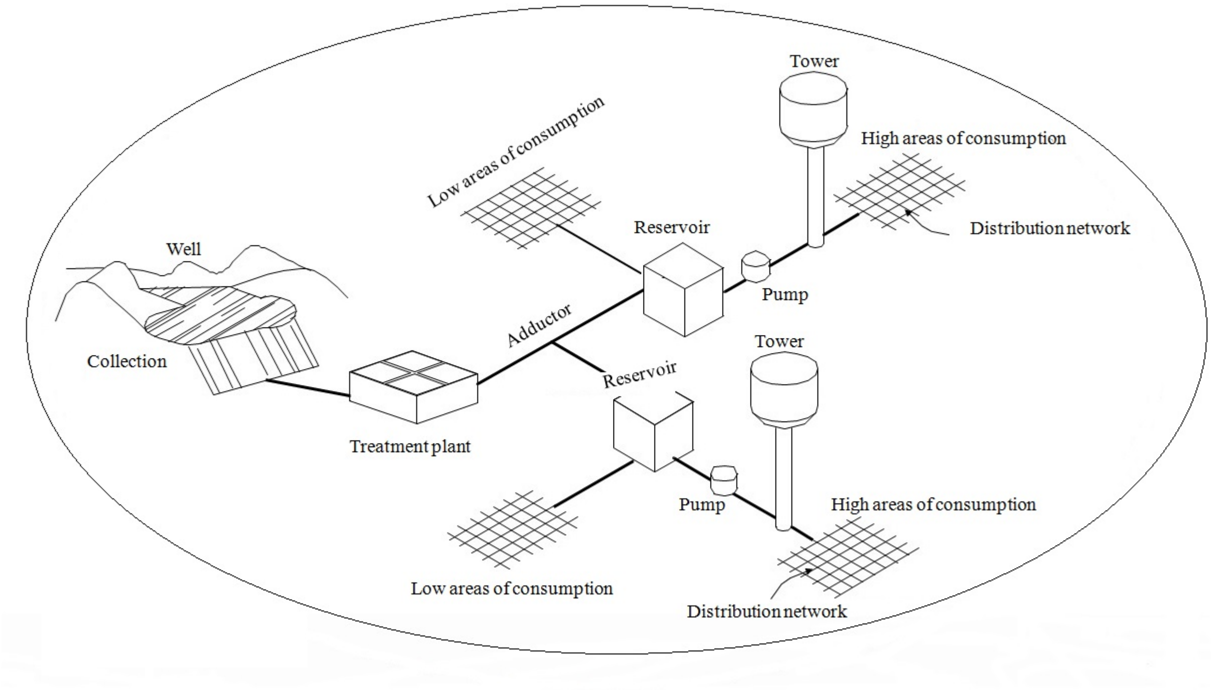

Fig. 1: An example of a water supply system topology. Adapted from SABESP (2004).

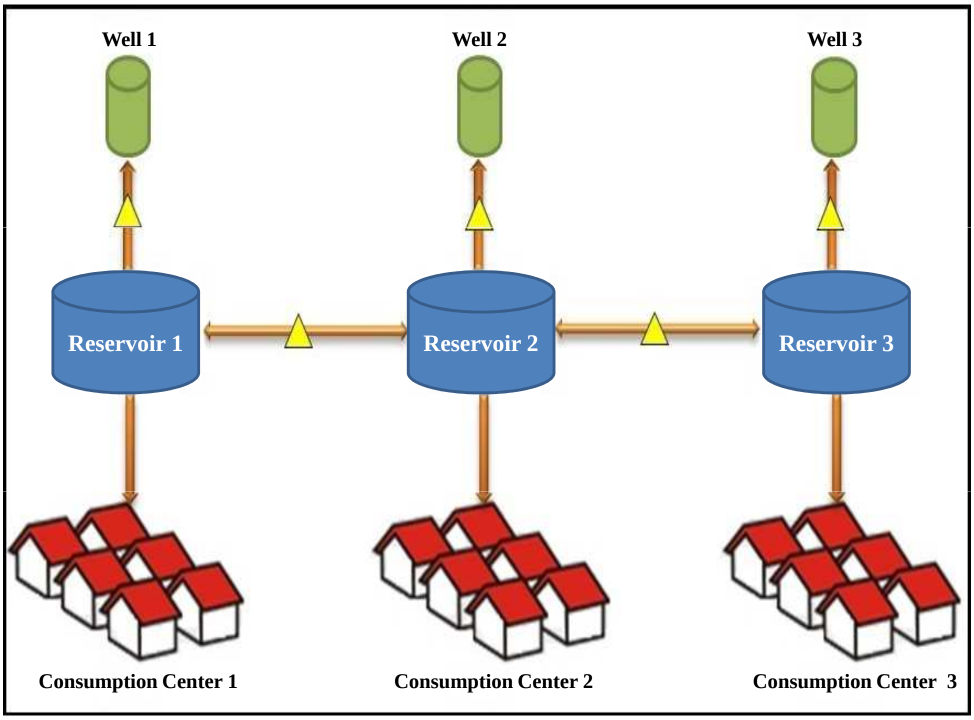

Fig. 2: Example of a Network of Urban Water Supply. Adapted from Toledo et al (2008).

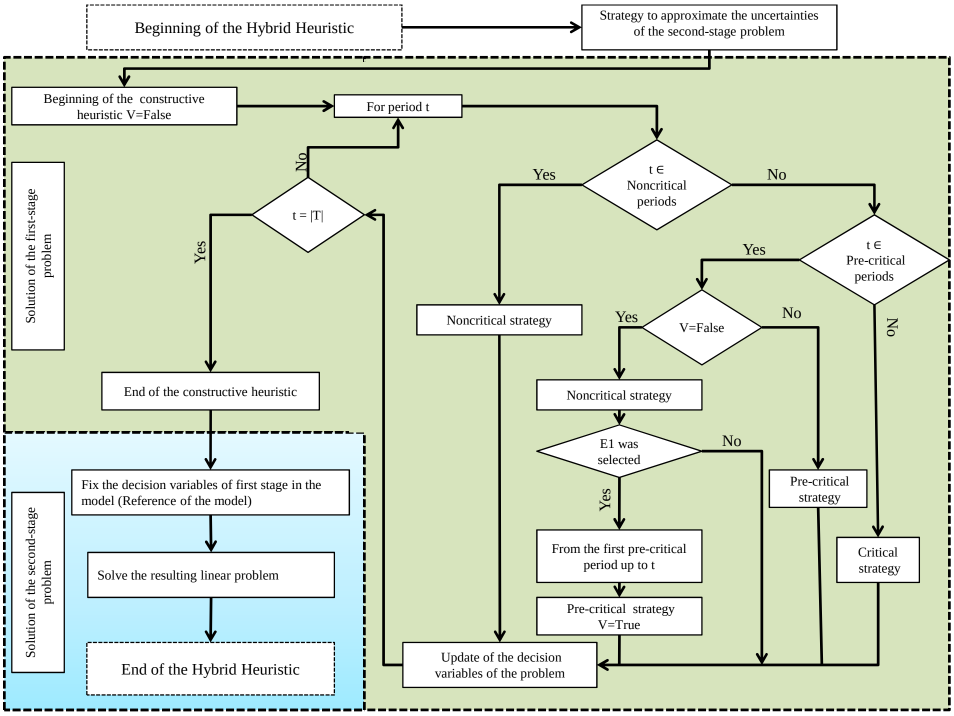

Fig. 3: Flowchart of the proposed hybrid heuristic to solve ISPSP.

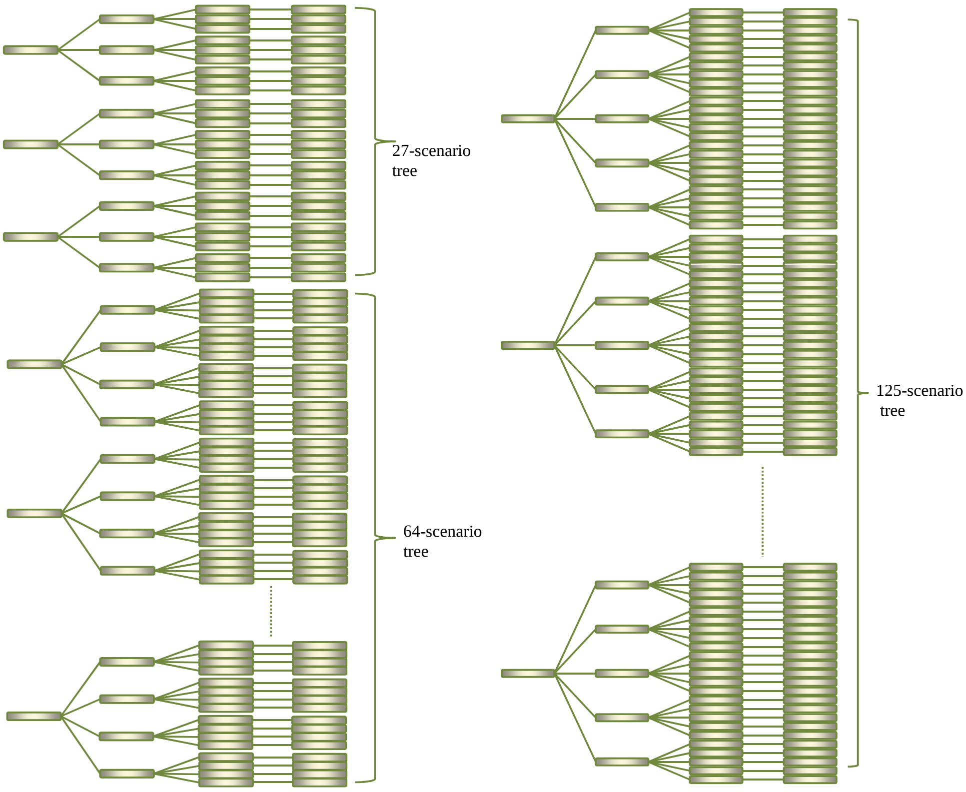

Fig. 4: Scenario trees composed by 27, 64 and 125 scenarios.

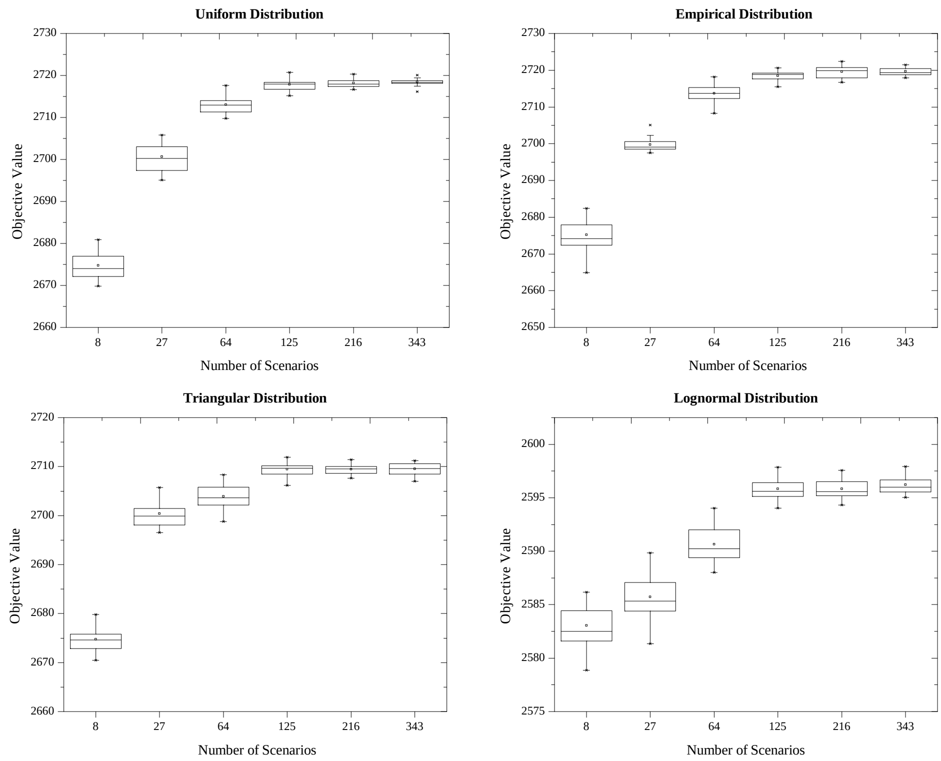

Fig. 5: Boxplot graphs to show the variation of the objective value for each probability distribution and each scenario tree size.

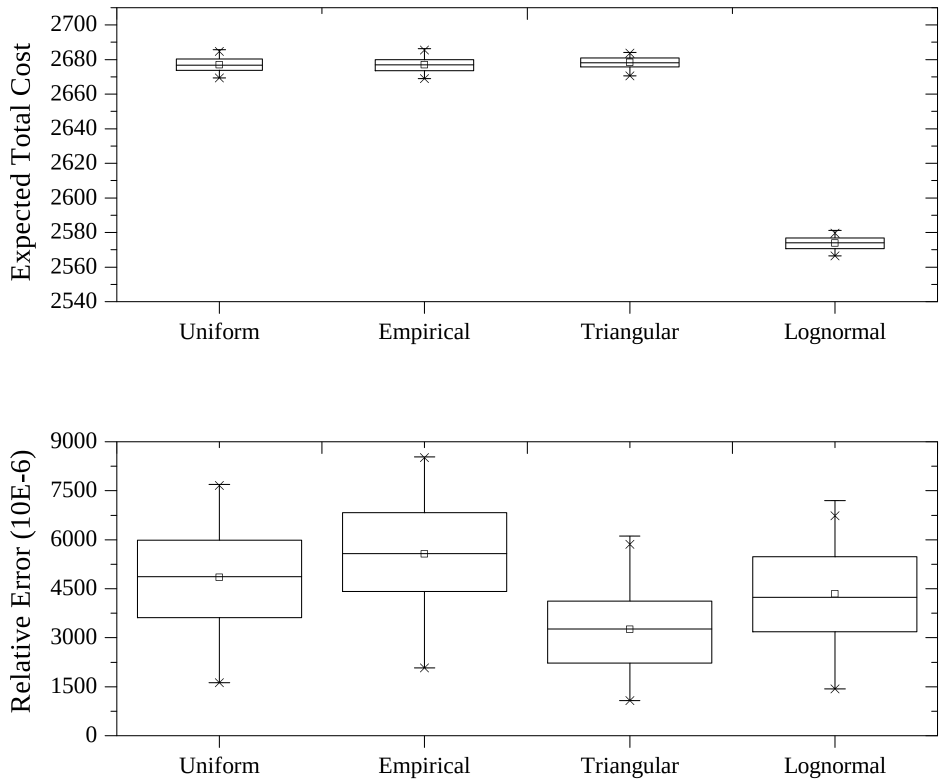

Fig. 6: Boxplot graph to evaluate the in-sample stability (chart above), and out-of-sample stability (chart below) of the proposed scenario generation method.

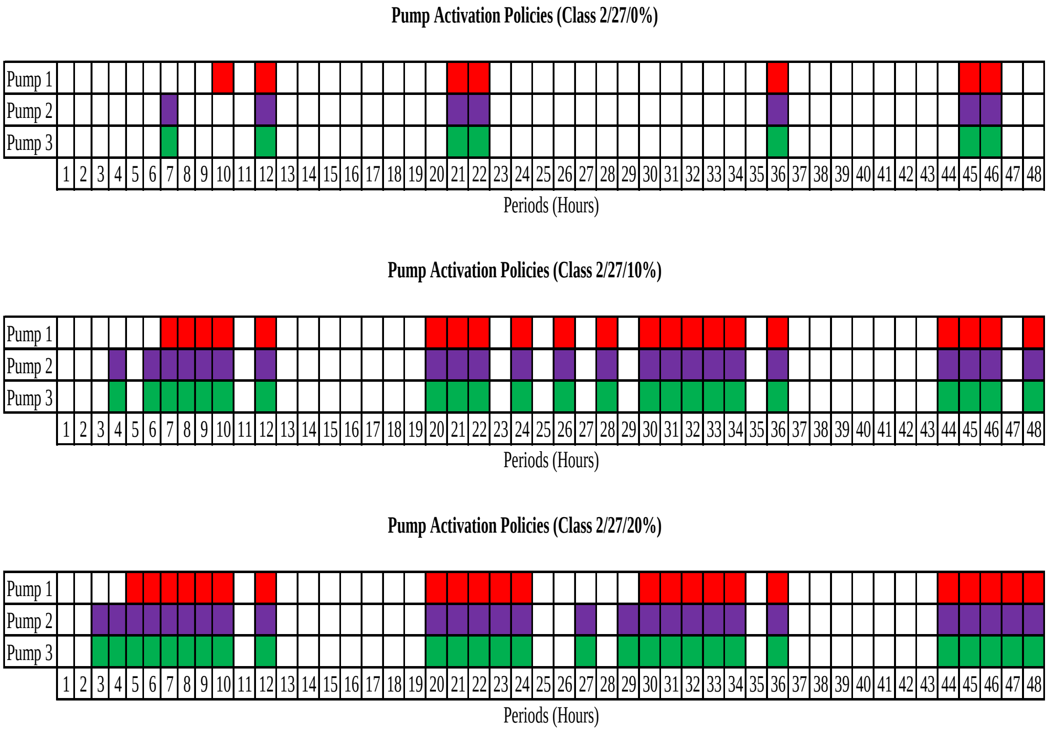

Fig. 7: Gantt Charts associated to the pump activation policies for illustrative instances of classes 2/27/0%, 2/27/10% and 2/27/20%.

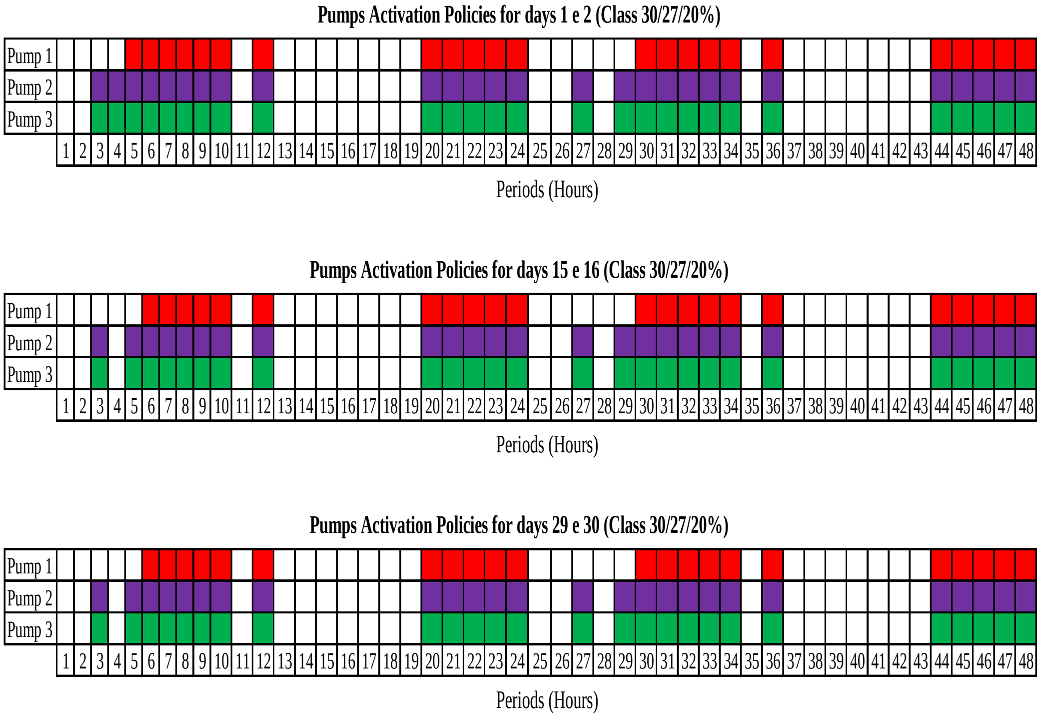

Fig. 8: Gantt diagrams to illustrate the pump activation policies for days 1, 2, 15, 16, 29 and 30 of the corresponding class 30/27/20% .

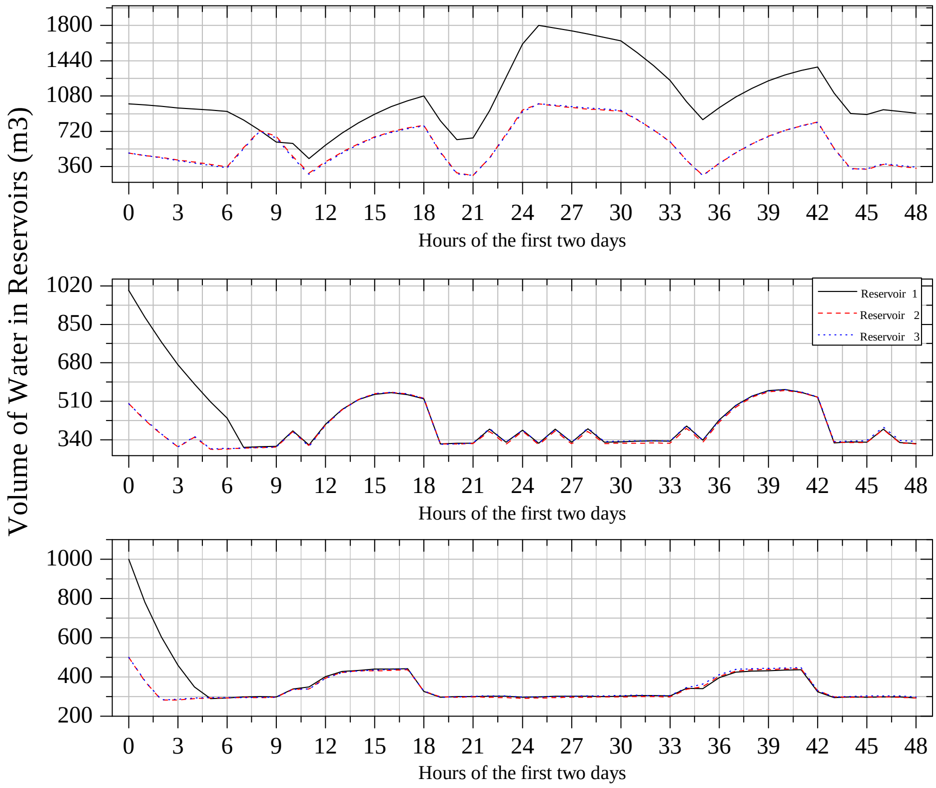

Fig. 9: Volume of water storage in the reservoirs during the scheduling period for one instance of each classe 2/27/0%, 2/27/10% and 2/27/20%.

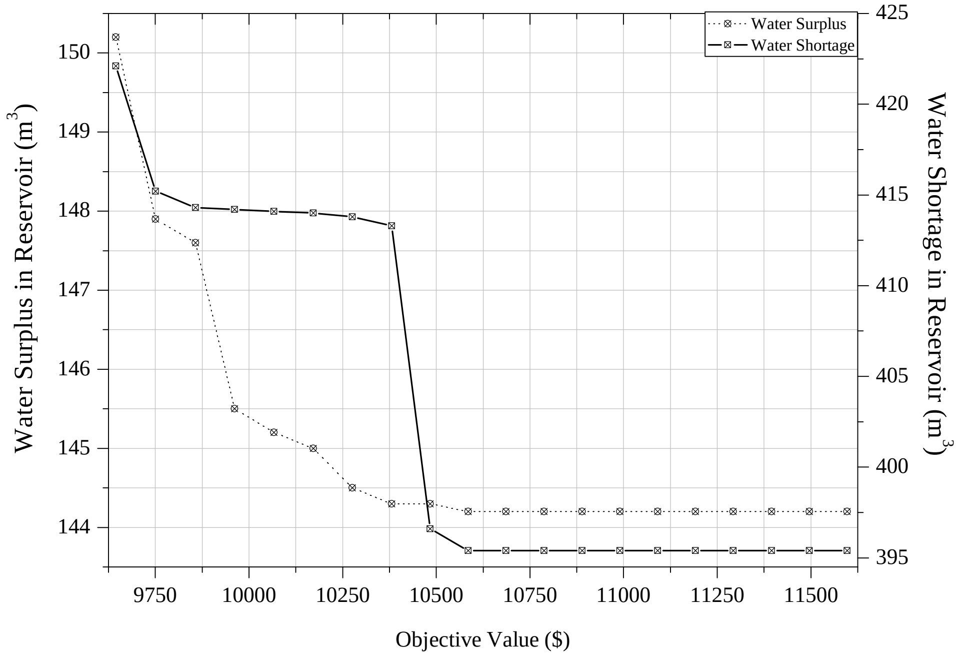

Fig. 10: Tradeoff between the total expected cost and the surplus/shortage of water as robustness is enforced for a fixed $\lambda = 0.05$.

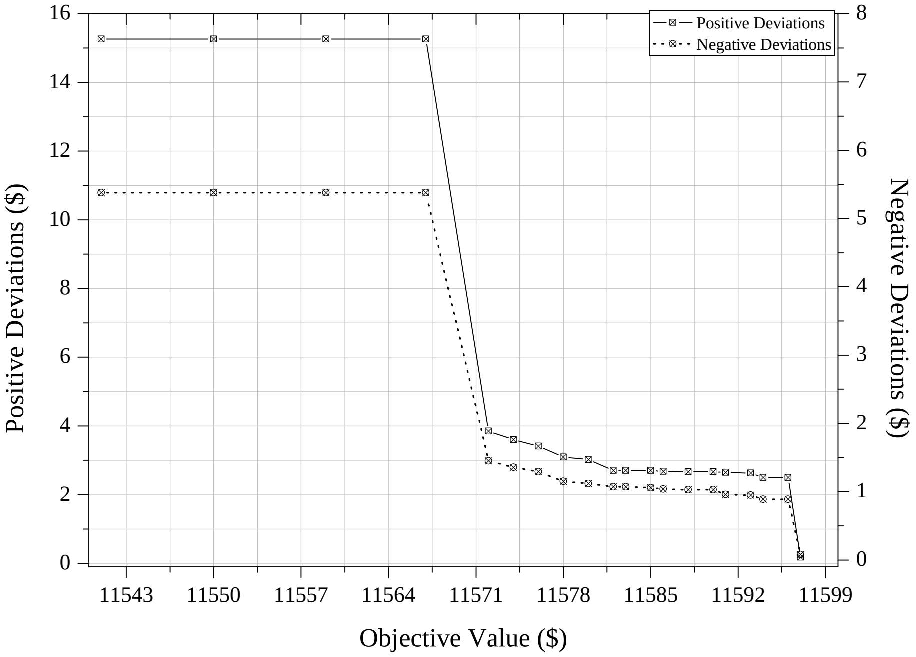

Fig. 11: Tradeoff between the total expected cost and the total positive/negative deviations as risk-aversion is enforced for a fixed $\gamma = 0.075$.

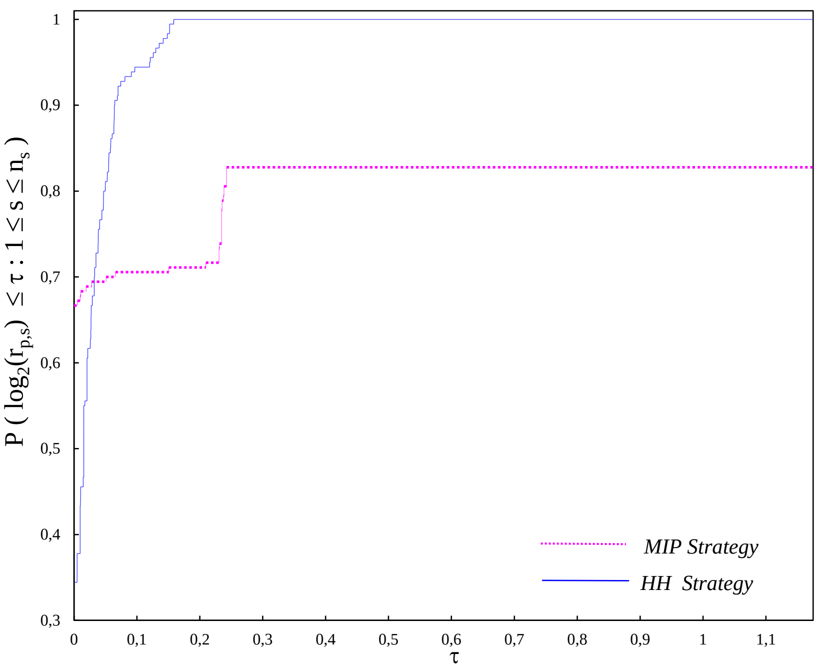

Fig. 12: Performance profile over 270 instances of the two solution strategies in relation to the optimality gap.