Robust optimal design of a tree-based water distribution network with intermittent demand

F. Babonneau[*], D. Gilbert[†], O. Piller[‡] and J.-Ph. Vial[§]

Abstract

This paper discusses the design of a tree-shaped water distribution system for small, dispersed rural communities. It revisits the topic that was discussed in [2] and is nowadays implemented in the field [1]. It proposes a new approach to pipe selection based on robust optimization to account for the uncertainty inherent in intermittent demands. It also proposes a fast projected reduced Newton method of calculating stationary flows to test the performance of the networks thus designed by Monte-Carlo simulation. Numerical experiments conducted on real study cases have shown promising results both in terms of quality and performance of the generated robust solutions and in terms of computation time for simulations.

Keywords Gravity-driven distribution systems, Optimal design, Robust optimization, Simulation.

1 Introduction and outline

This paper discusses the design and analysis of a tree-shaped water distribution system for small, dispersed rural communities. It revisits the topic that was discussed in [2] and was implemented in *NeatWork*, a free access software [1] which has been used in the field for years by the NGO *Agua para la Vida*[1] operating in Nicaragua. The goal of *NeatWork* is to build simple and robust networks that provide stable flows to water outlets. In the context of very poor communities, each outlet typically consists of a single faucet to serve all inhabitants of a household. This configuration is unusual in developed areas and not so much studied in the literature.

In this context, a least-cost design of the water distribution system must take into consideration the users' behavior at each individual faucet, whereas almost all studies of distribution networks focus on systems where distribution points gather a fairly large number of users making their cumulated demand uncertain but smooth and relatively stable. This difference has important implications. In our case of interest, the demand is essentially intermittent, depending on whether the user at the end of the line opens or closes the faucet. The difficulties involved in designing such a low-cost network have been discussed in a previous article. The main

[*]KEDGE Business School, Bordeaux, France and ORDECSYS, Switzerland. frederic.babonneau@kedgebs.com
[†]INRAE, Bordeaux, France.
[‡]INRAE, Bordeaux, France.
[§]UNiversity og Geneva, Switzerland and ORDECSYS, Switzerland.
[1]https://aguaparalavida.org

challenge is, of course, that the system once designed and implemented will have to operate under extraordinarily varied scenarios of open and closed faucets.

The approach in [2] has two facets : a design module, based on a heuristic, and a simulation module to predict the expected behavior in the field. In the present paper we propose a new approach for the design module that better accounts for the inherent uncertainty via Robust Optimization [4] and Chance-Constraint Programming [7]. In the simulation module, we exploit the tree structure of the network to implement the projected reduced Newton method [6]. It considerably speeds up the simulation phase allowing extensive studies by the users.

Let us briefly examine these contributions. The design module seeks a compromise between cost and friction. The larger is the pipe diameter, the greater the cost but the less friction. And the less friction there is, the less the pressure varies as the flow rate varies. The analysis in the least cost design looks at each path from the reservoir to the individual faucets separately. The laws of physics state that the total friction losses induced by the steady-state flow in pipes and end devices (faucets, orifices) exactly match the gravity potential. If the flows were known as deterministic values, it would be easy to find a least-cost design compatible with the physics. This is not the case when demand is intermittent.

In [2] the authors recognize the uncertain character in the number of open faucets downstream of a node along a path to a faucet. They chose to replace it with an ad-hoc deterministic approximation and, at the same time, assumed that the flow rate at each open faucet is exactly the target flow rate. In doing so, they neglect the law of conservation of mass. This is what makes their approach heuristic. In this paper, we propose to deal with the uncertain number of open faucets through Robust Optimization. This approach introduces a safety factor between the deterministic estimate of friction losses and the gravity potential. This safety term is present in the literature on water distribution systems in the form of additional demand requirement, head loss gradient, or overpressure requirement at each outlet point (e.g., [10, 3]). In practice the amount of overpressure is an experimental value, the same for all outlets and often the same for all networks to be designed. Robust Optimization allows it to be determined on an analytical basis. The overpressure is therefore adapted to each faucet and depends on the network, which is a significant improvement. Finally, we show that Chance-Constraint Programming provides the same solution as Robust Optimization. We are thus able to assign a probability to the satisfaction of the constraints.

The second contribution of the present study concerns the simulation module. The approach uses the formulation of the steady-state flow as the minimization of a convex integral function of the energy [12, 9] subject to the linear network mass conservation constraints. By taking advantage of the tree structure of the network it is possible to specialize and simplify the content-based approach [9] used in [1] for computing steady-state flows. The optimization algorithm, i.e., a projected reduced Newton method, is very much the same as in [6]. It proves to be very efficient and provides considerable speedup.

Robust optimization has been used for the least cost design of water distribution systems [13] but not in the context of tree-shape networks with intermittent demands. The literature on methods for solving the water distribution equations is extensive. Efficient algorithms based content-based, or least action principle, and on active sets have been proposed and successfully implemented [11]. We show that in the case of a tree-shaped network, the content-based approach leads to great simplifications and efficiency.

The paper is organized as follows. In Section 2, we introduce the robust optimal design model and derive probabilistic results based on a chance constraint interpretation. In Section 3 we present the projected reduced Newton algorithm used for computing steady-state flows in simulations. In Section 4, we report numerical results on real study cases in Nicaragua. Finally, Section 5 concludes.

2 The optimal design module

The problem is to equip with appropriate pipes a network whose three-dimensional coordinates of the junction and withdrawal points are fixed and whose graph is a tree. If the demand is intermittent, the problem of sizing a tree network with minimum investment cost is a matter of stochastic programming. However the formulation in this framework faces enormous difficulties, if only because of the need to quantify the value of the service provided to the users. Indeed, the water withdrawal management behavior of a user is likely to influence the flow of other users. Quantifying users' dissatisfaction to flow fluctuations at their withdrawal point is difficult, but necessary to enter the framework of stochastic programming because this methodology weighs the objective, the cost of the investment and the uncertain constraints, some of which concern user satisfaction.

The challenge of the design module is to build a network that provides a roughly constant flow at users' outlets while those users withdraw water intermittently. This challenge is arduous because the prevailing context for these rural water distribution systems makes it inappropriate to rely on external flow control systems, whether human or material. On top of that, the objective is to achieve a least cost design, where costs pertain to pipes exclusively.

Given the randomness of the demand, the problem seems to belong to the realm of stochastic programming. However, an intermittent demand that manifests itself by a random opening or closing of the faucets leads to an exponential multiplicity of configurations whose description implies random binary variables. A stochastic programming would also require to quantify the degree of failure when the flow rate at a faucet is too small or too large in certain situation. We propose two approaches to address the random demand challenge. The first simply replaces the cumulative effect of downstream demands on an arc with an average of these demands. This is the approach that was followed in *NeatWork* [1, 2]. The second one, which is new, makes further use of the probabilistic properties of the demands through robust optimization and chance constraint programming. First, it is necessary to specify the probabilistic model of the demand.

2.1 Model of the probabilistic demand

The model is based on several assumptions.
i) Each housing unit is equipped with a single faucet;
ii) The housing units have the same number of inhabitants (typically 6) and their daily water requirements are identical;
iii) A user draws water intermittently. The duration of the withdrawal corresponds to the filling of a standard container. The number of filling sequences during the peak period corresponds to the total requirements of the housing unit during this period;
iv) Users have the same needs and behave identically. Their demands are statistically independent.

When designing a water distribution system, it is often assumed that the total demand during the peak period is a baseline. Let us further assume that the system is capable of delivering

(almost) the same flow rate to each faucet. Given the assumptions i)–iv), we can deduce and quantify the proportion of time each faucet is open during the peak period.
In addition, we ask:
v) The probability that a given faucet is open at any point of time is the same during the period and is equal to the proportion of time the faucet is known to be open during the peak period.

Probabilistic demand model *At any time, the faucet openings are independent and identical Bernoulli random variables, with parameter p.*

This model provides a reasonably sound basis for accounting for the uncertainty of intermittent demands. A similar probabilistic demand model has been already proposed [2, 8]. The assumption that the flow rate at the faucet is close to a fixed target value is a consistency requirement. If a simulation study shows that this is not the case, it means that a user with a lower average flow rate will spend more time filling his containers. The likelihood that his faucet will be opened will be greater. On the other hand, the model is considered valid if the simulation shows that the flows are on average close to the target value.

In the following analyses, the model will be used to define scenarios of demand configurations. A scenario is a set of open faucets (the others remain closed). The number of open faucets downstream of the node is a sum of Bernoulli variables, i.e., a binomial variable with parameter p.

2.2 Deterministic approximation for the probability model

To overcome the challenge of intermittent demands, *NeatWork* in its early release [1] uses a heuristic for the design issue. The idea is to replace the actual flows with a deterministic approximation reflecting the average flow in each pipe as a function of the average number of open faucets downstream of that pipe. This heuristic does not model reality because it does not reflect the essential condition of mass preservation, but it provides a basis for estimating and controlling the head losses in the pipes so as to ensure satisfactory flows at the outlets. Because the design is based on heuristic arguments, it is mandatory to couple the network sizing it generates with a Monte-Carlo simulation of real flows. This is the subject of a later section.

In this section we start with a description of the heuristic exposed in [2] and we will continue with an extension based on robust programming for a better handling of the uncertainty. We first give some notations in Table 1.

In the numerical experiments we used the following values[3] for the physical constants in IS units: $\beta = 8.94 \times 10^{-4}$, $\gamma = 0.53 \times 10^{8}$, $\lambda = 1.781$ and $\mu = 4.781$. The Hazen-Williams formula [11, 14], more widely used in civil engineering, was also tested, with similar results but different physical constants.

The heuristic for the design is the following linear programming problem. In this formulation, the problem is to determine for each arc which piece of each pipe available in the database will be used. The decision variable x_{kd} is the fraction of the length L_k of the arc k that will be filled by the pipe d of diameter Φ_d at the unit cost C_d.

[2]In a tree-based network, such a node is terminal to one incident arc, with a one-to-one relationship between terminal nodes and arcs.

[3]These values have been calibrated for ground conditions in Nicaragua.

$\mathcal{N}_f$	:	Set of terminal faucet nodes ;
$\mathcal{N}_b$	:	Set of branching (intermediary or transit) nodes ;
$\mathcal{N}$	:	Set of nodes downstream the source node; It corresponds to the set of arcs[2];
	:	We have $\mathcal{N} = \mathcal{N}_b \cup \mathcal{N}_f$;
P_t^i	:	Path from node $i \in \mathcal{N}_b$ to the terminal node $t \in \mathcal{N}_f$;
P_t	:	Path from source node S to the terminal node $t \in \mathcal{N}$;
P_{t-}^i	:	Subpath from node $i \in P_t$ down to the immediate predecessor of $t \in \mathcal{N}_f$;
$\mathcal{D}$	:	Set available pipes;
$\bar{Q}$	:	Target flow at each faucet ;
p	:	Probability that a faucet is open ;
L_k	:	Length of the arc ending at node $k \in N_f \cup \mathcal{N}_b$;
Φ_d	:	Internal diameter of pipe $d \in \mathcal{D}$;
C_d	:	Cost per unit of length of pipe $d \in \mathcal{D}$;
Θ_k	:	Number of nodes downstream of node $k \in \mathcal{N}_b$;
θ_k	:	Estimated number of open faucets downstream of node $k \in \mathcal{N}_b$ ($\theta_k \leq \Theta_k$) ;
β	:	Scalar parameter for the pressure loss in the pipes ;
γ	:	Scalar parameter for the pressure loss at the faucet ;
x_{kd}	:	Fraction of the length of arc k that uses pipe d ;
λ	:	Exponent of the flow in the formula for the loss of pressure in the pipes ;
μ	:	Exponent of the diameter in the formula for the loss of pressure in the pipes.
$h_0 - h_t$	:	Static piezometric head, i.e., altitude difference between source node and node $t \in \mathcal{N}$.

Table 1: Notations

$$\min_x \sum_{k\in N}\sum_{d\in\mathcal{D}} L_k C_d x_{kd} \tag{2.1a}$$

$$\sum_{k\in P_t}\sum_{d\in\mathcal{D}} \beta(\theta_k\bar{Q})^\lambda \frac{L_k}{\Phi_d^\mu} x_{kd} + \gamma\bar{Q}^2 \leq h_0 - h_t,\ \forall t \in \mathcal{N}_f \tag{2.1b}$$

$$\sum_{k\in P_t}\sum_{d\in\mathcal{D}} \beta(\theta_k\bar{Q})^\lambda \frac{L_k}{\Phi_{kd}^\mu} x_{kd} \leq h_0 - h_t,\ \forall t \in \mathcal{N}_b \tag{2.1c}$$

$$\sum_{d\in\mathcal{D}} x_{kd} = 1,\ \forall k \in N \tag{2.1d}$$

$$x_{kd} \geq 0,\ \forall k, d. \tag{2.1e}$$

Constraint (2.1b) expresses the condition that the sum of the resisting forces consisting of head losses upstream of the faucet and within the faucet itself match the driving force of gravity. The condition mimics reality because the actual flows upstream of the faucet are not known but just estimated at each node (and pipe incident to it) as a multiple of the target flow. This can be seen in the double sum on the left of (2.1b) in which the the quantity $\theta_k\bar{Q}$ is an estimate of the essentially random flow through arc k. The second term in the left is the impact of the faucet itself at the desired target flow. The slack between the left- and right-hand sides is the excess pressure at the faucet. How this excess pressure is managed in practice will be discussed later.

Constraint (2.1c) concerns intermediary nodes. It aims to protect against possible leakage because in case of leakage the excess of potential gravity over upstream head losses would induce an outward leak, preventing dirt and pollution entering the system. This constraint is not essential. Qualitatively, the optimization goes as follows. The objective promotes a choice of x for the least cost pipe. The cost of a pipe increases as the inner diameter increases, but smaller diameters increase the head losses in the first sum of (2.1b) and (2.1c). The formulation by linear programming allows to find a compromise between the cost and the physical constraints.

In the above formulation the parameter θ_k is a deterministic approximation of the uncertain number of open faucets downstream of arc (node) k. We also denote n the total number of faucets downstream of k, *not including the faucet* t associated with (2.1b). In order to simplify the notations, we drop the index k in the discussion to follow. Recall that the constraint (2.1b) is relevant only when the faucet is open. Hence the number of open faucets (at node k on the path to faucet t) is $\Theta = 1+X$, where $X \backsim \text{Bin}(n, p)$ is a binomial variable. The expected number of open faucets downstream of k is $E(1+X) = 1+np$, with variance $V(1+X) = V(X) = npq$.

Let us explicit the random variable $\Theta = 1 + X$ at the node k along the path leading to faucet $t \in \mathcal{N}_f$. We can write it

$$\Theta_k^t = \sum_{j \in P_t^k} \nu_j$$

where the $\nu_j \backsim \text{Bin}(n_j, p)$ and ν_j is the number of open faucets downstream of $j \in P_t^k$ that connect to the path exactly at j. Note that $\nu_t = 1$ is the only deterministic component since (2.1b) is a relevant inequality at t only when faucet t is open. Thus

$$\Theta_k^t = 1 + \sum_{j \in P_{t-}^k} \nu_j.$$

Clearly,
$\Theta_k^t - 1 \backsim \text{Bin}(\sum_{j \in P_{t-}^k} n_j, p)$ and $E(\Theta_k^t) = 1 + \sum_{j \in P_{t-}^k} n_j p$, and $V(\Theta_k^t) = \sum_{j \in P_{t-}^k} n_j p(1-p)$.

We conclude this section with a discussion of how the expected excess pressure at faucet t is handled in practice. This excess pressure is the slack in constraint (2.1b) of model (2.1). We denote it s_t for convenience. In the physical network it must be absorbed by some device. Remember that the basic assumption on the networks we are interested in is that there is no human or automatic equipment tuning intervention. Users just open and close their faucets. Excess pressure must therefore be absorbed by a fixed and passive mechanism. This is done through orifices. An orifice is a disc of the same diameter as the pipe in which it is placed and which has a small hole in the center. It is placed just upstream of the faucet. It obstructs the pipe, but allows water to pass through its central hole, imposing a pressure loss given by the formula $(\frac{0.59}{\phi})^4 \bar{Q}^2$. It is a very sensitive to variations of the internal diameter and at must be carefully machined. If $s_t > 0$ is the optimal slack for (2.1b), then the orifice has internal diameter

$$\phi = 0.59\sqrt{\frac{\bar{Q}}{\sqrt{s_t}}}.$$

It is a common practice in the design of water distribution system to impose an excess pressure at each water outlet. The reason is intuitive, but hard to establish as a general principle: The higher the residual excess pressure at the outlet, the lower should be the upstream head losses. Smaller head losses are obtained through larger pipe diameters. Finally, the head loss in a larger pipe is less sensitive to the relative variation of flow, a feature that favorizes flow stability. In this paper, we use robust optimization to quantify for each faucet the excess pressure by means of a safety term in (2.1b).

2.3 Robust Optimization

The deterministic approach to the design problem has proven reasonably effective, but simulations sometimes reveal excessive flow variability and even failure for some episodes. Although there is little hope of expressing the problem (2.1) in a tractable stochastic programming formulation, it is worth exploring the alternatives of chance-constraint programming and robust optimization. The idea is ensure that the design will permit constraint (2.1b) to be satisfied with a large enough probability.

In this analysis we face two major difficulties. The first one stems from the uncertain nature of flow that results from the stochastic behavior of the users. The flow is a deterministic function for each scenario of open or closed faucets. This dependence can be made explicit by a computational scheme (see the next section on simulation) but it is not sufficient to work out the chance-constraint formulation. To get around this difficulty we posit that the design will be efficient enough to entail little variability around the target flow $\bar{Q}$. Simulation will justify or invalidate this working assumption on each problem instance. The second major difficulty is that the uncertain factor appears in the nonlinear power function with the exponent $\lambda > 1$. We propose to substitute to the power function a linear function that majorizes the power function in the domain of interest of the uncertain variables ν. Linearization is used in [13] to overcome the appearance of uncertain parameters in nonlinear expressions, but in a different context.

Let us focus on the uncertain components of (2.1b). Define $y_i = \sum_{d \in \mathcal{D}} \frac{\beta L_i \bar{Q}^\lambda}{\Phi_d^\mu} x_{id}$. The constraint is now written (recall that faucet t in (2.1b) is open so $\nu_t = 1$)

$$\sum_{i \in P_{t-}} (1 + \sum_{j \in P_{t-}^i} \nu_j)^\lambda y_i + y_t + \gamma \bar{Q}^2 - (h_0 - h_t) \leq 0 \quad \forall t \in \mathcal{N}_f. \tag{2.2}$$

Our goal is to propose an alternative formulation of (2.2) that would guarantee that the original (2.1b) is satisfied for all values of interest of the uncertain variables ν. Robust optimization [5, 4] is the tool of choice to achieve the goal. The solution is elegant. It adds a deterministic term to the left-hand side of (2.2) that enforces the original inequality in the worst case under consideration. Since the uncertainty stems from independent random variables with known binomial distribution, it will be shown later on that it is possible to assess a probabilistic statement to the constraint satisfaction, in the spirit of Chance-Constraint Programming [7].

2.3.1 Linearization scheme

Our solution starts with a linearization of the power term $f_i(\nu) = (1 + \sum_{j \in P_{t-}^i} \nu_j)^\lambda$ around a nominal value ν^{nom}. We propose the following linear substitute

$$f_i^L(\nu; d) = f_i^{nom} + \sum_{j \in P_{t-}^i} (\nu_j - \nu_j^{nom}) d_j^i, \tag{2.3}$$

with $f_i^{nom} = (1 + \sum_{j \in P_{t-}^i} \nu_j^{nom})^\lambda$. The goal is to give the d-vectors a value that ensures an upper bound on $f_i(\nu; d)$ for the ν-variables within the relevant interval $[\nu^{nom}, \nu^{max}]$. We suggest $\nu_j^{nom} = E(\nu_j) = n_j p$ and $\nu_j^{max} = E(\nu_j) + \kappa\sqrt{V(\nu_j)}$, where $V(\nu_j) = n_j p(1-p)$ is the variance of ν_j. The larger is the parameter κ, the larger the probability that $\nu_j \leq \nu_j^{max}$. However the parameter must satisfy $\kappa \leq \sqrt{n_j \frac{1-p}{p}}$ to ensure $\nu_j^{max} \leq n_j$. A value $\kappa = 1$ already yields a probability around 0.8 for the cases of interest.

We suggest to use the gradient at $\nu = \nu^{nom}$. By convexity of f_i ($\lambda \geq 1$)

$$d_j^i = \frac{d}{d\nu_j}(1 + \sum_{j \in P_{t-}^i} \nu_j)^\lambda\Big|_{\nu=\nu^{nom}} = \lambda(1 + \sum_{j \in P_{t-}^i} \nu_j^{nom})^{\lambda-1} \tag{2.4}$$

provides a lower bound

$$f_i^L(\nu; d^i) \leq (1 + \sum_{j \in P_{t-}^i} \nu_j)^\lambda.$$

Using the linearization scheme and (2.4) we have

$$\begin{aligned}
\sum_{i \in P_{t-}} f_i^L(\nu; d^i) y_i &= \sum_{i \in P_{t-}} f_i^{nom} y_i + \sum_{i \in P_{t-}} y_i \sum_{j \in P_{t-}^i} (\nu_j - \nu_j^{nom}) d_j^i \\
&= \sum_{i \in P_{t-}} f_i^{nom} y_i - \sum_{i \in P_{t-}} \sum_{j \in P_{t-}^i} \nu_j^{nom} d_j^i y_i + \sum_{i \in P_{t-}} y_i \sum_{j \in P_{t-}^i} \nu_j d_j^i \\
&= \sum_{i \in P_{t-}} f_i^{nom} y_i - \sum_{i \in P_{t-}} y_i \sum_{j \in P_{t-}^i} \nu_j^{nom} d_j^i + \sum_{j \in P_{t-}} \nu_j \sum_{i \in P_j} y_i d_j^i.
\end{aligned}$$

For the sake of more compact notation, we shall write the uncertain inequality (2.2) as

$$a_0 + \sum_{j \in P_{t-}} a_j \nu_j \leq 0, \tag{2.5}$$

with

$$\begin{aligned}
a_0 &= \sum_{i \in P_{t-}} f_i^{nom} y_i - \sum_{i \in P_{t-}} y_i \sum_{j \in P_{t-}^i} \nu_j^{nom} d_j^i + y_t - (h_0 - h_t) + \gamma \bar{Q}^2 \\
&= \sum_{i \in P_{t-}} y_i (f_i^{nom} - \sum_{j \in P_{t-}^i} n_j p d_j^i) + y_t - (h_0 - h_t) + \gamma \bar{Q}^2
\end{aligned}$$

and

$$a_j = \sum_{i \in P_j} y_i d_j^i.$$

Note that $d_j^i > 0$ and $y_i \geq 0$. So $a_0 < 0$ is a necessary condition for (2.2) to hold true.

2.3.2 Certainty equivalent for robust optimization

Robust optimization is based on the concept of uncertainty set. In the present context, this set encapsulates all values of the uncertain parameters ν for which it is desired that constraint (2.2) remains satisfied. The set is denoted U_t.

The robust counterpart of (2.2) is

$$a_0 + \max_{\nu_j \in U_t} \sum_{j \in P_{t-}} a_j \nu_j \leq 0, \tag{2.6}$$

in which the uncertain parameters are now variables in the maximization problem. The range of interest for the variables is $0 \leq \nu \leq \nu^{max}$. We propose the budget uncertainty set

$$U_t = \{\nu \mid 0 \leq \nu_j \leq \nu_j^{max}, j \in P_{t-}, \text{ and } \sum_{j \in P_{t-}} \sigma_j^{-1} \nu_j \leq K\}$$

where $\sigma_j = \sqrt{n_j p(1-p)}$ is the standard deviation of ν_j whose distribution is known to be binomial. K is a "safety parameter" chosen by the user. The larger K, the less the chances for constraint (2.2) to be violated.

The worst case analysis for the uncertain constraint is given by the linear program

$$\max_{\nu} \Big\{ \sum_{j \in P_{t-}} a_j \nu_j \mid 0 \leq \nu_j \leq \nu_j^{max}, j \in P_{t-}, \text{ and } \sum_{j \in P_{t-}} \sigma_j^{-1} \nu_j \leq K \Big\}. \tag{2.7}$$

Its dual is

$$\min_{u \geq 0, w \geq 0} \Big\{ \sum_{j \in P_{t-}} u_j \nu_j^{max} + Kw \mid \sigma_j u_j + w \geq \sigma_j a_j, \ \forall j \in P_{t-} \Big\}. \tag{2.8}$$

Therefore, the robust counterpart of (2.2) can be replaced in the original minimization problem by the set of inequalities

$$\begin{aligned} a_0 + Kw + \sum_{j \in P_{t-}} u_j \nu_j^{max} &\leq 0 \\ \sigma_j u_j + w &\geq \sigma_j a_j, \ \forall j \in P_{t-} \\ u \geq 0, \ w \geq 0. & \end{aligned}$$

In the final formulation, we make the inequality stronger by replacing a_0 by

$$\begin{aligned} \alpha_0 &= a_0 + \sum_{j \in P_{t-}} n_j p a_j \\ &= \sum_{i \in P_{t-}} f_i^{nom} y_i + y_t + \gamma \bar{Q}^2 - (h_0 - h_t) - \sum_{i \in P_{t-}} \sum_{j \in P_{t-}^i} n_j p d_j^i y_i + \sum_{j \in P_{t-}} n_j p a_j. \end{aligned}$$

We have

$$\sum_{i \in P_{t-}} \sum_{j \in P_{t-}^i} n_j p d_j^i y_i = \sum_{j \in P_{t-}} n_j p \sum_{i \in P_j} y_i d_j^i = \sum_{j \in P_{t-}} n_j p a_j$$

Hence

$$\alpha_0 = \sum_{i \in P_{t-}} y_i f_i^{nom} + y_t + \gamma \bar{Q}^2 - (h_0 - h_t) > a_0.$$

Finally, the robust deterministic equivalent of (2.1b) is

$$\sum_{i \in P_{t-}} y_i f_i^{nom} + y_t + \gamma \bar{Q}^2 - (h_0 - h_t) + Kw + \sum_{j \in P_{t-}} u_j \nu_j^{max} \leq 0 \tag{2.9a}$$

$$\sigma_j u_j + w \geq \sigma_j a_j, \ u_j \geq 0, v \geq 0, \qquad \forall j \in P_{t-}. \tag{2.9b}$$

At the optimal solution of the global problem, constraint (2.9a) may appear to be saturated while the left-hand side of the robust counterpart (2.6) is strictly negative. Recall that the variable u and w are associated with the dual (2.7) of the inner maximization problem in (2.6). If (2.9b) is satisfied, the component $Kw + \sum_{j \in P_{t-}} u_j \nu_j^{max}$ in (2.9a) provides an upper bound to the inner maximization problem in (2.6), but u and w are not necessarily the optimal values for the dual (2.8).

2.4 Chance constraint interpretation

In the Chance-Constraint Programming approach, we focus on the random variable $Z = a_0 + \sum_{j \in P_{t-}} a_j \nu_j$ and look for a condition under which the probability of constraint violation is small enough probability

$$\text{Prob}(Z > 0) \leq \epsilon.$$

Theorem 2.1. *Under the assumption that the ν_j are independent random variables following binomial distributions with parameters (n_j, p), the constraint*

$$\sum_{i \in P_{t-}} y_i (1 + \sum_{j \in P^i_{t-}} n_j p)^\lambda + y_t + \gamma \bar{Q}^2 - (h_0 - h_t) + \sqrt{\frac{card(P_{t-})}{2pq} \ln \epsilon^{-1}} \max_{j \in P_{t-}} \sigma_j a_j \leq 0 \tag{2.10}$$

ensures that the original uncertain inequality (2.2) is satisfied with a probability at least equal to $1 - \epsilon$. In this formulation $\sigma_j = \sqrt{n_j pq}$ is the standard deviation of ν_j.

Proof. The derivation of the result is based on the observation that $Z > 0$ is equivalent to $sZ > 0$ for all $s > 0$. Let $I(.)$ be the indicator function of a, i.e., $I(a) = 0$ if $a \leq 0$ and $I(a) = 1$ if $a > 0$. Then $\text{Prob}(sZ > 0) = E[I(sZ)]$. Since $I(sZ) \leq e^{sZ}$, we conclude

$$\text{Prob}(sZ > 0) \leq \inf_{s>0} E[e^{sZ}].$$

Let $\Psi(s) = E[e^{sZ}]$ be the moment generating function of Z. In view of the stochastic independence of the ν_j

$$\Psi(s) = E[e^{sa_0} \prod_{j \in P_{t-}} e^{sa_j \nu_j}] = e^{sa_0} \prod_{j \in P_{t-}} E[e^{sa_j \nu_j}].$$

The variable ν_j are binomial variables, taking the integer values $k = 0, 1, \ldots, n_j$. So

$$\begin{aligned} \Psi_j(s) = E[e^{sa_j \nu_j}] &= \sum_{k=0}^{n_j} \pi_k e^{sa_j k} \\ &= \sum_{k=0}^{n_j} \binom{n_j}{k} (e^{sa_j} p)^k (1-p)^{n_j - k} \\ &= (pe^{sa_j} + q)^{n_j} \end{aligned}$$

where $q = 1 - p$. It is convenient to work with $\psi_j(\tau) = n_j \ln(pe^\tau + q)$, with $\tau = sa_j$. Consider its exact Taylor expansion up to order 2

$$\psi_j(\tau) = \psi_j(0) + \psi_j'(0)\tau + \frac{1}{2}\psi_j''(\bar{\tau})\tau^2,$$

for some $0 \leq \bar{\tau} \leq \tau$. Here, $\tau = sa_j$, $\psi_j(0) = 0$, $\psi_j'(\tau) = n_j p \frac{e^\tau}{pe^\tau + q}$ (so $\psi_j'(0) = n_j p$) and $\psi_j''(\tau) = n_j pq \frac{e^\tau}{(pe^\tau + q)^2} \geq 0$ (and $\psi_j''(0) = n_j pq$). We claim that $\psi_j''(\tau)$ achieves an optimum at $\tau_0 = \ln(q/p)$. Indeed, the third derivative

$$\psi_j'''(\tau) = n_j pq \frac{e^\tau (pe^\tau + q)^2 - 2pe^{2\tau}(pe^\tau + q)}{(pe^\tau + q)^3} = n_j pq \frac{e^\tau (q - pe^\tau)}{(pe^\tau + q)^2}$$

is zero at τ_0 with $\psi_j'''(0) = n_j pq(q - p)$. If $p \geq q$, $\tau_0 \leq 0$, and $\psi_j'''(\tau) < 0$ for all $\tau \geq 0$; consequently, $\psi_j''(\tau) \leq \psi_j''(0) = n_j pq = \sigma_j^2 \leq n_j/4$. If $q > p$ (i.e., $p < 1/2$, the more relevant case for our problems of interest), it is easy to see that $\psi_j'''(\tau) > 0$ for $\tau < \ln(q/p)$ and negative otherwise. Consequently $\psi_j''(\tau)$ achieves a maximum value at $\tau_0 = \ln(q/p)$, and for all $\tau \geq 0$:

$$\psi_j''(\tau) \leq \psi_j''(\ln(p/q)) = n_j \frac{q^2}{4q^2} = \frac{1}{4} n_j.$$

Recall that $\tau = sa_j$ so we have

$$\psi_j(sa_j) \leq n_j psa_j + \frac{1}{8} n_j s^2 a_j^2.$$

So

$$P(Z > 0) \leq \Psi(s) \leq e^{(a_0 + \sum_{j \in P_{t-}} n_j pa_j)s + \frac{1}{8}\left(\sum_{j \in P_{t-}} n_j a_j^2\right)s^2}. \tag{2.11}$$

Let us use the notation introduced in the previous subsection

$$\alpha_0 = \sum_{i \in P_{-t}} y_i f_i^{nom} + y_t + \gamma \bar{Q}^2 - (h_0 - h_t) = a_0 + \sum_{j \in P_{t-}} n_j pa_j.$$

We posit[4] $\alpha_0 < 0$.

Since the inequality (2.11) holds for all $s > 0$,

$$\Psi(s) \leq \inf_{s>0} \left\{ e^{\alpha_0 s + \frac{1}{2}\left(\sum_{j \in P_{t-}} \frac{n_j}{4} a_j^2\right)s^2} \right\}.$$

Since $\alpha_0 < 0$, the infimum is a minimum that occurs at $s = -4\alpha_0 / \sum_{j \in P_{t-}} n_j a_j^2$. Consequently: $\Psi(s) \leq \exp(-\frac{1}{2}\alpha_0^2 / (\sum_{j \in P_{t-}} \frac{n_j}{4} a_j^2))$. The equivalence below ensures that $\Psi(s) \leq \epsilon$

$$e^{-2\alpha_0^2 / (\sum_{j \in P_{t-}} n_j a_j^2)} \leq \epsilon \Longleftrightarrow \alpha_0^2 \geq \frac{1}{2} \ln \frac{1}{\epsilon} \sum_{j \in P_{t-}} n_j a_j^2.$$

Therefore the following condition holds

$$\alpha_0 + \sqrt{\frac{1}{2} \ln \epsilon^{-1}} ||z||_2 \leq 0$$

with $z = \{z_j\}_{j \in P_{t-}} = \{a_j \sqrt{n_j}\}_{i \in P_t^-}$ implies $\text{Prob}(a_0 + \sum_{j \in P_{t-}} \nu_j a_j > 0) \leq \epsilon$. The dimension of z is $\text{card}(P_{t-}) \leq N - 1$, that is at most the total number of faucets minus one. In view of it, $||z||_2 \leq \sqrt{\text{card}(P_{t-})} ||z||_\infty$. Since $z_j = a_j \sqrt{n_j} = a_j \sigma_j / \sqrt{pq} \geq 0$, then $||z||_2 \leq \sqrt{\text{card}(P_{t-})} \max_{j \in P_{t-}} a_j \sigma_j / \sqrt{pq}$. Hence the condition

$$\sum_{i \in P_{t-}} y_i (\sum_{j \in P_{t-}^i} n_j p + 1)^\lambda + y_t + \gamma \bar{Q}^2 - (h_0 - h_t) + \sqrt{\frac{\text{card}(P_{t-})}{2pq} \ln \epsilon^{-1}} \max_{j \in P_{t-}} a_j \sigma_j \leq 0$$

ensures that the original uncertain inequality is satisfied with a probability at least equal to $1 - \epsilon$.

$\square$

[4] Recall that $\alpha_0 \leq 0$ is the deterministic substitute of the uncertain constraint (2.1b) in the heuristic approach described in the second section. We could add it as an extra constraint in the chance constraint formulation so as to validate the argument.

The certainty equivalent (2.10) of the original uncertain inequality can be written:

$$
\begin{aligned}
\sum_{i\in P_{t^-}} y_i(\sum_{j\in P^i_{t^-}} n_j p+1)^\lambda + y_t + \gamma\bar{Q}^2 - (h_0-h_t) + \sqrt{\frac{\text{card}(P_{t^-})}{2pq}\ln\epsilon^{-1}}\, w &\leq 0 \\
a_j\sigma_j \leq w, \forall j\in P_{t^-}.&
\end{aligned}
$$

If we define the uncertainty set as

$$U_t = \{\nu_j,\, j\in P_{t^-} \mid \sum_{j\in P_{t^-}} \nu_j\sigma_j^{-1} \leq K\}$$

with $K = \sqrt{\frac{\text{card}(P_{t^-})}{2pq}\ln\epsilon^{-1}}$, we can use the robust counterpart formulation

$$\sum_{i\in P_{t^-}} y_i(\sum_{j\in P^i_{t^-}} n_j p+1)^\lambda + y_t + \gamma\bar{Q}^2 - (h_0-h_t) + \max_{\nu\in U_t}\sum_{j\in P_{t^-}} a_j\nu_j \leq 0.$$

The uncertainty set can be reduced by adding the following condition: $\nu_j \leq \nu_j^{max}$. The relaxed uncertainty equivalent (2.10) of the original uncertain inequality becomes

$$
\begin{aligned}
\sum_{i\in P_{t^-}} y_i(\sum_{j\in P^i_{t^-}} n_j p+1)^\lambda + y_t + \gamma\bar{Q}^2 - (h_0-h_t) + Kw + \sum_{j\in P_{t^-}} \nu_j^{max}u_j &\leq 0 \qquad (2.12)\\
u_j \geq 0,\, \sigma_j\, u_j + w \geq \sigma_j\, a_j, \forall j\in P_{t^-}.&
\end{aligned}
$$

How to choose ν_j^{max}? Considering that the variable is binomial $B(n_j,p)$, the choice

$$\nu_j^{max} = \inf_x\{\text{Prob}(\nu_j \leq x) \geq 1-\eta\}$$

with η small enough, is very sensible. A simple scheme is to take $\nu_j^{max} = n_j p^{max}$ and choose p^{max} to enforce

$$n_j p^{max} = n_j p + \kappa\sqrt{n_j pq}.$$

We want the smallest κ that keeps $\text{Prob}(\nu_j > n_j p^{max})$ small enough.

3 Algorithm for the computation of steady-state flows

Steady-state flow rates in the system depend on how users adjust their valves. We assume that each user adjusts the pressure drop across the valve to achieve the desired flow rate. If the excess pressure is not sufficient to achieve the desired flow rate, i.e., the valve is wide open but the actual flow rate is less than the desired flow rate, the user must accommodate this deficient situation. This is modelled as a nonlinear optimization problem in the flows at the outlets with simple upper bounds on each flow. The solution of that problem may involve flows strictly less than their target values. If the default is small, e.g., less than 10% of the target, the solution might be accepted by the user. Otherwise, the design may be declared inappropriate.

3.1 The content-base formulation of the stationary flows in a tree network

We recall here the formulation of the steady-state flows in a tree network as the solution of a convex optimization problem with simple bound constraints. To this end, we change notation to fit the more conventional notation of optimization problem.

The basic variable is the flow $x \in R^{n_f}$, with $n_f = \mathrm{card}(\mathcal{N}_f)$. The flow in the intermediary branches is denoted $y \in R^{n_b}$, with $n_b = \mathrm{card}(\mathcal{N}_b)$. If S the $n_b \times n_f$ matrix, whose columns are made of 0 and 1 and which determines which branching node is visited by the path from the source to the faucet associated with this column, then $y = Sx$. It is convenient to use u^λ to denote the vector $\{u_i^\lambda\}$ and $\mathbb{U} = \mathrm{diag}(u)$ to denote the square matrix with main diagonal u. Finally, we denote $\mathbf{1}_b$ and $\mathbf{1}_f$ the vectors of all ones with dimension n_b and n_f respectively.

The objective function in the content-base formulation is the integral of the energy. The function involves the following parameters. The first one $\beta \in R^{n_b+n_f}$ is associated with the friction, a quantity that is computed in the design phase. The friction loss in the segment i is $\beta_i y_i^\lambda$, $i \in \mathcal{N}_b$ and $\beta_i x_i^\lambda$, $i \in \mathcal{N}_f$. Their integrals are $\frac{\beta_i}{\lambda+1} y_i^{\lambda+1}$ and $\frac{\beta_i}{\lambda+1} x_i^{\lambda+1}$, respectively. The objective function may include an additional term associated with a passive pressure reduction valve with parameter $\gamma \in R^{n_f}$ and pressure drop $\gamma_i x_i^2$ at the terminal node i. The integral is $\frac{\gamma_i}{3} x^3$. For the sake of condensed matrix notation we denote $\mathbb{G} = \mathrm{diag}(\gamma)$ and $\mathbb{B} = \mathrm{diag}(\beta)$.

The content-base formulation [9] (see also [14]) is the following optimization problem.

$$\min\{f(x) = F_b(y(x)) + F_f(x) + G(x) + E(x) \mid 0 \le x \le \hat{x}\}, \tag{3.1}$$

where

$$F_b(y(x)) = \sum_{i \in \mathcal{N}_b} \frac{b_i}{\lambda+1} y_i(x)^{\lambda+1} = \sum_{i \in \mathcal{N}_b} \frac{b_i}{\lambda+1} (Sx)_i^{\lambda+1} = \frac{1}{\lambda+1} \beta_{\mathcal{N}_b}^T (Sx)^{\lambda+1},$$

$$F_f(x) = \sum_{i \in \mathcal{N}_f} \frac{\beta_i}{\lambda+1} x_i^{\lambda+1} = \frac{1}{\lambda+1} \beta_{\mathcal{N}_f}^T (x)^{\lambda+1},$$

$$G(x) = \sum_{i \in \mathcal{N}_f} \left(\frac{\gamma_i}{3} x_i^3\right) = \frac{1}{3} \gamma^T (x)^3$$

and

$$E(x) = \sum_{i \in \mathcal{N}_f} (h_i - h_0) x_i = \Delta h^T x.$$

with $\Delta h < 0$. The first and second order of the objective are

$$f'(x) = \frac{d}{dx} f(x) = S^T \mathbb{B}_b \mathbb{Y}^\lambda \mathbf{1}_b + \mathbb{B}_f x^\lambda + \mathbb{G} x^2 + \Delta h$$

and

$$f''(x) = \frac{d^2}{dx^2} f(x) = \lambda S^T \left[\mathbb{B}_b \mathbb{Y}^{\lambda-1}\right] S + \lambda \mathbb{B}_f \mathbb{X}^{\lambda-1} + 2\mathbb{G}\mathbb{X}.$$

The Hessian matrix is positive semi-definite because the diagonal matrices $\mathbb{B}_b \mathbb{Y}^{\lambda-1}$, $\mathbb{B}_f \mathbb{X}^{\lambda-1}$ and $\mathbb{G}\mathbb{X}$ are positive.

3.2 Projected Reduced Newton algorithm

Using the compact formulation $\min\{f(x) \mid 0 \leq x \leq \hat{x}\}$,we provide a handy sketch of the projected and reduced Newton method of [6]. This approach is based on active sets. To this end, for a feasible x, we define the partition (I, J) of $\mathcal{N}_f$ and $I = (I^+, I^-)$ according to

$$I^+(x) = \{i \in \mathcal{N}_f \mid x_i = \hat{x}_i \text{ and } \frac{d}{dx_i} f(x) \leq 0\},$$

$$I^-(x) = \{i \in \mathcal{N}_f \mid x_i = 0 \text{ and } \frac{d}{dx_i} f(x) \geq 0\},$$

and $J = \mathcal{N}_f \backslash (I^+ \cup I^-)$. $I(x)$ is the set of active (constraint) indices at x satisfying complementary slackness conditions. I^- is characterised by the fact that the flows are zero and the local steepest descent given by the opposite of the gradient points to negative values of x and is therefore infeasible. Similarly, the descent direction for $i \in I^+$ is infeasible because x_i is at its upper bound $\hat{x}_i$. Denote $H = f''(x)$ be the Hessian matrix of f. Suppose that the rows and columns of H have been reordered so that the first set of columns (and rows) is associated with the active set J. We define the Hessian restricted to the inactive constraints as

$$H^r_{ij} = \begin{cases} H_{ij} & \text{if } i \text{ and } j \in J \\ 0 & \text{if } i \neq j \text{ and } i \text{ or } j \notin J \\ 1 & \text{if } i = j \in I^- \cup I^+. \end{cases}$$

After reordering, the matrix looks like

$$H^r = \begin{pmatrix} H^r_J & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \end{pmatrix}$$

where $\mathbf{0}$ is the zero matrix and I is the identity matrix, both with appropriate dimensions. The reduced Newton direction in the active set solves the equation

$$H^r dx = -\frac{d}{dx} f(x).$$

Clearly, $dx_J = -(H^r_J)^{-1} \frac{d}{dx_J} f(x)$ and $dx_{I^-} = -\frac{d}{dx_{I^-}} f(x) \leq 0$ and $dx_{I^+} = -\frac{d}{dx_{I^+}} f(x) \geq 0$. The projected Newton step is

$$x^{new} = \min\{\hat{x}, \max\{0, x + dx\}\}.$$

Therefore the two active sets $I^-(x^{new})$, $I^+(x^{new})$ and the inactive set J are updated,

$$x^{new}_{I^-} = 0 \text{ and } x^{new}_{I^+} = \hat{x}_{I^+}.$$

Necessary and sufficient optimality condition. *As the problem is convex the first order optimality condition* $\frac{d}{dx_J} f(x) = 0$ *is necessary and sufficient.*

In order to avoid jamming or zizagging, Newton's step is only accepted if it leads to a significant improvement of the objective function f. Otherwise, a backtracking search is performed along the projected Newton direction according to a variant of the Armijo criterion. Since in each iteration we compute the gradient and the hessian, we have a quadratic approximation of f which is likely to be of better quality than the linear approximation. Indeed, the strongly convex function f is quasi-cubic. The implemented variant consists in decreasing the step a in

$x^+ = \min\{\max(0, x + adx), \hat{x}\}$ by a fraction, such as 0.9 until we obtain an improvement of the objective function $f(x) - f(x^+) > \tau(f_Q(x) - f_Q(x^+))$ where f_Q is the quadratic approximation in x and $\tau < 1$ a parameter to be chosen. (In the implementations we chose $\tau = 0.7$.) In practice, the full step 1 is accepted in the immense majority of cases, but the line-search rules out exceptional failures of convergence.

The inactive status for a variable is not necessarily permanent. Akin, in order to avoid jamming or zigzagging, it is desirable to enlarge the set I^- and I^+ to

$$I^- = \{i \in \mathcal{N}_f \mid 0 \ \leq x_i \leq \epsilon \text{ and } \frac{d}{dx_i} f(x) \geq 0\},$$

$$I^+ = \{i \in \mathcal{N}_f \mid \hat{x}_i - \epsilon \leq x_i \leq \hat{x}_i \text{ and } \frac{d}{dx_i} f(x) \leq 0\},$$

for some small enough $\epsilon > 0$.

For convex problems, Bertsekas [6] proved global and superlinear convergence under the condition that the Hessian is bounded.

4 Numerical results

The design and validation process of a water distribution network is done by interaction between the design module and the simulation module. If the statistics of the simulation results show insufficient performances, a new design is sought with more demanding robustness parameters. It is therefore essential that reliable simulations, i.e. with a large sample size, can be performed quickly. A first set of results concerns the behavior of the projected-reduced Newton method. The second set of results displays statistics on a number of selected criteria and discusses the price of robustness to obtain a satisfactory design. The choice of evaluation criteria contains a degree of arbitrariness that requires discussion before implementation and interpretation.

The two sets of experiments were performed on 4 networks that were built in Nicaragua to serve 61 housing units for the smaller network and 269 for the larger one (see Table 2). To begin with, we tested the simulation procedure, which includes scenario generation and calculation of steady-state flows, on each of the dimensioned networks without concern for robustness. The four networks were designed to meet a target flow of 0.12 l/s and a probability of having an open faucet probability $p = 0.2$. The code is written in Matlab R2021b/15 and the computation are performed on an iMac (3.3 GHz Intel core i5-6600K).

Network	number of nodes		iterations			total time
	faucets	transit	min	average	max	seconds
Las Pinares	61	74	2	3.09	5	2.8
Red Mesa	138	172	2	2.92	7	4.7
Ceiba	222	174	2	3.22	7	6.5
Wany	269	289	3	2.98	6	7.1

Table 2: Performance over 10,000 Monte-Carlo simulations. Probability $p = 0.2$.

Table 2 reveals that the projected Newton algorithm is well suited to the problem that it solves in about three iterations for each of the instances. It also shows that Matlab is very efficient, since the solution times per scenario are measured in fractions of a thousandth of a second. Indeed, it was possible to formulate the various operations of the algorithm in the form of vectorized commands, such as vector term-by-term products and the use of efficient Matlab libraries such as Cholesky factorization. It is also possible to verify that these operations take about 70% of the computation time and that the time progression for each of them is linear (product of vectors) or quadratic (product of matrices). The verification was done on the fourth problem by doubling the probability of open faucets. Thus with $p = 0.2$ the problem dimension for a scenario is on average $0.2 * 269 \approx 54$ and with $p = 0.4$ it is about 108.

The quality of service is evaluated on each faucet. The criteria used are the mean of the distribution and two measures on the tails of the distribution: the conditional averages of the 5% smallest flows, thereafter named cVar5% and of the 5% largest flows, thereafter named cVar95%. The situation is all the more favorable as these three quantities are close to the target flow 0.12 l/s. More precisely, the service at the faucet will be considered satisfactory if the conditional average of the 5% smallest flows is higher than half of the target flow, i.e., 0.06 l/s. We also display two additional measures: the ratio between the mean and the standard deviation and the excess pressure at faucets.

The base case corresponds to the deterministic solution without robustness. The goal assigned to the robust solution is to eliminate all cases of cVar5% value lower than the tolerance 0.06 l/s, and of course to do it at the lowest investment cost. Table 3 displays the statistics for the base case and the best robust solution. The increased cost to achieve the assigned goal is the price of robustness.

The average values hide a fairly large disparity between faucets. For example, we can see that the average of the cVar5% on all the faucets is satisfactory for the basic solution, whereas a sometimes quite important number of faucets do not meet the satisfiability criterion. For this reason, we present the results for this category of defective faucets only. Table 3 displays this information.

Numerical results show the contribution of robust optimization on the stabilization of the faucet flows. The failures have disappeared with a reduced variability both globally and on the critical faucets, i.e., the coefficients of variability are divided by a factor 2 globally and by a factor 3 for the critical faucets. As a result, cVar5% are increasing and cVar95% decreasing. The price of robustness varies from 4% to 20% depending on the specific network topology of each case study.

5 Conclusion

The problem addressed in this paper concerns the minimum cost design and simulation of a loopless pressurized drinking water distribution network. The application that motivates this study is service to very poor rural communities where quality drinking water was previously unavailable. This framework requires that attention be paid to installation costs and meeting the demand for housing units. To take into account in the design of the network that individual demands are essentially intermittent, we have developed a robust optimization approach and derived probability results based on a chance constraint interpretation. The quality of the robust solutions is estimated through a Monte-Carlo simulation process that consists in solving non-linear steady-state flow optimization problems for different scenarios. To solve the simulation problem efficiently, we have developed a fast projected reduced Newton method. Numerical

Network	cost	failures	average values: cVar5%	mean	cVar95%	coeffvar	pressure	cost+
				Las Pinares				
Determinist	2257	16	0.074	0.124	0.195	0.234	34.00	
(Critical faucets)		16	0.031	0.127	0.262	0.442	7.80	
Robust	2353	0	0.091	0.121	0.151	0.125	65.47	4.2%
(Critical faucets)		0	0.079	0.121	0.165	0.177	45.04	
				Red Mesa				
Determinist	3547	18	0.083	0.120	0.165	0.166	12.41	
(Critical faucets)		18	0.027	0.122	0.230	0.413	1.98	
Robust	4014	0	0.098	0.120	0.140	0.087	24.57	13.1%
(Critical faucets)		0	0.083	0.120	0.153	0.145	15.11	
				Ceiba				
Determinist	4624	40	0.079	0.121	0.180	0.203	14.39	
(Critical faucets)		40	0.033	0.136	0.295	0.470	3.10	
Robust	4836	0	0.092	0.121	0.151	0.121	26.72	4.6%
(Critical faucets)		0	0.081	0.121	0.161	0.165	18.94	
				Wany				
Determinist	24678	23	0.091	0.121	0.158	0.136	4.19	
(Critical faucets)		23	0.030	0.124	0.232	0.419	3.42	
Robust	29713	0	0.103	0.120	0.134	0.064	52.57	20.4%
(Critical faucets)		0	0.085	0.120	0.148	0.132	24.08	

Table 3: Price of robustness. Statistics based on the full set of faucets and on the subset of critical faucets. Sample size 10,000. Critical faucets are faucets whose cVar5% is less than 0.06 in the deterministic base case.

experiments conducted on real study cases have shown promising results both in terms of quality and performance of the generated robust solutions and in terms of computation time for simulations.

The tools that have been developed for the niche problem of water distribution for very poor rural and small communities turn out to be relevant to address problems of a more industrial nature such as loopless pressurized irrigation systems where demand, and perhaps other factors, are uncertain. This is the subject of current work in progress.

Acknowledgement We thank our colleague Gilles Corcos for his enriching comments and suggestions.

References

[1] F. Babonneau, G. Corcos, L. Drouet, and J.-P. Vial, *NeatWork: Software and user guide*. APLV and Ordecsys, 2002. 1, 2, 3, 4

[2] F. Babonneau, G. Corcos, L. Drouet, and J.-P. Vial, NeatWork: A tool for the design of gravity-driven water distribution systems for poor rural communities, *Informs Journal on Applied Analytics*, 49(2):129–136, 2019. 1, 2, 3, 4

[3] A. Babayan D. Savic, G. Walters, and Z. Kapelan, Robust Least-Cost Design of Water Distribution Networks Using Redundancy and Integration-Based Methodologies, *Journal of Water Resources Planning and Management*, 133(1):67-77, 2007. 2

[4] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009. 2, 7

[5] A. Ben-Tal and A. Nemirovski, Robust solutions of uncertain linear programs, *Operations Research Letters*, 25(1):1–13, 1999. 7

[6] D. P. Bertsekas, Projected Newton methods for optimization problems with simple constraints, *SIAM J. Control and Optimization*, 20:221–246, 1982. 2, 14, 15

[7] A. Charnes, W. Cooper, and G. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil., *Management Science*, 4:235–263, 1958. 2, 7

[8] R. Clément, Calcul des débits dans les réseaux d'irrigation fonctionnant à la demande, *La Houille Blanche*, 5:553-575, 1966. 4

[9] M. Collins, L. Cooper, R. Helgason, R. Kennington, and L. Leblanc. Solving the pipe network analysis problem using optimization techniques. *Management Science*, 24(7):747–760, 1978. 2, 13

[10] E. Creaco, M. Franchini, and E. Todini (2016), Generalized Resilience and Failure Indices for Use with Pressure-Driven Modeling and Leakage, *Journal of Water Resources Planning and Management*, 142(8):04016019-10. 2

[11] S. Elhay, O. Piller, J. Deuerlein, and A. Simpson. A robust, rapidly convergent method that solves the water distribution equations for pressure-dependent models. *Journal of Water Resources Planning and Management*, 142:04015047-12, 2016. 2, 4

[12] J. J. Maugis, Etude de réseaux de transport et de distribution de fluide, *RAIRO Recherche Opérationnelle/Operations Research*, 11(2):243–248, 1977. 2

[13] L. Perelman, M. Housh, and A. Ostfeld, Robust optimization for water distribution systems least cost design, *Water Resources Research*, 49:6795–6809, 2013. 2, 7

[14] O. Piller, B. Bremond, and M. Poulton. Least action principles appropriate to pressure driven models of pipe networks. In *World Water and Environmental Resources Congress*. DOI: 10.1061/40685(2003)113, 2003. 4, 13