NeatWork, a tool for the design of gravity-driven water distribution systems for poor rural communities[1]

Frédéric Babonneau[a], Gilles Corcos[b], Laurent Drouet[c], Jean-Philippe Vial[d]

[a]*Ordecsys, Switzerland and Business School, University Adolfo Ibanez, Chile*
[b]*University of California, Berkeley and APLV*
[c]*Fondazione Eni Enrico Mattei (FEEM)*
[d]*University of Geneva and Ordecsys*

Abstract

NeatWork is an advanced optimization and simulation tool for the design of purely gravity-driven water distribution systems aiming at delivering clean water to poor rural communities. The exclusion of any adjustable devices, such as pumps and valves, for controlling pressures and flows is motivated by two main reasons: firstly, the system should be as simple and as robust as possible to operate; secondly, the operating, maintenance and investment costs must be kept as low as possible, an absolute requirement in very poor rural communities. The challenge of designing such systems at minimum material cost is amplified by the random behavior of users. NeatWork proposes an original heuristic approach to the design of least cost purely gravity-driven systems under stochastic water demands. The concept of gravity driven distribution system has been used with success by the NGO APLV (Agua Para La Vida) for more than 29 years in the center of Nicaragua. The installed systems are still all in use and serve clean drinkable water to over 90 communities encompassing over 40,000 people. The user-friendly version of NeatWork first appeared in 2003 and is commonly used by the local technicians to design distribution systems at lowest material cost. In 2015 NeatWork and its successful use have been selected by the organization Water for Life (CMA) as the winner of the seventh edition of the Water and Sanitation Prize sponsored by the Inter-American Development Bank (IDB) and the FEMSA Foundation.

Keywords: Gravity-driven distribution systems, Design, Simulation, Optimization.

1. Introduction

NeatWork is an advanced computer tool for the optimal design and simulation of water distribution systems that is currently being used by the NGO APLV (Agua Para La Vida) in Nicaragua. It aims at delivering clean water to rural communities with very limited financial and technical competence. The choice of purely gravity-driven water distribution systems excluding any adjustable devices, such as pumps and valves, for controlling pressures and flows, faces two main requirements: Firstly, the system should be as simple and as robust as possible to operate; secondly, the operating, maintenance and investment costs must be kept as low as possible, an absolute requirement in very poor rural communities. NeatWork includes two modules. The design module proposes a least cost

combination of commercial pipes for a predefined topology of a tree-shape network aiming to satisfy the demand at peak hour, taking into account that the localization of open faucets depends on the independent decisions of the users and is thus uncertain. The second module is a simulation tool that permits the user to test a large number of conditions of use resulting form the random user choice of opening or closing the faucets they control.

To meet the assigned goals, the authors had to depart from the large body of existing literature on optimal design of water distribution systems. At the time when NeatWork started being used, and, to our knowledge, up to now, no contribution seems to consider the special problem of ensuring relatively stable deliveries in gravity driven systems with randomly chosen configurations of open/closed faucets. The NeatWork solution method for the design is based on a heuristic that approximates the average flows in the tree network and enables the pipe selection by linear programming. The resulting design hopefully ensures satisfactory deliveries, but this has to be checked by after-the-fact Monte-Carlo simulations. This second module exploits a formulation of the steady state flow solution as the result of the minimization of the total rate of dissipated energy. This leads to a strongly convex minimization programming problem which is solved very efficiently by the embedded commercial code MOSEK [10].

There is a large body of literature on the optimal design of water distribution systems. The common practice is to use a mathematical model based on the continuity equations (conservation of masses) and on the continuity of the hydraulic grade line[2] to represent the steady state flow in a network. The grade line follows the experimental Colebrook law of energy loss of a steady state flow in a pipe. As early as 1936, Hardy Cross [4] proposed a method to compute the steady state flows in a general network, i.e., a network including loops, as the solution of a system of linear and nonlinear equations. Alternative and more efficient methods have been proposed since. For instance, the freely distributed software EPANET [12, 13, 6] is based on the gradient method of [14]; it analyzes and simulates the hydraulic behavior of large networks including loops, but it does not deal with the optimal design problem. In most papers dealing with the optimal design problem, the pipe material cost is the major, if not unique cost component in the objective function. Models and methods have been proposed to deal with the interaction between the cost minimization and the mathematical representation of steady state flows. Paper [1] formulates the joint problem of minimizing cost and serving a prescribed demand as a nonlinear programming problem. The variables are the flows and the pipe diameters; the constraints are the mass conservation and the grade line equations. The latter are nonlinear; they involve the absolute value of flows and, as such, are nondifferentiable. This nonconvex nondifferentiable problem is hard to solve and the solution methods generate local optima, possibly suboptimal. Paper [1] proposed the so-called Linear Programming Gradient (LPG). In [5] the authors developed an efficient branch and bound scheme to compute a global solution.

In all the quoted studies the outlets are given with fixed demands or possibly demands varying around some value. None of these studies consider the case of open/closed faucet configurations, a situation that the systems conceived by APLV are facing. This has been the main motivation for developing the dedicated tool NeatWork.

In [9], a paper devoted to the related problem of gas transmission, the author noticed that integrating the Colebrook law with respect to the flow yields an expression whose interpretation is the total energy dissipation in a pipe as a function of the flow. This opens the possibility of computing the steady state flows in an existing network as the solution of a strongly convex

---

[2]The hydraulic grade line differs from the energy line by the dynamic head. In problem instances considered with NeatWork we consider that difference negligible. We shall use one term or the other indifferently.

programming problem whose objective is the steady state total energy per unit of time and the only constraints are the mass conservation. The embedding of the total energy approach into the optimal design has been done in [3], but it is not clear that it can be used in the context of stochastic open/closed faucet configurations. In the present paper we use the the total energy approach only to compute flows in a simulation module.

The paper is organized as follows. In the next section, we discuss at some length the context where APLV operates and explain why it leads to the development of a dedicated optimization and simulation software. We then proceed with a description of the two modules, design and simulation. We point out that our approach based on the principle of minimum action allows the simulation of steady state flows on networks with an arbitrary number of loops. In the next section we discuss some implementation issues and give an example of a real project. In the last section we conclude with a discussion of suitable extensions and the opportunity for NeatWork to be used in different but related contexts. In the appendix we detail the mathematical models for the design and the simulation.

2. Water distribution for very poor communities in Nicaragua

While the growing scarcity of fresh water is a grave contemporary problem, the need for its availability to all of earth's inhabitants is an age-old one: Its global resource is getting short but its fair distribution remains a fundamental right, conspicuous by its absence. And for human consumption, this water needs to be not only fresh but also potable. The number of deaths, principally of infants and babies, due to polluted water exceeds by far those due to natural catastrophes and rivals those due to man himself.

To remedy this crying need, it is unrealistic in poor countries to depend either on their governments –national or local– or, in rural areas, on private water distribution companies; the first, because both their tax revenues are too meager and local competence is often lacking; the second, because the investments required cannot be justified by revenues derived from exceedingly poor populations.

This is truest for the small isolated villages of Nicaragua, which are the "clients" of the NGO Agua Para La Vida (APLV). APLV [2] has been active in Nicaragua since 1987, designing and supervising the construction of drinking water systems and associated sanitary facilities in those small villages. It has also created a technical school focused on imparting to selected young farmers the knowledge required to design and supervise these projects without the help of engineers. Moreover it has developed and made available without charge to other similar NGOs original computer tools that greatly facilitate and improve the design of key components of these water systems.

If we focus on the population to whom access is almost completely denied, the rural inhabitants of small and destitute villages, there are a number of ways to bring it water: thru wells, rain harvesting, desalination, and convection from distant sources. The choice of a particular method will depend strongly first on the geomorphology of the land surrounding the village, then on the local rainfall distribution, and importantly as well on the relative initial and maintenance costs and on the simplicity of the system. If we grant that separate individual supply to each home is rarely a rational choice, though occasionally the only possible one (rain harvesting requiring almost continuous rainfall), no matter how the water is conveyed to the site of the village, it will have to be distributed from some accumulation device such as a tank to the sites of consumption. It is with this aspect of the distribution problem that this presentation is concerned.

Assuming that i) Water has to be distributed in the simplest and most economical possible way to a rural population which needs to make only the most elementary use of that water (to fill containers or use the water at points of delivery for washing self or clothes) and ii) Maintenance

should be performed by the beneficiaries and so needs to be as simple as possible, what is the most satisfactory means of engineering that type of distribution? It has to be dependable the year around; it should deliver flow rates within prescribed limits to individual faucets under variable conditions of use; and it should be as free of maintenance and inexpensive as possible. For small NGOs with limited available funds like APLV the latter consideration is particularly important. The costs components consist of investment and operations. The investment costs are essentially material costs (primarily the pipes), especially, for the specific Nicaragua context, for which, housing is particularly dispersed: A village of 150 inhabitants may require more than 10 kms of piping for its distribution system. The cost of manpower is negligible because it is made up of the un-paid beneficiaries and the costs of operations should be minimized because the fund providers to whom NGOs like APLV have access are reluctant to fund them.

We concluded that gravity, excluding the use of pumps and pressure regulating devices, best answered this search by almost eliminating maintenance and operations costs. However, regulating flows on such networks under variable conditions of use is a real challenge. Even at peak hours not all users open their faucet at the same time. Still, one would like the flow at each open faucet to remain about the same, independently of the user behavior at the other faucets. NeatWork proposes an original approach to the optimal cost design problem under the constraint of satisfactory faucet independence. It is based on a heuristic. The actual performance of the design is checked by simulation. The two modules, design and simulation, are discussed in the next sections.

3. Flow simulation on general networks

The simulation tool in NeatWork is aimed at checking the actual behavior of a system produced by the design tool. For the designer this module comes in second position, to test a proposed design. However, we start with its description because it sheds light on the physics of the system.

The steady-state flow in a predefined water distribution system is fully determined by its physical characteristics: the topography, the node elevations, the link lengths, the pipe diameters, and, additionally the hydraulic properties of the faucets at the outlets. The flow in pipes induces head losses in the pipes and faucets. On any path from the inlet to an outlet the head losses are cumulated. The total headloss on such a path represents the rate of total energy dissipation per unit of time and per unit of flow. The law of energy conservation implies that the dissipation energy rate matches the gravity energy rate. The latter is equal to the difference in elevation at the two extremities of the path (we disregard here the headloss induced by the faucet). The standard mathematical representation of the headloss in a pipe is proportional to the length of the pipe, to a power p of the flow ϕ and an inverse power q of the diameter D. To account for the uncertain sense of the flow on loops, one considers that the graph is oriented. The headloss on the oriented link (i, j) takes the form

$$\gamma_{ij} L_{ij} \frac{\phi_{ij} |\phi_{ij}|^{p-1}}{D_{ij}^q} = \pi_i - \pi_j. \tag{1}$$

The constant γ depends on the pipe characteristics, e.g., the degree of roughness. The quantities π_i and π_j are the head pressures at the extremities of the link. Equation (1) expresses the steady state condition as an equilibrium state.

Accordingly, the steady state flows in a network (possibly involving loops) are solutions of two systems of equations: a set of nonlinear equations which enforce the conservation of energy in loops and along paths to the inlet and all outlets, and a set of linear equations enforcing conservation of masses. Solving these joint systems of equations is not an easy task, because all loops have to be identified and the sense of the flows has to be guessed. Fortunately enough, there is an alternative to the problem formulation. Both sides of equation (1) can be integrated with respect to the flow.

As pointed out in [9], the integral on the left hand side is the rate of energy dissipated by friction in the pipe, while the right hand is the negative of the gravity energy. Summing the two terms yields the total energy

$$\mathcal{E}_{ij}(\phi_{ij}) = \gamma_{ij} \frac{L_{ij}}{D_{ij}^q} \frac{|\phi_{ij}|^{p+1}}{p+1} + (h_i - h_j)\phi_{ij}, \tag{2}$$

where h_i and h_j are the elevations of the end nodes of the pipe. If there is a faucet at one of the end node a term in the power $p+1$ of the exiting flow has to be added.

Summing the rates of energy losses in the network, due to friction in the pipes and faucets, and gravity, one obtains the total rate of energy dissipated in the network in a steady state. The principle of minimum energy applies and the search for the steady state solution can be formulated as a nonlinear programming problem of energy minimization under the constraint of mass conservation (see Appendix). It is straightforward to check that the first order optimality conditions yield the very same systems of equations that those used in most published works. This approach converts the equilibrium problem of solving a set of nonlinear equations into an optimization problem. It has several advantages. The problem is strongly convex, so the solution is unique. The problem is easy to solve. Finally, the sense of the flows in a loop need not be guessed by trial and errors: It directly results from the optimization process.

4. Optimal design on an arborescent network

If the pipes diameters are given on each link, the flows are uniquely determined by physical laws for each configuration of open/closed faucets. Since the configuration of open/closed faucets depends on the independent random choice of the users, the resulting flows in the network is stochastic. For some configurations, the flows at the outlets may fall within an admissible range; for some others, the flows may be out of bounds. Some faucets may dry up, some others may overspill. The challenge is to find a least cost stable design. By stable we mean that the system ensures "acceptable flows" at each faucets "most of the time". If the total pipe cost is not a priority, as is the case in non-so-poor countries, stability can be achieved by a drastic reduction of the headloss ahead of all faucets. For example, this can be achieved by imposing very large pipe diameters everywhere in the network and installing regulating devices, named orifices[3], in front of each faucet. The headloss is thus concentrated at the faucet and can be adjusted to meet the gravity pressure drop. In this way, the flow at each faucet becomes independent of the flows at all other faucets, and is thus independent of the open/closed faucets configuration. If the material cost of the pipes is a major concern, as it is the case in a very poor environment, the above approach is not viable. One must select cheaper pipes with smaller diameters. The headloss in the pipes ceases to be negligible and the faucet flows become more and more interdependent. Stability is lost. The whole purpose of the design module is to find a reasonable compromise between the material cost and the stability.

Contrary to the simulation module, the design module applies to arborescent networks only. In such a tree, the knowledge of the flows at the leafs (faucets) uniquely determines the knowledge of the flows in each link of the tree. An easy, but unrealistic, version of the design problem would be to consider only the deterministic case when all faucets are simultaneously open and each faucet delivers the required target flow. In such a case the flow in a given link should be equal to the faucet target flow multiplied by the number of faucets downstream the link. The headloss on that

---

[3]An orifice is device consisting of a disk with a small hole at its center. The disk diameter matches the diameter of the pipe in which it is inserted. The device induces a headloss proportional to the fourth power of the diameter of the inner hole. Choosing a small enough inner diameter makes it possible to absorb the headloss at the flows considered in the systems under consideration.

link is given by (1) as an explicit function the pipe diameter. Physics tells that the sum of the link headlosses along the path from the source to a faucet should match the static head at the faucet, which is given by the change in elevation. Since the total headloss is an explicit function of the pipe diameters, the constraint on the path can be given an explicit mathematical formulation in the optimal design problem. The objective is the total investment cost.

The unrealistic case of all faucets simultaneously open does not yield the solution to the design problem facing random configurations of open/closed faucets. Nonetheless it provides a hint for a heuristic approach to its stochastic version. If we were to maintain the objective of delivering the exact target flow at each faucet in a now stochastic environment, we would have the flow on each link directly proportional to the number of open downstream faucets, a now random variable. The heuristic replaces the stochastic value of the flow in the link by a deterministic one. Our probability model is that the individual faucets are open or closed according to an independent identically distributed binary process. If n is the number of faucets downstream of a given link, the heuristic will assign a flow $\lambda \bar{\psi}$ instead of $n \bar{\psi}$ with $\lambda < n$. We name λ the load factor. The higher λ, the greater is the probability that the flow $\lambda \bar{\psi}$ be sufficient to match the requirement of delivering $\bar{\psi}$ at each open downstream faucets. We can attach to the load factor the probability that the number of open faucets be less than λ. We interpret this probability as a quality of service. The main feature in the heuristic consists in assigning appropriate load factors to each link so as to achieve the same quality of service everywhere. The quality of service becomes a simple one-dimensional design parameter. We then mimic the logic of the deterministic case with all faucets open to compute the total headloss along each path to a faucet. But instead of computing in a branch a headloss associated with the flow $n \bar{\psi}$, we use the one given by the load factor $\lambda \bar{\psi}$. The constraint on the headloss is an explicit function of the pipe diameters, and thus amenable to a mathematical programming formulation.

In practice, pipes are to be selected among a finite set of available commercial pipes with known diameters and known cost per unit length. A branch can be split into segments of pipes with different commercial diameters. The investment cost on the branch becomes a linear combination of the unit pipe costs with coefficients equal to the lengths of the segments. A similar observation can be made on the headlosses. Equation (1) shows that the headloss is proportional to the pipe length. The optimal design problem is thus converted into simple linear programming. The precise formulation as a mathematical programming problem is given in the appendix.

It is worth mentioning another set of constraints that are inserted into the optimization problem. The idea is to protect the network against damage caused by a leakage in some pipe. If the inner pressure in the pipe is less than the atmospheric pressure, there is a risk that surrounding polluted water penetrates the network. To enforce higher pressure in the pipe, we should impose at each point of a branch a smaller gradient loss than the static pressure. It is quite sufficient to introduce the constraint at the end node of the branch only. This precautionary constraint is not so important in itself. However, it turns out to be extremely useful in the tuning phase of a design. Let us explain why and how. The constraint on intermediary (non-faucet) nodes is all the more stringent as the node elevation is high. When the constraint becomes active, it enforces a lower gradient loss on the path from the reservoir to the node. Since the gradient loss from the reservoir to a faucet must match the faucet static pressure, reducing the gradient loss upstream of the intermediary node induces an increase of the gradient loss from the intermediary node to the terminal nodes attached to it. The faucets downstream of the intermediary node become less dependent on the actual flows in the upper part of the network. They are more "stable". This nice feature can be exploited in the fine tuning phase of the design.

Based on the heuristic, the design module computes the least cost design. At this stage one cannot predict the behavior of this design under the variety of open/closed faucets configuration. Only simulation can provide this information. The simulation module does this job. It performs an

extensive analysis of the proposed design and collects information on the actual flows at each faucet as well as the pressures at the various nodes. In particular, out of range variations are signaled. If the design turns out to be unsatisfactory at some faucets, typically because of too high an occurrence of flows below a tolerance level, the user can rerun the design module with alternative parameters. The tool of choice is to make the headloss constraint at the intermediary nodes more stringent. For instance, by stating that the headloss should be less than a fraction of the gravity pressure[4]. The inverse of this fraction plays the role of a typical engineering *safety factor*. The simplest approach is to impose the same safety factor at all intermediary nodes. As will be explained in the case study a satisfactory factor is found after very few trials and errors. The user can also play on other design parameters, but impacts are far less conclusive.

We conclude the presentation of the design module with a word on a particular feature of the output. Because the collection of available pipes is finite, it is not possible to ensure the exact match between the total headloss on a path to a faucet and the faucet elevation itself. Pipes with very small diameters may be missing for the needed final adjustment. This translates into the fact that in the mathematical formulation (D.2)–(D.6) given in the appendix, some constraints are inactive at the optimal solution. To compensate for possible gaps in headloss, NeatWork proposes an adjustment by means of orifices just ahead of faucets.

5. Case study

We present the design phase of a typical project done by APLV. The project, named *La Luna*, involves to a network with 129 faucets, for over 700 hundred inhabitants. The faucet elevation range is $[-128.7, -1.8]$ meters. There are 117 branching nodes[5], just as many intermediary branches, and 129 terminal branches. The total length is 4837 meters. Finding a satisfactory design is certainly a challenge, in view of the number of faucets, the network length and the large range of elevations. Significantly larger projects have been successfully handled by NeatWork at APLV. For instance, the project for the *Wany* community involved 269 faucets and a total length above 25 kilometers. However, the design process is exactly the same for *Wany* as for *La Luna*.

The design was produced by a local technician, who was taught how to use NeatWork. In a first run, the technician specifies the three main design parameters: the target flow, the quality of service (which determines the load factor via the design heuristic) and the proportion of open faucets at peak hours, according to the norms in use in the organization. The performance of the design proposed by NeatWork is simulated. If the performance is not deemed satisfactory, another design is generated, mainly by modifying the safety factor in the headloss constraints at the branching nodes. (The headloss constraints at the terminal or faucet nodes are kept unchanged in this process.)

The established norm at APLV is a target flow at 0.12 l/s (liter per second). The recommended fraction of open faucets is 0.28 at peak hours and the quality of service is usually 65%. A simulation produces for each faucet an empirical distribution of the flow when the faucet is open. The simulation output is an empirical distribution of flows and pressure at all nodes. It is summarized by statistics such as minimum, mean, maximum and various fractiles. It turns out that the most important statistics is the proportion of times the flow at a faucet is below a threshold value. The acceptability

---

[4]In the mathematical programming formulation given in the appendix, the right-hand side of constraint (D.4) would be divided by the safety factor.

[5]Very few branching nodes are adjacent to exactly two nodes. Clearly, this in-between node could be suppressed and the two adjacent branches should be replaced by a single one with length the sum of the two lengths. This simplification does not alter the results of the design and simulation modules. The intermediary nodes were introduced by the technicians to spot critical points in the topography. The advantage of keeping this information is that NeatWork also computes the pressure at those nodes.

criterion at APLV is that the flow be above 0.06 l/s[6] at least 98% of the times. The sample size to capture information on the tail of the distributions must be large. In this study it was taken to be 10,000.

A first design was built using a quality of service 65%, with a total cost[7] of c$71,553. The overall average flow in a first simulation was 0.123, an adequate value. Unfortunately, the second goal of flows being rarely below the threshold value 0.06 was not met on 11 out of the 129 faucets. On these critical faucets the flow was below the threshold value more than 3% of the times. (See Table 1.) A new design was built with a safety factor of 1.2 on the constraints at the intermediary nodes. A quick simulation on a moderate sample showed that the problems at the faulty faucets disappeared at the expense of a cost increase. Lower values for the safety factor were tried. Eventually, the design with a 1.05 safety factor was deemed satisfactory. A simulation with the large sample size confirmed the observation. The cost of this final design is c$72,784, less than a 2% increase. The computer time is not an issue. The design takes about 4 seconds. A 10,000 simulation runs in a little more than 4 minutes. Computations were performed on Windows 7 running on a MacBook Pro.

Table 1 displays information on the faulty faucets in the initial design and their improved behavior in the final design. For each design the table gives the proportion of time the flow was below the threshold value 0.06 as well as the coefficient of variability[8]. All figures are percentages. The table does not reports on the other 116 faucet node because none flowed below 0.06 in the final design.

Table 1: Critical faucets in the initial and final designs.

	Initial design 0.123 l/s 71,553		Final design 0.122 l/s 72,784	
Average flow Design cost				
Faucet	% in the simulation with flow < 0.06	Flow variability in %	% in the simulation with flow < 0.06	Flow variability in %
P89	**20.47**	59.46	0.52	23.55
P127	**12.96**	49.86	0.39	22.01
P121	**12.25**	44.81	0.04	19.12
P62	**10.91**	47.8	0	18.64
P109	**9.45**	41.31	2.39	24.6
P120	**8.64**	41.52	0.03	15.33
P83	**6.9**	38.29	0	16.84
P84	**4.17**	38.48	0	18.06
P125	**4.12**	36.72	0.17	19.66
P105	**3.14**	30.61	0	13.55
P88	2.66	36.44	0	19.24

Simulation sample size: 10,000.
Tolerated fraction of flows below the 0.06 l/s threshold: 3%

The figures support the argument that the safety factor on the constraints at the intermediary nodes contributes an indirect way to a decrease of the variability. On the critical faucets the

---

[6]This figure can be interpreted as follows filling a one liter bottle would with a flow 0.06 l/s takes 17 seconds, twice the time for the target value 0.12 l/s.

[7]The cost is given in Cordoba, the local currency. One hundred Cordobas, denoted c$100, amounts to US$3.3.

[8]The coefficient of variability is the ratio between the experimental standard deviation and the experimental mean.

variability is more than halved in average. To further illustrate the benefit of our approach based on cost minimization, we run the same case with a safety factor 10, that is, the (negative) elevations of the intermediate nodes are divided by 10. NeatWork proposes a design with almost perfect stability of faucet flows. Variability is at most 1%, and only on very few faucets. The total cost jumps to c$134,319, nearly twice the cost of the final design. The compromise between the costs and the variability displayed by the final design is a good illustration of NeatWork achievement.

Significantly larger projects have been successfully handled by NeatWork at APLV. For instance, the project for the *Wany* community involved 269 faucets and a total length above 25 kilometers. However, the design process is exactly the same for *Wany* as for *La Luna*. In *Wany* final design, one faucet out of the 269 ones experienced flows below the threshold, but it occurred only 4% of the time. At all other faucets, the flow was never below the threshold. The total cost is c$2,991,645. In that example, it was necessary to resort a safety factor 2.3 to improve the initial design and eliminate problems at 6 faulty faucets with flows below the threshold between 16% and 28% of the time. Variability at all faucets was drastically reduced in the final design, but the cost increased by 2.7% only. Just as for *La Luna* a fully satisfactory design is obtained at a the expense of a slightly higher material cost.

6. Conclusion and desirable enhancements

NeatWork is a computer tool to support the design of self-regulated gravity driven water distribution systems. The tool is designed to meet the very specific and restrictive requirements of water systems in hilly regions, with dispersed and poor inhabitants, and a lack of local administrative support for adequate maintenance. The total investment cost is the driving force. Low cost networks are more sensitive to open/closed faucets configurations. The design problem is far from trivial, and it must be performed by local technicians with limited training. The experience over 70 projects designed and executed in the last 17 years of use has been highly successful. The success of an earlier experimental version of the software is included in that period.

6.1. What in the end is the point of NeatWork?
- First, NeatWork solves a problem that, as far as we know, has never been discussed in the literature but is nevertheless a real, practical problem in many areas of the World: How to size a network that guarantees a sufficient self-regulation of the flow rates when the users have no control and only the option to open or close their faucets at will. While the peak consumption of the group of user is specified, the combination of open faucets is a random variable.

- Second, NeatWork's dimensioning of the network's conduits minimizes the material cost, an important factor for the poor communities that appeal to agencies with ever-limited funding.

- Third, the use of NeatWork by designers is both simple and rapid. Neatwork's technicians none of whom could conceivably graduate from one of our own secondary schools have been using it for more than 17 years on more than 70 different projects. The computing time in the two modules is very limited, less than half a minute for the design and less than 5 minutes for a simulation with a very large sample size. As a result, the technicians manage to produce a satisfactory design in a very short laps of time.

- Finally the confrontation of the actual measurements of individual flow rates with the predictions of the simulator confirms that the design has met its objectives.

6.2. Enhancements and extensions

NeatWork may turn out to be useful in other contexts. In less hilly and/or in dryer regions, gravity-actuated distribution systems between an elevated tank and a lower set of habitations are common. Moreover, the ease with which one obtains a design with NeatWork, makes it possible to perform a tradeoff between the elevation of the reservoir and the cost of pipe material since the higher the reservoir, the higher the pression and the lesser the need for large pipe diameters.

The current version of NeatWork was released in 2010. An important update should be undertaken, firstly to make the software compatible with the more recent versions of Microsoft operating systems. This would be an opportunity introduced new functionalities, such as 2-dimensional displays of the network or a built-in device for handling the safety factor in the intermediary nodes. (At present the adjustment is performed by an external excel macro directly on the NeatWork ascii files.) NeatWork's efficient execution owes much to Mosek powerful optimizing engine. Mosek has been made graciously available to us.

Appendix

In this appendix, we give the mathematical formulation of the Design and Simulation modules of Neatwork. We also discuss the underlying physical laws and technical specificities.

Appendix A. The distribution network

The distribution network is a standard one with one source and several sinks. The network is described by a set of nodes $\mathcal{N} = \{1, 2, \ldots, J\}$ (branching nodes, reservoir node and faucet nodes), and a set of links $\mathcal{L} = \{1, 2, \ldots, l\}$. A link k joins two nodes, an origin $i_k \in \mathcal{N}$ and an extremity $j_k \in \mathcal{N}$. We consider that the links are oriented.

The water distribution network is modeled as a graph $G = (\mathcal{N}, \mathcal{L})$, with certain arc and node characteristics. The arcs support flows in either direction, so it is convenient to put an arbitrary orientation on the graph. We denote $\mathcal{N}_o \subset \mathcal{N}$, $\mathcal{N}_f \subset \mathcal{N}$ and $\mathcal{N}_t \subset \mathcal{N}$, the subsets of nodes at which there is an inflow (supply nodes), an outflow (faucet nodes), and no inflow and outflow (branching nodes), respectively. The subsets form a partition of $\mathcal{N}$. Let $A \in R^{|\mathcal{N}| \times |\mathcal{L}|}$ be the incidence matrix of the graph and $B \in R^{|\mathcal{N}| \times |\mathcal{N}|}$ a matrix with off-diagonal elements equal to 0 and diagonal elements B_{ii} equal to -1 if $i \in \mathcal{N}_o$ to 1 if $i \in \mathcal{N}_f$; and 0 otherwise. Flows through the network are represented by a vector $\phi \in R^{|\mathcal{L}|}$ of flows on the arcs and a vector $\psi \in R^{|\mathcal{N}|}$ of inflow (at the reservoir) and (outflows at the faucets). The component of F on the branching nodes are irrelevant as there is no inflow or outflow there. We keep them to ease the notation. Flows on the network must satisfy the mass balance equation $A\phi - B\psi = 0$. Note flows along the links and at the inlet and outlets are not restricted in sign. The supporting graph may include loops (cycle).

Appendix B. Energy dissipation

Appendix B.1. Approximation of the Colebrook formula for the steady state flow

The quantity of interest for calculations by Neatwork of head losses by friction is H_f/L the (non dimensional) head loss per unit length of pipe for all the pipes utilized. That quantity is related to three others by: $\frac{H_f}{L} = 0.0826 f \frac{\phi^2}{D^5}$, where ϕ, the flow rate is expressed in m³/sec and D, the inner diameter of the pipe in meters. The numerical constant 0.0826 has the units of

acceleration (m/sec^2). The undefined variable f appearing in the above formula is dimensionless; it is a function of two other dimensionless parameters, the Reynolds number R_e and the ratio ϵ/D of an *equivalent pipe wall roughness* to the pipe diameter. The corresponding energy dissipation obtains by integration over ϕ. The Reynolds number depends on the flow ϕ, the diameter D and the temperature. This dependence on these quantities leads to an implicit formulation for the headloss which is approximated in NeatWork (for PVC or other plastic pipes) by $\frac{H_f}{L} = \beta(T)\frac{\phi^{1.781}}{D^{4.781}}$, where $\beta(T)$ is given by an explicit formula of T. The formula accounts for the PVC roughness. For $T = 20^o$ Celsius, as it is the case in Nicaragua, $\beta(T) = 0.00213$. (For a detailed justification of these formulas, we refer to the NeatWork user guide [11] which can be downloaded from APLV and Ordecsys sites.)

By direct integration, we obtain that, in steady state, the energy dissipated by friction in a branch (i, j) is

$$\mathcal{E}_f(ij) = \frac{\gamma_{ij}L_{ij}}{2.781}\frac{|\phi_{ij}|^{2.781}}{D_{ij}^{4.781}}.$$

Appendix B.2. Energy dissipation through faucets and orifices

The headloss through a faucet is of the form $\frac{\psi^2}{\alpha}$, measured in meters. The faucet coefficient α takes a value that depends both on the size and on the construction of the faucet. We associate to this headloss the dissipated energy at faucet j :

$$\mathcal{F}_j = \frac{\psi_j^3}{3\alpha_j}.$$

The default value for the faucet coefficient is $1.83 \ 10^{-8}$. (Recall that $\psi_i \geq 0$.)

A similar formula holds for orifices. Dimensional analysis indicates that α should be nearly proportional to D^4, where D is the internal diameter of the orifice. The energy dissipated through the orifice is

$$\mathcal{O}_j = (0.62)^4\frac{\psi^3(i)}{3D_j^4}.$$

The factor 0.62 results from extensive experimental studies. A more detailed analysis is given in [11]. Note that $\mathcal{O}_j$ decreases sharply with D. In the formulation we can assume that there is always an orifice in front of a faucet, possibly virtual with a large diameter.

Appendix B.3. Potential energy

It applies to the flows at the terminal node j according to the formula where $h_j < 0$ is the node elevation. (A similar term exists at the source node, but the elevation there is zero.) At the faucet node $j \in \mathcal{N}_f$ we have at $\mathcal{P}_j = -ah_j\psi_j$, where a is the pressure induced by a one-meter high column of water. It depends on the gravity parameter g and of the local density of water.

Appendix C. Convex programming for the simulation module

It is well-known [9, 3] that any compatible pair of flows and pressures in a network (without compressor and regulator) can be interpreted as an optimal primal-dual solution of a strictly (actually, strongly) convex problem.

$$\min \quad \sum_{k\in\mathcal{L}}\frac{\gamma_{i_kj_k}L_{i_kj_k}}{p+1}\frac{|\phi_{i_kj_k}|^{p+1}}{D_{i_kj_k}^q} - \sum_{j\in\mathcal{N}_o\cup\mathcal{N}_f}h_j\psi_j + \sum_{j\in\mathcal{N}_f}\frac{\psi_j^3}{3\alpha_j} + \sum_{j\in\mathcal{N}_f}\left(\frac{0.62}{D_j}\right)^4\frac{\psi_j^3}{3} \quad \text{(C.1a)}$$

$$\text{s.t.} \quad A\phi - B\psi = 0 \quad \text{(C.1b)}$$

$$\psi_i \geq 0, \forall i \in \mathcal{N}_i. \quad \text{(C.1c)}$$

11

The nonnegativity constraint on the outflow simply says that the faucets can become inlets: if the gradient loss ahead of the faucet is larger than the difference of elevation, the faucet cannot outflow. The cannot become negative ; so it must be zero. Note in this formulation, the variables ψ_i, $i \in \mathcal{N}_t$, do not enter in the objective and their value in the network constraints because $B_{ii} = 0$ for all transit node $i \in \mathcal{N}_t$.

Appendix D. Linear programming for the design module

The design module applies to arborescent network only, that is a tree with no cycle (loop). The design module aims at satisfying two goals : minimize the material cost and ensure, at each faucet, a flow as close as possible to a specified target flow $\bar{\psi}$. The second goal is difficult to attain, in view of the uncertainty in the actual use of the network –each faucet can be open or closed by a user in a non predictable way.

Appendix D.1. Probability model and load factor

The water consumption by users is a certain stochastic process. Faucets are open, then closed, perhaps a certain number of times during the peak hour. The opening and closing times are random according to independent identically distributed distributions, the same for all users. During the peak period, the user opens his faucets during a fraction π of the peak period duration. The goal is that during consumption time the flow be exactly the target flow $\bar{\psi}$. Under the scheme, we can say at any time during the peak period there a probability π that a given faucet be open, independently of all other faucets.

Consider now a branching node with n downstream faucets. We are interested in the number N of open downstream faucets at some point of time. The random variable is N is binomial with parameters (n, π). It can be used to describe the desired flow $N\bar{\psi}$ in the upstream branch adjacent to the node. Note that when a faucet is closed, the characteristic of the network is of no concern for this node. Therefore the case $N = 0$, which occurs with probability $(1 - \pi^*)^n$, is not relevant in the computation of the load factor. The relevant quantity is the conditional random variable is the number of open downstream faucets under the condition that least one downstream faucet be open. Using a simple argument on conditional expectation, it is easy to show that the variable of interest is $\tilde{N} = 1 + X_{n-1,\pi}$, where $X_{n-1,\pi}$ is the binomial distribution and 1 stands for the contribution of the open faucet.

Let F be the cumulated distribution of $\tilde{N}$. Suppose we assign to the pipe the flow $\lambda\bar{\psi}$, with $\lambda < n$. The resulting quality of service is $F(\lambda) = \text{Prob}\{\tilde{N}\bar{\psi} \leq \lambda\bar{\psi}\}$. Now, if the desired quality of service is a given value α, the load factor is $F^{-1}(\alpha)$. Since $\tilde{N}$ takes integer values only, we typically have

$$F(F^{-1}(\alpha) - 1) < \alpha \leq F(F^{-1}(\alpha)).$$

Instead of taking $\lambda = F^{-1}(\alpha)$, we interpolate between to the two values

$$\lambda = F^{-1}(\alpha) - \frac{F(F^{-1}(\alpha)) - \alpha}{F(F^{-1}(\alpha)) - F(F^{-1}(\alpha) - 1)}. \tag{D.1}$$

Appendix D.2. Optimization model

From the above discussion, the choice of a certain quality of service induces at any branching node j the load factor λ_j defined by (D.1). The target flow passing through the node $j \in \mathcal{N}$ is $\lambda_j\bar{\psi}$. On any link $k = (i_k, j_k) \in \mathcal{L}$, the target flow is $\lambda_{j_k}\bar{\psi}$. In this way we can predict the total headloss on any path from a faucet to the root node (reservoir) of the tree network as a function of the chosen diameters on each link and the orifice ahead of the faucet. To this end we denote P_i the

set of nodes that are on the path from the reservoir to node i. (Node i can be either a branching node, or a terminal/faucet node.)

$$\min_{\ell} \quad \sum_{k \in \mathcal{L}} \sum_{\nu=1,\ldots,n} \ell_{k\nu} C_\nu \tag{D.2}$$

$$\sum_{k \in P_t} (\lambda_{j_k} \bar{\psi})^p \sum_{\nu=1}^{n} \ell_{k\nu} (\frac{\gamma_\nu}{D_\nu^q}) + \theta_f \bar{\psi}^p \leq h_o - h_t, \forall t \in \mathcal{N}_f \tag{D.3}$$

$$\sum_{k \in P_i} (\lambda_{j_k} \bar{\psi})^p \sum_{\nu=1}^{n} \ell_{k\nu} (\frac{\gamma_\nu}{D_\nu^q}) \leq h_o - h_i, \forall i \in \mathcal{N} \setminus \mathcal{N}_f \tag{D.4}$$

$$\sum_{\nu=1,\ldots,n} \ell_{k\nu} = L_{i_k j_k}, \ \forall k \in \mathcal{L}, \tag{D.5}$$

$$\ell_{k\nu} \geq 0, \forall k, \nu. \tag{D.6}$$

In view of the restriction on the availability of pipes with arbitrarily large inner diameters, the problem may be infeasible. Note also that an optimal solution may involve a strict inequality in constraint (D.3). In that case, the slack can be met with an orifice, a regulator that is placed just before the faucet. An ideal orifice is one with an inner diameter that ensures the desired headloss at the flow $\bar{\psi}$. For practical reasons, the choice of orifices is restricted to commercial items. NeatWork will choose the commercial orifice which ensures the closest headloss to the desired value. The key point of the above analysis is the knowledge of the exact flow in each segment of the network. In the above formulation we added a set of constraints (D.4) that hold at each branching nodes. It resembles contraints (D.3) and ensures, that in case of a leak at the node, the pressure at the node remains large enough to prevent an inflow of surrounding polluted water.

References

[1] E. Alperovits and U. Shamir. Design of optimal water distribution systems. *Water Resources Research*, 13(6):885–900, 1977.

[2] APLV, *Agua para la Vida, Home page.* http://www.aplv.org.

[3] F. Babonneau, Y. Nesterov, and J.-P. Vial, "Design and operations of gas transmission networks," *Operations Research,* doi:10.1287/opre.1110.1001, vol. 60, no. 1, pp. 34–47, 2012.

[4] H. Cross, "Analysis of flow in networks of conduits or conductors," University of Illinois at Urbana Champaign, College of Engineering, Engineering Experiment Station. Bulletin; no. 286, University of Illinois at Urbana Champaign, College of Engineering, 1936.

[5] G. Eiger, U. Shamir, and A. Ben-Tal. Optimal design of water distribution systems with networks. *Water Resources Research*, 36(9):2637–2646, 1994.

[6] EPANET, "Software That Models the Hydraulic and Water Quality Behavior of Water Distribution Piping Systems", *https://www.epa.gov/water-research/epanet*

[7] R. Farmani, G. Walters, and D. Savic. Trade-off between total cost and reliability for any-town water distribution network. *Journal of Water Resources Planning and Management*, 131:161–171, 2005.

[8] A. Klesser and U. Shamir. Analysis of the linear programming gradient method. *Water Resources Research*, 25(7):1469–1480, 1989.

[9] J.-J. Maugis, "Etude de réseaux de transport et de distribution de fluide," *RAIRO Recherche Opérationnelle/Operations Research*, vol. 11, no. 2, pp. 243–248, 1977.

[10] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).* http://docs.mosek.com/7.1/toolbox/index.html, 2015.

[11] NeatWork: Software and user guide available at NeatWork home page: http://neatwork.ordecsys.com/.

[12] Rossman, L. A., "the EPANET Water Quality Model" in B. Coulbeck, ed., Integrated Computer Applications in Water Supply, Vol. 2, Research Studies Press Ltd., Somerset, England, 1993.

[13] Rossman, L.A., EPANET 2, Users Manual, US Environmental Protection Agency, Cincinnati, Ohio, 2000.

[14] E. Todini and S. Pilati, "A gradient algorithm for the analysis of pipe networks," in *Computer Applications in Water Supply: vol. 1—Systems Snalysis and Simulation* (B. Coulbeck and C.-H. Orr, eds.), pp. 1–20, Research Studies Press Ltd. Letchworth, Hertfordshire, UK., 1987.

[15] K. Varma, S. Narashiman, and S. Bhalamudi, "Optimal design of water distribution systems using an NLP method," *Journal of Environmental Engineering*, vol. 123, pp. 381–388, 1997.