

## PLANNING REHABILITATION ACTIONS ON WATER DISTRIBUTION NETWORKS IN CONDITION OF DATA SCARCITY

C. Arena, A. Fortunato and M.R. Mazzola

Dipartimento di Ingegneria Idraulica ed Applicazioni Ambientali, Università degli Studi di Palermo  
Palermo (Italy)

E-mail Address [arena@idra.unipa.it](mailto:arena@idra.unipa.it), [fortunato@dra.unipa.it](mailto:fortunato@dra.unipa.it), [mazzola@idra.unipa.it](mailto:mazzola@idra.unipa.it)

### Abstract

*The paper shows a viable approach to the selection of water networks rehabilitation actions, in condition of data scarcity and uncertainty, based on the use of four concise performance literature indices, allowing the analysis and identification of efficient, reliable and robust rehabilitation alternatives, even if assessed on the basis of a relatively small amount of data.*

*A promising statistical sampling method of water consumptions is also described, that may be fruitfully employed within the proposed approach for drafting water balances at network district scale in order to quantify the extent of leakages.*

*The illustrated case study shows that the selected indices are capable to describe suitably the performance the network would achieve, with respect to water demand satisfaction and to service efficiency, if the proposed rehabilitation options (e.g. leakage detection and reduction, cleaning and replacement of mains) were undertaken, as well as the reliability and robustness of the options themselves, and hence provide a good selection criterion of alternative rehabilitation options.*

*The application of the statistical sampling method to water consumptions of a small municipality in Sicily, (Italy), shows that it may be applicable in practice and that it provides results having the accuracy level predictable by the theory.*

### Keywords

Water distribution networks, rehabilitation actions, data scarcity, performance indices, resiliency, water balances, network sampling

## 1. INTRODUCTION

The correct management of a water distribution network and the definition of the most appropriate rehabilitative actions should be based on the monitoring of both network condition and operation hydraulic parameters, as well as on records of maintenance, rehabilitation and substitution works done in the past years. However, utilities often have to manage infrastructures with inadequate performance levels, bad maintenance state and high water losses, having in addition a limited knowledge of network and plants characteristics and operation, and small budgets to execute service-improving works. This motivates the need of investments planning tools which may cope with the problem of data scarcity and uncertainty, leading to robust solutions.

In order to identify network districts and pipes that show the greatest criticalities, typology and scheduling of actions to undertake, several decision support systems (DSSs) are already available. Among those developed in Europe the well-known CARE-W (Computer Aided Rehabilitation for Water Networks, 2005) could be mentioned, that has been developed within a European Union co-funded project.

DSSs provide valuable indications on infrastructure's state and of rehabilitation works to carry out, suggesting investment needs both in the long one and short term, and the optimal strategic and financial planning of the interventions. Although they are undoubtedly useful, they require a great amount of input data, regarding network's topography, characteristics of pipes, joints, valves and measure devices,

mechanical actions, historical data on pipes breaks, leakages and service failures. Even the more efficient utilities often do not have such a detailed knowledge; it is therefore necessary to develop strategies based on more pragmatic approaches. From this standpoint, there is a growing scientific literature dealing with rehabilitation planning under uncertainty (Savic, 2005), frequently resorting to genetic algorithms-based optimization models (Alvisi and Franchini 2006, Alvisi and Franchini, 2009). However, these models often face the problem of inherent uncertainty that consists of random fluctuations within the studied problem. This type of uncertainty is irreducible and may be characterised using statistical approaches; on the other hand, the problem of statistical epistemic uncertainty given by lack of sufficient data (Savic, 2005) may be tackled by attempting to find a way to limit the amount of necessary data, by using information-parsimonious performance indices, thus minimizing time and economical efforts to gain non readily available data.

## **2. A POSSIBLE STRATEGY UNDER CONDITIONS OF LACK OF DATA**

The decision-making process on the investments to undertake cannot however ignore a minimal level of knowledge, which must be acquired in any case. In this regard, the main problem is the availability of enough resources and time. Under conditions of shortage and uncertainty of information on the consistence, the state and the operation of networks and plants, it is therefore necessary to find rehabilitation actions' prioritization and ranking criteria based on concise indices, computable on the basis of a limited amount of information and capable of leading to robust decisions, optimizing the available resources.

In the first place a field survey of network, plants and equipment is required to acquire information on pipe length, dimensions, materials, etc.

If systematic data are missing, it is then necessary to retrieve information (even if simply qualitative) on pipes' age, when available, and on frequency of breaks and maintenance in the different parts of the network.

It is advisable to gauge pressures levels in order to establish whether to carry out pressure-reducing interventions based on valves: these interventions result in a reduction of the rate of bursts and losses, but make them harder to detect (Mazzola and Bazzurro, 2008).

At the meantime, it is necessary to quantify water losses in the different zones of the network through the execution of water budgets or the analysis of minimum night flows. To this end, it is useful to split the network in metered districts, making also easier to regulate pressures and allowing a better service management (Cascetta et al, 2004).

Areas in the network with rehabilitation priority are commonly the oldest (supposing that they are in bad state) and those with the greatest frequency of breaks, the greatest level of water losses and the greatest exercise pressures.

Working on the areas of the network first identifiable as the best options for rehabilitation on the basis of water losses, maintenance frequency and date of laying or substitution, certainly allows to decrease lost volumes, number of pipes breaks and, consequently, maintenance costs, but does not necessarily ensure the improvement of service levels. To this end, it is necessary to asses the general performance the system would be able to express after the execution of rehabilitation or substitution actions also in areas of the network that were not included in the initial screening, according to a holistic approach (Dandy and Engelhardt, 2001).

Selection of actions to be undertaken (e.g. leakages search and reduction, mains cleaning or relining, mains replacement with same or different size and material pipes) will be therefore made on the basis of criteria of network efficiency (extent of water losses compared with the volumes entering the network), reliability (probability that the system will meet minimum service requirements) and robustness (probability that the system will perform acceptably even under conditions far from those assumed in the project).

Among the several technical literature performance indices (Alegre et al, 2000), the use of indices of

reliability  $R_s$ , efficiency  $E$ , and global performance of the system  $P$  (Bertola and Nicolini, 2006) is particularly advisable, together with the resiliency index  $I_r$  (Todini, 2000) that has been slightly rearranged in this work to make it suitable for the assessment of rehabilitation measures (see numerical application at paragraph 4). This set of indices can provide insight into four of the most important issues in network rehabilitation, i.e. meeting demand, losses, reliability and robustness.

## 2.1 Indices of reliability, of efficiency and of global performance of the network

System reliability,  $R_s$ , is defined as the capability of the network to meet specific water demands at network nodes under various operation conditions in a given period of time. It is assessed as the ratio of the volume actually delivered to the users to total demanded volume:

$$R_s = \frac{\sum_i V_{ero,i}}{\sum_i V_{req,i}}$$

where  $V_{ero,i}$  is the water volume actually supplied at the  $i$ -th node and  $V_{req,i}$  is the required volume at that node.

Efficiency index  $E$ , which takes into account the of effect water losses on system's performance, is defined as the fraction of total volume entering the network actually delivered to the users:

$$E = \frac{ADV}{ADV + TLV}$$

wherein  $ADV$  is the volume actually delivered to costumers, and  $TLV$  is the total volume of water losses. System global performance is therefore evaluated through the product of efficiency and reliability:

$$P = E \cdot R_s$$

This index is useful for comparing rehabilitation alternatives that give rise to discordant variations of  $E$  and  $R_s$  values: high reliable systems are not necessarily efficient and *vice versa*.

## 2.2 Resiliency index

Reliability index, though very meaningful, is a measure of costumers demand satisfaction limited to the period and operation scenarios considered. Although in the approach originally proposed by Bertola and Nicolini a long term simulation (dozens of years) is adopted and the probability density functions of the times of break and reparation of mains and pumps are taken into account, the index does not provide information about system ability to face stressful future situations, apart from those considered in calculations. This aspect can be taken into account through the resiliency index  $I_r$ , that measures network intrinsic capability of tackling stressful operational conditions (e.g. exclusion of a main from the service because of a burst, maintenance or substitution, or a change in water demands), when energy dissipation increases, avoiding or limiting performance shortfalls. In order to increase service reliability and the robustness of design options it is therefore better to have greater heads and hydraulic power than those strictly necessary.

The total hydraulic power entering the network is given by:

$$P_{tot} = \gamma \sum_{k=1}^{n_r} Q_k H_k$$

where  $\gamma$  is water's specific gravity,  $Q_k$  and  $H_k$  are respectively the flow entering the network and the head corresponding to the  $k$ -th reservoir, and  $n_r$  is the number of reservoirs supplying the network. Denote with  $P_{int}$  the hydraulic power dissipated within the network and with  $P_{ext}$  the hydraulic power totally provided to the users, then:

$$P_{int} = P_{tot} - P_{ext}$$

where  $P_{ext}$  is given by:

$$P_{ext} = \gamma \sum_{n=1}^{n_n} q_i h_i$$

where  $h_i$  and  $q_i$  are respectively the actual head and the supplied flow at the  $i$ -th node, and  $n_n$  is the network number of nodes.

$P_{int\ max}$  is the maximum hydraulic power which can be dissipated compatibly with minimum head nodes constraints  $h_{i\ min}$  to supply required node demands  $q_{i\ req}$  and is given by:

$$P_{int\ max} = P_{tot} - \gamma \sum_{n=1}^{n_n} q_{i\ req} h_{i\ min}$$

Resiliency index  $I_r$  is defined as:

$$I_r = \frac{P_{int\ max} - P_{int}}{P_{int\ max}} = 1 - \frac{P_{int}}{P_{int\ max}}$$

Resiliency index, as defined and used in Todini's original formulation, can only take on positive values and range within the interval [0, 1): it can never be equal to 1 as that would imply the total absence of energy dissipation. This index had originally been proposed to characterize hydraulic reliability of alternative network design solutions in the context of a multiobjective (cost minimization and resilience maximization) optimization problem with demand driven hydraulic simulations, hence the case of supplied flows lower than demanded (and actual heads lower than minimum ones) had not been considered, as heads lower than minimum admissible ones were treated as infeasibilities in the optimisation process.

In this work a slightly different formulation of the index is therefore adopted: in computing  $P_{int}$ , actual flows are used instead of demands; in critical conditions they can be lower than the corresponding values required by users, and consequently  $I_r$  can also be negative. From this it follows that:

- a) a value of  $I_r$  greater than zero does not necessarily imply meeting flow and pressure targets at each network node, but indicates that the system has enough energy to face a stressful future situation and is able to avoid or minimize the onset of performance failures or their worsening, should the system have already experienced crises;
- b) if  $I_r = 0$ ,  $P_{int} = P_{int\ max}$ , there could obviously be nodes in critical conditions, and the network does not have any energy reserve;
- c) finally, when  $I_r < 0$ , then  $P_{int} > P_{int\ max}$  but this does not necessarily imply that required flows and pressures are not met at all network nodes.

### 3. SAMPLING OF WATER USE FOR LEAK ASSESSMENT VIA WATER BUDGETS

Developing water budgets is a basic step to assess the level of water losses in district-metered water distribution networks and is hence preliminary to the selection of asset management options such as leak detection and removal, pipe rehabilitation or substitution. It requires knowledge of water use of individual customers.

Albeit conceptually straightforward, the water budget methodology has its major drawback in the metering of water use, as the metering system can be (heterogeneously) old (or even absent in some instances) or it can prove difficult to organise extensive and detailed surveys of consumption data, whose format and information attached are in most cases designed for accounting purposes, rather than for technical ones. In addition, the amount of time required to implement such procedures may be incompatible with the requirements of the asset planning (and subsequent decision-making) process and with a real-time strategy of monitoring losses.

For such reasons, water budgets are often given up in the diagnosis of water losses in a water distribution network in favour of bottom-up approaches such as continuous night flow measurement (Farley and Tow, 2003).

It can be hence of some interest to explore the option to provide only few districts, or customers, with a perfectly working and updated metering system and to extend the results obtained to similar districts or users. In this paper the focus will be on this latter option, i.e. working with microdata of residential water use to assess water use of the whole district. Curiously enough, the idea has received little attention both from the research community and from water utilities: after the seminal paper by Hanke and Mehrez (1979) on microdata of peak hour water use, although the importance of sampling domestic water use has been recognised and illustrated (Mays, 2004), rather little research and field work has been produced on the subject.

After a brief review on the theoretical background of sampling, two cases study are introduced.

#### 3.1 Some results from sampling theory

Statistical sampling allows to obtain information on a given characteristic  $\eta$  of a population  $\Omega$  of size  $N$  ( $\eta$  may be the water use of  $N$  domestic users) through a subset of  $\Omega$ , i.e.  $n$  sampling units, with  $n < N$ . Sampled values of the characteristic will be denoted with  $y$ , in order to distinguish them from population values  $\eta$ . Information is conveyed through a synthetic value  $f(\eta)$ , such as the mean of the characteristic or its total, whose estimator will be denoted with  $h(y)$ . The problem is to assess the number of sampling units to be drawn such that the estimated synthetic value  $h(y)$  may be assumed with a certain degree of confidence  $1-\alpha$  to approximate the “true” value with a given precision  $\varepsilon$ . In statistical terms this implies that  $f(\eta)$  has a probability of  $1-\alpha$  to fall within interval  $[h(y) - \varepsilon, h(y) + \varepsilon]$ .

Sampling techniques vary according to how sampling is performed, i.e. to how the  $n$  units are selected out of the population. Simple random sampling, for instance, refers to a method of selection of the  $n$  sampling units such that the distinct samples have an equal chance of being drawn. In stratified sampling, the population is first divided into  $M$  non-overlapping subpopulations  $\Omega_k$  (called groups or *strata*) of size  $N_k$  with  $1 \leq k \leq M$  and samples of size  $n_k$  are then drawn from each stratum. If each stratum is sampled randomly, then the sampling procedure is described as stratified random sampling. Some standard definition and results from sampling theory are reported in the following (Cocchi, 2006).

Denote with  $m(\eta)$ ,  $v^2(\eta)$  and  $s^2(\eta)$  the population mean, the population variance and the commonly used unbiased estimator of population variance.

In stratified sampling are also needed strata's means  $m(\eta_k)$  and variances  $v^2(\eta_k)$ , within-strata variance  $v^2_e(\eta)$  and between-strata variance  $v^2_b(\eta)$ :

$$v_e^2(\eta) = \frac{1}{N} \sum_{k=1}^M N_k v^2(\eta_k)$$

$$s_e^2(\eta) = \frac{N}{N-M} v_e^2(\eta)$$

$$v_t^2(\eta) = \frac{1}{N} \sum_{k=1}^M N_k [m(\eta_k) - m(\eta)]^2$$

$$s_t^2(\eta) = \frac{N}{M-1} v_t^2(\eta)$$

$v_e^2(\eta)$  is hence the weighted average of the group variances, group size being the weight, and  $v_t^2(\eta)$  is a measure of the distance of the mean of each group from the mean of population  $\Omega$ .

It can be shown that the following additive decomposition holds:

$$v^2(\eta) = v_e^2(\eta) + v_t^2(\eta)$$

The strategy of stratified sampling consists of maximizing the distance among different groups (the variance between strata) thus minimizing within-stratum variance. As within-stratum variance drives the sample size to be used, reducing it through appropriate stratification will determine significant reductions of the overall sample size: in random sampling the number  $n_{random}$  of sampling units necessary to obtain an estimation of the population mean with precision  $\varepsilon$  at a confidence level  $1-\alpha$  is given by:

$$n_{random} = \left( \frac{(\varepsilon/z_{\alpha/2})^2}{s^2(\eta)} + \frac{1}{N} \right)^{-1}$$

where  $s^2(\eta)$  is the population variance that can be substituted by its unbiased sample estimator  $s^2(y)$  and  $z_{\alpha/2}$  is a given percentile of the (assumed Gaussian) distribution of standardized  $h(y)$ .

In stratified sampling  $n_{strat}$  is instead given by:

$$n_{strat} = \left( \frac{N(\varepsilon/z_{\alpha/2})^2}{\sum N_k s^2(\eta_k)} + \frac{1}{N} \right)^{-1}$$

where within-stratum variances appear instead of total variance (as in simple random sampling), hence  $n_{strat}$  will always be less than  $n_{random}$ .

If sampling fractions  $f_k$  in each stratum are the same, i.e. if  $f_k = n_k/N_k = n/N = f$  then stratified sampling is called proportional. Optimal stratified sampling consists instead of subdividing  $n_{strat}$  among the different strata in such a way that the  $n_k$ 's minimize estimators' variance (on which depends  $\varepsilon$ ), thus maximizing estimations precision.

In the case of the mean or the total of population (the total being characterized by an error of  $\varepsilon N$ ), solving the minimization problem, one obtains:

$$n_k = n \frac{s(\eta_k)N_k}{\sum_{k'=1}^M N_{k'}s(\eta_{k'})}$$

The sample size within each stratum is proportional to the subpopulation variance  $s(\eta_k)$ .

It should be pointed out that any sampling procedure assumes population variance to be known, implying that a suitable number of preliminary data should be available in order to obtain estimates of population variance. In the case of stratified random sampling, such preliminary information should be larger than for random sampling, as the available data will be split into different groups. In addition, stratified sampling requires that appropriate stratification criteria be adopted. Such stratification criteria should be applied to the set of preliminary data to estimate group variances.

In the case of domestic water use, reasonable stratification criteria may be household typology (e.g. detached houses or semidetached houses, with or without garden, flats, etc.), number of users per connection (which seem to be the most meaningful variable) and probably also income.

### 3.2 A case study

The above ideas were first applied, as a preliminary theoretical study, to a set of connections (around 10,600) in a hypothetical network district, shaped after a small group of homogeneous users located in a council house district in Palermo, Italy (Arena et al, 2009). All households share the same typology and inhabitants belong to the same socio-economic level; nevertheless, daily water use per connection in the district features a considerable variability (average daily water use per connection: 351 l, standard deviation of daily water use per connection: 225 l) than can be ascribed to the different number of users in each connection, which was adopted as only stratification variable. Sample data belonging to a given stratum were used to estimate within-stratum variances of the district, assuming that  $k$ -th subpopulation variance was equal to the variance of the available sample.

Assuming one wishes to assess the average daily use per connection with a precision of  $\pm 25$  l/day and a confidence level of 95%, was obtained  $n_{strat} = 16$  and  $n_{rand} = 303$ . It is clear how stratified sampling allows a considerable reduction of sample size.

A sensitivity analysis was then performed to assess, even roughly, the range of variability of the required sample size for different values of precision, confidence level and model parameters (group variances). From the results obtained, sample size seems to be quite sensitive to sample variance in the case of stratified sampling, and much less in the case of random sampling, as should be expected. The procedure is however even more sensitive to precision, and quite less to confidence levels.

The method practical applicability has been then checked for a small town in Sicily of about 5,000 inhabitants, in a full-field real data case study. Annual water uses of each residential connection (totally 2,074), from 1995 to 2008, have been gathered and from them the average annual water use per connection have been calculated (Table 1).

In this application, it was possible to link each connection with the corresponding number of users, without considering other stratification variables. There are only individual connections since every single real estate of each apartment building is equipped with its own meter.

As can be seen from Table 1, water consumptions show very little variability with the years; the most relevant aspect regarding sampling method application is however the population internal variability of water consumptions which is nearly constant, as proved by  $CV$  values in each year, ranging from 0.522 to 0.533, as well as by water uses of each single connection during the observation period (data not shown), ranging overall from 0.025 to 0.049.

For this reason, stratification has been performed with reference to the statistical population of daily average water uses of each connection (2,074) between 1995 and 2007 (Table 2), using the corresponding numbers of components as the only stratification variable, leaving 2008 data to validate the sampling

procedure, as described in detail below.

Table 1. Average daily water uses of connections in a small town in Sicily, Italy

Year	Population water use total, $t(\eta)$ [m <sup>3</sup> /day]	Population water use mean, $m(\eta)$ [l/conn.day]	Population water use standard deviation, $s(\eta_k)$ [l/conn.day]	Population water use variance, $s^2(\eta_k)$ [l <sup>2</sup> /conn. <sup>2</sup> day <sup>2</sup> ]	Population water use variation coeff., $CV$
1995	787.8	379.9	201.8	40,706.1	0.531
1996	778.0	375.1	195.9	38,379.2	0.522
1997	775.5	373.9	198.7	39,499.0	0.531
1998	745.4	359.4	190.8	36,398.0	0.531
1999	769.3	370.9	196.9	38,766.0	0.531
2000	779.0	375.6	199.3	39,737.8	0.531
2001	746.9	360.1	191.7	36,732.7	0.532
2002	777.2	374.7	199.6	39,853.0	0.533
2003	777.3	374.8	199.2	39,663.4	0.531
2004	782.0	377.1	200.8	40,337.5	0.533
2005	796.1	383.9	203.9	41,556.6	0.531
2006	819.4	395.1	209.8	44,022.7	0.531
2007	837.2	403.7	214.4	45,957.2	0.531
2008	835.2	402.7	213.8	45,702.8	0.531

Table 2. Residential connections water use data of a small town in Sicily – stratification has been performed on the basis of the statistical population of daily average uses of each connection (years 1995 to 2007).

Number of users per connection	Connection number per stratum, $N_k$	Stratum water use standard deviation, $s(\eta_k)$ [l/conn.day]	Stratum water use variance, $s^2(\eta_k)$ [l <sup>2</sup> /conn. <sup>2</sup> day <sup>2</sup> ]	Stratum water use mean, $m_k$ [l/conn.day]	Stratum water use variation coeff., $CV_k$
1	655	1.0	1.0	155.5	0.0064
2	544	7.1	50.5	311.8	0.0228
3	359	3.8	14.1	468.4	0.0080
4	395	4.1	17.0	624.9	0.0066
5	107	8.7	75.8	774.5	0.0112
≥ 6	14	7.7	59.9	933.1	0.0083
Population	$N$ 2,074	$s(\eta)$ 200.6	$s^2(\eta)$ 40,240.5	$m(\eta)$ 377.2	$CV$ 0.5318

If one then wishes to assess the average daily use per connection in 2008 from data in Table 2 with the quite conservative precision of  $\pm 1.5$  l/day and a confidence level of 99%, using  $t$ -student distribution, (which yields larger samples than those provided by the standardized normal), in order to account for the small sample size and the actually unknown exact distribution of sample mean, he or she would obtain that 70 meters must be read. This figure, optimally divided among strata, rounded off to the nearest integer number, increases to 75, i.e. to a sampling fraction  $n/N$  of 3.6%.

Randomly drawing 1,000 different (without repetitions) samples of 75 water use measures from the 2008



data set ( $m(\eta) = 402.7$  l/conn.day, Table 1),  $m(y)$  ranges between 400.0 l/conn.day and 404.5 l/conn.day, with a mean value (estimation of sample mean  $m(y)$  expected value) equal to 402.7 l/conn.day and standard deviation of 0.75 l/conn.day; consequently symmetric confidence interval of  $m(y)$  estimator, with a confidence level of 99%, turn out to be  $402.7 \pm 1.9$  l/conn.day. For each draw the variance of  $m(y)$  and the corresponding confidence intervals, with a confidence level of 99%, have been further calculated:  $m(y)$  confidence intervals include the “true”  $m(\eta)$  value in 98.7% of cases.

The width of the confidence interval of  $m(y)$  and the accuracy of single estimates confidence intervals are partly due to the number and the specific selection of samples, but also to the distribution of sample means, which in this case study is not be perfectly symmetric. It has in fact been checked that neither the normal distribution nor the log-normal one properly fit the set of sample means.

#### 4. NUMERICAL APPLICATION OF $R_s$ , $E$ , $P$ AND $I_r$ INDICES

In order to explain the use of reliability  $R_s$ , efficiency  $E$ , and total system performance  $P$  indices, together with resiliency index  $I_r$ , for the selection of rehabilitations options, a numerical application was set up only containing the amount of information strictly necessary to carry out hydraulic simulations, i.e.: network topology, lengths, diameters, materials and roughness of mains, node demands and leakages.

In a real-world application, information on network topology, diameters and materials of mains should stem from a field survey, roughness could be assessed on the basis of mains age and/or inspection, nodal demands and leakages would be quantified resorting to the sampling method described above: nodal demand would then be characterized by an appropriate dimensionless curve demand and by the number of users supplied, while leakages would be evaluated via water balances. Actually, pressure measures would be also needed to perform a basic hydraulic model calibration, acting for instance on roughness and number of users served by each node, and to link leakage levels to pressures.

Analysis have been done referring to a unit duration operation period (e.g. peak hour of yearly maximum consumption day), thus adopting flows instead of volumes for evaluating indices. The hypothetical cast iron network, showed in Figure 1, has been considered. Its characteristics are reported in Table 3.

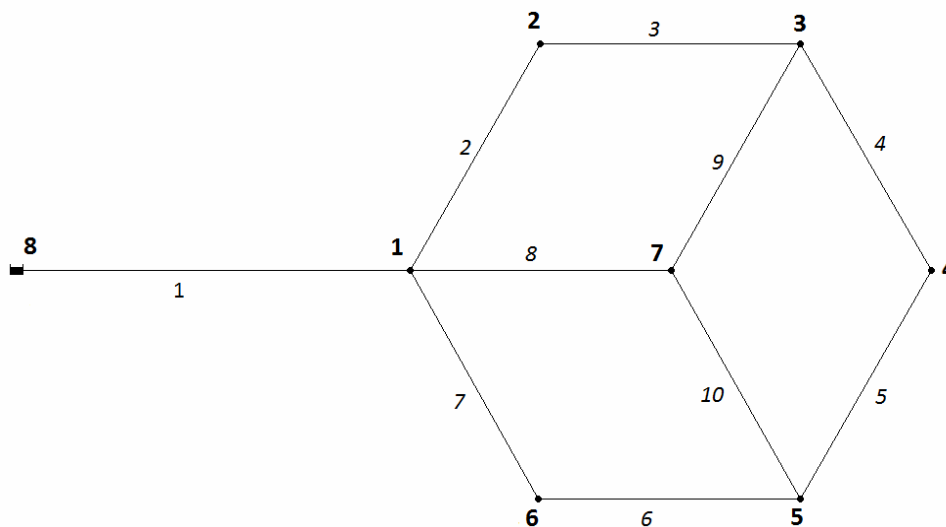


Figure 1. Topological scheme of examined network.

Nodal demands are not imposed, but head-driven simulations are performed, in order to simulate the likely dependence of nodal flows on actual pressures by means of EPANET-2 software, adopting the approach proposed by Bertola e Nicolini (2006) which is based on the use of *emitters*, EPANET’s devices associated with junctions that model the flow through a nozzle so that the demand varies in proportion to the actual junction pressure, connected to network nodes through a pressure reducing valve (PRV) in

order to model the effectively delivered flow,  $Q_{ero}$ , on node pressure  $p$ .

Table 3. Characteristics of the examined network.

Main	Length $L$ [m]	Diameter $D$ [mm]	Roughness $\varepsilon$ [mm]	Junction	Elevation $z$ [m]	Demand $Q_{req}$ [l/s]
1	850	200	0.15	1	455	5
2	300	150	0.15	2	465	5
3	300	100	0.15	3	490	5
4	300	100	0.15	4	490	5
5	300	100	0.15	5	490	5
6	300	100	0.15	6	470	5
7	300	150	0.15	7	480	7
8	300	150	0.15	Reservoir Elevation		
9	300	100	0.15			
10	300	100	0.15	8	530	

In particular, a required pressure  $p_{req} = 35$  m (meter of water column) to fully meet the required flow at nodes  $Q_{req}$ , and a minimum pressure of supply  $p_{min} = 0$  m, have been adopted. The relationship between  $Q_{ero}$  and  $p$  is expressed by the following equations:

$$Q_{ero} = Q_{req} \sqrt{\frac{p - p_{min}}{p_{req} - p_{min}}} \text{ if } p_{min} < p < p_{req}$$

$$Q_{ero} = Q_{req} \text{ if } p \geq p_{req}$$

$$Q_{ero} = 0 \text{ if } p \leq p_{min}$$

The supplied flow by an emitter, whose ground elevation is equal to the elevation of real node increased by  $p_{min}$ , is given by:

$$Q_{ero} = c(p - p_{min})^\gamma$$

where exponent  $\gamma$  is set equal to 0.5 and emitter coefficient,  $c$ , is given by:

$$c = \frac{Q_{req}}{\sqrt{p_{req} - p_{min}}}$$

The function of the PRV valve, characterized by null local head loss and setting (i.e. the maximum pressure allowed on its downstream side) equal to  $p_{req} - p_{min} = 35$  m, is to limit supplied discharge value,  $Q_{ero}$ , to required discharge,  $Q_{req}$ , when actual pressure  $p$  is greater than required one,  $p_{req}$ .

It has been further assumed that leakages along mains can be concentrated in correspondence to the end nodes of the mains, and that such localized leakages are linear functions of the corresponding actual node pressure; in particular, the following scattered leakages at reference pressure of 30 m have been hypothesized: 1.05 l/s/km for main 1, 1.95 l/s/km for mains 2, 7 and 8, 1.65 l/s/km for mains 3, 6, 9 and 10, and finally 1.35 l/s/km for mains 4 and 5.

Leakages can be made dependent on nodal pressures in the following way: after imposing leakages at some reference pressure (for example, the average pressure measured in the network part to which the specific node belongs), a hydraulic head-driven simulation is performed and leakage values corresponding to the simulated nodal heads are calculated; these new leakages values are then imposed, a simulation is performed again and leakages are recalculated until convergence is achieved (it was verified that convergence is reached rapidly even if the reference pressure is rather different from the actual one). This iterative procedure is needed because Epanet software does not allow to use different exponent values for each emitter (or at least for some emitter categories, i.e. emitter modelling uses and leakages); if this were possible, or if the Epanet code were modified to that end, it would be enough to add a further emitter at each real node (with exponent equal to 1), linked to node through a PRV valve with null local head loss and very high setting (so as not to limit leakages value when pressure increases). These emitters would have ground elevation of the real node,  $p_{min} = 0$  m, and coefficient  $c$  equal to:

$$c = \frac{Q_{leak,ref}}{p_{ref}}$$

where  $Q_{leak,ref}$  is the leakage value at reference pressure  $p_{ref}$ .

An approximate way to proceed, which however provides good results, is to appropriately calculate emitters coefficients to take into account that emitters exponent has been set equal to 0.5, according to the following formula:

$$c = \frac{Q_{leak,ref}}{p_{ref}} \sqrt{p'}$$

where  $p'$  is the assumed actual node pressure.

The approximation is due to the fact that pressure  $p$  is not known *a priori*, and therefore a presumed value  $p'$  has to be assumed (for example the pressure value measured in correspondence to the specific node). Obviously one may proceed iteratively with this second approach as well.

As far as the calculation of resiliency index is concerned, it should be remarked that the total hydraulic power entering the network  $P_{tot}$  is not entirely available for users since a part of it is due to, and adsorbed by, water losses; in this application, featuring a single reservoir,  $P_{tot}$  can be more appropriately assessed as:

$$P_{tot} = \gamma H \sum_{n=1}^{n_n} q_i,$$

where  $q_i$  is the actual supplied flow at  $i$ -th node and all quantities,  $P_{int}$  and  $P_{int max}$ , may be evaluated as previously described (subsection 2.2).

If the network features more than one reservoir, as it is impossible to assess how much of the flow entering the network in correspondence to each reservoir is due to leakages, it should be necessary to take some precautions in calculating  $P_{int}$  and  $P_{int max}$  by using the following formulas:

$$P_{int} = \gamma \sum_{k=1}^{n_r} Q_k H_k - \gamma \sum_{n=1}^{n_n} (q_i + l_i) h_i$$

$$P_{int max} = \gamma \sum_{k=1}^{n_r} Q_k H_k - \gamma \sum_{n=1}^{n_n} (q_{ireq} h_{i min} + l_i h_i)$$

where  $l_i$  is the leakage at  $i$ -th node and  $Q_k$  includes both supplied flows and leakages. Hydraulic simulation results, shown in Table 4, point out that:

- a) the overall flow into the network  $ADV + TLV$  is 43.96 l/s, overall flow delivered to users  $ADV$  is 35.71 l/s and water losses  $TLV$  amount to 8.25 l/s, equal to 18.77% of the flow entering the network;
- b) nodes 3, 4 and 5 are in crisis, as they feature a pressure value lower than the required one. Resiliency index is reported as infeasible since the fraction of total power entering the network due to nodal flows is lower than  $\sum_i q_i req \cdot h_{i min}$  and then  $P_{int max} < 0$ ; in this case it makes little sense speaking about resilience, which is a measure of global energy redundancy.

Table 4. Nodal pressures, supplied discharges and indices values for the network base pattern.

Junction	$p$ [m]	$Q_{ero}$ [l/s]	$R_s$	0.9651
1	66.63	5.00	$E$	0.8123
2	55.80	5.00	$P$	0.7840
3	29.43	4.58	$I_r$	infeasible
4	29.00	4.55		
5	29.43	4.58		
6	50.81	5.00		
7	40.11	7.00	$ADV$	35.71
$ADV + TLV$		43.96	$TLV$	8.25

Starting from this critical configuration, the following possible intervention options on the network have been considered; their effects are summarized in Table 5.

- Option 1: water leakage detection and reduction allowing a reduction of the assumed leakages of about 40% of the original ones (at the reference pressure of 30 m).
- Option 2: water leakage detection and reduction and cleaning of all mains, which has the effect of reducing their roughness to an “as new” roughness value (i.e. 0.05 mm).
- Option 3: substitution of main 1, increasing its diameter to 250 mm, and of mains 3, 6, 9 and 10, with 150 mm diameter pipes. The roughness of substituted mains (made of the same material) will be obviously that corresponding to new pipes state ( $\varepsilon = 0.05$  mm). Considering that substitution of the mains results in cutting down a part of the losses (namely those due to failures of the joints and to fractures in the pipes, although leaks in users connections and secondary network remain), it has been assumed that losses along main 1 are negligible and those along mains 3, 6, 9 and 10 decrease of about a third of the original value.
- Option 4: substitution of main 1 with a 250 mm diameter main made in HDPE ( $\varepsilon = 0.01$ mm), and of mains 2, 7 and 8 with the same diameter HDPE mains.

By inspecting table 5, the following remarks can be made:

- Option 1 could entail a reduction of flow entering the network, an increase in delivered flows and water loss reduction ( $TLV / (ADV + TLV) = 10.54\%$ ), thus giving an increase of all indices.  $R_s$  could however remain lower than 1 due to the presence of nodes in crisis that supply a flow lower than the required one, while the negative value of  $I_r$  indicates that the network would not have any energy resource to face further future crises.
- Option 2 could involve a raise in all indices with respect to option 1 but efficiency  $E$ . Reduction of mains roughness causes an increase of heads that leads to a raise of both supplied flows at nodes in crisis and losses ( $TLV / (ADV + TLV) = 10.59\%$ ). This notwithstanding,  $R_s$  is lower than

1 and  $I_r$  is positive because the network still has an energy reserve, albeit not very high (about 9.6%). This example shows how the revised  $I_r$  index can be positive even in critical conditions, signalling network capability to cope with further operational stress.

- Option 3 is characterized by  $R_s = 1$  (showing that demands are fully met) and by  $I_r$  much higher than in the previous case: there are no longer critical nodes and the network has an energy reserve of about 81% which makes this option particularly reliable and robust. It is nevertheless worth noticing that overall network performance index  $P$  is lower than in previous two cases, since efficiency index  $E$  is lower.
- Finally, option 4 is essentially equivalent to option 3: the network could be just a little more efficient and a little less resilient. This shows that, when demand is fully met (i.e.  $R_s = 1$  and  $ADV = \sum_i Q_{req,i}$ ,  $i = 1, 2, \dots, N$  nodes, constant with node heads),  $I_r$  and  $E$  could have variations of opposite sign, since a higher  $I_r$  is due to higher node heads which in turn cause higher losses ( $TLV$ ) and thus a decrease of  $E$  (which is given by  $ADV / (ADV + TLV)$ ). This last remark can show the effectiveness of the conjunctive use of the four indices in the selection process of rehabilitation alternatives.

Table 5. Hydraulic simulation results and performance indexes of intervention options.

Option 1			Option 2					
Junction	$p$ [m]	$Q_{ero}$ [l/s]	$R_s$	$E$	$P$	$I_r$	$ADV$	$TLV$
1	67.93	5.00	0.9749	0.8946	0.8721	-0.8650	36.07	4.25
2	57.18	5.00						
3	30.92	4.70						
4	30.52	4.67						
5	30.92	4.70						
6	52.18	5.00						
7	41.55	7.00						
$ADV + TLV$		40.32						

Option 3			Option 4					
Junction	$p$ [m]	$Q_{ero}$ [l/s]	$R_s$	$E$	$P$	$I_r$	$ADV$	$TLV$
1	72.74	5.00	1.0000	0.8601	0.8601	0.8114	37.00	6.02
2	61.74	5.00						
3	36.48	5.00						
4	35.97	5.00						
5	36.48	5.00						
6	56.74	5.00						
7	46.53	7.00						
$ADV + TLV$		43.02						

From the above remarks, it could be concluded that the best solution in term of overall performance is the option 2, that well balances demand satisfaction and efficiency, but it is not resilient and hence not robust. As far as reliability and robustness are concerned, the best solution seems to be the option 3, which has the highest  $I_r$  and  $R_s = 1$ , albeit less efficient than option 2. This approach lends itself to be integrated within an optimization procedure. The application shown is actually rather simple, as far as network topology and number of mains is concerned, since its aim was to assess the usefulness of proposed

performance indices in the process of rehabilitation alternatives selection. In a real-world situation, featuring networks with several dozens of mains, resorting to multi-objectives optimization procedures may be indispensable to explore effectively the space of possible solutions, constituted by the combination of rehabilitation actions and pipes.

## **5. CONCLUSIONS**

The paper has described an approach to the selection of water network rehabilitation options based on the assessment of indices of efficiency, reliability and global performance of the network (Bertola e Nicolini, 2006) and of resiliency (Todini, 2000). The index of resiliency has been readapted to the context of rehabilitation and may be considered as measure of both reliability and robustness.

The case study shows the capability of such indices, when used together, to describe suitably the performance that the network would achieve in terms of demands met and service efficiency, should the proposed intervention options be enacted, and to provide information on the corresponding levels of reliability and robustness. These indices can be seen as good selection criteria of rehabilitation alternatives, while still requiring a relatively small amount of data.

In addition, the case study has also shown the usefulness of joining the resiliency index, which characterizes distribution service reliability and design robustness, to the other proposed performance indices.

It has been also observed that the adopted generalized formulation of the resiliency index is suitable to describe a network in critical conditions, and to describe the capability of the rehabilitation options to enable the network to face future operational stresses without further jeopardizing service levels.

In real world applications, the choice of the strategy to undertake should be obviously based on a trade-off between costs and performance, reliability and robustness levels of alternative interventions, as well as economic considerations regarding treatment and transportations costs of water and availability of supplying sources of adequate quality, which could lead to place more emphasis on efficiency issues rather than on the full satisfaction of water demands or on service reliability.

The paper has also described a sampling method of water uses for the draft of water balances at network district scale, in order to assess losses level. The application of the method to costumers uses of a small town in Sicily, Italy, shows that it may be applicable in technical practice and that it provides results close to the ones predictable in theory.

## **References**

Alegre, H., Hirner, W., Baptista, J. M., and Parena, R. (2000). *Performance indicators for water supply services*, IWA Manual of Best Practice, ISBN 9002222723.

Alvisi, S., and Franchini, M. (2009). "Multiobjective optimization of rehabilitation and leakage detection scheduling in water distribution systems", *Journal of water resources planning and management*, 135(6), 426 – 439.

Alvisi, S., and Franchini, M. (2006). "Near optimal rehabilitation scheduling of water distribution networks based on multi – objective genetic algorithm", *Civ. Eng. Environm. Syst.*, 23(3), 143 – 160.

Arena, C., Criminisi, A., Fortunato, A., and Mazzola, M. R. (2009). "Sampling of residential water use for leak control via water budgets". *Proceedings of World Environmental and Water Resources Congress 2009: Great Rivers*, ASCE, 346-355.

Bertola, P., and Nicolini, M. (2006). "Evaluating Reliability and Efficiency of Water Distribution Networks". *Management of Water networks. Proceedings of the Conference "Efficient Management of water networks design and rehabilitation techniques. Ferrara, May, 2006"*, Fondazione AMGA, Genova, 7-23.

Cascetta, F., Di Nardo, A., and Di Natale, M. (2004). "Distrettualizzazione delle reti idriche: riflessioni e

- indicazioni metodologiche”. *La gestione delle perdite idriche: la situazione in Italia*, Fondazione AMGA, Genova, 71-92.
- Cocchi, D. (2006). *Teoria dei Campioni - Corso base*, Dipartimento di Scienze Statistiche, Università di Bologna.
- Dandy, G.C., and Engelhardt, M. (2001). “Optimal scheduling of water pipe replacement using genetics algorithms”. *Journal of water resources planning and management*, ASCE, 127 (4), 214 – 223.
- Farley, M., and Trow, S. (2003). *Losses in water distribution networks – A Practitioner’s Guide to Assessment, Monitoring and Control*, IWA Publishing, UK.
- Hanke, S., and Mehrez, A. (1979). “An optimal sampling procedure for the collection of residential water use data”. *Water Resources Research* 15(6), 1343-1348.
- Mays, L.W. (2004). *Urban water supply management tools*, Mc Graw-Hill, New York.
- Mazzola, M. R., and Bazzurro, N. (2008). “Aspetti economici nelle strategie di gestione delle perdite idriche”. *Ricerca e controllo delle perdite nelle reti di condotte. Manuale per una moderna gestione degli acquedotti*, Città Studi ed., Novara, 3-42.
- Sægrov, S. (2005). *CARE-W: Computer Aided Rehabilitation for Water Networks*, IWA Publishing, London.
- Savic, D., (2005). “Coping with Risk and Uncertainty in Urban Water Infrastructure Rehabilitation Planning”. *Acqua e Città - I Convegno Nazionale di Idraulica Urbana. Sorrento 28-30 settembre 2005*.
- Todini, E. (2000). “Looped water distribution networks design using a resilience index based heuristic approach”. *Urban Water* 2.