European Water 58: 391-398, 2017.

Using bayesian networks for risk assessment of real losses in water distribution systems

M. Tabesh[1*], N.R. Faghihi[2], A. Roozbahani[3], R. Heidarzadeh[2] and B. Roghani[2]
[1] *Center of Excellent for Engineering and Management of Civil Infrastructure, School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran*
[2] *School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran*
[3] *Department of Irrigation and Drainage Engineering, College of Abouraihan, University of Tehran, Tehran, Iran*
[*] *e-mail: mtabesh@ut.ac.ir*

Abstract: Severe restrictions on water resources and rapidly increasing water demands as well as the threats arising from the design and the operation of water distribution networks (WDN) are unavoidable. Deciding on which alternative to choose amidst conflicting multiple criteria and different interests of stakeholders is a rather challenging task. Therefore, the recognition of major drivers of Non-Revenue Water (NRW) and its components in WDNs is an effective strategy that addresses water scarcity. Real losses are challenging components of NRW and lead to loss of about 10 percent of water inflow; hence it is crucial to identify and reduce the effective factors on them. This paper aims to identify effective factors on real losses and score their level of presence in the pilot through questionnaires. Then it presents a Bayesian Network (BN) to demonstrate the probability relationships between factors influencing real losses. To the authors' knowledge, this is the first attempt to model factors influencing real losses in WDN using BN. The methodology properly handles the unavoidable uncertainties in available information and lack of data. To demonstrate the application of the proposed model, District 4 of Tehran Water and Wastewater Company has been studied. Results indicate that "wrong and non-standard installation of pipes and other devices" and "poor training for workers and experts" have the most influence on real losses, respectively.

Key words: Bayesian network, questionnaire, real losses, water distribution networks

1. INTRODUCTION

One of the major challenges facing water utilities is the high level of water losses (Mutikanga et al., 2010). Water loss reduction is more cost effective than developing new water sources, which not only requires large amounts of financial resources, but also poses high costs and threats to the environment (Adams and LutzLey, 2012). Non-Revenue Water (NRW) is defined as the difference between total inflow and measured consumptions that has two components, apparent and real losses. Apparent losses are caused by unauthorized consumption, personnel errors, management and operational errors, data-handling errors, and unbilled authorized consumptions. Real losses comprise leakage from system parts and overflows of storage tanks. Losses are also caused by poor operations and maintenance, lack of active leakage control, and poor quality of underground assets and so on. Real losses contain visible leakage and invisible leakage.

Although it is widely acknowledged that NRW levels in developing countries are very high, in fact, very few data are available in the literature regarding the actual figures (Kingdom et al., 2006). This is largely because most water utilities in developing regions do not have adequate monitoring systems for assessing water losses and there is a great deal of uncertainty about data available (Kingdom et al., 2006). In the absence of adequate data and a proper methodology, most developing countries use default values. For instance, unauthorized consumption is computed as 0.5% of the total system input for NRW components (District 4 of Tehran Water and Wastewater Company, 2015). Moreover, since lack of accurate data regarding NRW and its components leads to data uncertainties, it is crucial to use appropriate probabilistic tools for modelling. In the context of a probabilistic framework, the Bayesian Network (BN) is naturally suited to handle such data cases and update the description with new information (Nannapaneni et al., 2016).

According to National Water and Wastewater Engineering Company of Iran reports, the level of average NRW is in the range of 25-30 percent of total inflow. Moreover, according to these reports around 40-60 percent of NRW is lost through pipelines and other parts of the urban water systems in Iran; so, consumers do not receive that amount. As a matter of fact, in order to preserve the water resources and prevent financial losses, it is crucial to gain a better understanding of all of the reasons and factors influencing real losses by clearly identifying design and performance parameters affecting the emergence of visible and invisible leakage.

Tabesh et al. (2009) focused mainly on leakage and proposed a methodology for assessment of NRW and water losses in WDN. In this research they applied some approaches and methodologies to estimate the nodal and pipe leakage as the main components of NRW. A methodology was presented by Mutikanga et al. (2011) for assessment of different components of apparent losses. They tried to determine the most significant components of apparent losses. In a separate research, Van den Berg (2015) assessed the major drivers of NRW. He proposed a water loss function that describes the relationship between NRW and physical characteristics of the water infrastructure, management characteristics of the utility, country effects and area in which the utility is located. Zyoud et al. (2016) applied MCDA (Multi Criteria Decision Analysis) in order to prioritize a set of water loss management strategies. They introduced the integration of fuzzy AHP (Analytic Hierarchy Process) and Fuzzy TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) to assign the importance of different options considering several stakeholder groups' preferences. Abu-Mahfouz et al. (2016) proposed a real-time DHM (Dynamic Hydraulic Model) connected to near real-time sensing and actuation capability on the WDN. This model addresses the limitation of current design and operation of most WDNs which are based on steady-state hydraulic models.

This paper mainly focuses on the evaluation and assessment of all of the factors influencing real losses and its components. The methodology is applied for District 4 of Tehran Water and Wastewater Company (DTWWC). The proposed methodology is a combination of designing a questionnaire to gather information and using BN to evaluate the probabilistic relationships between the factors. In the aggregate, influencing parameters will be ranked based on their extent of impact.

This paper is organized as follows. Section 2 outlines the concepts and frameworks of designing questionnaires and BN. Section 3 demonstrates the application of the proposed methodology in a case study. Finally the conclusions and discussion are drawn in section 4.

2. METHODOLOGY

2.1 Assessment of factors

The initial objective of this research is gaining a profound perception of factors influencing components of real losses. Therefore, the first step is bound to identify the threats and factors which increase the level of real losses. Using the concepts and information in global standard IWA (International Water Association) and taking advantage of the experiences of researchers and experts, all of the factors are collected for both visible and invisible leakage.

2.2 Preparing questionnaire

Owing to the qualitative nature of the specified factors and lack of quantitative data, it is necessary to find an approach to gather the required information of the factors. Quantifying real losses and its components using various computing methods mostly presented by IWA is insufficient to address the drivers of real losses. Therefore, a questionnaire was designed to collect qualitative data regarding the level of presence of all factors. In the questionnaire for each of the effective factors three states of “high”, “low” and “not available” were designed. On the one hand, this category makes the questionnaire convenient to understand and answer. On the other hand, the impact of each factor on its child factor would be clearly demonstrated.

2.3 Bayesian network

Bayesian belief network is a graphical model that permits a probabilistic relationship among a set of variables (Pearl, 1988). In the BN, there are two types of nodes: parent nodes and child nodes. One of the favourable characteristic of this network is the potential to transform qualitative data into equivalent quantitative measures. In other words, BN has the capability to combine the knowledge of expertise and available data.

For illustration, consider the BN shown in Fig. 1, which presents the joint probability distribution of three random variables A, B, C (B and C are parent nodes and A is child node). All three variables have two possible values, T (for true) F (for false). The joint probability function is (Nannapaneni et al., 2016):

$$P(A,B,C)=P(A).P(B|A).P(C|A,B) \tag{1}$$

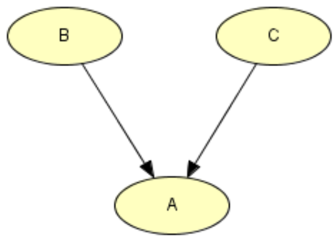

Figure 1. An example of a Bayesian Network

The first task in the development of BN is structure learning; but this step has been skipped owing to the fact that the cause-effect network of BN is fixed and defined in this article. The next step is the development of the conditional probability table (CPT) which determines the probability distribution of the value of a variable starting from the probability distribution of the parent variables (Pagano et al., 2014). In this article the Bayesian model was developed by aggregating the experts’ knowledge from questionnaires as input data.

2.4 Sensitivity analysis

In order to use the probability distribution of variables in BN model, sensitivity analysis is carried out. In this regards, for each parameter, the Sensitivity Index (SI) is computed as:

$$SI=(\Delta C_i/\Delta P_i)\times 100 \tag{2}$$

where ΔPi: is the variation of parent node and ΔCi: is the variation of child node. To apply this approach, one specific state of all of the parameters increases to the same level. Considering the variation of the parent node, for all of the influencing factors on real losses, the SI would be computed. It is worth mentioning that the higher the SI, the higher the priority of a factor for real losses risk management.

3. CASE STUDY

The proposed methodology is applied on the (DTWWC), the area covered by tank no.2. Special features of this area distinguishes it from other areas; for instance, its old water network, the high level of groundwater that contributes to decay of pipes and repair and maintenance problems of urban design. Based on the reports from DTWWC, NRW was 22% and real losses account for about 9.5% of the total NRW in 2015 (District 4 of Tehran Water and Wastewater Company, 2015).

3.1 Data collection

Following the collecting all of the probable factors affecting real losses in the case study, their level of presence are asked through questionnaires. The designed questionnaire is shown in Table 1.

Table 1. A questionnaire on factors influencing real losses

	Factors	**State**		
		High	**Low**	**Not available**
Visible leakage	Inappropriate quality selection for pipe and other devices			
	Wrong and non-standard installation			
	Failure to collect and complete details of events			
	Low speed and quality of repair			
	Lack of regular inspection and navigation of networks			
	Poor movement of pipes and other devices			
	Poor construction of pipes and other devices			
	Wrong design of network			
	Lack of timely replacement of devices			
	Lack of pressure management			
	Lack of investigating the cause of fractures			
	Poor management of overflow of tanks			
	Lack of modern leakage management technologies (telemetry-SCADA)			
	Lack of prevention maintenance (PM)			
	Financial shortage in w/w companies			
	Poor training for workers and experts			
	Poor management (lack of updated maps and GIS)			
Invisible Leakage	Lack of step testing			
	Lack of District Metering Areas (DMA)			
	Lack of trying to find the cause of leakage in the network			
	Poor management (lack of updated maps and GIS)			
	Failure to purchase and use of precise measurement and calibration tool			
	Failure to purchase and use of leakage detection tools and modern technologies in leakage management			
	Inappropriate quality selection for pipes and other devices			
	Incorrect and non-standard installation			
	Lack of prevention maintenance (PM)			
	No justification for leakage detection because of the low price of water			
	Poor movement of pipes and other devices			
	Poor construction of pipes and other devices			
	Incorrect design of network			
	Lack of timely replacement of devices			
	Lack of pressure management			
	Lack of timely and continuous leakage detection			
	Financial shortage in w/w companies			
	Poor training for workers and experts			
	Poor operational procedures and contracts			

3.2 Summary questionnaire data

In order to decrease the uncertainty of responses, knowledgeable and experienced personnel in Tehran Water and Wastewater Company and university professors were asked to fill out the questionnaires based on their area of expertise. In the aggregate, 36 questionnaires were returned completed. Analysing questionnaires data leads to identifying the final factors in BN modelling. In addition, the personnel's response to each factor is considered as Bayesian model input.

3.3 Application of Bayesian Network

Base on the questionnaires data, some of the factors which have similar states combine as a single factor. Therefore, the final list of factors influencing real losses comprise of 23 parameters that affect visible leakage or invisible leakage or both. The next step is defining the probability relationship of these parameters using BN. In this article an available software has been used; Hugin Lite 8.4, for constructing the BN graphical model (Fig. 2). In this network all of the nodes have two possible states which are determined as H (High) and L (Low). Besides, selected labels for each node are shown in Table 2.

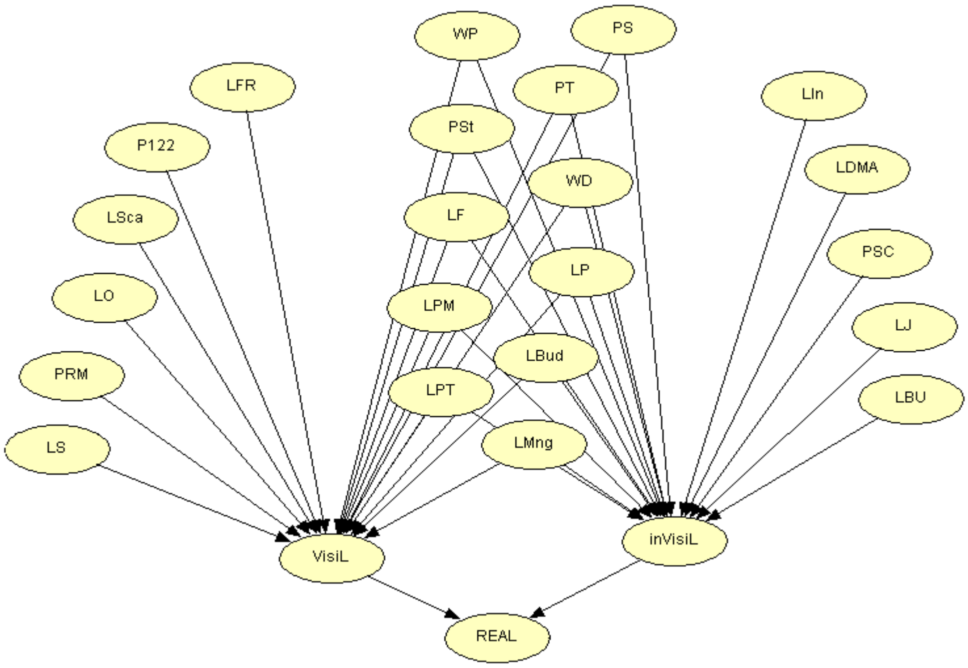

Figure 2. Bayesian Network

The next important step is parameter learning. In this article, expectation-maximization algorithm (EM) has been used for parameter learning. EM algorithm is particularly useful when probability distributions of the variables in a BN are unknown (HUGIN EXPERT A/S, 2016).

Since the prepared questionnaires merely asks about the state of the identified factors, information relative to the state of real “visible leakage”, “invisible leakage” and “real losses” were required. Therefore, for each parameter, the weighting approach was applied by assigning 7 to “high” state, 5 to “low” state and 1 to “N/A” state. As a result the percentage of visible and invisible leakage in each questionnaire is determined. In order to consider real condition of the WD in the pilot, the percentage of real losses was derived from annual water balance reports of DTWWC from 2011 to 2015. Using the percentage of visible and invisible leakage of these reports indicates that visible and invisible leakage account for 70% and 30% of real losses, respectively. Finally, for each questionnaire, the state of real “visible leakage”, “invisible leakage” and “real losses” were defined. Eventually, the BN runs and the probabilities of a node in a certain state are shown in Fig. 3.

According to sensitivity analysis, a specific state of all of the nodes is changed to a certain extent and the changes in child nodes will be evaluated. Generally, the more parent nodes change, the more accurate sensitivity analysis is done. Nevertheless, the node “LBud” is restrictive and the factors have to be increased up to 25%. To do so in the software, the “high” and “low” states in Conditional Probability Table (CPT) have been altered for each node manually. Then, the sensitivity index is calculated for all of the nodes. For instance; for lack of investigating the cause of fractures (LFR) node, the “high” state is 57.58% and it would change into 71.98%. This alteration contributes to the modification of its child nodes; real losses, from 35.04% to 35.07%. So, the SI for the LFR node is computed as:

$$SI = (71.98-57.58)/(35.04-35.07) = 0.002084 \quad (3)$$

Table 2. Labels of each factor in Bayesian Network

Labels	Factors
VisiL	visible leakage
inVisiL	invisible leakage
LS	low speed and quality of repair
PRM	poor management of overflow of tanks
LO	lack of regular inspection and navigation of networks
LSca	lack of modern leakage management technologies
P122	failure to collect and complete details of events
LFR	lack of investigating the cause of fractures
WP	incorrect and non-standard installation
PS	inappropriate quality selection
PT	poor movement
PSt	poor construction
WD	incorrect design of network
LF	lack of timely replacement of devices
LP	lack of pressure management
LPM	lack of prevention maintenance
LBud	financial shortage in w/w companies
LPT	poor training for workers and experts
LMng	poor management
LIn	lack of trying to find the cause of leakage in the network
LDMA	lack of step testing and district metering areas
PSC	poor operational procedures and contracts
LJ	no justification for leakage detection because of the low price of water
LBU	failure to purchase and use of precise measurement and calibration tool, modern technologies in leakage management failure to purchase and use of leakage detection tools and

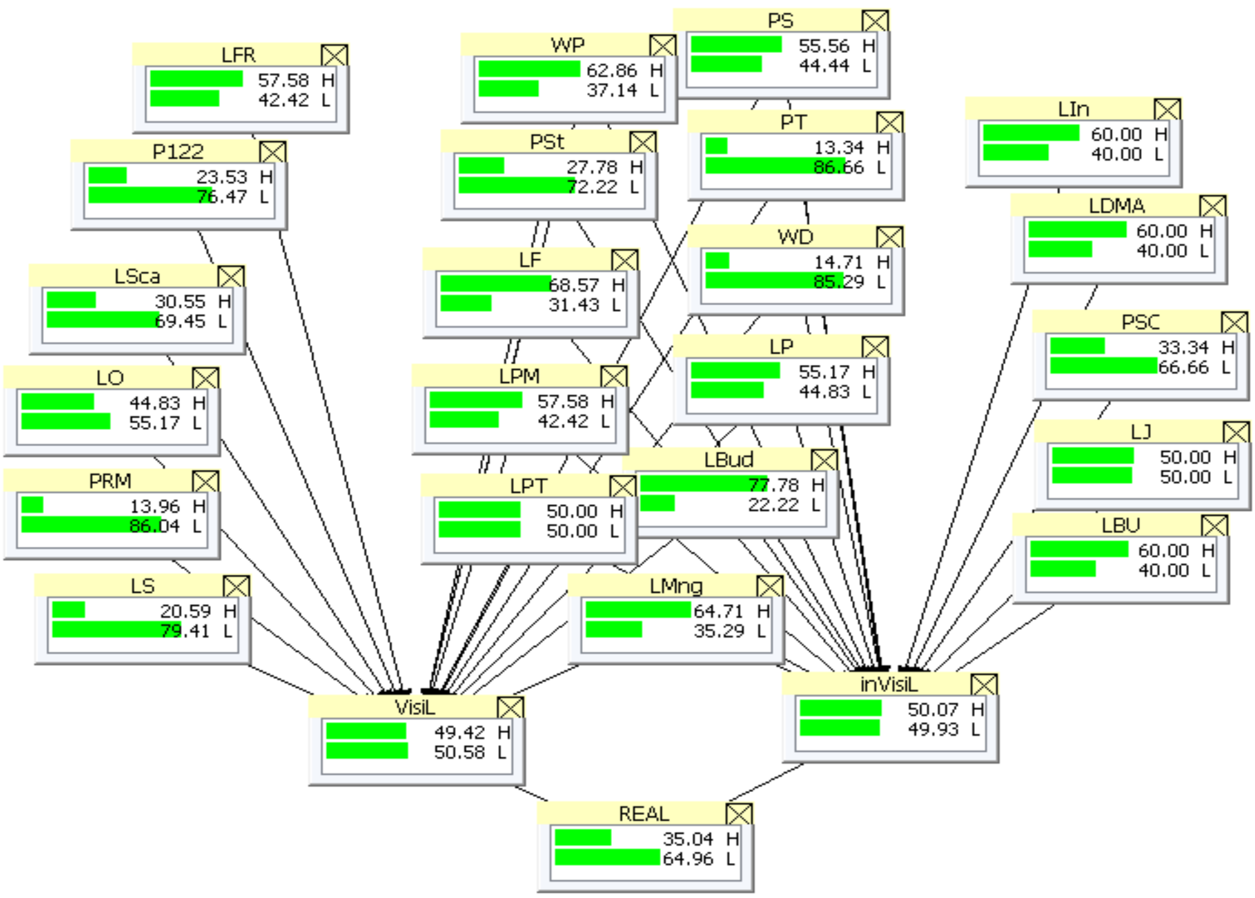

Figure 3. Trained Bayesian Network

Table 3. Ranking factors influencing real losses

Rank	Factors influencing real losses	Sensitivity Index (%)
1	Incorrect and non-standard installation	0.891
2	Poor training for workers and experts	0.88
3	Lack of timely replacement of devices	0. 875
4	Inappropriate quality selection	0.864
5	Poor management	0.433
6	Poor construction	0.43
7	Failure to collect and complete details of events	0.34
8	Poor movement	0.3
9	Lack of prevention maintenance	0.278
10	Incorrect design of network	0.272
11	Lack of regular inspection and navigation of network	0.267
12	Lack of pressure management	0.218
13	Lack of investigating the cause of fractures	0.208
14	Low speed and quality of repair	0.194
15	Financial shortage in w/w companies	0.103
16	Lack of trying to find the cause of leakage in the network	0.067
17	Lack of step testing and District Metering Areas Failure to purchase and use of precise measurement and calibration tool and Failure to purchase and use of leakage detection tools and modern technologies in leakage management Poor management of overflow of tanks Lack of modern leakage management technologies Poor operational procedures and contracts No justification for leakage detection because of the low price of water	0

4. CONCLUSION

Since authorities in water companies seek to find an optimal solution to achieve an economic balance between the costs of physical water losses control and the benefits that accrue, it is important to prioritize the most effective factors of real losses components. In this study, initially, effective factors on real losses were identified and a questionnaire was designed to ask the state of these factors (high or low) in the case study, DTWWC. Then, BN, a graphical model, was constructed to evaluate the conditional probabilities of the factors. After that, by sensitivity analysis of the model output, effective factors were prioritized in order to find out which of them have the most impact on NRW. The results showed that among all of the factors influencing real losses, “incorrect and non-standard installation of pipes and other devices”, “poor training for workers and experts” and “lack of timely replacement of devices” were respectively the significant components of real losses in this case study. It is worth mentioning that the results of the BN and sensitivity analysis particularly depend on the questionnaire data and their accuracy; hence, if respondents change their answers, the priority of strategies for reducing real losses will definitely be altered. The result of this research/article will be a valuable reference for senior managers and decision makers, who may require an overview of the principles and procedures for controlling real losses.

REFERENCES

Abu-Mahfouz, A. M., Hamam, Y., Page, P. R., Djouani, K., & Kurien, A., (2016) Real-time dynamic hydraulic model for potable water loss reduction. Procedia Engineering, 154, 99-106.

Adams, A. S., & LutzLey, A. N., (2012) An empirical study of factors affecting water loss in Mexican cities. Journal of Water Sustainability, 2(3), 167-178.

District 4 of Tehran Water and Wastewater Company, (2015) Annual Water Balance reports from 2012 to 2015. Tehran, Iran.

HUGIN EXPERT A/S, (2016) Hugin Researcher Userguide, version 7.8. HUGIN EXPERT A/S, Denmark.

Kingdom, B., Liemberger, R., & Marin, P., (2006) The challenge of reducing non-revenue water (NRW) in developing countries. How the Private Sector Can Help: A Look at Performance-based Service Contracting. Water Supply and Sanitation Board Discussion Paper Series, Paper No. 8, The World Bank Group, Washangton DC, USA.

Mutikanga, H. E., Sharma, S. K., & Vairavamoorthy, K., (2011) Assessment of apparent losses in urban water systems. Water and Environment Journal, 25(3), 327-335.

Mutikanga, H. E., Sharma, S. K., & Vairavamoorthy, K., (2011) Multi-criteria decision analysis: a strategic planning tool for water loss management. Water Resources Management, 25(14), 3947-3969.

Nannapaneni, S., Mahadevan, S., & Rachuri, S., (2016) Performance evaluation of a manufacturing process under uncertainty using Bayesian networks. Journal of Cleaner Production, 113, 947-959.

National Water and Wastewater Engineering Company of Iran, (2015)

National Water and Wastewater Engineering Company of Iran, (2015) Water and wastewater industryat a glance report. Tehran, Iran.

Pagano, A., Giordano, R., Portoghese, I., Fratino, U., & Vurro, M., (2014) A Bayesian vulnerability assessment tool for drinking water mains under extreme events. Natural Hazards, 74(3), 2193-2227.

Pearl, J., (1988) Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Francisco: Morgan Kaufmann Publishers, Inc.

Tabesh, M., Yekta, A. A., & Burrows, R., (2009) An integrated model to evaluate losses in water distribution systems. Water Resources Management, 23(3), 477-492.

van den Berg, C., (2015) Drivers of non-revenue water: A cross-national analysis. Utilities Policy, 36, 71-78.

Zyoud, S. H., Kaufmann, L. G., Shaheen, H., Samhan, S., & Fuchs-Hanusch, D., (2016) A framework for water loss management in developing countries under fuzzy environment: Integration of Fuzzy AHP with Fuzzy TOPSIS. Expert Systems with Applications, 61, 86-105.