

RESEARCH

Open Access



# Adaptation of multiple regression analysis to identify effective factors of water losses in water distribution systems

Dongwoo Jang , Gyewoon Choi and Hyoseon Park\*

\* Correspondence:  
hyoseondw@gmail.com  
Incheon National University,  
Incheon, South Korea

## Abstract

It is important to manage leaks in water distribution systems by smart water technologies. In order to reduce the water loss, researches on the main factors of water pipe network affecting non-revenue water (NRW) are being actively carried out. In recent years, research has been conducted to estimate NRW using statistical analysis techniques such as Artificial Neural Network (ANN) and Principle Component Analysis (PCA). Research on identifying factors that affect NRW in the target area is actively underway. In this study, Principle components selected through Multiple Regression Analysis are reclassified and applied to NRW estimation using PCA-ANN. The results show that the principal components estimated through PCA are connected to the NRW estimation using ANN. The detailed NRW estimation methodology presented through the study, as a result of simulating PCA-ANN after selecting statistically significant factors by MRA, forward method showed higher NRW estimation accuracy than other MRA methods.

**Keywords:** Smart water management, Non-revenue water ratio, Water distribution systems, Principal component analysis, Multiple regression analysis, Artificial neural networks

## Introduction

Smart water grids (SWGs) are required for water supply systems for use in water management platforms, which integrates information and communication technology (ICT) into a single water management scheme. SWG technology is seen as a promising solution for resolving recent critical water problems in water distribution systems (Lee et al., 2015).

Water distribution systems are subject to deterioration over time, it is usually leads to problems like decreased capacity of water supply facilities, water loss, service disruption and lower water quality (Saldarriaga et al., 2010). To overcome pressure management problems and ensure continuous, efficient and economic operation of water distribution systems, an effective rehabilitation strategy is required. (Engelhardt et al., 2000). Since the economic resources available for the rehabilitation of water distribution systems are scarce, assistance in prioritization of investment is important (Halhal et al., 1997).

The International Water Association (IWA) has acknowledged this problem and established the Water Loss Task Force (WLTF). The WLTF examined international best practices and developed a standardized terminology for non-revenue water (Frauendorfer and Liemberger, 2010).

Non-revenue water (NRW) includes physical (leaks) and commercial losses (illegal connections, unmetered public use, meter error, unbilled metered water and water for which payment is not collected) (Wyatt, 2012). IWA has proposed performance indicators (Alegre et al., 2000; Lambert and Hirner, 2000). Also a percentage indicator was suggested not to being used in performance comparison, especially where target areas see large differences in consumption per service area (Lambert, 2002).

In this study, a methodology for NRW ratio estimation for smart water management was studied. NRW was estimated using multiple regression analysis, principal component analysis (PCA), and artificial neural network (ANN). In particular, the main parameters of the water pipe network for predicting the NRW are set as input data, which is expected to be helpful in selecting the factors affecting leakage in smart water management. And various statistical analysis techniques were used to predict NRW. There are various studies of estimating NRW using ANN was performed by Jang, et al. (2017). It is proved that ANN show better results than MRA in NRW estimation (Jang & Choi, 2017, Jang 2017). In particular, Jang (2017, 2018) and Jang et al. (2018) suggested that the combination of PCA and ANN is the optimal method for estimating NRW using statistical methods.

In this study, various selected cases by MRA were reclassified to select optimal PCA factors for ANN analysis. Therefore, we prove that PCA-ANN with pre-application using specific MRA method is applied sequentially to achieve optimal NRW estimation. Statistical analysis methodology for estimating NRW was presented, and NRW observations and estimated values were compared in real site.

**Previous researches for non-revenue water analysis**

**Evaluation of water balance in water supply systems**

As components of water balance in Korea and their definitions are little different from those of IWA, they were rearranged as shown in Table 1 (Chung et al., 2004). Metering and under-registration was re-calculated and the remaining amount of the recalculation was added to ineffective water, which was considered equivalent to the real losses of IWA's water balance.

Because of the different definitions and lack of well-documented procedures for several components (e.g., supplier's official use, public use and metering under-registration) select

**Table 1** Components of Water Balance (Lambert, 2002)

System input volume	Authorized consumption	Billed authorized consumption	Billed metered consumption (including water exported)	Revenue water
			Billed non-metered consumption	
		Unbilled authorized consumption	Unbilled metered consumption Unbilled non-metered consumption	Non-revenue water
Water losses		Apparent losses	Unauthorized consumption Metering inaccuracies	
		Real losses	Leak on transmission and/or distribution mains Leaks and overflows at utility's storage tanks Leaks in service connections up to customer meters	

data with their own inaccuracies. Mean hydraulics pressure and location of customer meters were estimated from limited samples, possibly causing variations (Jang, 2017).

This study focused on physical parameters related to water distribution systems. Physical parameters were selected and measured data was also used for estimating NRW. Table 1 shows the components of water balance in water distribution systems by IWA.

Combined water balance in the network could be calculated by real measured data but doing so in real water distribution systems should be difficult because of unconstructed DMAs (District Metered Areas) and the design error of water distribution systems. Also, periodically operational management is an essential element in water distribution systems such as finding leaky pipes, management of hydraulic pressure and proper pump operation.

### **Calculation of NRW ratio in water distribution systems**

For NRW estimation, governments and institutes around the world are estimating leaks using those occurring in infrastructure. To calculate NRW, a formalized system is needed that calculates the NRW ratio by introducing physical parameters that reflect regional characteristics (Jang, 2017).

The world produces around 33 billion cubic meters of NRW every year, mostly caused by leak in water supply systems until 2006. Furthermore, around 16 billion cubic meters are delivered to customers but not paid for. Nearly 55% of NRW occurs in developing countries, where financing for the maintenance and expansion of water supply and sanitation systems are urgently needed, and bad water quality causes disease (Kingdom et al., 2006).

To perform reliable analyses of NRW and leaks, the management history of each system and District Metered Area (DMA) should be separately supervised. In addition, when analyzing the effect of the NRW project, analysis using the minimum flow rate at night is needed. The Process of NRW analysis can be divided into three stages. First, Design of DMA will be established in the determination of the initial NRW for local waterworks at the beginning of the improvement project. Second and third, the NRW analysis stage is divided into before and after the building of the DMA system (Park, 2014).

NRW analysis can improve water supply system by performing detailed leak analysis when DMA is established after selecting the initial NRW by main parameters of water distribution system analysis.

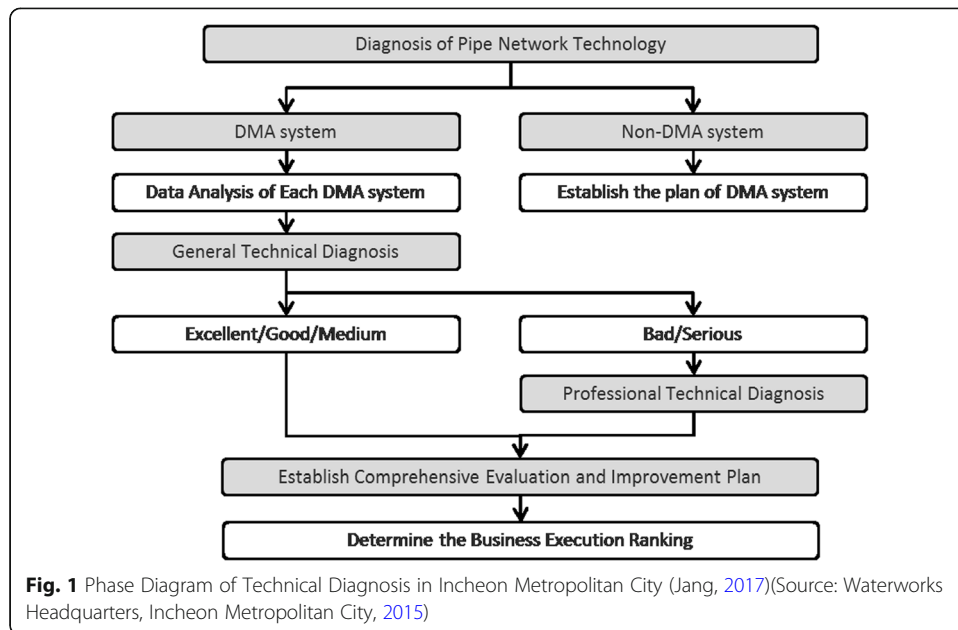
### **Methodology for NRW estimation**

#### **Phase diagram of technical diagnosis**

The technical diagnosis in the water distribution systems is conducted to rationally investigate and diagnose the status of the operation and management of water supply facilities to identify problems such as those of water quality and consumption of water demand at service area. The phase diagram of technical diagnosis in water distribution systems is shown in Fig. 1 (Jang, 2017).

#### **Classification of Main parameters for NRW ratio analysis**

Statistical methods was conducted on operational and physical parameters, and main factors were extracted by PCA. The operational and physical parameters for prediction of the NRW were categorized and the expected NRW ratio was compared with



measured NRW. Also, a statistically significant group was selected through multiple regression analysis (MRA).

Figure 2 shows the parameter classification process for calculation of the NRW ratio. Based on the basic statistical analysis, correlation between the main parameters of water distribution systems were analyzed. The main parameters of water distribution systems were classified as independent variables for simulation of ANN.

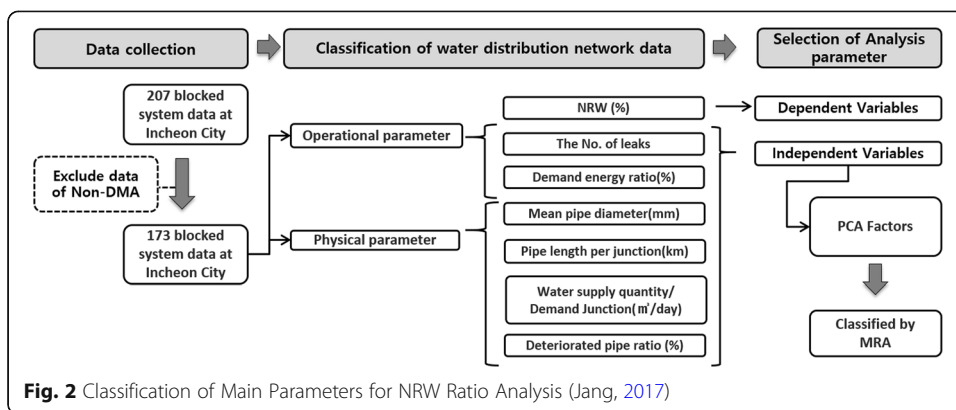
The principle components were converted through basic statistical analysis, data standardization. MRA was used to generate the independent parameters with conditions satisfying significant probability (Jang, 2017).

MRA selects independent variables according to statistical significance. The selected independent variable is described by a linear equation based on a combination of specific coefficient values and it is used to verify statistical significance with the dependent variable (NRW).

In this study, the PCA factors calculated from Jang 2017 study were applied to various MRA techniques. Therefore, the factor by PCA was reclassified and finally applied to ANN. This can be used as a basis for determining which of the MRA methods is suitable for PCA-ANN compared to the method of eliminating factors according to the existing statistical significance of MRA.

**Statistical analysis procedure**

Figure 3 shows the multiple regression analysis (MRA) procedures and its research methodology proposed by Jang (2017). In this study, PCA factors were reclassified by MRA based on the analysis sequence in Fig. 3. In the previous study, the basic factors shown in Fig. 2 were generated by PCA with 6 factors and the method of eliminating the factor with low statistical significance of MRA was used. This study differs from the previous studies in that six PCAs are newly constructed by various MRAs. Various data of DMA used to confirm the main parameters. MRA performed with selected parameters, and optimal parameters related to NRW ratio were estimated according to the



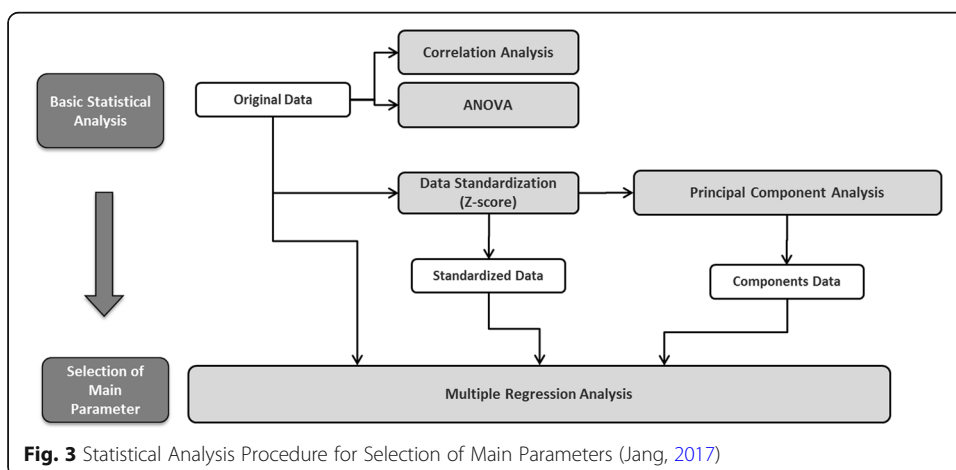
results of significant probability. A multiple regression equation using six independent parameters was derived by using the Enter, Stepwise, and Elimination method for estimating the NRW ratio using ANN.

In the previous study, the basic factors shown in Fig. 2 were generated by PCA with 6 factors and the method of eliminating the factor with low statistical significance of MRA was used. This study differs from the previous studies in that six PCAs are newly constructed by various MRAs.

### Verification of developed methodologies

The test bed for this study was the administrative area of Incheon, S. Korea. The data were surveyed on the status of the area, waterworks facilities and their operational rules, and the water supply indicators of Incheon waterworks (Incheon Metropolitan City, 2015). In addition, data from water distribution systems and simulation were collected (Jo, 2017 and Jang, 2017).

The input variables used were applied to the MRA using the six main components selected from the previous study (Jang, 2017). There are five statistical methods in MRA. In this study, Enter, Elimination, Backward, Stepwise and Forward methods of MRA were applied to select input variables for ANN. As a result of applying these five



MRA methods to six main parameters, Principle components were selected from 6 main components in Input, Delete, and backward methods.

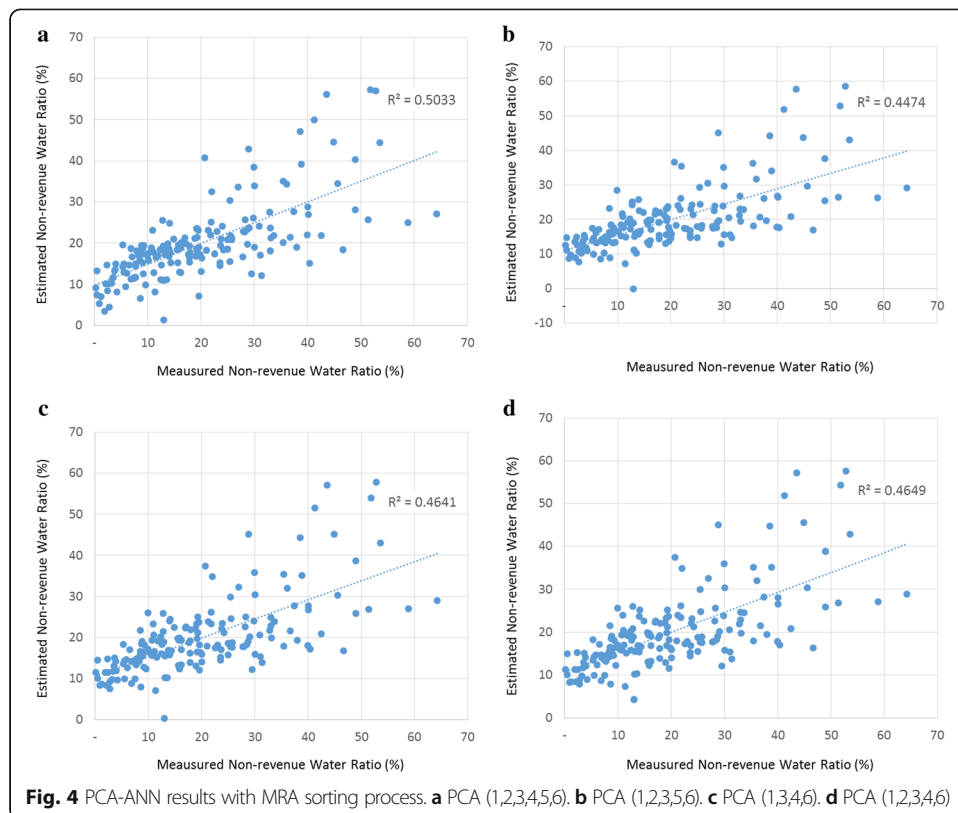
In the Stepwise method, four variables (principle components - 1, 3, 4 and 6 were selected as statistically significant. Forward method of MRA, five variables (principle components - 1,2,3,4 and 6 were selected as statistically significant. Figure 4 shows the NRW estimation results between actual and estimated values using ANN simulation.

Among the four simulation results, the comparison result that is similar to the measured NRW is the result of using six main components. The factors were statistically sorted by MRA in order of significance, but  $R^2$  was the highest for all six principle components.

Thus, even though the factor selection by MRA may be statistically significant, it was concluded that the selected six principle components were all involved in NRW estimation. In addition, the stepwise, one of MRW method is more accurate than other MRW method in the case of reclassifying factors through MRA among the results of NRW estimation using PCA-ANN.

As a result of the study, it is found that the case using the six factors is the most accurate in NRW estimation as in the study of existing Jang (2017, 2018). For the other five factors, the condition using the factor selected by the Forward Method of MRA was the second optimal method.

In addition, NRW estimation accuracy by stepwise and forward method was similar to each other. The selection of the proposed MRA method through this study needs to be applied to various regions and additional factors are required. In particular, it is expected that research to improve reliability by applying  $R^2$  to regions with high accuracy in NRW estimation should be given priority.



## Conclusions

The NRW estimation method for leak management for smart water system was analyzed. Statistical methods were used for this and NRW was estimated after re-selecting the factors by MRA in the conventional PCA-ANN method. A methodology for estimating the NRW ratio using newly suggested PCA-ANN with MRA was suggested by selected parameters for analyzing leaks in water distribution systems. This study drew the following conclusions.

NRW estimation method for smart water management is proposed. A variety of statistical techniques have been used from the factor selection to the NRW estimation using ANN. In particular, the six principal component factors selected through previous studies were re-selected as statistically significant factors through MRA and applied to NRW estimation.

As a result of simulating PCA-ANN after selecting statistically significant factors by MRA, forward method showed higher NRW estimation accuracy than other MRA methods. In this study, six principal components were used and the PCA-ANN results showed that all six major components were closely related to NRW prediction. In the future, additional studies are required to collect the data from new test-bed areas and verify that they are applicable in selected region.

The forward method of MRA showed the best performance, but reliable estimation of area and factor data is required because overall estimation accuracy by ANN is not high. Although the increase in accuracy is not high, MRA can play a role in improving the accuracy slightly.

## Acknowledgements

This research was supported by the Smart Water Journal, 2018.

## Funding

Our paper was invited from SWGIC 2017, and the code numbers are as follows.  
Code: SB-SKY-SMRT-1017.

## Availability of data and materials

We allow sharing of the Data and Materials used in this study.

## Declarations

We have confirmed that there is no potential competing interests. All authors have verified the submitted manuscript and this paper has not been published in any other journals.

## Authors' contributions

Dongwoo Jang carried out the smart water management studies, and wrote manuscript. Hyoseon Park carried out the statistical analysis. Gyewoon Choi participated in the design and draft of manuscript the study. All authors read and approved the final manuscript.

## Competing interest

The authors declare that they have no competing of interest.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 September 2018 Accepted: 12 December 2018

Published online: 08 January 2019

## References

- Alegre H, Hirner W, Baptista JM, Parena R (2000) Performance indicators for water supply services. IWA Publishing
- Chung SH, Lee HK, Koo JY, Yu MJ (2004) Characterization of the ratio of revenue water in the 79 cities by principal component analysis and clustering analysis. 2004 joint conference of KSWQ and KSWW, the Korean Society of Water and Wastewater. Republic of Korea, pp:133–142. [http://210.101.116.28/W\\_files/kiss3/07702915\\_pv.pdf](http://210.101.116.28/W_files/kiss3/07702915_pv.pdf).
- Engelhardt MO, Skipworth PJ, Savic DA, Saul AJ, Walters GA (2000) Rehabilitation strategies for water distribution networks: a literature review with a UK perspective. *Urban Water* 2(2):153–170. [https://doi.org/10.1016/S1462-0758\(00\)00053-4](https://doi.org/10.1016/S1462-0758(00)00053-4)

- Frauendorfer R, Liemberger R (2010) The Issues and Challenges of Reducing Non-Revenue Water. Asian Development Bank, Philippines
- Halhal D, Walters G A, Ouzar D, Savic DA (1997) Water Network Rehabilitation with a Structured Messy Genetic Algorithm, *Journal of Water Resources Planning and Management*, 123(3), pp. 137–146. <https://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%281997%29123%3A3%28137%29>
- Jang DW (2017) Estimation of Non-Revenue Water Ratio Using PCA and ANN in Water Distribution Systems, Incheon National University, Republic of Korea, Ph.D. thesis, Incheon
- Jang DW (2018) A parameter classification system for nonrevenue water management in water distribution networks. *Advances in Civil Engineering* 1(10):1–10. <https://doi.org/2018/2018/3841979ht>
- Jang DW, Choi GW (2017) Estimation of non-revenue water ratio for sustainable management using artificial neural network and Z-score in Incheon, Republic of Korea. *Sustainability* 9(11):1–15. <https://doi.org/10.3390/su9111933>
- Jang DW, Park HS, Choi GW (2018) Estimation of leakage ratio using principal component analysis and artificial neural network in water distribution systems. *Sustainability* 10(3):1–13. <https://doi.org/10.3390/su10030750>
- Jo, H. G. (2017) Study on Influence Factors of Non-revenue Water for Sustainable Management of Water Distribution Networks, Ph.D. Thesis, Incheon National University, Republic of Korea.
- 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. The World Bank, USA
- Lambert, A. O., 2002, International report on water losses management and techniques, Water Sci Technol Water Supply, IWA Publishing. 2(4), pp.1–20. DOI: <https://doi.org/10.2166/ws.2002.0115>
- Lambert AO, Hirner WH (2000) Losses from water supply system: standard terminology and performance measure, IWA the blue pages, vol 1-13. international water association, London
- Lee SW, Sarp S, Jeon DJ, Kim JH (2015) Smart water grid: the future water management platform. *Desalin Water Treat* 55(2): 339–346. <https://doi.org/10.1080/19443994.2014.917887>
- Park, C. S., 2014, A Case Study on Establishment of Block System for the Increase of Revenue Water in Distribution Systems, Master's Thesis, Chonnam National University, Republic of Korea. (In Korean)
- Saldarriaga JG, Ochoa S, Moreno ME, Romero N, Cortes OJ (2010) Prioritized rehabilitation of water distribution networks using dissipated power concept to reduce non-revenue water. *Urban Water J* 7(2):121–140. <https://doi.org/10.1080/15730620903447621>
- Waterworks Headquarters, Incheon Metropolitan City, 2015, Basic Plan of Waterworks Maintenance in Incheon, Incheon Metropolitan City. (In Korean)
- Wyatt AS, Shafei M (2012) Non-revenue water: financial model for optimal Management in Developing Countries. *Water Science & Technology Water Supply* 12(4):451–462. <https://doi.org/10.2166/ws.2012.014>

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---