Determinants of non-revenue water for a water utility in California

Gamze Güngör-Demirci, Juneseok Lee, Jonathan Keck, Robert Guzzetta and Paul Yang

ABSTRACT

Physical, technical, managerial, and environmental factors are all known to influence non-revenue water (NRW) volume, so a better understanding of these factors is important if we are to intervene in water loss problems more effectively. This study therefore identified determinants of NRW for a water utility in California by applying fixed effects panel regression analysis incorporating uncertainty. Network length, connection density, and net operating revenue per cubic meter of water sold were found to be negatively correlated with NRW while a positive relationship between number of leaks and NRW was identified. These findings will inform the water utility's management team/decision-makers regarding the specific impacts of NRW's critical factors and guide them to focus on these factors to further reduce NRW as well as set long-term benchmarks.

Key words | California, fixed effects panel regression, non-revenue water, uncertainty, water loss, water utility

Gamze Güngör-Demirci
Juneseok Lee (corresponding author)
Department of Civil and Environmental
   Engineering,
San Jose State University,
San Jose, CA,
USA
E-mail: *juneseok.lee@sjsu.edu*

Jonathan Keck
Robert Guzzetta
Paul Yang
California Water Service Company,
San Jose, CA,
USA

INTRODUCTION

As of 2016, California was the most populous state in the United States, with the current population of over 39.4 million projected to rise to 50 million by 2055 (State of California 2017). This is also a drought-sensitive region, with some areas being classed as semi-arid. These factors combine to make California one of the most water-stressed regions in the US, hence developing effective water conservation and water loss prevention strategies have become a priority. California Senate Bill 555, passed in October 2015, requires urban water suppliers in the state to submit a completed and validated water loss audit report annually to the California Department of Water Resources (DWR) beginning in October 2017. This new requirement is an indication of how seriously the problem of water loss is being taken at the state level.

Water loss from distribution networks is becoming a serious problem for natural resource management worldwide, especially given the impact of changing climate (Kanakoudis & Tsitsifli 2010; van den Berg 2015). The high cost of water production and delivery is also ensuring that utilities pay more attention to water loss issues (González-Gómez et al. 2012). According to the World Bank, more than $14 billion is lost every year by water utilities around the world due to water loss (Kingdom et al. 2006). Ultimately, the financial burden of this inefficiency is borne by paying customers (AWWA 2016a). As one of the major challenges water utilities are facing, the problem of water loss requires a solid and effective management strategy built on a fundamental understanding of the influencing factors (Kanakoudis & Tsitsifli 2010).

According to a survey carried out in 2012 across 48 cities in both Organisation for Economic Co-operation and Development (OECD) countries and emerging economies, in well-operated water utilities in OECD countries, water loss can be as high as 37%, while in developing countries this can climb to 65% (OECD 2016). This suggests that it is not simple to reduce the volume of lost water even in highly developed countries (Thornton et al. 2008; Wu

2011). Understanding the factors affecting water loss and what makes its reduction so difficult is crucial to implementing an effective water loss reduction program (van den Berg 2015). Physical, technical, managerial, sociological, and environmental factors such as the age of the systems, the length and type of networks, pressure in the systems, soil conditions, topography, traffic loading, and density of connections are all known to have an impact on the volume of water lost in distribution systems (Rajani & Kleiner 2001; González-Gómez *et al.* 2011). If we can develop a better understanding of these factors, which may or may not be under the utilities' control, more realistic target levels for water loss can be established in the long term.

Non-revenue water (NRW) is the term recommended by the International Water Association (IWA) Task Forces on Water Losses and Performance Indicators to represent the water loss in distribution systems. It is defined as the difference between the volume of water put into a water distribution system and the volume that is billed to customers (Farley & Trow 2003; Alegre *et al.* 2006). In practice, the water industry measures NRW using several indicators. Although the most common indicator remains the percentage of the system input volume (SIV), because of possible issues associated with its use the IWA instead recommends alternative indicators such as the NRW per connection and NRW per network length (Alegre *et al.* 2006). Unfortunately, these three indicators are not necessarily very strongly correlated and a utility showing excellent performance with one indicator can have a much lower performance when measured with a different one (van den Berg 2014).

The currently available literature on NRW focuses primarily on either the methodologies that can be applied to reduce NRW via practical applications (e.g., Farley & Trow 2003; Mutikanga *et al.* 2012; AWWA 2016a, 2016b; Gonelas & Kanakoudis 2016) or on the determination of economic levels of leakage (Deidda *et al.* 2014). Among the few studies that have examined the fundamental factors affecting NRW, González-Gómez *et al.* (2011) broached the question of why NRW is so high in many cities around the world from a sociological point of view, concluding that the lack of incentives for management units, the defense of private interests due to corruption, the lack of awareness of citizen-users of the water service, and the lack of political willingness were the main causes. For 133 municipalities in

Spain, González-Gómez *et al.* (2012) performed a regression analysis relating NRW to various different explanatory variables. Population growth, percentage of population outside the main village, population density, minimum water percentage of the storage reservoir, type of distribution system (gravity fed or pumped), tariff structure, indebtedness of the utility, and different forms of management (private sector or outsourcing of management) were all found to be important factors affecting NRW.

In a similar vein, van den Berg (2014) conducted a study covering the period between 2006 and 2011 for utilities in 69 countries via a regression analysis. One of the main findings of this study was that the key drivers of water losses were largely outside the control of the utilities. Connection density, population served per connection, size of the network, opportunity cost of water losses, staff productivity, and metering were some of the factors found to be having an impact on NRW levels. The same researcher then went on to apply the same methodology to a larger data set with similar results, confirming that the key drivers of water losses were at least partly linked to the physical characteristics of the water supply system (van den Berg 2015).

The first objective of this study was therefore to examine the relationship between three types of NRW indicator, namely NRW as percentage of SIV, NRW per connection, and NRW per network length, using the data set for a water utility in California. According to the literature survey summarized above, it is clear that there remains a substantial knowledge gap regarding the key factors affecting NRW. Thus, the second objective of this study was to identify the determinants of NRW for a California water utility. For most water utilities, collecting accurate data on water production and use has long been an issue because of metering/accounting inaccuracies, so uncertainty in the NRW data must also be integrated into the panel regression in the context of nonparametric uncertainty distributions. To the best of the authors' knowledge, this study represents the first application of an uncertainty embedded panel regression model for NRW determinant analysis in California. The results will help guide future efforts to reduce NRW and set realistic benchmarks not only for the utility under investigation but also for the water industry as a whole.

THE STUDY AREA

As the largest subsidiary of the California Water Service Group, the California Water Service Company (Cal Water) provides high quality regulated and non-regulated utility services to approximately 1.7 million people located in 83 communities across California. Cal Water has 24 districts, extending from Chico in the north to Palos Verdes in the south, and supplies water to roughly 478,000 connections. The company's main service elements are the production, purchase, storage, treatment, testing, distribution, and sale of water for domestic, industrial, public, irrigation, and fire protection purposes. The key infrastructure components supporting this service chain include about 9,200 kilometers of water main, 134,400 line and control valves, 970 booster stations, 650 wells, 7 surface water treatment plants, 420 water storage facilities, and 450 supervisory control and data acquisition (SCADA) transmitting units (Keck & Lee 2015).

METHODOLOGY

Data

The data considered in this study were obtained from company-specific files and databases covering the period from 1998 to 2014 for all the individual metered districts of Cal Water. During the study period, some connections were still unmetered in several Cal Water districts, which makes acquiring accurate water use data for these districts challenging and they were therefore eliminated from the study. NRW data may also be subject to uncertainty because of possible errors in measuring water production and metered water sales data. However, as Cal Water has been conducting a long-term sales meter repair/replacement program, it is expected that the sales data for Cal Water's metered districts will be fairly accurate. Districts depending primarily on purchased water as a water source are also expected to have reasonably precise production data since wholesale water agencies tend to have accurate water metering. However, the production data for districts utilizing groundwater supply are of unknown accuracy. Since water losses in California are generally low, typically less than 10%, it was decided to use data known to be the most accurate to ensure that the analytical results would not be interpreted incorrectly due to questionable data quality. As a result, only the data for five districts where more than 80% of the distributed water was purchased were considered in this study.

The data used for this study are considered to be panel data as they contain time series observations of a number of individual water districts (Hsiao 2003). This means that the NRW data structure involves two dimensions: a cross-sectional dimension and a time series dimension (Tanverakul & Lee 2015, 2016). Since some of the districts do not have data for some years, the study data set consists of a total of 76 year-district combinations.

Explanatory variables

The network length (NETLEN as km of network) is a known factor affecting leakage. Previous studies have indicated either a significant positive relationship (Alkasseh *et al.* 2013; van den Berg 2014; Hussein *et al.* 2017) or an insignificant relationship (González-Gómez *et al.* 2012) between network length and NRW. A positive relationship would likely be because a larger network is more costly to maintain to preserve functionality (van den Berg 2014).

Connection density (CON_DENS), expressed as number of connections per km of network, is also considered in the analysis, but this variable is beyond the utility's control and the expected sign of the relationship is ambiguous. Although a negative relationship might be expected as less water is lost in high-density areas because of the lower network maintenance cost per connection (González-Gómez *et al.* 2012), higher pressure must be maintained in higher-density networks to provide water to the upper floors in apartment buildings, thus creating a positive relationship between connection density and NRW as high pressure is one of the main causes of pipe breaks and system deterioration (González-Gómez *et al.* 2012; van den Berg 2015).

The number of pipe failures per year (LEAK) has been added to the model as a proxy for the quality of a network's physical integrity. A positive correlation between LEAK and NRW is expected since fewer pipe failures indicate a higher quality of maintenance and network integrity and hence a lower level of NRW. LEAK is also considered an indicator for the level and variation of pressure in the network since there is some evidence of a positive relationship between burst

frequency and pressure in the literature (Lee *et al.* 2012; Lambert *et al.* 2013) and pressure is thought to be the most important factor adversely affecting NRW (Thornton *et al.* 2008; Wu 2011).

In addition to the physical characteristics of the system, utility management practices and financial performance may have an effect on NRW (Güngör-Demirci *et al.* 2018a, 2018b). A measure of financial management of the utility, defined as the difference between operating revenue and operation and maintenance cost per cubic meter of water sold (NET_OPREV), is therefore also included in the model. An inverse relationship between this variable and NRW is expected since access to greater financial resources increases a utility's ability to deal with NRW (van den Berg 2014, 2015). Table 1 presents the descriptive statistics of the four explanatory variables, as well as the dependent variable NRW_CON_DAY.

Regression model specification

For the purposes of this study, a fixed effects panel regression model was applied to the data set. The fixed effects model attempts to control for unobservable factors among each district by assigning unique time-invariant identifiers. A simplified form of a fixed effects model is as follows (Hsiao 2003):

$$y_{it} = \alpha_i + \beta x_{it} + \mu_{it} \tag{1}$$

where y_{it} is the dependent variable observed for district i ($i = 1....N$) at time t ($t = 1...T$); α_i is the unknown intercept for each district (i.e., N entity-specific intercepts); x_{it} is the independent variable; β is the coefficient for that independent variable; and μ_{it}, is the error term. In this case, the intercept value, α_i, depends on omitted factors specific to

each district i that are possibly correlated with the chosen independent variables, x_{it}. Any time-invariant variables that may have an effect on NRW are thus absorbed into the intercept term. The error term, μ_{it}, represents effects from unique district factors that were not accounted for or are uncorrelated with identified independent variables. In this study, district heterogeneity is assumed to have an influence on water loss, hence a fixed effects model is adopted. In addition, Hausman test was performed to justify the adoption of fixed effects model over random effects model (Tanverakul & Lee 2015). The statistical program R was used to perform the analysis (Croissant *et al.* 2016).

In the model NRW, expressed as m^3 of water lost/connection/day (NRW_CON_DAY), is selected as the dependent variable and four explanatory (i.e., independent) variables, identified based on data availability, are included on the right-hand side of the panel regression equation (Equation (2)). The calculated coefficients measure the elasticity of NRW, revealing how much NRW varies in response to a change in the various drivers. The final regression equation utilized is as follows:

$$
\begin{aligned}
(\text{NRW\_CON\_DAY})_{it} = {} & \alpha_i + \beta_1(\text{NETLEN}) \\
& + \beta_2(\text{CON\_DENS}) + \beta_3(\text{LEAK}) \\
& + \beta_4(\text{NET\_OPREV}) + \mu_{it} \tag{2}
\end{aligned}
$$

Uncertainty in NRW indicator

As mentioned earlier, NRW data may suffer from uncertainties due to possible errors in water production and/or sales measurement. Since the uncertainty in production numbers directly affects the NRW indicator, it was deemed necessary to consider involved uncertainty in the regression analysis. Although only five districts with accurate production data

Table 1 | Descriptive statistics of dependent and independent variables, $n = 76$[a]

Variable	Description	Unit	Mean	St. dev.	Min	Max
NRW_CON_DAY	m^3 of NRW per connection per day	m^3/connection/ day	0.13	0.07	0.01	0.34
NETLEN	Network length	km	414.95	155.23	165.42	629.65
CON_DENS	Connection density	1/km	50.24	14.60	36.28	77.26
LEAK	Number of reported pipe failures	–	19	18	0	65
NET_OPREV	Difference between total operating revenue and total operation and maintenance costs per m^3 of water sold	$/m^3$	0.11	0.07	−0.02	0.35

[a]The values in this table represent the statistics used in the original data set ($n = 76$).

were selected for inclusion in this analysis, it was considered possible that even these data may be subject to error to a certain extent. Based on discussions with Cal Water engineers, a ±5% uncertainty was assumed to be present.

In the proposed fixed effects panel regression model (Equation (2)), the objective is to find values for all the coefficients β_1 through β_4 (i.e., the deterministic model). As we are considering the intrinsic uncertainty of NRW_CON_DAY, Equation (2) becomes the stochastic model in which all coefficients (β_1 through β_4) are to be estimated. In this study, nonparametric uncertainty distributions for NRW_CON_DAY are utilized in order to account for the inherent metering uncertainty. This is accomplished by synthetically generating random numbers for each of the 76 year-district combinations by assuming that the error bounds follow a normal distribution (Loganathan & Lee 2005). Hence, the total number of NRW_CON_DAY (dependent variable) combinations considered in Equation (2) is 7,600. The mean value of the randomly generated hundred NRW volume data is equal to the annual NRW volume data obtained from Cal Water files for each district; the standard deviation of these hundred data points is 5%. Figure 1 shows a graphical representation of the NRW volume data distribution for two sample districts.

RESULTS AND DISCUSSION

Cal Water's water balance using IWA standard

The water loss volume in a network can be estimated by conducting the IWA Water Balance Standard. Figure 2 shows the standard international approach developed by the IWA Task Forces on Water Losses and Performance Indicators to calculate the water balance of a water distribution network (Farley & Trow 2003; Alegre *et al.* 2006; Kanakoudis & Tsitsifli 2010). The water balance calculation is done through actual measurements on the network or simply by estimating the water volumes produced, consumed, and lost (Kanakoudis & Tsitsifli 2010). For the current study, there are no available measurements performed on the system level. Therefore, an estimation based upon literature values was done.

To safely calculate the water balance, the quality/integrity of the utilized data should be as high as possible. The reliability and the accuracy of the data are among the most important issues in water balance calculations in literature since water utilities do not always keep the necessary data records (Kolbl *et al.* 2007; Kanakoudis & Tsitsifli 2010). The only available data for Cal Water districts are billed authorized consumption and SIV. There are no data regarding unbilled authorized consumption, apparent losses and real losses and their individual components (Figure 2). Therefore, these water balance elements can only be assumed based on literature values. Unbilled authorized consumption is assumed to be 0.5% of SIV based on Charalambous & Hamilton (2011). Apparent losses are assumed to be 1% based on the worldwide values given in Kanakoudis *et al.* (2013) and McKenzie & Lambert (2008). Real losses are calculated by subtracting unbilled authorized consumption and apparent losses from NRW. Figure 3 presents the WB estimation for five Cal Water districts for the year 2014. The figure shows that the majority of the water losses in the four districts are real losses. However, it should be noted that the water balance estimation can be subject to error since it is based on several assumptions, as explained above.
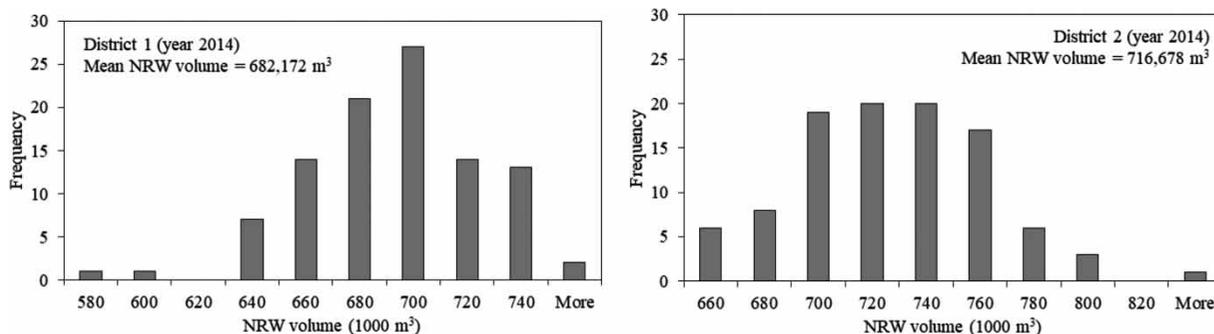

Figure 1 | Two sample histograms of the randomly generated NRW volume data (*n* = 100).

Figure 2 | IWA Standard International Water Balance (adopted from Farley & Trow 2003).
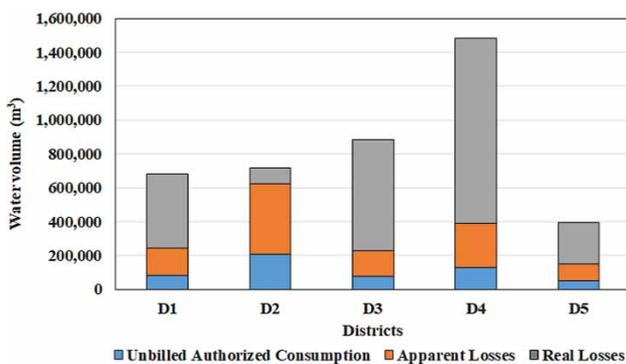

Figure 3 | WB estimation for five Cal Water districts for the year 2014.

NRW performance indicators

Different NRW indicators are not necessarily very strongly correlated (van den Berg 2014). A utility showing an excellent performance in one NRW indicator can appear to have a much lower performance with a different NRW indicator. This is especially true when percentage of the SIV is compared with the other NRW indicators as percentage does not provide any information about lost water volume, which is the most important parameter in water loss assessments. A water utility which performs better than another in percentage of SIV may actually be losing the same or more water when the actual NRW volume is taken into account with the other NRW indicators (AWWA 2012). As a single NRW indicator may not show the whole picture and can sometimes even be misleading, it is important to

calculate more than one NRW indicator for any system being assessed, as van den Berg (2014) noted. In this study, three different NRW indicators, namely, percentage of SIV, m^3 of water lost/km of network/day, and m^3 of water lost/connection/day, were calculated for the panel data set and the correlational relationships among these three indicators were analyzed using a linear regression model. Figure 4 shows the results of the linear regression analysis including all five districts' data between 1998 and 2014.

According to Figure 4(a), for the Cal Water utilities included in this study, NRW as a percentage of SIV is correlated with NRW as m^3 of water lost/km of network/day with an R^2 of 0.81. However, Figure 4(b) and 4(c) show that the correlations between the NRW indicators % of SIV and m^3 of water lost/connection/day, and between m^3 of water lost/km of network/day and m^3 of water lost/connection/day are both weaker, with R^2 values of 0.58 and 0.72, respectively.

The linear regression analysis was repeated for each district between 1998 and 2014. The results are presented in Table 2. It is seen that most of the correlation results for each district are above 90%. The results of the correlational analysis conducted on the five-district data set altogether and on single-district data sets demonstrate that depending on the available data/location of districts, different NRW indicators can either be very well or weakly correlated. For a water utility, it is therefore wise to calculate different NRW indicators simultaneously and then evaluate and compare the results.
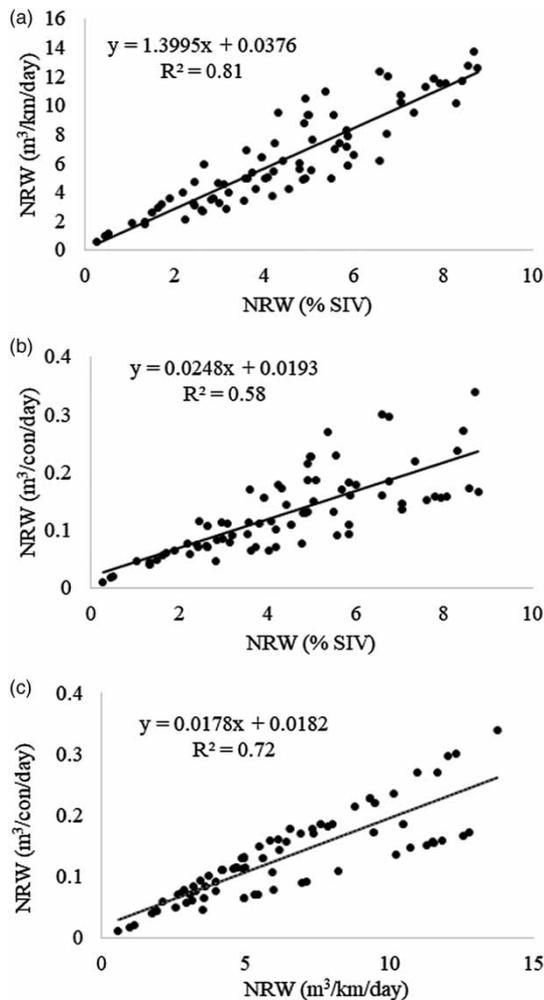
Figure 4 | Results of the linear regression between (a) NRW as % of SIV and NRW as m³ of water lost/km of network/day, (b) NRW as % of SIV and NRW as m³ of water lost/connection/day, (c) NRW as m³ of water lost/km of network/day and NRW as m³ of water lost/connection/day.

Fixed effects model

The results for a fixed effects panel regression model applied to the panel data set with embedded uncertainty and with

NRW measured by NRW_CON_DAY are shown in Table 3. All variables are statistically significant and the overall fit of the model (i.e., R^2 value) is 0.40.

Having a negative correlation between NRW_CON_DAY and NET_LEN is an unexpected result based on the literature findings. Normally, the opposite is deemed more likely as the larger networks will have more probability to leak. However, for our data set, there is a significant linear relationship between NET_LEN and number of connections. Although having longer network with more connections physically means more possibility of having leakage, it also shows the financial capability of a utility to deal with NRW as more connections generally mean more revenue. In our case, NET_LEN is a measure of the size of the districts as it is correlated with the NO_CON. When the size of districts gets bigger, the financial capability to deal with NRW increases due to economies of scale.

The negative correlation between NRW_CON_DAY and CON_DENS is explained as the loss of less water in more densely connected areas because of the lower network maintenance cost per connection (González-Gómez *et al.* 2012). The positive sign of LEAK indicates that the number of pipe failures each year is correlated with NRW per connection per day as fewer pipe failures indicate a higher quality of maintenance and network integrity and hence a lower level of NRW. The last variable, NET OPREV, has an inverse relationship with NRW_CON_DAY. A higher net operating revenue per cubic meter of water sold means that the district has a greater financial capacity to manage water loss problems as they arise. Therefore, this negative relationship is corresponding to the expected engineering judgments.

The uncertainty embedded panel regression model used in this study has provided the advantage of overcoming data uncertainty while applying fixed effects model that considers the influence of district heterogeneity. The resulting

Table 2 | Results of the correlational analysis for each district between 1998 and 2014

	% SIV vs. m³/km/day	% SIV vs. m³/con/day	m³/km/day vs. m³/con/day
District 1	y = 0.9764x + 0.048; R² = 0.941	y = 0.026x + 0.0021; R² = 0.930	y = 0.0268x + 0.0003; R² = 0.997
District 2	y = 2.1845x − 0.326; R² = 0.988	y = 0.0391x − 0.0039; R² = 0.992	y = 0.0178x + 0.0021; R² = 0.999
District 3	y = 1.6188x − 1.3274; R² = 0.979	y = 0.0222x − 0.0215; R² = 0.972	y = 0.0138x − 0.0037; R² = 0.999
District 4	y = 1.2766x + 0.0334; R² = 0.982	y = 0.0296x + 0.0006; R² = 0.982	y = 0.0232x − 0.0002; R² = 1
District 5	y = 1.6987x + 0.2155; R² = 0.938	y = 0.0419x + 0.0043; R² = 0.941	y = 0.0246x − 0.0007; R² = 0.999

Table 3 | Results of the panel data fixed effects regression ($n = 7,600$)

Variable	Coefficient	*p*-value
NETLEN	−0.0011	0.0000
CON_DENS	−0.0176	0.0000
LEAK	0.0020	0.0000
NET_OPREV	−0.1872	0.0000
R^2	0.40	

Ln(NRW_KM_DAY) was used as the dependent variable.

equation is obviously applicable to the specific data set belonging to Cal Water's five districts and can be used to predict their approximate NRW level under different conditions. Other water utilities in California, or any other utility in the world, may get a different correlation result due to their unique physical/mechanical characteristics of water distribution systems as well as socioeconomic issues/status, utility's private–public structure, political situations, etc. However, we believe that the current approaches/methods can be applied to any water utilities to further understand the fundamental nature/determinants of NRW.

SUMMARY AND CONCLUDING REMARKS

This study considered the fundamental determinants of NRW in the context of uncertainty embedded fixed effects by applying a panel regression model to a water utility in California. The correlational analysis between three NRW indicators (i.e., NRW as % of SIV, as m^3 of water lost/km of network/day, and as m^3 of water lost/connection/day) showed that depending on the available data, different NRW indicators can either be well or weakly correlated. Consequently, for a water utility, an analysis based on the calculation of more than one NRW indicator provides more useful results. The results of the panel regression revealed that all variables are statistically significant in terms of explaining variations in NRW for the case study considered here. Network length, connection density, and net operating revenue per cubic meter of water sold were found to be negatively correlated with NRW. A positive relationship between number of leaks and NRW was identified for the five districts in California included in the case study.

At present, Cal Water's maturity level, approach, and management of NRW is evolving from an ad hoc state to one that is more optimized. The utility is committed to managing water as a precious resource that is vital to the communities and customers they serve. Cal Water is also enhancing the business processes that support their Water Audit and Loss Control Program. These improvements will provide a more accurate picture of the components that make up NRW, including unbilled authorized consumption, apparent losses, and real losses. With the ongoing research efforts, the utility will be able to optimize their efforts to efficiently and effectively reduce NRW. Cal Water continues to test and calibrate production meters and routinely replace customer meters. Water loss stemming from water system leakages has been controlled through the utility's Main Replacement Program and by implementing timely main break repairs. NRW will continue to be closely monitored in the future and improvements to the utility's infrastructure renewal program, business processes, maintenance practices, and technology will be considered and applied to optimize the overall management of water supply and delivery.

ACKNOWLEDGEMENTS

REFERENCES

Alegre, H., Baptista, J. M., Cabrera Jr, E., Cubillo, F., Duarte, P., Hirner, W., Merkel, W. & Parena, R. 2006 *Performance Indicators for Water Supply Services*, 2nd edn. IWA Publishing, London.

Alkasseh, J. M. A., Adlan, M. N., Abustan, I., Aziz, H. A. & Hanif, A. B. M. 2013 Applying minimum night flow to estimate water loss using statistical modeling: a case study in Kinta Valley, Malaysia. *Water Resources Management* **27** (5), 1439–1455.

AWWA 2012 *Water Loss Control: Water Loss Control Terms Defined*. American Water Works Association. Available at: http://www.awwa.org/portals/0/files/resources/water20knowledge/water20loss%20control/water-loss-control-terms-defined-awwa-updated.pdf.

AWWA 2016a *The State of Water Loss Control in Drinking Water Utilities*. A white paper from the American Water Works Association. Available at: https://www.awwa.org/Portals/0/files/resources/water20knowledge/water20loss20control/State20of20Water20Loss20Control202016.pdf.

AWWA 2016b *Water Audits and Loss Control Programs: M36.* American Water Works Association, Denver, CO.

Charalambous, B. & Hamilton, S. 2011 Water balance – the next stage. *Water Utility Journal* **1**, 3–10.

Croissant, Y., Millo, G., Tappe, K., Toomet, O., Kleiber, C., Zeileis, A., Henningsen, A., Andronic, L. & Schoenfelder, N. 2016 *Linear Models for Panel Data.* Available at: https://cran.r-project.org/web/packages/plm/vignettes/plm.pdf.

Deidda, D., Sechi, G. M. & Zucca, R. 2014 Finding economic optimality in leakage reduction: a cost-simulation approach for complex urban supply systems. *Procedia Engineering* **70**, 477–486.

Farley, M. & Trow, S. 2003 *Losses in Water Distribution Networks: A Practitioner's Guide to Assessment, Monitoring and Control.* IWA Publishing, London.

Gonelas, K. & Kanakoudis, V. 2016 Reaching economic leakage level through pressure management. *Water Science and Technology: Water Supply* **16** (3), 756–765.

González-Gómez, F., García-Rubio, M. A. & Guardiola, J. 2011 Why is non-revenue water so high in so many cities? *International Journal of Water Resources Development* **27**, 345–360.

González-Gómez, F., Martínez-Espiñeira, R., García-Valiñas, M. A. & García-Rubio, M. A. 2012 Explanatory factors of urban water leakage rates in southern Spain. *Utilities Policy* **22**, 22–30.

Güngör-Demirci, G., Lee, J. & Keck, J. 2018a Assessing the performance of a California water utility using two-stage data envelopment analysis. *Journal of Water Resources Planning and Management* **144** (4), 05018004.

Güngör-Demirci, G., Lee, J. & Keck, J. 2018b Measuring water utility performance using nonparametric linear programming. *Civil Engineering and Environmental Systems* 1–15. Available at: https://www.tandfonline.com/doi/full/10.1080/10286608.2018.1425403.

Hsiao, C. 2003 *Analysis of Panel Data.* Cambridge University Press, New York.

Hussein, M., Yoneda, K., Othman, N. A., Zaki, Z. M. & Yusof, M. H. M. 2017 Effects of number of connections and pipe length to the water losses in Melaka. *Journal Teknologi* **79** (3), 45–59.

Kanakoudis, V. & Tsitsifli, S. 2010 Results of an urban water distribution network performance evaluation attempt in Greece. *Urban Water Journal* **7** (5), 267–285.

Kanakoudis, V., Tsitsifli, S., Samaras, P. & Zouboulis, A. 2013 Assessing the performance of urban water networks across the EU Mediterranean area: the paradox of high NRW levels and absence of respective reduction measures. *Water Science and Technology: Water Supply* **13** (4), 939–950.

Keck, J. & Lee, J. 2015 A new model for industry – university partnerships. *Journal of the American Water Works Association* **107** (11), 84–90.

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 Sector Board Discussion Paper Series, Paper No. 8, The World Bank, Washington, DC.

Kolbl, J., Theuretzbacher-Fritz, H., Neunteufel, R., Perfler, R., Gangl, G., Kainz, H. & Haberl, R. 2007 Experiences with Water Loss PIs in the Austrian Benchmarking Project. In: *Water Loss 2007, Conference Proceedings*, Vol. 1, pp. 176–187.

Lambert, A., Fantozzi, M. & Thornton, J. 2013 Practical approaches to modeling leakage and pressure management in distribution systems: progress since 2005. In: *Proceedings of the 12th International Conference on Computing and Control for the Water Industry-CCWI2013*, Perugia, Italy.

Lee, J., Lohani, V., Dietrich, A. & Loganathan, G. V. 2012 Hydraulic transients in plumbing systems. *Water Science and Technology: Water Supply* **12** (5), 619–629.

Loganathan, G. V. & Lee, J. 2005 Decision tool for optimal replacement of plumbing systems. *Journal of Civil and Environmental Engineering Systems* **22** (3), 1–16.

McKenzie, R. & Lambert, A. 2008 *Benchmarking of Water Losses in New Zealand Manual.* Water and Wastes Association, New Zealand.

Mutikanga, H. E., Sharma, S. K. & Vairavamoorthy, K. 2012 Methods and tools for managing losses in water distribution systems. *Journal of Water Resources Planning and Management* **139** (2), 166–174.

OECD 2016 *Factors Shaping Urban Water Governance. Water Governance in Cities*, OECD Publishing Paris.

Rajani, B. & Kleiner, Y. 2001 Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water Journal* **3**, 161–164.

State of California 2017 *Department of Finance, E-1 Population Estimates for Cities, Counties and the State with Annual Percent Change – January 1, 2016 and 2017.* Sacramento, CA, May.

Tanverakul, S. & Lee, J. 2015 Impacts of metering on residential water use in California. *Journal of the American Water Works Association* **107** (2), 78–79.

Tanverakul, S. & Lee, J. 2016 Decadal review of residential water demand analysis from a practical perspective. *Water Practice and Technology* **11** (2), 433–447. doi: 10.2166/wpt.2016.050.

Thornton, J., Sturm, R. & Kunkel, G. 2008 *Water Loss Control.* McGraw-Hill, New York.

van den Berg, C. 2014 *The Drivers of non-Revenue Water: how Effective are non-Revenue Water Reduction Programs?* Policy Research Working Paper 6997, World Bank Group, Washington, DC.

van den Berg, C. 2015 Drivers of non-revenue water: a cross-national analysis. *Utilities Policy* **36**, 71–78.

Wu, Z. 2011 *Water Loss Reduction.* Bentley Institute Press, Exton, PA.