IoT-based water demand forecasting and distribution design for smart city
Lakshmi Kanthan Narayanan and Suresh Sankaranarayanan

ABSTRACT
The percentage of fresh water resource availability in the world is diminishing every year. According to a world economic forum survey, the increase in water demand will result in high scarcity globally in the next two decades. The eradication of the water demand increase and reducing the losses during the transportation of water is challenging. Thus accordingly, an IoT-based architecture integrated with Fog for underground water distribution system has been proposed. Towards designing an IoT water distribution architecture for a smart city, we need to first forecast the water demand for consumers. Hence, accordingly, water demand forecasting has been carried out on a daily basis for a period of three months as a case study using ARIMA and regression analysis. Based on water demand forecasting analysis, a water distribution design for an IoT-based architecture has been carried out using hydraulic engineering design for proper distribution of water with minimal losses which would result in the development of a smart water distribution system (SWDS). This has been carried out using EPANET.

Key words | ARIMA, EPANET, Fog, IoT, SWDS

INTRODUCTION
The International Water Association (IWA) states that water loss management has achieved increased attention. The recommendations of IWA have proposed new methods for modelling leakage detection and loss management components (Gupta et al. 2017).

The common problem that occurs during the transportation of water from the source via underground pipes to consumers is transportation loss. The losses are mainly due to the fittings of the pipe network, leakages, breaks and cracks in the pipe, overflow in the main tanks/sub-tanks, pressure loss and obstruction due to sediments or blocks in pipes.

The construction of SWDS architecture for a smart city is done with storage reservoirs, booster pumping stations, fire hydrants and consumer service lines and redundancy of the network is provided via smart water grids and loops (Bibri & Krogstie 2017). In the past, the water distribution models used fuzzy-based decisions in the estimation of flow rate (Zischg et al. 2018).

The impact of Cloud computing has played a vital role in bringing the concept of IoT into reality. But at the same time, Cloud computing cannot be integrated into all the IoT-based systems. The data acquired from the sensors need to be processed in real time for providing quick control action for the industrial IoT devices. Fog computing has achieved many more advantages over Cloud computing, such as low latency, less computational delay and less bandwidth operation (Bonomi et al. 2014; Veeramanikandan & Sankaranarayanan 2019).

Thus, accordingly, an IoT-based underground water distribution architecture integrated with Fog computing and Cloud has been proposed (Narayanan & Sankaranarayanan 2019) where the real-time processing happens at the Edge, also called Fog computing, based on the sensor data
captured from underground pipes for control action at the substation. The Cloud is responsible for storing all historical information and other pertinent information for Big Data analyses.

Towards designing such an IoT-based underground water distribution architecture for a smart city with minimal transportation losses, we first need to understand and study consumer behaviour towards water consumption based on the historical data available. Thus, the water demand prediction was made for daily consumption over a three-month period as a case study using ARIMA and linear regression with comparative analysis. Based on the demand prediction, an effective and efficient water distribution system has been designed on the basis of day-wise water demand prediction, which will resolve issues related to water distribution. Figure 1 shows a sample water distribution system (www.epa.gov).

The main contributions of the paper are as follows.

1. Statistical analysis of water consumption data and demand forecasting using ARIMA and regression.
2. Comparative analysis on statistical model towards demand forecasting.
3. Designing of water distribution in IoT-based water distribution architecture using EPANET-based on the demand forecasting.

The remaining section of the paper is organized as follows. The next section presents a complete literature survey on various technologies adapted in water distribution and demand forecasting methods. This is followed by a section discussing in detail the construction of IoT-based WDS design with the integration of Fog and Cloud computing along with demand forecasting-based WDN construction. The statistical methods adopted for the demand forecasting are presented in the next section. Then a section deals with the comparative analysis between ARIMA and linear regression and water distribution design based on the forecast using EPANET. The final section presents the conclusions and future work.

**LITERATURE REVIEW**

Much research has been carried out on designing an IoT-based WDS using sensors in order to monitor the supply
and quality of water. This WDS will display real-time water consumed by customers (Perera et al. 2017).

For a block of apartments, the WDS is built with a series of interconnected sensors, which are deployed in the pipe network in order to measure the flow and consumption in the system. The consumed data are sent to the Cloud which will perform the task of intimating the customer regarding their consumption (Amatulla et al. 2017).

Research has been carried out in constructing an effective distribution system for rural distribution networks with a combination of ultrasonic and conductivity sensors for assessing the water level and quality of water in the distribution tanks, respectively. In this system, the measured data are sent to a mobile-based application for further assessment (Chanda et al. 2017).

An IoT-based distribution prototype was designed by Amatulla (Amatulla et al. 2017) for smart cities. This system is purely a microprocessor-based monitoring system where the periodically monitored data are transmitted through Wi-Fi to the web-based system which is remotely located (Varma et al. 2015). Ultrasonic sensors are also used in the water management system (WMS) for identifying the level of water in the tanks (Candelieri & Archetti 2014).

Similarly, in the WDS, the sensors are also deployed in the external pipes that connects the apartment or community in order to monitor the amount of water distributed to each individual household. These data are transmitted through Wi-Fi to the Cloud platform for further analysis and graphical report generation (Laspidou et al. 2015).

The above-discussed water distribution systems are only IoT-based monitoring systems in which no intelligence and control automation is involved.

The following paragraphs will discuss in detail research work that has been carried out by employing machine learning (ML) and deep learning (DL).

The usage of time series analysis plays a vital role in forecasting and prediction. Data clustering is executed by using time series analysis. Support vector machine regression is implemented over hourly water consumption data observed from test-bed setup. This characterization pattern method is validated over the urban demand of the water distribution network (WDN) of the city of Milan (Gwaivangmin & Jiya 2017).

Feature extraction of water demand pattern is done by the construction of Kohonen self-organizing maps. This analysis is purely dependent upon consumers’ bills. This characterization enables autonomous classification on the basis of consumption of water by the consumers. This consumption pattern analysis was carried out with the consumers of the Greek island of Skiathos (Herrera et al. 2014).

An artificial neural network is deployed over a distribution and control design based on the supervisory control. This system is designed on the basis of a case study done in Nigeria. The ultimate aim of this system is to provide an optimal solution towards WDN in order to solve the issues that arise due to water scarcity (Chen & Boccelli 2014).

Water consumption and demand forecasting of a residential tower in Korea was used as a case study analysis. The average consumption over the period of 2012 to 2014 was carried out and used as a training input for the prediction model. This resulted in the construction of a decision-making model for water distribution (Rinaudo et al. 2015).

Environmental variables such as humidity, rainfall, temperature, and atmospheric pressure were left out by implementing multiple kernel regression algorithms for daily water demand prediction in order to improve the accuracy and computational efficiency of the system which resulted in the design of a decision-making system for a smart city (Herrera et al. 2014).

A computational frame for the analysis of water demand in urban areas is implemented by using a machine learning approach called regression analysis. It is implemented in two stages. The first stage is clustering of consumption data on the basis of consumption pattern or behaviour.

A time series forecasting framework was implemented for statistical forecasting of short-term water demands using fixed seasonal autoregressive model and adaptive seasonal autoregressive model for predicting real demand of consumers in a real-world scenario using prototype models (Candelieri & Archetti 2014).

Multivariate statistical modelling and microcomponent modelling are the two adaptive techniques that were utilized in long-term forecasting of water demand prediction (Laspidou et al. 2015).

From the above examples, there were some developments and research pertaining to an IoT-based monitoring system for water distribution without integration of Fog...
and intelligence. Also, those systems were not designed for an underground water distribution system for a smart city.

In addition, regarding water demand forecasting, long-term predictions were done based on water consumption data collected from various sources like data source organizations, water meters or from test-bed setup.

In none of these systems discussed above, had water demand forecasting been carried out on a daily basis for a short term based on available daily water consumption historic data. In addition, water demand forecasting was not integrated with water distribution system design for an efficient and optimal distribution of water to the consumers in a smart city concept without wastage of water.

Accordingly, we will be discussing in the forthcoming sections, the proposal of an IoT-based water distribution architecture integrated with Fog and Cloud, followed by water demand forecasting and water distribution design in detail.

**IOT-BASED WATER DISTRIBUTION ARCHITECTURE**

In the real-world scenario, there are many independent systems which are used for demand prediction, water distribution automation, monitoring the distribution, estimation of supply, quality evaluation, etc. However, all these systems are independently designed for specific applications.

In India, currently, there are no concrete systems in practice since they are more than 100 years old. There are only a few systems in practice in order to perform underground pipe monitoring such as in oil refineries, mines, etc.

Taking these points into consideration and with the advent of IoT and Fog computing, an IoT-based water distribution architecture integrated with Fog and Cloud is proposed (Narayanan & Sankaranarayanan 2019) with the ultimate aim of providing cost-effective, highly reliable architecture for water distribution and underground pipe health monitoring based on the predicted water demand and determined hydraulic parameters. There are various variable components, such as climate, season, rainfall, family type, which can have a huge impact on the decision-making system in the determination of consumer demand. The following are the essential goals that need to be satisfied for a water distribution system:

1. The water distribution system (WDS) should understand the consumer demands and hydraulic parameters which are dependent on geographical topology.
2. The WMS should be aware of the basics of supply requirements as well as demand. It should also investigate the purpose of the application.
3. The WMS should be thoroughly trained with the intake structure, source of supply and also WDN (WDN) structure.
4. The WDS should be fed with complete information about the material types and controlling mechanism used in the system.

**Figure 2** illustrates an IoT-based architecture for water distribution and underground pipe health monitoring system. **Figure 3** shows a sample community-based WDS.

**Water distribution system**

Based on the IoT-based architecture for a WDS, the design of the water distribution system is important and a key task in distribution of water to consumers. This design in the architecture is based on demand prediction which is based on historical data of water consumption and, accordingly, supply of water will be initiated by the local SCADA engineer from the substation.

Hence, predicting the water demand for consumers with regards to water distribution design, we need to look into methodologies pertaining to water demand prediction based on historical water consumption data which are discussed below.

**WATER DEMAND FORECASTING**

The water demand forecast is performed using statistical models of forecasting algorithms, namely, ARIMA and linear regression. For the prediction of water demand and analysis, a dataset of day-to-day water consumption of people residing in the city of Austin (Texas, USA) is taken into account.

The dataset contains the aggregated reading of water consumption on a daily basis. The dataset also consists of two maximum consumptions, namely, Peak-1 and Peak-2 consumption during peak hours of the day. The dataset
detailed account is provided by US government data (www.data.gov). The dataset contains eight years and three months of readings from 1 January 2010 to 31 March 2018. Figure 4 illustrates the consumption of water by the people of Austin over the period January 2010 to March 2018.

We will now discuss in detail the methodology of the two algorithms – ARIMA and linear regression – regarding prediction and analysis on historical water consumption data over a period of eight years.

**ARIMA for demand prediction**

An autoregressive integrated moving average (ARIMA) algorithm has the ultimate aim of performing statistical analysis of seasonal data regarding data arrangement
based on the time pattern globally known as time series analysis. This analysis is done to determine the trend of data, seasonality of occurrences, moving average and to evaluate the algorithmic performance on the basis of determining and analysing error.

Box and Jenkins performed a wide range of analysis using ARIMA. There are three phases that are involved in the Box–Jenkins analysis model (Tom et al. 2019).

**Phase-I: data identification**

The primary and initial step of the Box–Jenkins model is preparation of data. In this phase, the data are transformed in order to make the data series stationary and to stabilize the variance.

Model selection is the continuous process of this phase in which the partial autocorrelation function (PACF) and ACF are carried out in order to find the optimal model.

**Phase-II: estimation of parameters and testing**

In this phase, the parameters are estimated in all the possible models and the best-fit model is chosen on the basis of optimal criteria. This step is followed by diagnostics, where the residuals are diagnosed by portmanteau test and PACF/ACF is also verified. The residuals are verified whether it is a white noise or not. If the residuals are found to be white noise, then Phase III is carried out, else the procedures in Phase I are repeated.

**Phase-III: data forecasting**

In this phase, the data is being forecasted. In the procedure of time series analysis, the pattern is decomposed into simpler form or should be divided into sub-patterns. The data is assumed as a function of seasonality, error and trend cycle.

**Phase-IV: data decomposition modelling**

In time series analysis, data are always defined as a function of error, trend cycle and seasonality component which is expressed as:

\[ Z_t = f(S_t, E_t, T_t) \] (1)

In general, the additive form of decomposition is determined by computing the sum of all the components. The additive form is represented as:

\[ Z_t = S_t + E_t + T_t \] (2)

This decomposition is applicable only if the seasonal functions maintain a constant pattern, i.e., no rapid decrease.
or increase in functional component over a period of time. Here, the water consumption is a seasonal dependent function where the consumption rate varies with respect to seasons of a year. Hence, the decomposition is multiplicative in nature.

\[ Z_T = S_T \cdot E_t \cdot T_t \]  

(3)

This is further transformed into additive model by taking the log of the series in order to derive the additive relationship from the multiplicative relationship.

\[ \log Z_t = \log S_t + \log E_t + \log T_t \]  

(4)

Smoothing of data is done to obtain trend-cycle by reducing the random variations. The simple way of smoothing the time series is moving average (MA). The expression of MA is as follows:

\[ T_t = \frac{1}{K} \sum_{j=0}^{m} Z_{t+j}, \text{ where } m = (K + 1) \div 2 \]  

(5)

Here, MA is determined as first order of k, where k is an integer. It is suitable for calculating an odd number of observations. For computing MA for six-month observations, the following equations are derived:

Example of 2 × 6 MA:

\[ T_{2.5} = \frac{(Z_1 + Z_2 + Z_3 + Z_4 + Z_5 + Z_6)}{6} \]  

(6)

\[ T_{3.5} = \frac{(Z_2 + Z_3 + Z_4 + Z_5 + Z_6 + Z_7)}{6} \]  

(7)

On averaging these two 6 MA smoothers:

\[ T'_5 = \frac{(T_{2.5} + T_{3.5})}{2} \]  

(8)

\[ T'_3 = \frac{(Z + 2Z_2 + 2Z_3 + 2Z_4 + 2Z_5 + 2Z_6 + 2Z_7)}{12} \]  

(9)

Figure 5 illustrates the 7 MA, i.e., 7 day moving average applied over the water consumption dataset in order to analyze the trend for the next 7 days and how it is used in the process of decomposition. This is beneficial for water demand prediction on a daily basis over a three-month period.

Figure 6 describes the trend and seasonality pattern. A good prediction or data forecasting model will have residues or forecast errors as described in the following equation:

\[ Z_t = e_t + C \]  

(10)

**Statonarity testing**

Dickey–Fuller testing is widely used in practice to test the stationarity in time series data (Figure 7). It is highly essential to satisfy the unit root testing method. The primary step is estimation of the regression model.

\[ Z_t = \varnothing Z_{t-1} + b_1 Z_{t-1} + b_2 Z_{t-2} + \ldots + b_p Z_{t-p} \]  

(11)

where \( \varnothing \) value is −4.68, which means \( Z_t \) is stationary. Hence, \( Z'_t \) is denoted as differential series of \( Z_{t-1} \).
Figure 6 | The decomposition of time series: (a) trend, (b) seasonality, (c) error component.
Prediction/forecasting

In order to perform prediction, this model used the seasonal ARIMA: (SARIMA) as

\[
\text{SARIMA} = (p, d, q)(P, D, Q)_s
\]  

where \( s \) is no. of periods per season, \( p \) is order of auto-regression (AR), \( d \) is degree of data differencing (I) invoked and \( q \) is order of moving average (MA).

The general ARIMA \((1, 1, 1)(1, 1, 1)_{12}\) model is expressed as follows:

\[
(1 - \theta_1 B^{12})(1 - \Theta_1 B^1)(1 - B)(1 - B^{12})Z_t = (1 - \theta_1 B)(1 - \theta_1 B^{12})\epsilon_t
\]  

Here \( B \) is the lag factor, the values of \( \Theta_1, \theta_1 \) are parameters of autoregressive and moving average and used for forecasting. These steps are applied to the daily water consumption dataset for a period of eight years in order to predict their consumption pattern. This work aims to predict the water demand on a daily basis for the next three months. The demand forecast analysis and accuracy is discussed in the Results and discussion section.

Least square linear regression

Linear regression is a statistical model which is used to derive the relationship between two variables and also helps to determine the impact of one variable over the other. Here, the water consumption is one variable, which is impacted by the change in date/month of year. The regression model is expressed in terms of association between variables \( X \) and \( Y \). The algorithmic representation of this association is expressed in terms of a straight line as follows:

\[
Y = a + bX
\]  

Here, \( a \) is the intercept of \( Y \) and \( b \) is the slope of the line.

If the line is a fair line which intercepts the variables \( X \) and \( Y \), and also if the line obtained is a fairly flat line, then the slope will be small and also the rise of the line obtained will be smaller than the run. The expression to determine slope is:

\[
\text{slope of line} = \frac{\text{rise}}{\text{run}}
\]  

\[
slope = b = \frac{Y_2 - Y_1}{X_2 - X_1}
\]  

Forecasting through least square regression lines

In the water demand forecasting model, the \( X \) axis is assumed as an independent variable called date and the consumption variable is assumed as \( Y \). In this dataset, there is no implication to say that \( Y \) is the resultant of \( X \). The following expression is used to determine the regression line:

\[
Y = a + bX
\]
Here, $X$ is the observed value of independent variable called date and $\hat{Y}$ is the dependent variable of $Y$ called consumption.

The least square criterion is determined through the following algebraic equation. Here, consider there are ‘$n$’ observed points for both $X$ and $Y$ variables with a range from 1 to $n$. Then for the ‘$i$’ observations, the difference or deviation between the predicted and observed value can be formulated as:

$$e_i = Y_i - \hat{Y}_i$$  \hspace{1cm} (18)
In order to determine the best-fit points to the demand prediction curve, the positive and the negative errors get cancelled and the error $\sum e_i = 0$.

Figure 8 shows the decomposition of linear regression on water consumption dataset illustrating the trend, residual water usage, normalized residuals and plot of error versus residual.

Figure 9 shows the best fit regression line for water demand prediction on a daily basis for the period of three months which is January–March 2018. The result is 0.2698 which shows that it is a good fit. Best fit should lie between 0 and 1 (www.ncss.com).

**RESULTS AND DISCUSSION**

Based on ARIMA and least square linear regression method applied for demand forecasting on water consumption dataset, this section discusses demand forecast result analysis in terms of error rate accuracy. Based on the demand forecast, the WDS will be designed for IoT architecture using EPANET.

**Measurement of forecast accuracy**

The statistical estimation method of determining error is discussed previously in detail. The following algebraic
expression is used in this error calculation which plays an important role in determining the efficiency of the forecast model. Figure 10 and Table 1 show the statistical forecast using the LSLR model. Figure 11 and Table 2 show the statistical forecast using ARIMA model:

\[ e_t = O_t - F_t \]  

(19)

where \( O_t \) is the actual consumption and \( F_t \) is the forecasted demand.

Mean error

\[ ME = \frac{1}{n} \sum_{t=1}^{n} e_t \]  

(20)

Mean absolute error

\[ MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t| \]  

(21)

Table 1 | Forecast for three months using LSLR

<table>
<thead>
<tr>
<th>Date</th>
<th>2018 Consumed (y)</th>
<th>2018 Actual predicted (P)</th>
<th>(P-Y)/Y – Error</th>
<th>Mod (2018 error)</th>
<th>Percentage error</th>
<th>Actual error (Y-P)</th>
<th>2018 Sqr. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Jan</td>
<td>13,953.60</td>
<td>13,247.67</td>
<td>-0.05</td>
<td>0.95</td>
<td>94.94</td>
<td>705.93</td>
<td>498,343.99</td>
</tr>
<tr>
<td>2-Jan</td>
<td>13,861.20</td>
<td>13,204.14</td>
<td>-0.05</td>
<td>0.95</td>
<td>95.26</td>
<td>657.06</td>
<td>431,732.93</td>
</tr>
<tr>
<td>3-Jan</td>
<td>15,435.60</td>
<td>13,936.75</td>
<td>-0.10</td>
<td>0.90</td>
<td>90.29</td>
<td>1,498.85</td>
<td>2,246,555.86</td>
</tr>
<tr>
<td>29-Mar</td>
<td>14,768.40</td>
<td>13,583.30</td>
<td>-0.08</td>
<td>0.92</td>
<td>91.98</td>
<td>1,185.10</td>
<td>1,404,471.19</td>
</tr>
<tr>
<td>30-Mar</td>
<td>13,683.60</td>
<td>13,077.66</td>
<td>-0.04</td>
<td>0.96</td>
<td>95.57</td>
<td>605.94</td>
<td>367,165.20</td>
</tr>
<tr>
<td>31-Mar</td>
<td>13,449.60</td>
<td>12,968.19</td>
<td>-0.04</td>
<td>0.96</td>
<td>96.42</td>
<td>481.41</td>
<td>231,752.04</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>87.62%</td>
<td></td>
<td></td>
<td></td>
<td>9.47 \times 10^5</td>
</tr>
</tbody>
</table>

Figure 10 | Prediction vs consumption using LSLR model.
The percentile error PAE is calculated as follows, in which the relative measures are frequently used:

\[ PE = \left( \frac{Y_t - F_t}{Y_t} \right) \times 100 \]  \hspace{1cm} (22)

Mean absolute percentage error

\[ (MAPE) = \frac{1}{n} \sum_{t=1}^{n} |Pe_t| \]  \hspace{1cm} (23)

Mean absolute square error

\[ MASE = \left( \frac{1}{n} \sum_{t=1}^{n} e_t \right)^2 \]  \hspace{1cm} (24)

Root mean square error

\[ RMSE = \sqrt{MASE} \]  \hspace{1cm} (25)
The forecast calculation accuracy between the two statistical models ARIMA and LSLR is compared and displayed in Table 3. From the comparative analysis it is evident that error in terms of MASE and RMSE is higher for ARIMA compared to LSLR in terms of predicted and actual value towards demand prediction. However, in regard to prediction accuracy, the ARIMA model forecast accuracy is better than the LSLR forecast model where MAPE is higher for LSLR as compared to ARIMA, as shown in Table 3. That is, higher MAPE signifies that percentage error is greater for LSLR which results in low accuracy as compared to ARIMA where MAPE is less. From the statistical analysis it is also found that in spite of the demand prediction, consumption is at quite a high rate. It is not necessary that the consumption is entirely utilized by the consumers. The consumption reading is inclusive of losses due to pipe breakages, joint materials and gate valves, ageing of WDN, etc., natural causes and theft of water. In order to minimize the losses, a highly efficient WDN is required. The following section discusses in detail the design of the WDN and simulating the hydraulic behaviour as a pre-requisite for IoT-based WDS.

**Design of WDN using EPANET**

Based on the demand prediction made using the ARIMA model for the first season (January–March) of 2019, the WDN is constructed on the basis of the following assumptions (Table 4).

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Name of parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No. of distribution points</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>Material of pipe</td>
<td>PVC</td>
</tr>
<tr>
<td>3</td>
<td>No. of junctions</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>Diameter of pipe</td>
<td>100 mm</td>
</tr>
<tr>
<td>5</td>
<td>Length b/w each junction</td>
<td>500 m</td>
</tr>
<tr>
<td>6</td>
<td>Roughness coefficient</td>
<td>140</td>
</tr>
<tr>
<td>7</td>
<td>Calculation method</td>
<td>Hazen–Williams formula</td>
</tr>
</tbody>
</table>

The forecast calculation accuracy between the two statistical models ARIMA and LSLR is compared and displayed in Table 3. From the comparative analysis it is evident that error in terms of MASE and RMSE is higher for ARIMA compared to LSLR in terms of predicted and actual value towards demand prediction. However, in regard to prediction accuracy, the ARIMA model forecast accuracy is better than the LSLR forecast model where MAPE is higher for LSLR as compared to ARIMA, as shown in Table 3. That is, higher MAPE signifies that percentage error is greater for LSLR which results in low accuracy as compared to ARIMA where MAPE is less. From the statistical analysis it is also found that in spite of the demand prediction, consumption is at quite a high rate. It is not necessary that the consumption is entirely utilized by the consumers. The consumption reading is inclusive of losses due to pipe breakages, joint materials and gate valves, ageing of WDN, etc., natural causes and theft of water. In order to minimize the losses, a highly efficient WDN is required. The following section discusses in detail the design of the WDN and simulating the hydraulic behaviour as a pre-requisite for IoT-based WDS.

**Design of WDN using EPANET**

Based on the demand prediction made using the ARIMA model for the first season (January–March) of 2019, the WDN is constructed on the basis of the following assumptions (Table 4).
The formula for calculating head loss occurring in pipe is determined using the Hazen–Williams formula as follows:

\[ h_{100ft} = 0.002083 \left( \frac{100}{c} \right)^{1.852} L \cdot \frac{q^{1.852}}{d_n^{1.852}} \]  

(26)

where \( h_{100ft} \) is friction head loss inside pipe per 100 feet, \( L \) is length of pipe in feet, \( c \) is Hazen–Williams friction constant, \( q \) is volume of flow in gallons per minute, \( d_n \) is hydraulic diameter inside the pipe.

The date of 17 January 2019 predicted data is taken for pipe network simulation. The water distribution on that particular date is taken as the average quantity for 24 hours’ distribution and it is illustrated in Figure 12.

The WDN network is constructed as per the details furnished below. EPANET network design consists of 50 distribution mains supplied by 1 reservoir.

- The average demand and the maximum demand for January 2019 is computed as 12,625 million gallons (MG) as per the prediction for 17 January 2019. 1,286 MG is the maximum requirement in the three months’ predicted analysis (January 2019–March 2019).
- The maximum demand is for a postal zone and it is assumed as 50 supplier tanks connected through 25 junctions.
- The variable length of PVC for the upper networks is 500 and the bottom network is 1,000 m.
- The diameter of the pipe is 100 mm.
- The base demand of each node is distributed equally across the 25 junctions.
- The base water level in the supply reservoir is assumed as the quantity of water required as per the prediction made by analysis.

Table 5 illustrates how the water demand for 17 January 2019 is assumed to be supplied and simulated along with the peak and normal duration supply.

The construction of the WDN and the hourly flow of water in the WDN is illustrated in the following figures.

From Figures 13–15, it is understood that the simulation of WDN is done using EPANET. This simulation helps in determining the hydraulic parameters such as flow volume, pressure, head-loss, etc. at each and every junction and node of the WDN. It is also helpful in predetermining the hydraulic nature of WDN for the already existing water distribution system. It is helpful in the design of a new hydraulic distribution system as well as being useful in determining the performance of the existing or already constructed WDN. EPANET analysis, as shown in Figures 14 and 15, describes the flow of water in the WDN in the course of time. It helps the SCADA engineer to map the flow in the pipeline. It also illustrates how data prediction can be integrated with the WDN design, which is done with the ultimate aim of reducing losses in water distribution. This analysis also helps to determine the exact flow including the losses that happen during the transmission in order to match the supply as per requirement.

<p>| Table 5 | Hourly supply of water on 17 January 2019 |</p>
<table>
<thead>
<tr>
<th>EPANET Supply 17 January 2019 (in hrs)</th>
<th>Supply (MG)</th>
<th>Multiplier factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>375</td>
<td>0.2</td>
</tr>
<tr>
<td>1</td>
<td>350</td>
<td>0.06</td>
</tr>
<tr>
<td>2</td>
<td>450</td>
<td>0.02</td>
</tr>
<tr>
<td>3</td>
<td>550</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>575</td>
<td>0.06</td>
</tr>
<tr>
<td>5</td>
<td>775</td>
<td>0.13</td>
</tr>
<tr>
<td>6</td>
<td>1,000</td>
<td>0.9</td>
</tr>
<tr>
<td>7</td>
<td>1,400</td>
<td>1.68</td>
</tr>
<tr>
<td>8</td>
<td>1,600</td>
<td>1.35</td>
</tr>
<tr>
<td>9</td>
<td>1,400</td>
<td>1.35</td>
</tr>
<tr>
<td>10</td>
<td>300</td>
<td>0.67</td>
</tr>
<tr>
<td>11</td>
<td>300</td>
<td>0.56</td>
</tr>
<tr>
<td>12</td>
<td>250</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>200</td>
<td>2.56</td>
</tr>
<tr>
<td>14</td>
<td>200</td>
<td>3.56</td>
</tr>
<tr>
<td>15</td>
<td>200</td>
<td>2.81</td>
</tr>
<tr>
<td>16</td>
<td>200</td>
<td>1.24</td>
</tr>
<tr>
<td>17</td>
<td>300</td>
<td>0.75</td>
</tr>
<tr>
<td>18</td>
<td>500</td>
<td>0.9</td>
</tr>
<tr>
<td>19</td>
<td>400</td>
<td>1.75</td>
</tr>
<tr>
<td>20</td>
<td>300</td>
<td>1.25</td>
</tr>
<tr>
<td>21</td>
<td>300</td>
<td>1.25</td>
</tr>
<tr>
<td>22</td>
<td>400</td>
<td>0.34</td>
</tr>
<tr>
<td>23</td>
<td>300</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Total predicted = 12,624.72  
Total supplied = 12,625
CONCLUSION AND FUTURE WORK

To conclude, an IoT-based architecture integrated with Fog for an underground WDS has been proposed. In addition to proposing an IoT-based architecture for WDS, water demand forecasting has been done for water distribution design. The water demand forecast has been carried out on a daily basis for a period of three months using ARIMA and regression analysis.

Based on water demand forecasting analysis, water distribution design for an IoT-based architecture has been carried out using hydraulic engineering design for proper
distribution of water with minimal losses which would result in the development of a smart water distribution system (SWDS). This has been carried out using EPANET.

In future, the system can extend to predicting water demand using LSTM which is an extension of recurrent neural networks and be similarly extended for water distribution design for varying types of consumers. Also, the system can be incorporated with intelligence for an underground pipe health monitoring system using intelligent agents for immediate action from SCADA engineers.

REFERENCES


https://www.epa.gov/dwsixyearreview/drinking-water-distribution-systems.


ip=103.4.223.69&id=3057266&acc=ACTIVE%
20SERVICE&key=045416EF4D9A69D9%
2EA1D9F3C769B82345%2E4D4702B0C3E38B35%
2E4D4702B0C3E38B35&__acm__=1556078150_f57b96d7
ab5f648ca15d09135d036


