"Science Stays True Here"
Advances in Ecological and Environmental Research (ISSN 2517-9454, USA) | Science Signpost Publishing

Developing Deterioration Prediction Model for the Potable Water Pipes Renewal Plan – Case of Jubail Industrial City, KSA

Ali Madan Al-Ali, Jean Laurent, Jean Philippe Dulot
Marafiq-Saur Operation & Maintenance Company.

Received: April 04, 2019 / Accepted: May 12, 2019 / Published: Vol. 4, Issue 12, pp. 359-373, 2019

Abstract: The aim of this study is to identify the appropriate parameters for predicting the potable water pipes deterioration. The study evaluated the strength of some variables related to pipe breaks probability of failure based mainly on logistic regression model. The independent variables included in the study are static variables such as pipe diameter, pipe length and pipe material in addition to some dynamic (time-based) variables such as pipe age, water pressure and water velocity. The pipe break history (number of pipe breaks) for each pipe segment is used as dependent variable to be predicted in the statistical model. The resulted prediction equation is then used to calculate the failure probability for each pipe in the potable water network. Finally, prioritization of pipes is performed and the annual renewal plan is developed for the city of Jubail Industrial City in KSA based on the model results.

Key words: GIS, Geographical Information Systems, Pipes, Renewal, Logistic, Regression

1. Introduction

The objective of the study is to identify priority pipes segments in community areas of Jubail industrial city for replacement program in the next five years. The study performed the screening process by evaluating all the pipes in the database of Jubail community areas. The statisical analysis such as Logistic Regression require data for at least 5 years in order to provide reliable prediction of the pipe failures (Ambrose, Burn, DeSilva, & Rahilly, 2008). The most cost-effective pipes replacement strategy gives approximately 2% annual return on investment (Moglia, Burn, & Meddings, 2005). The life cycle cost range from 100 years (Ambrose, Burn, DeSilva, & Rahilly, 2008) to 200 years (Grigg, Fontane, & Zyl, 2013), which means that at least between 0.5% to 1% of the network length need to be renewed every year. However, the US national median of 1.7% for city pipeline replacement was reported by the American Water Works Association from aggregate data related to

Corresponding author: Ali Madan Al-Ali, Marafiq-Saur Operation & Maintenance Company.

combined water utilities including transmission and distribution (AWWA;, 2017) which is more applicable to the study combined network.

2. Scope of Work

Total number of the pipe segments under considerations is 29,658 with total length of 928.25 km that were built during years 1980 to 2017. Total number of 1053 pipe break notifications and 847 affected pipe segments that were recorded during 01/01/2012 to 25/04/2018 in the study area. All key pipes information such as age, diameter size, material and length are recorded in the Geographical Information System (GIS). In addition, some of other support information such as average operating pressure and velocity are recorded. Other parameters such as soil types, customer complaints and water quality are currently out of the scope but gradually could be used in the future as input to the analysis. The study covered only the community area in Jubail industrial city including districts in Deffi, Fanateer, East Corridor, Jalmudah and southern part of Mutrafiah.

The study adopted the American Water Works Association target to renew 1.7% (AWWA;, 2017) of the whole PW pipes network in the Jubail community areas each year in order to meet target life cycle of around 59 years of the whole network. The study aimed to identifying the most critical pipe segments (8.5% of the total network) that needs to be replaced during the next 5 years. In other words, the study attempted to identify the most critical 78.9 km of the current PW community pipes network where around 15.7 km need to be replaced each year.

3. Requirements, Preparation and Methodology

a. Data Requirements

Table (1) shows the essential data requirement for the analysis where continuous variables are numeric and categorical variables are binomial (0 or 1):

Table 01: Pipes data sources and parameters needed in the study

Data Variable	**Remarks**	**Input in Logistic Regression**
General Required Information		
Pipe ID	Unique GISID to differentiate each pipe segment and used to connect to the maps in GIS	No
Pipe Length (km)	Length information of each pipe segment in kilometers	No
District Boundary	Used for risk analysis and criticality calculation	No
Dependent Variable for Logistic Regression Analysis		
No. of Pipe Breaks (PB)	Dependent categorical variable (0: no PB event; 1: PB event)	Yes
Independent Variables for Logistic Regression Analysis		
Pipe Age (years)	Continuous variable (Age = current year – installation year)	Yes
Pipe Diameter (mm)	Used to classify 30 independent categorical variables (DIA_20, DIA_25, DIA_32, DIA_40, DIA_50, DIA_63, DIA_65, DIA_75, DIA_80, DIA_90, DIA_100, DIA_110, DIA_150, DIA_160, DIA_200, DIA_225, DIA_250, DIA_280, DIA_300, DIA_315, DIA_350, DIA_400, DIA_450, DIA_500, DIA_600, DIA_800, DIA_900, DIA_1000, DIA_1200 and DIA_1400)	Yes
Pipe Material	Used to classify 8 independent categorical variables (M_AC, M_PVC, M_DI, M_FRP, M_GRP, M_uPVC, M_RCP and M_SCP)	Yes
Velocity (m/s)	Used to calculate continuous independent variable (absolute velocity). Extracted from hydraulic model. Blank records filled by average values.	Yes
Pressure (kPa)	Used to calculate 3 independent continuous variables (P_Mean, P_Max and P_Min). Extracted from field loggers.	Yes

b. Data Preparation

The following are the procedures for preparing the data:

1. Check for raw data completeness to be 100% for the study area
2. Data cleansing and maintenance
3. Ensure integrity and consistency of materials and diameters records and convert to bionomial parameters.
4. Calculate Age of the pipes based on the installation year and classify age groups
5. Calculate number of pipe breaks in each pipe segment and in each zone
6. Conduct pipe break analysis of the pipes characteristics such as size, age, zone risk and pipe material
7. Calculate absolute velocity values for the pipes
8. Process pressure data logs, identify pressure zones and associate values to related pipes
9. Recode some independent parameters to be categorical variables and make sure the dependent variable to be binary
10. Use Logistic Regression analysis to identify significance and weights for the parameters

c. Overall Process and Methods

The method is partially inspired from the Water Distribution System Risk Tool for Investment Planning by Water Research Foundation, EPA and WERF (Grigg, Fontane, & Zyl, 2013). This methodology has been customized according to the local situation of Jubail community network to accommodate local available data in GIS. The methodology is predictive method based on statistical analysis and ranking of multiple criteria from historical performance and failure. Logistic regression analysis has been selected to evaluate the strength of all parameters in predicting the occurrence of future pipe break events in all pipe segments. Following are main methodology steps:

1. Identify criticality of residential zones
2. Identify influence threat factors
3. Specify probability of failure values based on logistic regression result
4. Prioritization of critical pipes and plan renewal accordingly

4. Identify Criticality of Residential Zones

The criticality rate is calculated for each district in the community areas as in the following equation:

Pipe Breaks Rate = (Total # of Pipe Breaks in a District / Network Length of a District) / no. of monitoring years

Note that all districts of the strudy area have started to be monitored in the same year (2012) but many pipes are newely installed after 2012 which will have lower number of monitoring years. Therefore, it is required to divide by the number of monitoring year to get correct rate for all pipes segments. The result and thematic map showed the most critical residential districts in Jubail Industrial city that are facing highest rate of repeated pipe breaks per km of network length which are respectively: Huawailat (Camp 11) and Al Hijaz (B1), Al Kods (D2), Makkah (B2), JIC, Al Faiha (D3) and Camp 10. These are the areas, which got extreme risk and upper high risk. The result of the remaining districts can be seen in Figure (1) for the high, medium and low risk.

MaSa GIS

Rate: No. of Pipe Breaks / KM of Pipes for each District (2012 to 2018)

AREA NAME	AREA CODE	PB Qty	PW Pipes Length (KM)	Rate: PB/km
Al-Rayhan	F1	0	0.00	0
Al-Yasameen	F3	0	0.00	0
AL-ULA	A4	0	33.97	0
DAREEN	B6	2	41.31	0.04
RIYADH SECTOR	A1	2	41.86	0.04
Al-Rabea	F4	1	19.68	0.05
TAIF	A3	2	26.27	0.07
YANBU	A5	8	62.26	0.12
CAMP-7 & 8 EPS	C7 & 8EPS	1	5.57	0.18
Al-Khozamah	F2	13	56.10	0.23
SUDAYER	B3	17	44.99	0.37
TAIBAH	A2	16	39.48	0.4
CAMP 9	C9	8	18.15	0.44
NAJD	B4	33	61.49	0.53
AL AHSA	B5	24	41.94	0.57
AL ANDALUS	D1	29	38.56	0.75
SWC	SWC	4	4.87	0.82
AL FAROUK	D4	66	71.78	0.91
AL FERDOS	D5	48	48.61	0.98
CAMP 10	C10	25	21.36	1.17
AL FAIHA	D3	67	54.20	1.23
JIC	JIC	29	21.69	1.33
MAKKAH	B2	64	47.54	1.34
AL KODS	D2	50	31.07	1.6
AL HIJAZ	B1	132	57.18	2.3
HUAWALAT	C11	91	37.56	2.42

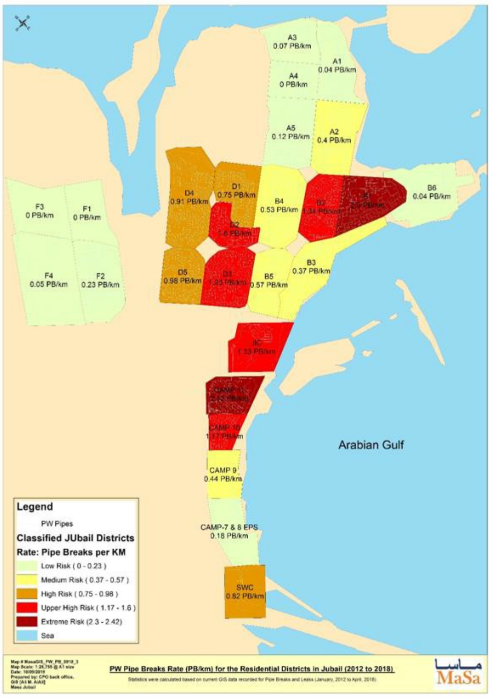

Figure 1: Table and Thematic map of the criticality analysis of pipe breaks per districts in Jubail Industrial city

5. Identifying Influence Threat Factors

The data analysis is based on five main factors where three are considered as assets data related to the pipe (age, diameter and material) and two are operational hydraulic factors (pressure and velocity). These factors are the predictors that are used to get the probability of pipe breaks occurrence.

a. Asset Data

i. Age of the Pipe

Some statistics related to the age of pipes were extracted from the database (Table 2). It has been noticed that around 61% of the network length with age more than 26 years. Overall, the average of 1.1 pipe breaks per km was calculated. However, some of particular ages (29, 33 and 35) have the highest rate of 2 or more of pipe breaks per km.

Table 2: Age groups and calculations of length and pipe break rates.

Pipes Age Group	Length (km)	Total Pipe Breaks	Rate (PB/km)
Age Group 1: *1 to 26 years*	363.02	115	0.31
Age Group 2: *30 to 38 years*	565.23	938	1.65
Grand Total	**928.25**	**1053**	***Avg = 1.13***

ii. Diameter size

Calculation and analysis of the diameter sizes of the pipes in the study area shows that diameters of smaller sizes (20mm to 110mm) represents around 21.6% of the network in the study area while pipes with the largest diameter sizes (450mm to 1400mm) only represents around 4% of the network length. However, the majority of the pipes in the network falls in the middle class of diameters (150mm to 400mm) which represents 74.4 % of the network length. This is reflected on the high number of pipe breaks (88% of the pipe breaks) in the diameters (150mm to 400mm) as seen in Figure (2).

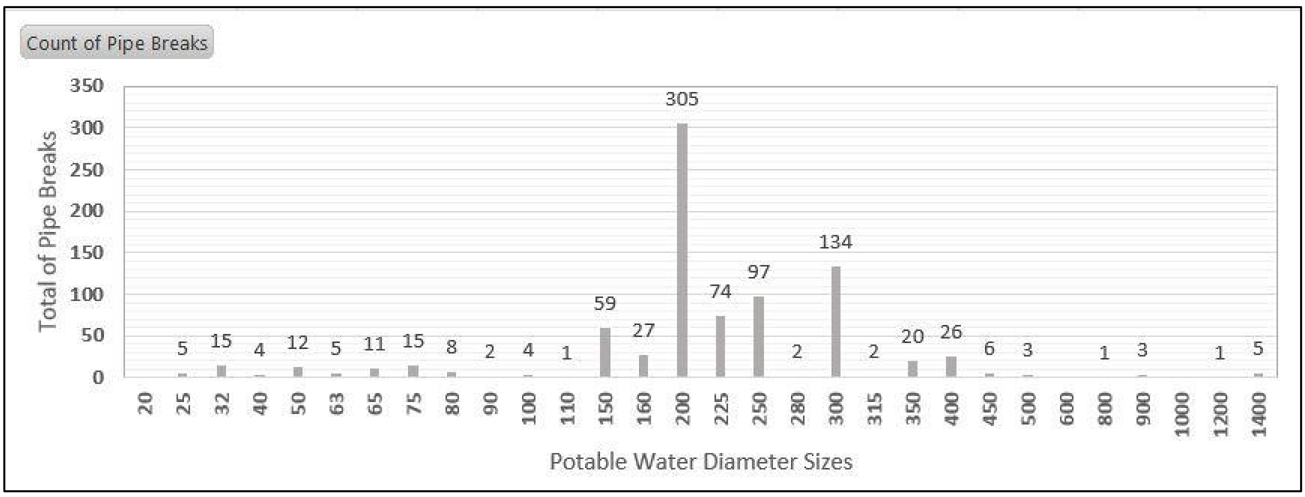

Figure 2: Total potable water network pipe breaks according to the diameter sizes of the pipes

iii. Material

The rate (PB/KM) indicates if certain type of pipe material is breaking more often than the other types of material such as PVC, which has the highest rate of (1.32 PB/km) among other types of materials (Table 3). Following comes the uPVC with approximately an average rate of 0.99 PB/km. On the other hand, FRP and RCP materials showed the lowest PB rates.

Table 3: Calculations of length and pipe breaks categorized by material type

Material Type	Length (km)	Total Pipe Breaks	Rate (PB/km)
AC	59.66	28	0.47
DI	41.21	15	0.36
FRP	0.09	0	0.00
GRP	52.01	25	0.48
PVC	129.24	171	1.32
RCP	20.17	1	0.05
SCP	20.14	9	0.45
uPVC	605.74	598	0.99
Grand Total	**928.3**	**847**	***Avg = 0.91***

b. Hydraulic Data

The hydraulic parameters includes pressure and velocity. Velocity data were extracted from the main lines in the hydraulic model while the pressure data are extracted from 23 field data loggers.

i. Velocity (M/S)

The velocity data were exported from the hydraulic model and processed inside the developed GIS. The current hydraulic model provided calculations of velocity for the main lines only which represent only 20% of the total network length (187.2 km). Furthermore, only 202 pipe break events (23.8%) occurred on these main lines as indicated in Table (4). Records of the other pipes (740.4 km) were filled with the average absolute velocity value (0.089626 m/s) in order to be able to run the statistical model.

Table 4: Length and pipe breaks of main lines categorized by velocity level

Velocity Group	Length (km)	Total Pipe Breaks	Rate (PB/km)
Low Velocity (< 0.10 m/s)	116	135	1.16
High Velocity (=> 0.10 m/s)	71.8	67	0.93
Grand Total	**187.8**	**202**	***Avg = 1.07***

ii. Pressure (kPa)

The pressure data was extracted from 23 field data loggers (Table 5) where a total of 6,543,563 logs for the period 17/07/2017 to 03/09/2018 were processed. The statistics were calculated for each data logger and Thiessen polygons were created using GIS for the position of each data logger in order to cover the network pipes in nearest area to each logger. The mean, minimum and maximum pressure was associated with each pressure zone created from these polygons. Then, the mean, minimum and maximum pressure was associated to each related pipe segment within each pressure zone. The three pressure measures (P_Mean, P_Max and P_Min) variables were used as predictors for fitting the logistic regression model for all pipes' segments.

Table 5: Network and Pump stations Data loggers used to measure Pressure (kPa) parameter

Seq.	Logger Location	Logger Type	P_Max	P_Min	P_Mean	Date_From	Date_To	No. of Logs
1	T-230, Near Petrokemya, FH-90	KPI Logger	393.90	188.06	299.84	01/10/2017	31/10/2017	89,280
2	T-154, WWPS-7, Near SAFCO	KPI Logger	400.36	182.07	309.84	01/10/2017	30/10/2017	89,280
3	Ferdaus, T-Ahzab, FH-32	KPI Logger	280.73	125.54	230.79	01/10/2017	31/10/2017	89,280
4	T-Dammam/Dammam 17, FH-16	KPI Logger	262.06	117.75	227.46	01/10/2017	31/10/2017	89,280
5	Makkah, T-Sarat/T-Zamzam, FH 1-1	KPI Logger	295.40	153.78	254.41	01/10/2017	31/10/2017	89,280
6	Sudayer, Hawiyah 2, FH-15	KPI Logger	264.75	128.54	222.37	01/10/2017	31/10/2017	89,280
7	RC Building, backside visiter building	Network Logger	483.05	-0.97	371.08	12/07/2017	09/11/2017	345,826
8	T-Dammam, near Dammam 26, FH-1/3	Network Logger	290.96	67.02	230.41	17/07/2017	22/11/2017	368,714
9	T-Andulus/Andulus 9, FH-33	Network Logger	273.03	-1.38	211.18	17/07/2017	09/10/2017	242,039
10	T-Ferdaus/Ferdaus 20, FH-5	Network Logger	301.16	-1.38	228.99	13/07/2017	21/11/2017	378,028
11	T-Faiha/T-Khamees, near Faiha 27, FH-201	Network Logger	290.68	-1.65	226.28	13/07/2017	27/11/2017	394,700
12	Camp 11, T-Huwaylat/T-Dairie, St. 46, FH-01	Network Logger	251.80	61.09	179.81	12/07/2017	09/10/2017	255,979
13	T-Najd/Najd 16, FH 1-8	Network Logger	349.15	75.43	227.60	17/07/2017	22/11/2017	368,662
14	T-Faiha/Faiha 7, FH-62	Network Logger	273.31	-8.27	152.87	13/07/2017	27/11/2017	394,642
15	Kods 8/T-Khalil, in front of fire station, FH-30	Network Logger	292.75	91.84	214.29	17/07/2017	21/11/2017	366,384
16	T-Andulus/Andulus 23, FH-43	Network Logger	294.27	-3.72	220.20	17/07/2017	09/10/2017	241,924
17	Farooq, T-Karamah/T-Batra, FH-327	Network Logger	279.24	-0.55	217.25	13/07/2017	21/11/2017	377,771
18	T-Farooq/T-Sedieg, FH-37	Network Logger	280.62	72.95	224.38	13/07/2017	21/11/2017	377,773
19	Fanateer PS, Discharge line A	PS Logger	441.61	48.61	221.47	18/07/2017	27/11/2017	380,257
20	Fanateer PS, Discharge line B	PS Logger	448.50	52.06	245.81	18/07/2017	27/11/2017	380,204
21	Deffi PS, Discharge line A	PS Logger	315.44	26.54	229.78	18/07/2017	27/11/2017	380,240
22	Deffi PS, Discharge line B	PS Logger	317.16	-2.07	231.25	18/07/2017	27/11/2017	380,191
23	Jalmudah, T-6, after EXTRA mall, FH	PS Logger / RTU	339.56	32.75	225.05	17/07/2017	03/09/2018	374,549

6. Statistical Analysis

a. Research Question

Failure predictions are thorough analysis of existing asset and failure data. Use of the failure predictions rather than just the historical performance when making pipes renewal decisions could reduce the predicted costs considerably. Statistical logistic regression analysis is required in order to get the prediction equation based on the explanatory variables. Therefore, the research question is: What is the impact of age, diameter, material, velocity and pressure on the probability of pipe breaks?

Overall Likelihood index of Failure = *f (age, diameter, material, velocity, pressure)*

b. Initial Logistic Regression Analysis

Multiple duplicate records were created for pipe breaks occurred more than once in a single pipe segment in order to have only 0 or 1 in the response variable for each record, which provide *Binary* response type using logistic function (*Logit*) model. Direct logistic regression was performed to assess the impact of all factors related to the function (*age, diameter, material, velocity and pressure*) on the likelihood that pipe break will occur. The model contained 43 independent variables as explained in Table (1). Result of the initial logistic regression analysis indicated that coefficients of 2 predictors (DIA_1400 and M_SCP) could not be defined by the model because of singularities. The low p-value out of the logistic regression model fitting result indicating that only the intercept and 7 independent variables are statistically significant suggesting a strong association between them with the probability of pipe break event.

c. Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) was also performed as statistical technique for investigating data by comparing the means of subsets of the data. The function compares the sequential logistic regression models which compares the smaller model with the next more complex model by adding one variable in each step. Each of those comparisons is done via a likelihood ratio test (LR test). Then, each coefficient against the full model containing all coefficients. ANOVA test of the 'main effect' for each independent variable which also explore the possibility of an 'interaction effect' among levels of independent variables on the dependent variable.

It has been noticed in the resulted analysis of deviance table which measure the goodness of fit that Reside. Dev is decreasing from 9113 (at intercept level) and every time when new independent variable added to the model until it reaches 7190.7 (at the full model level). The term were added sequentially from first to last where the deviance or the difference between null model and after adding the Age_Years variable = 670.18 was the largest deviance. The 2nd largest deviance was for DIA_150 = 509.14, followed by other DIA variables such as DIA_32, DIA_40 and DIA_50 with values (106.24, 111.34 and 107.74) sequentially. The probability of seeing a difference in Reside. Dev "Pr(>Chi)" indicated possible improvement in the model fit upon adding some variables is greater than what is expected by chance alone. These additional significant independent variables are DIA_63, DIA_65, DIA_90, DIA_110, DIA_160, DIA_225, DIA_250, DIA_300, DIA_400, DIA_450, M_AC, M_PVC, M_DI, M_GRP and P_Max.

d. Final Logistic Regression Analysis

Direct logistic regression was performed again to assess the impact of significant factors after performing ANOVA on the initial logistic regression model as these additional factors showed possible improvement in the model fit on the likelihood that pipe break will occur. (see Table 6). The low p-value out of the final logistic regression model indicated that the model fit improved and the significant predictors increased from 7 to 16 independent variables which are statistically significant suggesting a strong association between them with the probability of pipe break event.

Out of the statistically significant predictors, it has been noticed that the intercept and 9 diameter variables have negative coefficients suggesting that these variables being equal, the related pipe segments are less likely to have pipe breaks. In particular, the significant variables with negative coefficient are representing the relatively smaller diameter pipes as following: DIA_25, DIA_32, DIA_40, DIA_50, DIA_63, DIA_65, DIA_90, DIA_150 and DIA_160.

Additionally, one material variable (M_DI) showed negative coefficient indicating that DI material pipes are less likely to have pipe breaks compared to other types of materials.

On the other hand, other types of material variables (M_PVC and M_GRP) have positive coefficient suggesting that these types of materials are more vulnerable to pipe breaks. Also, the larger diameter variables (DIA_250 and DIA_300) along with Age_Years and pressure mean (P_Mean) have positive coefficient suggesting that all other variables being equal, the relatively old and large diameter pipes with high pressure mean are more likely to have pipe breaks. Finally, M_AC material variable along with the other diameters, maximum pressure and velocity variables showed high p-values in the logistic regression model fitting results which indicate that all remaining variables are not statistically significant.

Table 6: Model result of fitting logistic regression analysis in R

```
Call:
glm(formula = PB_Count.f ~ Renewal.data$Age_Years + DIA_25.f +
    DIA_32.f + DIA_40.f + DIA_50.f + DIA_63.f + DIA_65.f + DIA_90.f +
    DIA_110.f + DIA_150.f + DIA_160.f + DIA_225.f + DIA_250.f +
    DIA_300.f + DIA_400.f + DIA_450.f + M_AC.f + M_PVC.f + M_DI.f +
    M_GRP.f + Renewal.data$P_Max + Renewal.data$P_Mean, family = binomial(link = "logit"),
    data = model1data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2615  -0.2056  -0.1398  -0.1026   3.8441

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)            -5.8679067  0.2118612 -27.697  < 2e-16 ***
Renewal.data$Age_Years  0.0894120  0.0045638  19.592  < 2e-16 ***
DIA_25.fDIA = 25       -2.9881975  0.4525545  -6.603 4.03e-11 ***
DIA_32.fDIA = 32       -2.4958960  0.2680098  -9.313  < 2e-16 ***
DIA_40.fDIA = 40       -3.2727430  0.4533895  -7.218 5.26e-13 ***
DIA_50.fDIA = 50       -2.5612037  0.2960059  -8.653  < 2e-16 ***
DIA_63.fDIA = 63       -2.1198955  0.4535722  -4.674 2.96e-06 ***
DIA_65.fDIA = 65       -1.6706617  0.2864155  -5.833 5.44e-09 ***
DIA_90.fDIA = 90       -1.8579021  0.7148429  -2.599 0.009349 **
DIA_110.fDIA = 110     -2.4019933  1.0069747  -2.385 0.017063 *
DIA_150.fDIA = 150     -2.0850377  0.1400334 -14.890  < 2e-16 ***
DIA_160.fDIA = 160     -0.7561222  0.2006584  -3.768 0.000164 ***
DIA_225.fDIA = 225     -0.1349843  0.1239883  -1.089 0.276292
DIA_250.fDIA = 250      0.3030320  0.1144329   2.648 0.008094 **
DIA_300.fDIA = 300      0.7815801  0.0978849   7.985 1.41e-15 ***
DIA_400.fDIA = 400     -0.2992226  0.2292412  -1.305 0.191800
DIA_450.fDIA = 450      0.9825621  0.4434521   2.216 0.026711 *
M_AC.fAC Material       0.4279170  0.2267588   1.887 0.059147 .
M_PVC.fPVC Material     0.4424478  0.0851588   5.196 2.04e-07 ***
M_DI.fDI Material      -1.1501682  0.2630056  -4.373 1.22e-05 ***
M_GRP.fGRP Material     1.0838486  0.2364036   4.585 4.55e-06 ***
Renewal.data$P_Max     -0.0011517  0.0008658  -1.330 0.183472
Renewal.data$P_Mean     0.0048819  0.0011928   4.093 4.26e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9113.0  on 29863  degrees of freedom
Residual deviance: 7238.7  on 29841  degrees of freedom
AIC: 7284.7

Number of Fisher Scoring iterations: 9
```

e. Probability of Failure Prediction

The equation of the final prediction model (Variable Pipe_Breaks) is: Pred (Pipe_Breaks = 1) = exp(z) / [1 + exp(z)]

Where;

$z = b_0 + b_1x_1 + b_2x_2 + \ldots\ldots + b_nx_n$

b_0 = the intercept constant

b_n = the regression coefficient of the n variables

Then;

z = -5.8679067 + 0.0894120 X Age_Years + -2.9881975 X DIA_25 + -2.4958960 X DIA_32 + -3.2727430 X DIA_40 + -2.5612037 X DIA_50 + -2.1198955 X DIA_63 + - 1.6706617 X DIA_65 + -1.8579021 X DIA_90 + -2.0850377 X DIA_150 + -0.7561222 X DIA_160 + 0.3030320 X DIA_250 + 0.7815801 X DIA_300 + 0.4424478 X M_PVC + - 1.1501682 X M_DI + 1.0838486 X M_GRP + 0.0048819 X P_Mean

The statistics of the predicted pipe breaks probabilities are: N = 29,658, Mean = 0.047756, Min = 0.000381, and Max =0.530694. The final prediction model was tested on N = 837 pipes with previous real failure history where the mean of 0.047756 was used as decision boundary where values predicted above the mean will have 1 (predicted pipe break event) and prediction values less than the mean will have 0 (no predicted pipe break).

The results showed that 74.3% of the pipe breaks were predicted correctly as in reality. (see Figure 3).

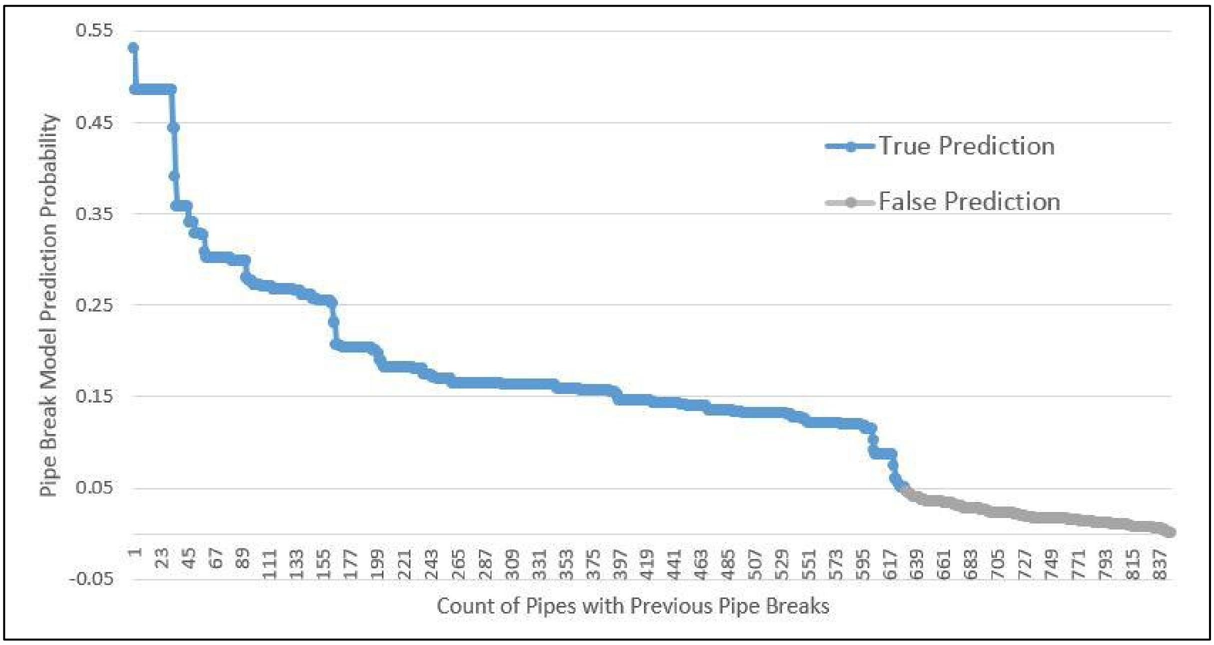

Figure 3: Graph of the model predicted probability (0 to 1) as result of logistic regression prediction equation tested on real sample.

7. Prioritization of Critical Pipes

The predicted pipe break probability values were calculated based on the final logistic regression model for each pipe segments in the whole network. Then, prioritization of the pipes was performed based on the highest probability values for the most critical 78.76 km of the complete potable water network. Table (7) and the map in Figure (4) provides more details about the critical pipes chosen by the model to renew as priority in the next five years plan.

Table 7: Priority levels for the annual critical pipes renewal plan

Priority Levels	Length (km)	Quantity of Pipes	Predicted PB Probability Range
Priority 1	17.28	110	0.358 to 0.530
Priority 2	15.69	150	0.302 to 0.358
Priority 3	14.65	305	0.267 to 0.302

Priority 4	15.68	259	0.251 to 0.267
Priority 5	15.46	206	0.204 to 0.251
Total	**78.76**	**1030**	**0 to 1 pipe break probability**

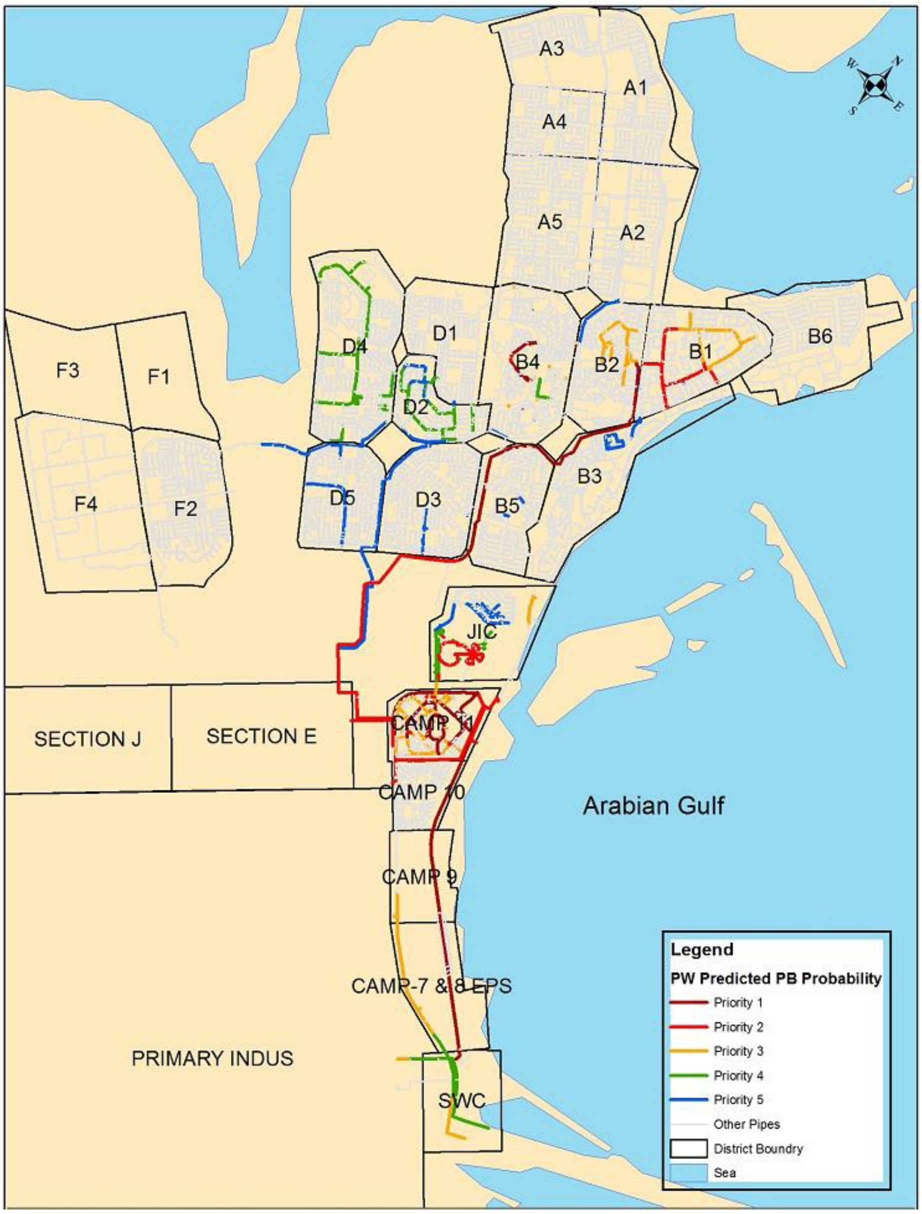

Figure 4: Map of the Potable Water critical pipes based on high predicted pipe break probability, *Pred (Pipe_Breaks = 1) = exp(z) / [1 + exp(z)]*.

8. Discussion and Conclusion

The methodology developed in this paper is essential for water utility companies in order to maximize utilization of all available asset and historical data to direct the huge pipes renewal investment in the right way. Out of the initial 43 independent variables, 16 predictors showed to have impact on the pipe break occurrence. In particular, the age, some diameter classes (250mm and 300mm), some material types (PVC and GRP) and the pressure mean showed positive correlation which could increase probability of pipe break events. Other variables showed tendency to decrease pipe breaks such as smaller diameter sizes and pipes made from DI material. Some litrature (K & Sagar, 2016; Achim, Ghotb, & McManus, 2007) found that pipe length has an important impact on the annual pipe break rate. Actually, the length was tested in the initial model and gave significant results as well with large positive coefficient but the authors decided to discard it from the model of this paper as its effect was clear on the final priority map covering only 121 main and long pipes on the network and consider them as most critical. The result of final model of this paper gave more detailed answer to the initial analysis of critical areas (Figure 1) and provided higher resolution plan for the most critical 78.76 km pipes in the network (Table 7 and Figure 4) as it can be seen that the priority 1 and 2 pipes are falling mainly on the most critical areas (B1 and Camp 11). Finally, the use of the GIS tool as a master repository for all key analysis information was very useful and efficient especially for detailed mapping and planning of the final results.

9. Future Study Improvement

The study used some assets and hydraulic parameters to estimate around the failure likelihood. However the study can be advanced in the future by improving some of the current parameters (such as more complete velocity based on GIS/hydraulic integration) and adding more explanatory variables. These parameters could include water temperature, ground water, weather condition, improper bedding, low stiffness, traffic vibration, water hammer, external vibration, corrosion issues, air pocket, operating condition, roots from trees, leakage and water loss, history of water quality complaints and bad joining. Root cause analysis findings and some previous studies/reports could help in addressing some of these additional factors. Furthermore, future studies could include estimation of the consequence of failures and getting the consequence rating scores (SAR) for each pipe segment. The cost of failure parameters could include number of affected facilities and customers, potential flooding, water loss, and cost of repair.

About Main Author: Ali Madan Al-Ali was born in Saihat, KSA, in 1982. He graduated in 2005 from King Fahd University of Petroleum & Minerals, with a BS degree in Management Information System. In 2008, he finished his degree of Master of Science in Geographical Information Systems (GIS) from the University of Leeds in UK. In 2014, he finished his degree of Master in Environmental Science from Trinity College Dublin in Ireland. He published a number of scientific papers in spatial and environmental sciences related to wind directions, mangroves and archeology. The author has professional work experience mainly in Oil & Gas exploration industry along with water utilities operation & maintenance.

References

[1]. Achim, D., Ghotb, F., & McManus, K. J. (2007). Prediction of Water Pipe Asset Life Using Neural Networks. *Journal of Infrastructure System*, 26-30.

[2]. Ambrose, M., Burn, S., DeSilva, D., & Rahilly, M. (2008). LIFE CYCLE ANALYSIS OF WATER NETWORKS. *Plastic Pipes Confrecnces Association*. Budapest: Plastic Pipes XIV.

[3]. AWWA;. (2017). *2017 AWWA Utility Benchmarking Performance Managment for Water and Wastewater*. Kissimmee: American Water Works Association.

[4]. Grigg, N. S., Fontane, D. G., & Zyl, J. v. (2013). *Water Distribution System Risk Tool for Investment Planning*. Denver: Water Research Foundation.

[5]. K, F. H., & Sagar, G. Y. (2016). Statistical Analysis of Pipe Breaks in Water Distribution Systems in Ethiopia, the Case of Hawassa. *IOSR Journal of Mathematics (IOSR-JM)*, 127-136.

[6]. Moglia, M., Burn, S., & Meddings, S. (2005, October). Parms-Priority: a methodology for water pipe replacement. *Pipes Wagga Wagga*, 17-21.