

Public Infrastructure Asset Assessment with Limited Data

Frederick Bloetscher¹, Lloyd Wander², Greg Smith², Nihat Dogon³

¹Department of Civil, Environmental and Geomatics Engineering, Florida Atlantic University, Boca Raton, USA

²Utility Sealing Services Inc. in Venice, Venice, USA

³ModelsbyToggles Inc., Phoenix, USA

Email: h2o_man@bellsouth.net

How to cite this paper: Bloetscher, F., Wander, L., Smith, G. and Dogon, N. (2017) Public Infrastructure Asset Assessment with Limited Data. *Open Journal of Civil Engineering*, 7, 468-487.
<https://doi.org/10.4236/ojce.2017.73032>

Received: July 31, 2017

Accepted: September 23, 2017

Published: September 26, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

When creating an asset management plan, missing data is perceived to be a huge problem, especially when the event data (breaks in water distribution pipes as an example) are not tracked. The lack of tracking makes it difficult to determine which factors are the critical ones. Many utilities lack the resources for examining buried infrastructure, so other methods of data collection are needed. The concept for this paper was to develop a means to acquire data on the assets for a condition assessment (buried pipe is not visible and in most cases, cannot really be assessed). What was found was that for buried infrastructure, much more information was known than anticipated. Knowing exact information is not really necessary. However, there was a need to track event-breaks, flooding etc.—what would indicate a “failure”. The latter would be useful for predicting future maintenance needs and the most at-risk assets.

Keywords

Bayesian, Information Theory, Asset Uncertainty, Infrastructure

1. Introduction

During the period 1973 to 1985, public net investment in the United States and Japan averaged 0.3% and 5.1% of gross domestic product, while their respective growth rates of real gross domestic output per employed person were 0.6% and 3.1% per annum (OECD National Accounts and Historical Statistics). Twenty-five years ago, Aschauer [1] and Munnell and Cook [2] found a strong positive relationship between infrastructure and growth. Aschauer [1] estimated that the productivity-raising power of infrastructure investment to be huge, as much as quadrupling that of private investment. Barro [3] noted that production is dri-

ven by its flow of productive expenditures toward infrastructure.

However, at present state and local governments spend about 2.4% of their GNP on infrastructure, as compared to 3.1% in 1970 [4]. A large portion of that amount is spent for growth as opposed to repair and replacement. As a result, it should come as no surprise that public infrastructure has been poorly rated by the American Society of Civil Engineers [5] [6] [7] [8] [9] and most public officials acknowledge the deterioration and risk of failure of the infrastructure relied on daily for economic and social systems. However, many jurisdictions have limited information about their systems, and little data to use to justify spending. Hence the infrastructure tends to deteriorate further each year as local officials opt to limit budgets in the absence of good needs data further, hence the need for better tools for asset management.

Asset management is a process of integrating design, construction, maintenance, rehabilitation, and renovation to maximize benefits and minimize cost. It is a plan for managing an organization's infrastructure through a decision-making process driven by a defined standard level of service. The term asset management refers to business principles aimed at balancing risk and minimizing life-cycle costs of the physical assets, such as pipes, roads, structures and equipment. Asset management is also used as a tool to help municipalities gauge the health of infrastructure [10]. A strategy should be continuously reviewed and revised to acquire, use and dispose of assets while optimizing service and minimizing costs over the life of the assets. An asset management plan (AMP) considers financial, economic, operational, and engineering goals in an effort to balance risk and benefits as they relate to potential improvement to the overall operation of the system [11].

Asset management plays a vital role to help minimize unnecessary or misplaced spending while meeting the health and environmental needs of a community [11]. Organizations that utilize asset management programs experience prolonged asset life by aiding in rehabilitation and repair decisions while meeting customer demands, service expectation and regulatory requirements [12]. The general framework of asset management programs involves collecting and organizing data on the physical components of a system and evaluating the condition of these components. The importance and the potential consequences associated with the failure of the individual assets are determined by this evaluation. Managers and operators can then prioritize what infrastructure is most critical to the operation of the system and furthermore which assets to consider for repair, rehabilitation or replacement. This strategy allows for funding, in terms of both repair and replacement (R & R) and operation and maintenance (O & M) dollars, to be distributed accordingly amongst the asserts that are the most vulnerable and most likely to fail. The goal is to provide strategic continuous maintenance to the infrastructure before total failure occurs. Of utmost importance is to define "failure" of the infrastructure. For example, for a storm water system, "failure" might mean that the community has areas that flood as a

result of rainfall, tidal impacts, sea level, or groundwater elevation. Costs should be well distributed over the life of the asset to help avoid emergency repairs, because emergency repairs can cost multiple times the cost of a planned repair.

The reliability of the assets within the area of interest starts with the design process. Decision-making dictates how assets will be maintained and means to assure the maximum return on investments. An inventory of assets needs to be established. Depending on the accuracy wanted, the data can be gathered in many ways ranging from on-site field investigation which could take a lot of time, to using existing maps, using maps while verifying the assets using aerial photography and video, or field investigations. Through condition assessment, the probability of failure can be estimated. Assets can also fail due to exceeding its maximum capacity. Prioritizing the assets by a defined system will allow for the community to see what areas are most susceptible to vulnerability/failure, which assets need the most attention due to their condition, and where the critical assets are located in relation to major public areas (hospitals, schools, etc.) with a high population.

But what if the infrastructure data is limited, which is often the case with buried pipelines? In such cases it is difficult to analyze the condition of the system and prioritize repair and replacement dollars. The goal of this paper is to outline a means to assess the system's condition, by evaluating what can be inferred from the known data, without the need to dig up the piping. The question is how to collect data that might be useful that does not involve destructive testing on buried infrastructure which is costly and inconvenient. The reality is that there is more data than one thinks.

2. Methodology

The problem with condition assessments is that for many of the infrastructure assets, determining the condition is nigh on impossible. Buried infrastructure is nearly impossible to determine without unburying it. Even infrastructure that is visible may provide a false assessment. A fire hydrant is partially buried. The foundations for bridges and the base of a roadway are not visible. Stormwater pipes may be visible only at outlets. Hence many asset management programs stall when there is a need to assess the condition of the assets. Many assume that since the buried infrastructure is unseen the condition cannot be determined. But this is rarely true. There is usually some information that is known, but the certainty of this information is the challenge. Uncertainty is a concept that many people, including engineers, operations staff and administrators are uncomfortable with. However statistical analysis is the mathematical means to address this uncertainty since even with some data, there is still uncertainty.

Several statistical methods have been developed to attempt to exploit limited information: resampling (bootstrap and jack-knife methods), fuzzy set theory, interval analysis, information theory and Bayesian methods [13]. Resampling

methods permit a sampling distribution to be created from limited data, by using the likelihood as a means to estimate the distribution parameters [13]. These parameters are then used to create probability distributions that are sampled. Sampling repeatedly from each distribution allows the investigator to simulate expected uncertainty and variability from the original set of data. As much data as the investigator wants can be created. The process is relatively straightforward for two variables but Haas [14] report that the method is tedious with multiple variables and the entire analysis is limited by the quality of the original data. Perhaps most importantly, the method offers no capability for using subjective information, which may be available [13].

Fuzzy set theory/logic was proposed as a paradigm shift in logic that involves a set of rules that define boundaries, and solves problems within those boundaries [15]. As the name suggests, fuzzy logic is the logic of underlying modes of reasoning that are approximate rather than exact, where everything is a matter of degree used to handle the concept of partial truths—between true and completely false. Fuzzy set analysis is usually applied to subjective, verbal information that is divided into two sets of data. Fuzzy set theory measures the intersection of the two sets. The intersection is the fuzzy set pairing value. The most obvious limitation noted in the literature is that the data cannot be mutually exclusive from the two sets. In addition, real data are not helpful, nor is new data.

Interval analysis is an approach to the analysis of systems when the value of the quantity measured is uncertain. Interval analysis defines the value of the quantity by specifying the interval that the value is guaranteed to fall within. The methodology provides a correct formal method for measuring the upper and lower bounds required for the worst possible case. Interval analysis is not as powerful as other statistical methods when empirical information is available. Some prior information or data to create the subjective opinion is required. Like the methods previously discussed, updating with new data is not feasible, and with limited data, the ability to refine the distribution with prior data is a desirable ability.

A solution to address at least some of these issues comes from Shannon's Information Theory and a theorem that he proved in 1948: "The probability distribution having maximum entropy (uncertainty) over any finite range of real values is the uniform distribution over that range" [16]. Information theory is a branch of mathematical theory within probability and statistics that can be applied to a variety of fields [16]. The roots of information theory are found within the concepts of disorder or entropy in thermodynamics and statistical mechanics [17]. Since the turn of the 20th century, significant literature has been devoted to the studying the mathematical form of information entropy [17]. The literature assumes a sample space S , with a series of events i characterized by individual probabilities p_i . Shannon [18] developed the following measure to indicate the information content of a probability distribution, which is in the form of the

negative of the information content of the data set:

$$H = -\sum_i p_i \ln p_i \quad (1)$$

By maximizing information entropy, the most conservative or broadest distribution consistent with the available information can be derived - such as the mean, variance and range [19]. In general, if the mean or expected value of any function $F(X)$ of a quantity is known, the entropy can be maximized subject to:

$$\sum_i p_i = 1 \quad (2)$$

and

$$\sum_i F(X) p_i = E(F(X)) \text{ for all functions of } F \quad (3)$$

The concepts of information entropy are a useful theoretical underpinning in the application of Bayesian methods, useful in many aspects of the analysis [20]. Definitive observations do not play an important part in information entropy theory since once the definitive observation is made; the underlying uncertainty is greatly reduced or eliminated [13]. The uncertainty defines the confidence in the observation [19]. The use of subjective or recorded data in the absence of absolute certainty is akin to the use of Bayesian methods in the field of statistics.

The selection of Bayesian methods assumes that the absolute or unconditional probability density function $p(x)$ on X is the underlying distribution found through curve-fitting. Its form, as defined by Aitchison and Dunmore [21] is:

$$p(x) = \int p(x|\theta) p(\theta) d\theta \quad (4)$$

where $p(x)$ and $p(\theta)$ are completely different, and independent functions and the function $p(\theta)$ is the prior distribution. The Bayesian approach is to assume that while the true value of θ is unknown, there are probabilities that can be assigned for a series of possible values of θ [21]. More precisely it is assumed that $p(\theta)$ is a density function.

For the purposes of infrastructure assessment, the “mean” value of θ , or $E(\theta)$, is akin to the expected age, material, soil, depth, traffic, groundwater table, or other factor that the assessor wishes to consider. And, like Bayesian statistical methods, the more data gathered for a given asset, even when unknown, the less likely any error in the estimate will frustrate the true condition. The more information, the more likely that outcome can be predicted. As a result, Bayesian methods permit the evolution of the prediction with added data—the shape of the distribution of results becomes more narrowly focused on the likely solution. Inference, in the absence of fact, is the key. The concept works for sizes of incidents, condition of infrastructure and likelihood of failure.

The only thing that is missing in gathering data is the need for an “event” or consequence. To be useful there must be some form of tracking consequences: breaks, flooding, etc. So, the agency must identify if there is data to indicate the events, such as work orders. If so, they may contain enough data needed to piece

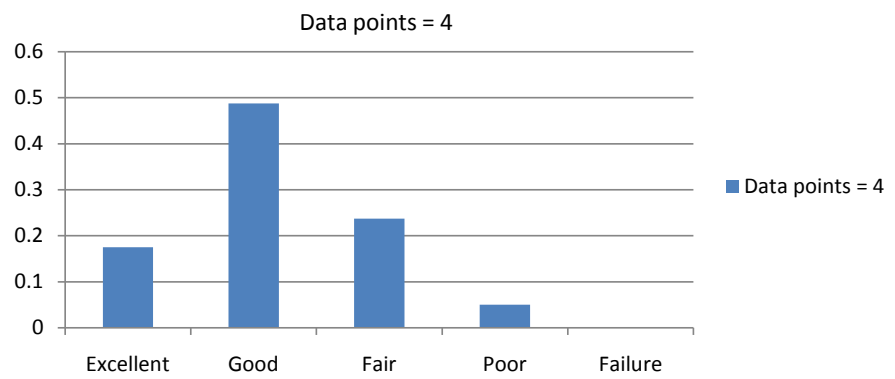
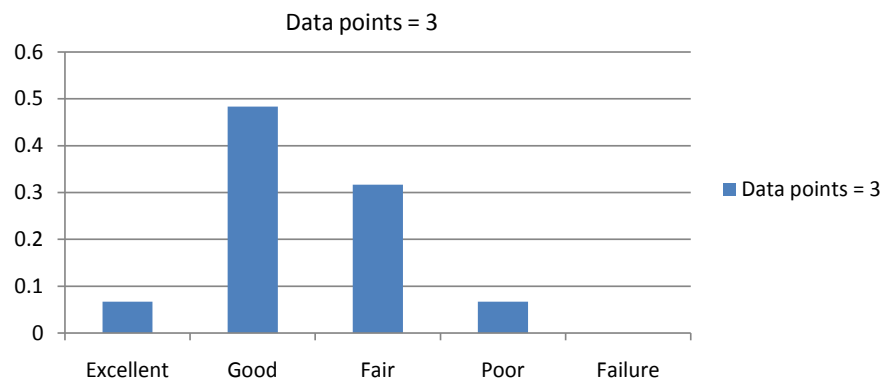
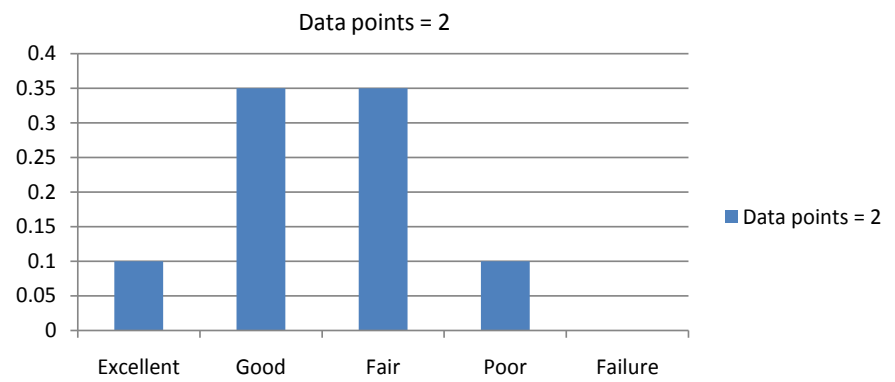
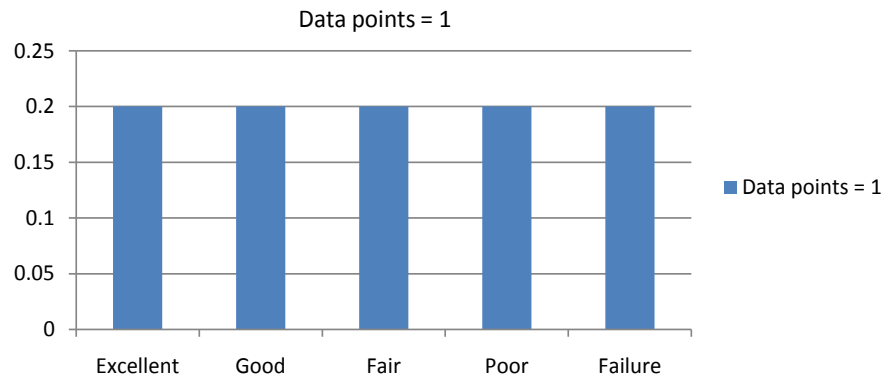
together missing variables that would be useful to add to the puzzle. Exact accuracy is not needed, but as much information as is available is helpful. An example is helpful.

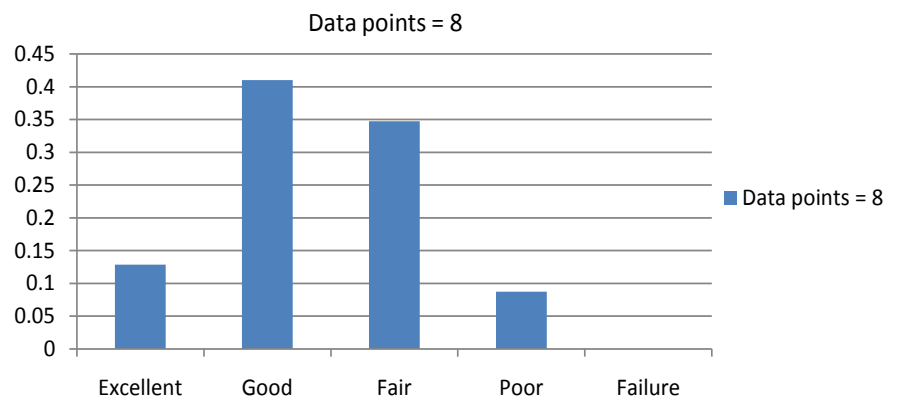
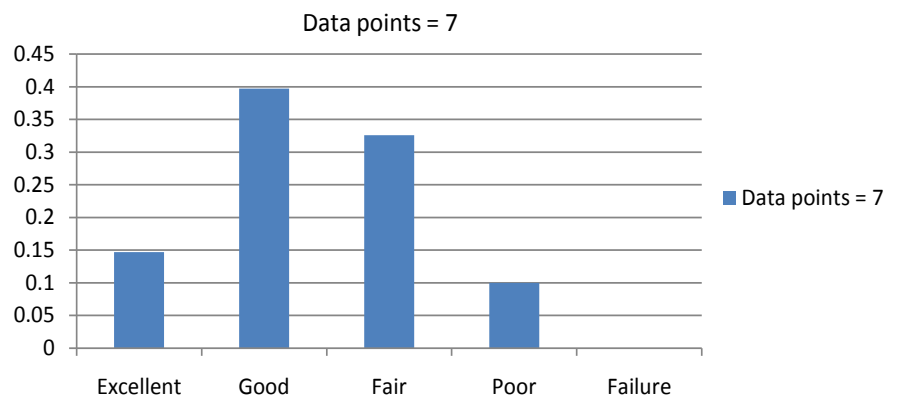
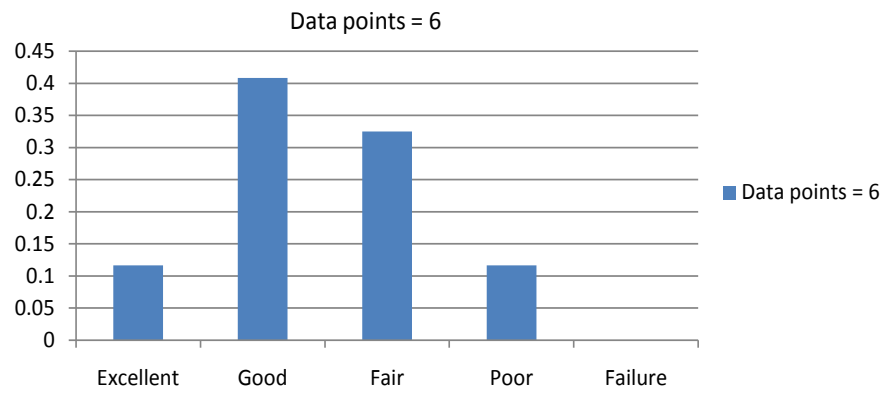
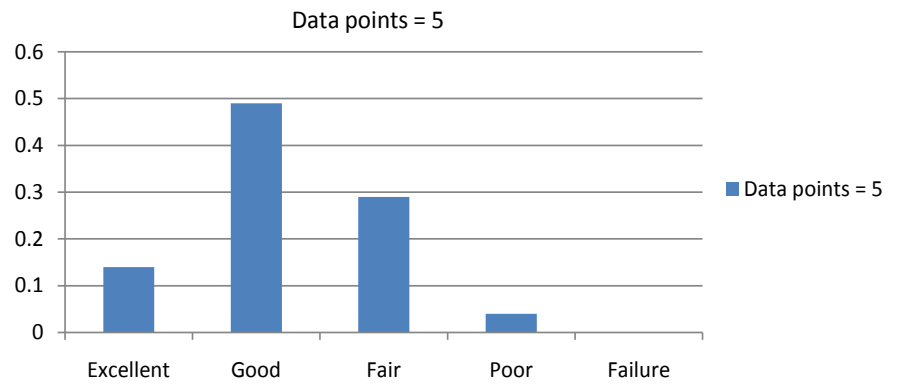
Assume that there are five arbitrary levels of condition available to analyze the asset—excellent/new, good, fair, poor and failed. If there is an asset and there is no information about it, the condition could be any one of these conditions. The probability is 0.2% or 20% for each. Assume that one data point is known—that would change the analysis considerably. Or what if the data were “sort of” known—say a probability that the asset was good or fair based on some factor? Then the probability would be altered toward the good/fair condition—less so to the poor, failed or excellent. Still there is uncertainty involved. This is precisely what Bayesian statistical methods are trying to get at. The assessor has a lot more data than one thinks even though much of it may not be known with complete certainty. The uncertainty is contained in the judgment of the assessor about certain factors.

Continuing the example, most utilities have a pretty good idea about the pipe materials. Worker memory can be very useful, even if not completely accurate. In most cases the depth of pipe is fairly similar—the deviations may be known. Soil conditions may be useful—there is an indication that that aggressive soil causes more corrosion in ductile iron pipe, and most soil information is readily available even if it is less specific per pipe or valve than desired. That can then be used as a predictive tool to help identify assets that are mostly likely to become a problem. Bloetscher *et al.* [11] are working on such an example now, but suspect that it will be slightly different for each utility. Also, in smaller communities, many variables (ductile iron pipe, PVC pipe, soil condition...) may be so similar that differentiating would be unproductive.

Construction may have altered the soils—for example muck and rock likely were replaced during construction with good fill. Likewise, tree roots will wrap around pipes, so their presence may indicate damage to the pipe. But no one can know this with certainty without digging the pipe up, something most communities would prefer to avoid. But the presence of trees is easily noted from aerials. Roads with truck traffic create more vibrations on roads, causing rocks to move toward the pipe and joints to flex. That brings up another possible variable—the field perception—what do the field crews recall about breaks? Are there work orders? If so do they contain the data needed to piece together missing variables that would be useful to add to the puzzle? With a little research there are at least 5 variables known.

Assume there are 9 variables that are developed. Each one has an assessment of adding to excellent, good, fair or poor condition of the pipe. These probabilities are added each time to build an understanding of overall condition (see Equation (4)—this is what this equation is trying to represent). **Figure 1** shows how the graph changes as more information is accumulated. There is a big change from 1 to 2 points, but notice how from 4 to 9 points the graph does not really





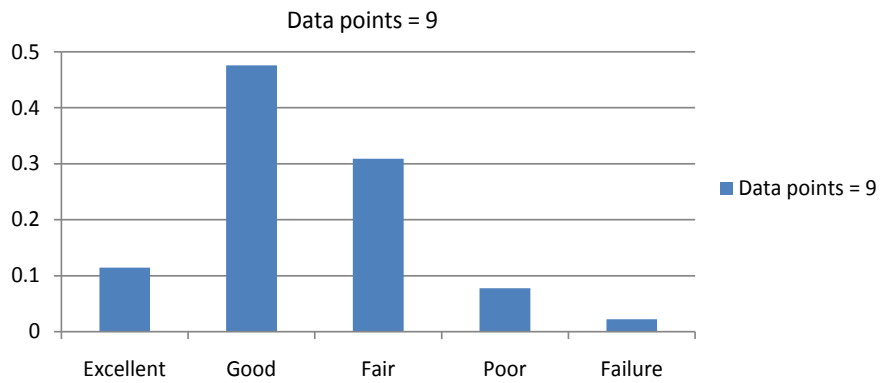


Figure 1. Distribution given additional amounts of information—note after 4 pieces of data, the distribution changes little.

change. This asset has a condition that is most probably good, maybe fair. It is probably not poor or excellent.

Consequences

Ultimately there is an interest to determine if these factors have an impact on a consequence. Determining that those consequences are is the issue, so one needs to know what that response is:

- Water main breaks?
- Sewer breaks?
- Sanitary sewer blockages or overflows?
- Stormwater system overflows?
- Roadway damage?

If the break history or sewer pipe condition is known, the impact of these factors can be developed via a linear regression model. The model would be developed

$$CI = w_1C_1 + w_2C_2 + w_3C_3 + w_4C_4 + \dots + w_iC_i \tag{5}$$

where:

- CI = Condition index
- w = weighting factor
- C is condition factor

If one knows the incident, the weights can be found:

$$f(x) = c_1x_1 + c_2x_2 + c_3x_3 + \dots + c_nx_n \tag{6}$$

where the values of C are real numbers and

$$x = [x_1 \quad x_2 \quad x_3 \quad \dots \quad x_n]^T \tag{7}$$

Are the factors line trees, materials, traffic, etc. If one assumes these constraints and linear variables in the matrices are non-negative. If there are negative values, they must be made positive as follows

$$x_i^+ = \begin{cases} x_i & \text{if } x_i \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$x_i^- = \begin{cases} -x_i & \text{if } x_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Based on the conceptual understanding of the “best guess” of data on the infrastructure, the following are the steps required to obtain a condition assessment with limited data, utilizing a series of assets gleaned from utility records for a water system for example purposes:

- Step 1: Create a table of assets (see **Table 1**, column 2-this is a small piece of a much larger table).
- Step 2: Create columns for the variables for which there is data (**Table 1**—age, material, soil type, groundwater level, depth, traffic, trees, etc.—columns 3 - 11).
- Step 3: Note that where there are categorical variables (type of pipe for example), these need to be converted to separate yes/no questions as mixing. Categorical and numerical variable do not provide appropriate comparisons; hence the need to alter the categorical variables to absence/presence variables. So descriptive variables like pipe material need to be converted to binary form—*i.e.* create a column for each material and insert a 1 or 0 for “yes” and “no” (see **Table 1**—columns 12 - 16).
- Step 4: Summarize the statistics for the variables. Note missing data is not permitted and known conditions should be entered directly (see **Table 2**).
- Step 5: Develop a linear regression to determine factors associated with each and the amount of influence that each exerts. The result will yield a series of coefficients (see **Table 3**).
- Step 6: Identify the predictive equation. In this case it is:

$$\begin{aligned} & \text{Likelihood of leaks} \\ & = 12.355 - 0.489 * \text{Dia} + 0.008 * \text{age} + 3.144 * \text{Sand} + 1.151 * \text{Lowtraffic} \\ & \quad + 5.96 * \text{heavyTraf fic} - 2.819 * \text{trees} + 0.297 * \text{trees} \\ & \quad + 0.58 * \text{Shallow bury under 6} + 0.194 * \text{presssure} + 2.34 * \text{Ductile} \\ & \quad + 5.473 * \text{GI} + 3.229 * \text{PVC} + 12.428 * \text{AC} \end{aligned} \tag{10}$$

Table 1. Summary of data known about the assets in **Table 1** so categorical variables are now presences/absence numerical variables.

Asset	breaks in 10 year	Dia	Age	Sand	Clay	Low Traffic	Heavy Traffic	Trees	no trees	shallow unde 6	deep bury	pressure	Ductile	GI	PVC	AC	HDPE
water main	17	2	45	1	0	1	0	1	0	1	0	55	0	0	0	1	0
water main	11	2	45	0	1	1	0	1	0	1	0	55	0	0	0	1	0
water main	12	2	45	1	0	1	0	1	0	1	0	55	0	0	0	1	0
water main	10	2	45	1	0	1	0	1	0	1	0	55	0	0	0	1	0
water main	2	4	50	1	0	1	0	1	0	1	0	55	1	0	0	0	0
water main	3	6	60	0	1	0	1	1	0	1	0	55	1	0	0	0	0
water main	1	6	60	0	1	0	1	1	0	1	0	55	1	0	0	0	0
water main	1	6	60	0	1	0	1	1	0	1	0	55	1	0	0	0	0
water main	0	6	20	1	0	1	0	1	0	1	0	55	0	0	1	0	0

Table 2. Summary statistics.

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Dia	93	0	93	1.000	16.000	5.011	3.255
Age	93	0	93	5.000	60.000	29.194	18.212
Sand	93	0	93	0.000	1.000	0.946	0.227
Clay	93	0	93	0.000	1.000	0.054	0.227
Low Traffic	93	0	93	0.000	2.000	0.968	0.231
Heavy Traffic	93	0	93	0.000	1.000	0.043	0.204
Trees	93	0	93	0.000	1.000	0.903	0.297
no trees	93	0	93	0.000	1.000	0.215	0.413
shallow unde 6	93	0	93	0.000	1.000	0.849	0.360
deep bury	93	0	93	0.000	1.000	0.032	0.178
pressure	93	0	93	55.000	65.000	55.323	1.616
Ductile	93	0	93	0.000	1.000	0.419	0.496
GI	93	0	93	0.000	1.000	0.054	0.227
PVC	93	0	93	0.000	1.000	0.247	0.434
AC	93	0	93	0.000	1.000	0.065	0.247
HDPE	93	0	93	0.000	5.000	1.075	2.065

Table 3. Model parameters: and statistics associated with model.

Source	Value	Standard error	t	Pr > t	Lower bound (95%)	Upper bound (95%)
Intercept	-12.355	6.805	-1.816	0.073	-25.901	1.190
Dia	-0.489	0.102	-4.795	<0.0001	-0.692	-0.286
Age	0.008	0.012	0.725	0.471	-0.015	0.032
Sand	3.144	0.891	3.528	0.001	1.370	4.918
Clay	0.000	0.000				
Low Traffic	1.151	1.546	0.744	0.459	-1.926	4.227
Heavy Traffic	5.961	2.107	2.830	0.006	1.768	10.154
Trees	-2.819	1.236	-2.280	0.025	-5.280	-0.359
no trees	0.297	1.134	0.262	0.794	-1.959	2.554
shallow unde 6	0.580	1.034	0.561	0.576	-1.478	2.639
deep bury	0.000	0.000				
pressure	0.194	0.119	1.625	0.108	-0.044	0.431
Ductile	2.342	0.640	3.658	0.000	1.068	3.617
GI	5.473	0.598	9.151	<0.0001	4.282	6.663
PVC	3.229	0.717	4.501	<0.0001	1.801	4.657
AC	12.428	0.757	16.408	<0.0001	10.920	13.935
HDPE	0.000	0.000				

Note that the variables with larger exponents generally have more impact on the number of leaks (see **Figure 2**), although the values (like age) might be a challenge.

- Step 7: The equation can then be used to predict the number of breaks going forward based on the information about breaks going back in time. **Figure 3** outlines how the equation worked for the 93 overall datapoints in this example.
- Step 8: Finally the data can be used to predict where the breaks might occur in the future based on the past (**Figure 4**).

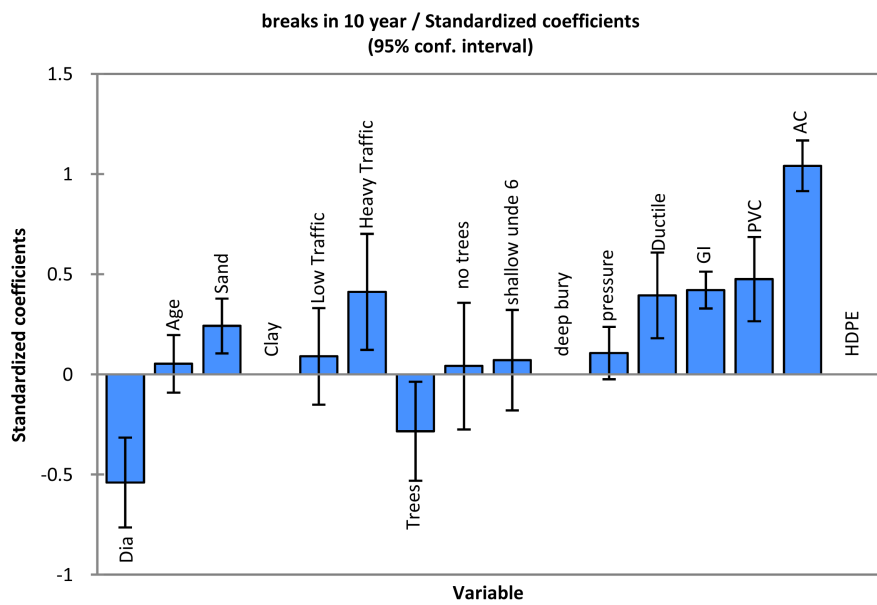


Figure 2. Impact of factors on leaks.

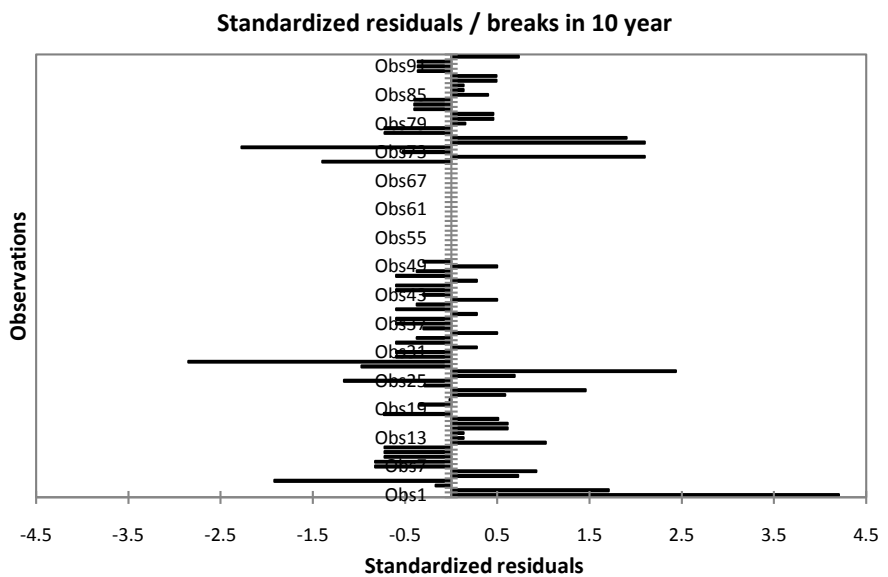


Figure 3. Residual breaks over 10 years based on the observation (asset).

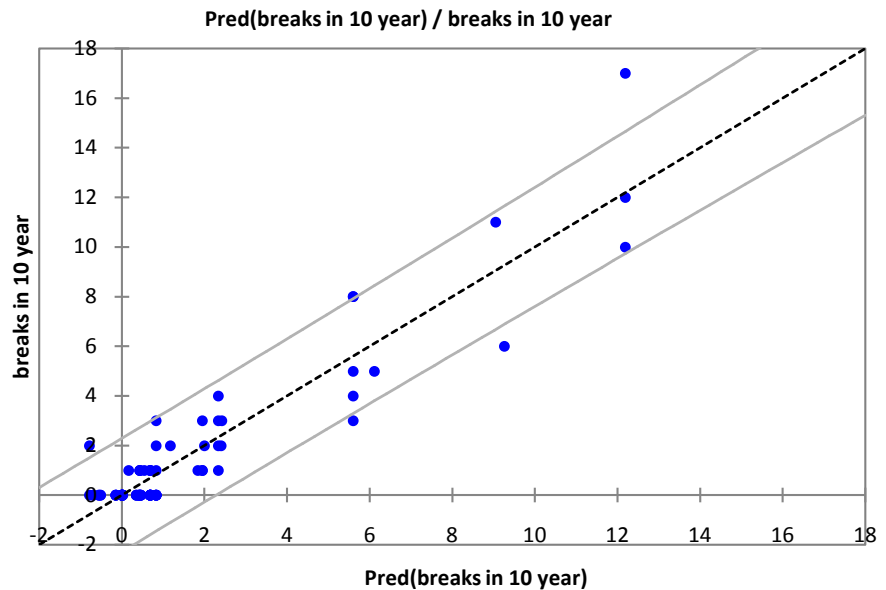


Figure 4. Comparison of predictive and actual breaks over 10 years (correlation desirable).

The hope is that these correlate well. The process is not time consuming but provides useful information on the system. It needs to be kept up as things change, but exact data is not really needed and none of this requires destructive testing.

3. Results

Conducting an exercise to develop the methodology was useful, but the next step was to do something to with the results. The Dania Beach, FL sewer system was used as an example given that actual data of failures (pipe breaks from sewer leak data) existed and an understanding of the system was available. The City of Dania is approximately 7.7 square-miles. The analyzed network includes approximately 1500 assets located within the public right-of-way (ROW). Two asset maps were acquired to aid in the data collection. The first map depicted the system as it existed in the early 2010s. The original intention of this map was to illustrate which pipes in the network were suspected of breakage based on a midnight monitoring exercise after sealing the system [11]. The original analysis utilized flow data to track areas within the network most likely to be suffering from infiltration. Once this analysis had been completed, the sewer lines in question were televised and the found breaks were recorded.

The original installation design records were obtained. Dates and materials were assigned for large sections of the City. Most of the pipe was vitrified clay installed in the 1960s and 1970s so an estimated install date within 5 years was assigned since the expected life of sewer pipe assets is expected to range from 80 to 100-years and ± 5 years was not deemed to be significant for the purposes of this analysis. Many other indicators of failure, for example, pipe diameter,

groundwater, soils, traffic, trees and pipe depth were included. A Geographic Information System (GIS) provided spatial analytics and asset data management, while Excel provided independent asset analysis and comparative analytics. Soils maps from the US Department of Agriculture and contractor and information developed by the authors was used to help with groundwater and soils data.

From the map created in **Figure 5**, a table of assets was created (see **Table 4** which shows a portion of this table). Columns were added for the variables for which there was data (age, material, soil type, groundwater level, depth, traffic, trees, etc.). For each asset, a value was assigned to the desired and undesirable attributes for each indicator of failure. For example, pipes greater than thirty-inches below surface grade were given a value of 1 and pipes with less than thirty-inches of cover were assigned a value of 0. These values were assigned independently to each asset for each indicator of failure analyzed. It is important to note that the values assigned do not provide a weight, or indicate importance. Categorical variables were converted to numerical ones. **Table 5** shows the summary statistics.

The statistical analysis tool XLStat® was utilized in the data analysis. XLStat® requires that each variable be represented as a numeral. **Table 6** shows the correlation matrix and **Table 7** the analysis of variance (ANOVA) results. One issue is that most of the pipe is vitrified clay, and the soils are similar in most of the City so there were not good differentiators. Likewise, little gravity pipe is under



Figure 5. GIS of the sewer system in Dania Beach.

Table 4. Assets for sanitary sewer system (partial).

Asset	Given Asset ID	Material	Age	Traffic Loading	Diameter	Depth	Length	Pipe Breaks
8" Gravity SS	1	1	0	0	0	1	1	1
8" Gravity SS	2	1	0	0	0	1	1	1
8" Gravity SS	3	0	1	0	0	1	1	0
8" Gravity SS	4	0	1	0	1	1	1	0
8" Gravity SS	5	1	0	0	1	1	1	0
8" Gravity SS	6	0	1	0	1	1	1	0
8" Gravity SS	7	0	1	0	1	1	1	0
8" Gravity SS	8	0	1	0	1	1	1	0
8" Gravity SS	9	0	1	0	1	1	1	0
8" Gravity SS	10	0	1	0	1	1	1	0
8" Gravity SS	11	0	1	0	0	1	1	0
8" Gravity SS	12	0	1	0	0	1	1	1
8" Gravity SS	13	0	1	0	0	1	1	1
8" Gravity SS	14	0	1	0	0	1	1	0

Table 5. Summary statistics for sewer system (complete table).

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Pipe Breaks	851	0	851	0	1	0.146	0.353
Material	851	0	851	0	1	0.831	0.375
Age	851	0	851	0	3	1.231	1.033
Traffic Loading	851	0	851	0	2	0.157	0.491
Diameter	851	0	851	0	1	0.114	0.318
Depth	851	0	851	0	1	0.155	0.362
Length	851	0	851	0	1	0.927	0.26

Table 6. Correlation matrix for sanitary sewer system (complete table).

Variables	Material	Age	Traffic Loading	Diameter	Depth	Length	Pipe Breaks
Material	1	0.101	0.119	-0.104	-0.603	-0.102	0.035
Age	0.101	1	-0.019	-0.166	-0.08	-0.099	-0.102
Traffic Loading	0.119	-0.019	1	0.556	0.008	-0.057	-0.038
Diameter	-0.104	-0.166	0.556	1	0.204	0.001	0.03
Depth	-0.603	-0.08	0.008	0.204	1	0.095	-0.011
Length	-0.102	-0.099	-0.057	0.001	0.095	1	0.116
Pipe Breaks	0.035	-0.102	-0.038	0.03	-0.011	0.116	1

Table 7. Analysis of variance for sewer system.

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	6	3.101	0.517	4.242	0
Error	844	102.831	0.122		
Corrected Total	850	105.932			

Computed against model $Y = \text{Mean}(Y)$.

heavy traffic areas so this was also not a good identifier. **Figure 6** is a varimax graph developed from principal component analysis of the data. **Figure 6** shows the correlation of any individual factor in relation to all other factors within the graph. Data points with a higher correlation have a smaller degree of separation within the chart; a probable correlation can be predicted if the factors are within 45° of each other. Therefore, the two factors of failure which are most closely correlated with pipe breakage are length and depth. Using this data, the weight of importance can be determined to complete the condition index.

A linear regression formula was developed from the factors and the amount of influence that each exerts to yield the predictive equation. In this case it is:

$$\begin{aligned} \text{Likelihood of breaks} = & 0.073 * \text{material} + 0.089 * \text{age} + 0.078 * \text{Loading} \\ & + 0.066 * \text{diameter} + 0.003 * \text{depth} + 0.011 * \text{length} \end{aligned} \quad (11)$$

This equation can then be used to predict the number of breaks using the consequences going back in time. **Table 7** shows the analysis of variance. The fit was good, but note the lack of variability in the variables. **Figure 7** shows the variance from actual to estimated breaks. Most pipes in good condition was noted as good in **Figure 8**. Note one issue here that only 1 break was noted in the pipe—there actually may have been more breaks but these could not be noted, an issue that requires follow-up video.

4. Conclusions

Many utilities have not implemented comprehensive asset management plans for their assets. In part, this is due to the belief that they cannot properly assess the assets or the cost to do so from traditional methods is too expensive or yields data of limited value. As a result, they have limited data to present to decision-makers about the condition of their assets, and the likelihood of failure, creating an atmosphere of hoping to avoid catastrophic failures. However, utilities with limited financial capability, and who might be most at risk if failure occurs, can develop an asset management program to help identify critical risks and provide data to decision-makers who need to provide the fiscal resources to properly manage and maintain a utility system.

In this exercise, an effort was made to develop a methodology to evaluate utility assets, buried and otherwise, to help identify financial resources needed to maintain a utility system. The concept was to create data on the assets for a condition assessment (buried pipe is not visible and in most cases, cannot really be

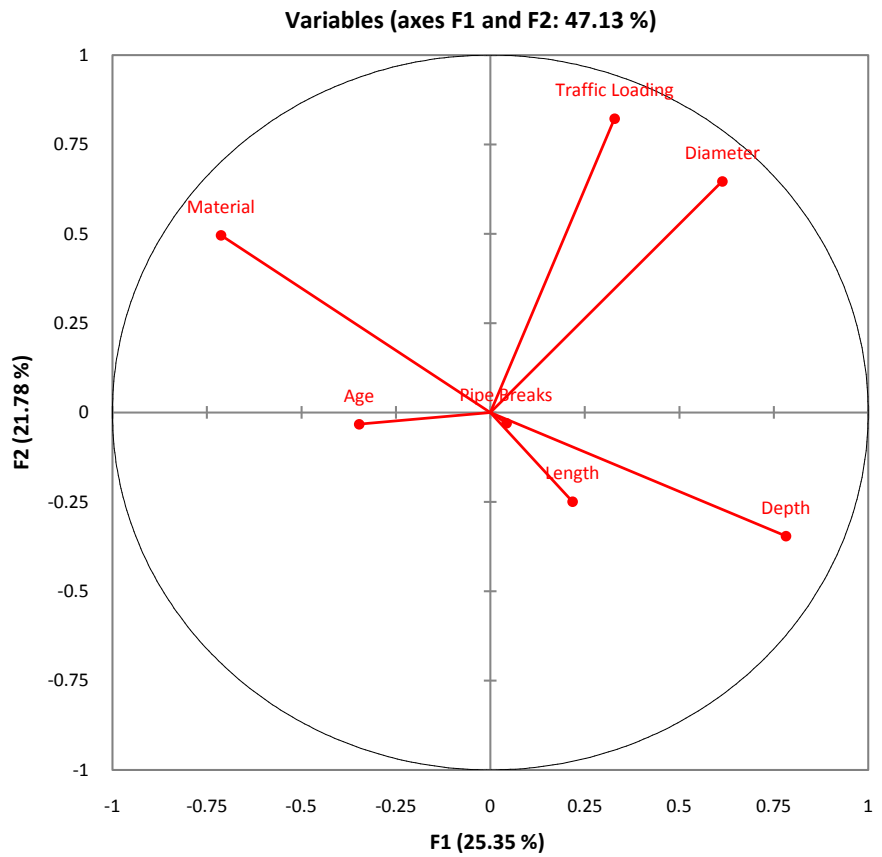


Figure 6. Factor graph indicating where there may be correlation between variables.

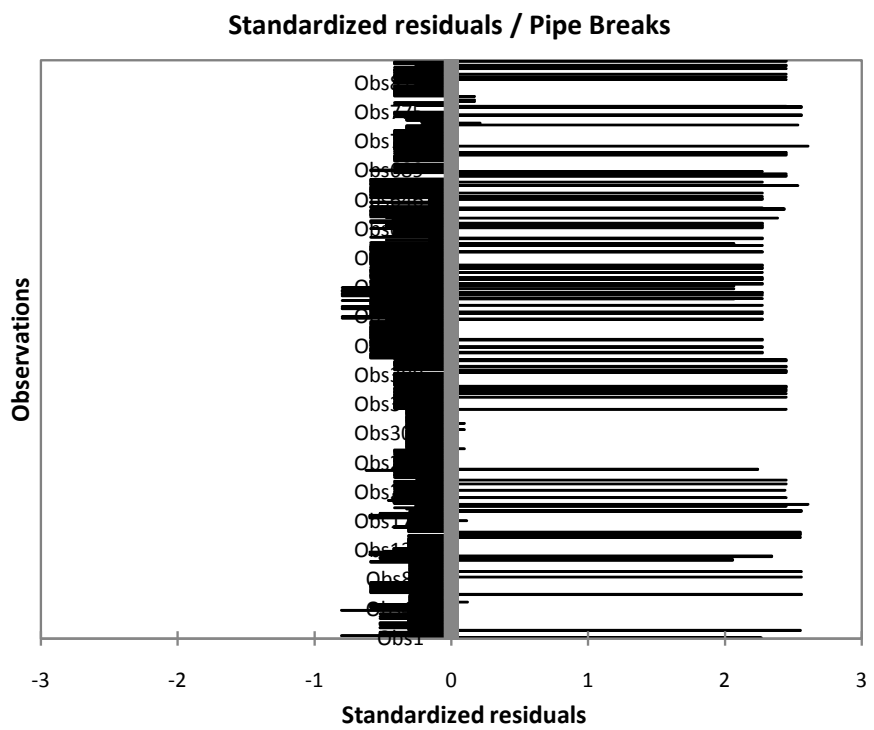


Figure 7. Variance from actual to estimated breaks.

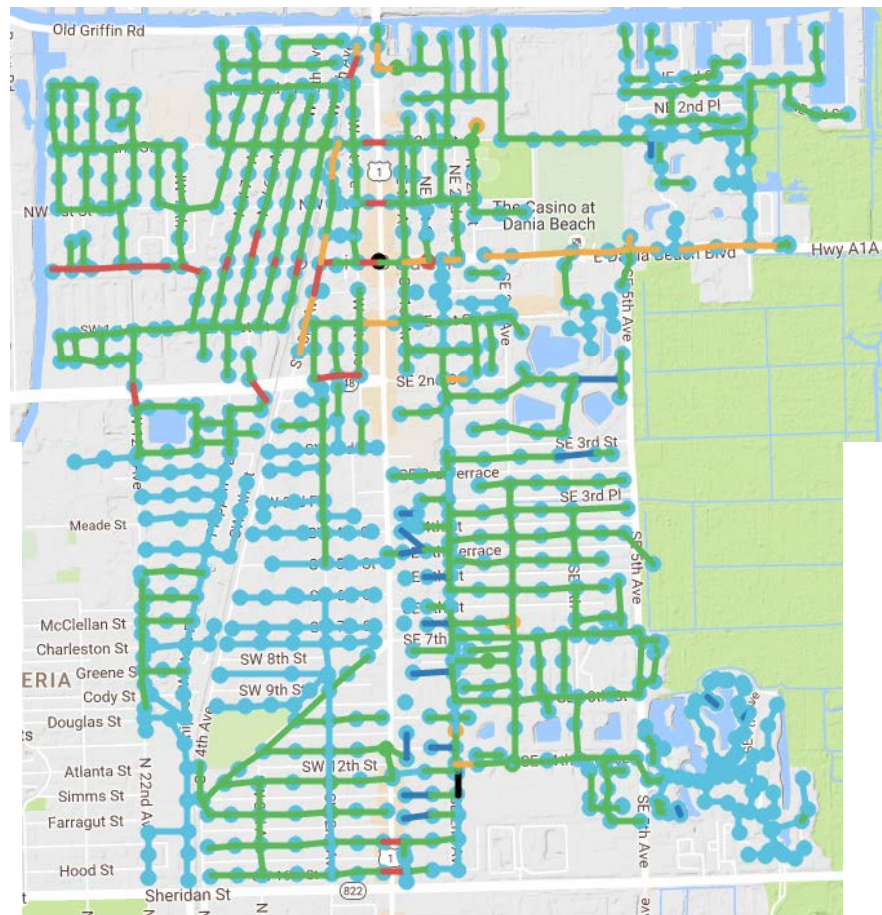


Figure 8. Shows that most pipes in good condition were noted as good. The failure (15% of the system) was less reliable.

assessed). A challenge is posed with buried infrastructure since many utilities lack the resources for examining buried infrastructure, so other methods of data collection are needed. However, much more information is known about buried infrastructure than one anticipates. This permits an assessment of likelihood of condition, using the parameters of Bayesian statistical methods applied in the field.

For predictive methods to work there needs to be a measurable consequence to be useful for predicting future maintenance needs—breaks, flooding etc.—that would indicate a failure. Unfortunately, many utilities do not do this or do not collect this data from work orders (if they use work orders). The lack of tracking makes it difficult to determine which factors are the critical ones. In this project, the effort was applied to a sewer system since that system tracked pipe damage. Pipe breaks is a consequence, but the number of breaks was found to be of greater value.

References

- [1] Aschauer, D. (1989) Public Investment and Productivity Growth in the Group of

Seven? *Economic Perspectives*, **13**, 17-25.

- [2] Munnell, A.H. and Cook, L.M. (1990) How Does Public Infrastructure Affect Regional Economic Performance? Federal Reserve Bank of Boston, Boston.
<https://www.bostonfed.org/-/media/Documents/conference/34/conf34c.pdf>
- [3] Barro, R. (1990) Government Spending in a Simple Model of Endogenous Growth. *Journal of Political Economy*, **98**, 102-125. <https://doi.org/10.1086/261726>
- [4] Congressional Budget Office (2015) Public Spending on Transportation and Water Infrastructure, 1956 to 2014, CBO, Washington, DC.
- [5] ASCE (2001) 2001 Report Card for America's Infrastructure, ASCE, Alexandria.
<http://ascelibrary.org/doi/book/10.1061/9780784478882>
- [6] ASCE (2005) 2005 Report Card for America's Infrastructure, ASCE, Alexandria
<http://ascelibrary.org/doi/book/10.1061/9780784478851>
- [7] ASCE (2009) 2009 Report Card for America's Infrastructure, ASCE, Alexandria.
<http://www.infrastructurereportcard.org/making-the-grade/report-card-history/2001-report-card/>
- [8] ASCE (2013) 2013 Report Card for America's Infrastructure, ASCE, Alexandria.
<http://www.infrastructurereportcard.org/>
- [9] ASCE (2017) 2017 Report Card for America's Infrastructure, ASCE, Alexandria.
<http://www.infrastructurereportcard.org/making-the-grade/report-card-history/2001-report-card/>
- [10] David, G. (2010) An Important Tool for Asset Management. *Utility Infrastructure Management: Journal of Finance and Management for Water and Wastewater Professional*.
<http://mail.uimonline.com/index/webapp-stories-action/id.378/archive.yes/Issue.2010-04-01/title.rehabilitation>
- [11] Bloetscher, F., Wander, L., Smith, G. and Dogan, N. (2016) Developing Asset Management Systems—You Know More Than You Think. *2016 Florida Section AWWA conference*, Orlando, November 27-December 1 2016.
- [12] Bloetscher, F. (2011) *Utility Management for Water and Wastewater Operators*. AWWA, Denver.
- [13] Bloetscher, F., Englehardt, J.D., Chin, D.A., Rose, J.B., Tchobanoglous, G., Amy, V.P. and Gokgoz, S. (2005) Comparative Assessment Municipal Wastewater Disposal Methods in Southeast Florida. *Water Environment Research*, **77**, 480-490.
<https://doi.org/10.2175/106143005X67395>
- [14] Haas, C.N., Rose, J.B. and Gerba, C.P. (1999) *Quantitative Microbial Risk Assessment*. John Wiley and Sons, New York.
- [15] Zimmermann, H.-J. and Zysno, P. (1980) Latent Connectives in Human Decision Making. *Fuzzy Sets and Systems*, **4**, 37-51.
[https://doi.org/10.1016/0165-0114\(80\)90062-7](https://doi.org/10.1016/0165-0114(80)90062-7)
- [16] Kullback, S. (1978) *Information Theory and Statistics*. Peter Smith, Gloucester.
- [17] Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana.
- [18] Shannon, C.E. (1949) A Mathematical Theory of Communication. *The Bell System Technical Journal*, **27**, 379-656. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [19] Englehardt, J. and Lund, J. (1992) Information Theory in Risk Analysis. *Journal of Environmental Engineering, American Society of Civil Engineers*, **118**, 890-904.

[https://doi.org/10.1061/\(ASCE\)0733-9372\(1992\)118:6\(890\)](https://doi.org/10.1061/(ASCE)0733-9372(1992)118:6(890))

- [20] Englehardt, J.D. (1995) Predicting Incident Size from Limited Information. *Journal of Environmental Engineering*, **121**, 445-464.
[https://doi.org/10.1061/\(ASCE\)0733-9372\(1995\)121:6\(455\)](https://doi.org/10.1061/(ASCE)0733-9372(1995)121:6(455))
- [21] John, A. and Dunsmore, I.R. (1975) *Statistical Prediction Analysis*. Cambridge University Press, Cambridge.



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ojce@scirp.org