

Review

Sewer Pipes Condition Prediction Models: A State-of-the-Art Review

Mohammadreza Malek Mohammadi ^{1,*}, Mohammad Najafi ², Vinayak Kaushal ² ,
Ramtin Serajiantehrani ², Nazanin Salehabadi ³ and Taha Ashoori ⁴

¹ Alan Plummer Associate, Inc., 1320 S University Dr # 300, Fort Worth, TX 76107, USA

² Center for Underground Infrastructure Research and Education (CUIRE), Department of Civil Engineering, The University of Texas at Arlington, Box 19308, Arlington, TX 76019, USA; Najafi@uta.edu (M.N.); vinayak.kaushal@mavs.uta.edu (V.K.); ramtin.serajiantehrani@mavs.uta.edu (R.S.)

³ Department of Computer Science and Engineering, The University of Texas at Arlington, Box 19308, Arlington, TX 76019, USA; nazanin.salehabadi@mavs.uta.edu

⁴ EnTech Engineering P.C., 17 State Street, 36th Fl, New York, NY 10004, USA; taha.ashoori@mavs.uta.edu

* Correspondence: mohammadreza.malekmohammadi@mavs.uta.edu; Tel.: +1-(469)-243-3736

Received: 4 September 2019; Accepted: 18 October 2019; Published: 22 October 2019



Abstract: Wastewater infrastructure systems deteriorate over time due to a combination of aging, physical, and chemical factors, among others. Failure of these critical structures cause social, environmental, and economic impacts. To avoid such problems, infrastructure condition assessment methodologies are developing to maintain sewer pipe network at desired condition. However, currently utility managers and other authorities have challenges when addressing appropriate intervals for inspection of sewer pipelines. Frequent inspection of sewer network is not cost-effective due to limited time and high cost of assessment technologies and large inventory of pipes. Therefore, it would be more beneficial to first predict critical sewers most likely to fail and then perform inspection to maximize rehabilitation or renewal projects. Sewer condition prediction models are developed to provide a framework to forecast future condition of pipes and to schedule inspection frequencies. The objective of this study is to present a state-of-the-art review on progress acquired over years in development of statistical condition prediction models for sewer pipes. Published papers for prediction models over a period from 2001 through 2019 are identified. The literature review suggests that deterioration models are capable to predict future condition of sewer pipes and they can be used in industry to improve the inspection timeline and maintenance planning. A comparison between logistic regression models, Markov Chain models, and linear regression models are provided in this paper. Artificial intelligence techniques can further improve higher accuracy and reduce uncertainty in current condition prediction models.

Keywords: sewer condition prediction; sewer pipe prioritization; deterioration models; pipe condition assessment; asset management

1. Introduction

U.S. infrastructure plays a critical role in urban communities, providing for the safe and efficient conveyance of water, sewer, gas, and other lifelines to protect human health and the environment. In the U.S., buried pipelines span thousands of miles and form a significant part of the total U.S. infrastructure [1]. Sanitary sewers, as a part of wastewater infrastructure systems, are designed to collect sewage from domestic, industrial, and commercial users and convey to treatment plants. Most sewer systems are gravity sewers, which transfer the flow based on a slope. There are over 800,000 miles of public sewer pipes and 500,000 miles of private sewer laterals in the United States.

Approximately 240 million Americans are connected to 14,748 treatment plants for wastewater treatment. By 2032, it is estimated that 56 million more people will use centralized treatment plants [2].

The majority of the U.S. wastewater infrastructure is over 100 years old and the combination of aging, chemical, and environmental factors cause at least 23,000 to 75,000 sanitary sewer overflows per year [3,4]. The latest infrastructure report card, published by the American Society of Civil Engineering (ASCE) in 2017, states a “D plus” grade for the wastewater infrastructure. ASCE indicated that water and wastewater systems in the U.S. are clearly aging and to keep up with the needs a capital funding gap of \$150 billion will be needed by 2025 [2]. Furthermore, the U.S. population is increasing and shifting geographically. This requires investment for new infrastructure and maintenance of existing infrastructure in areas of decreasing population with limited budgets [5].

Deterioration of sewer pipes is very complex process and several factors affect the condition of pipes rather than just a single factor. Thus, predicting the failure time of sewer pipes is a difficult task. Ana and Bauwens (2010) suggested that the best way to forecast pipe failure and deterioration time is development of probability-based condition prediction models based on actual inspection database [6].

According to AWWA (2012), municipalities spend a relatively smaller investment for sewer rehabilitation rather than expanding sewer systems to meet growth and treatment plant upgrades [7]. As sewer systems become older, the structural and operational performance degrade. The aging of sewer pipes increases the failure rates and can result in social, environmental, and economic impacts, such as water quality issues including chemical or biological contaminations, which may cause illness and extensive repair costs [8].

Maintenance and rehabilitation strategies are very important factors to keep the performance of the system at an acceptable level of service and to provide cost-effective solutions for avoiding unforeseen failures. In the past, repair or rehabilitation of sewer pipes were only done once a pipe collapsed or failed. However, the current trend is to maintain and manage pipe systems before failure time. To achieve this goal, municipalities and utilities have begun to implement asset management systems. Infrastructure asset management is a comprehensive and cost-effective tool to maintain pipeline system at desired conditions. Asset management programs can develop various strategies to help utility companies and municipalities to understand the timing and associated costs of maintenance, rehabilitation, or replacement of the pipes [9].

One of the main components of asset management system is condition assessment. Typically, inspection techniques are used to identify different type of defects along the pipe wall, and condition rating standards are employed to determine the condition status of sewer pipes. It is obvious that monitoring and inspection of all sewer pipes is almost impossible due to limited budget, time, and assessment technologies. Therefore, more attention is needed to develop deterioration models than can predict the current and future condition of sewer pipelines.

The objective of this paper is to present progress acquired over years in development of condition prediction models for sewer pipes and compare those models. Published papers were identified from various databases such as ProQuest, Engineering Village, ASCE Database, and Google Scholar that discussed prediction models over a period from 2001 through 2019. This paper illustrates and studies the most common statistical models used in predicting deterioration and condition states of sewer pipes.

2. Sewer Pipe Deterioration

Pipe systems require continuous inspection and maintenance, and risk of pipe deterioration can be raised if the asset management and condition assessment of the pipes are neglected. Najafi and Gokhale (2005) categorized pipe failures in two main structural and operational failures [10]. EPA (2009) grouped wastewater pipe failures into three categories, namely hydraulic restrictions (blockages), hydraulic capacity, and structural deterioration [11]. Opila (2011) considered both water and wastewater and classified failure modes into structural, operations, and maintenance, such as, hydraulic capacity, economic, and water quality [8].

According to previous studies and deterioration models, overall the mechanisms of sewer pipe deterioration can be generalized into structural, operational, and hydraulic capacity failure. Structural failure is caused by any kind of defects on pipe wall that reduce the structural integrity of pipe segment. Similarly, the soil surrounding the pipe has an essential role to failure time of pipes. In general, cracks, internal and external corrosion, pipe deflection, misaligned joints, and breaks are the most common type of defects associated with structural failure [11]. Operational failure is the most common failure in wastewater collection systems and generally occurs by a physical cause and can be resolved during a maintenance procedure and normally does not affect the structural integrity of the pipe.

Several type of defects, such as, debris, infiltration, root intrusion, sediment accumulation, obstruction, and grease build-up fall within operational failure category [8,11]. Hydraulic capacity failure occurs when flow is higher than pipe capacity. In other words, the pipe segment does not have adequate capacity to convey wastewater, without having any structural or operational problem. Hydraulic capacity failure may be the result of infiltration/inflow (I/I), where the groundwater and storm water enter the sewer system through connections, manholes, cracks, and defects. Hydraulic capacity failure is often a sign of other type of structural defects such as cracks, broken pipe, leaks, and other factors.

3. Factors Affecting Deterioration of Sewer Pipes

In recent years, numerous efforts have been exerted to evaluate the condition of sewer pipelines and to find the factors that influence deterioration and remaining useful life of sewer systems. Davies et al. (2001) provided a comprehensive review of previous studies on the factors that influence structural deterioration of rigid pipes and categorized them into three groups of pipe construction, operational, and environmental factors [12]. In other research, Al Barqawi and Zayed (2006) classified the factors that influence deterioration of water pipes into three categories of physical, environmental, and operational factors [13]. Ana and Bauwens (2010) considered physical factors, environmental factors, operational factors, and construction factors for sewer structural deterioration [6].

In general, agencies and municipalities know physical factors of pipes, but environmental and operational factors are often unavailable because collecting this kind of data is costly and time consuming. According to Kley and Caradot (2013), identification of influencing factors is very important due to decreasing number of required data during the data collection, reducing cost, and achieving high prediction accuracy for development of prediction models [14]. Table 1 provides a summary of features used in previous studies from 2000 to 2018 to identify the important factors affecting deterioration of sewer pipes.

Table 1. Factors affecting sewer pipe deterioration. (Adapted from Al Barqawi and Zayed, 2006).

Physical Factors	Environmental Factors	Operational Factors
Connections	Backfill type	
End invert elevation	Bedding material	
Installation method	Ground movement	Blockages
Joint type	Groundwater level	Burst history
Pipe length	pH	Debris
Pipe shape	Road type	Flow velocity
Pipe slope	Root interference	Hydraulic condition
Sewer age	Soil corrosivity	Infiltration/exfiltration
Sewer depth	Soil fracture potential	Previous maintenance
Sewer pipe material	Soil moisture	Sediment level
Sewer size	Soil type	Sewer function
Start invert elevation	Sulfate soil	Surcharge
	Surface type	

4. Deterioration Models for Sewer Pipes

Prediction models can perform an essential role to generate a comprehensive asset management program as they provide valuable information to forecast short-term and long-term behavior of sewer pipes. Most of the previous studies used the CCTV inspection data and condition rating standards to provide sewer deterioration models.

Several models have been developed in previous studies to predict deterioration or remaining useful life sewer pipes based on different condition rating standards and deterioration factors. Deterioration models can be used to predict condition rating of a sewer pipe by using information obtained from inspection databases. In general, utility companies and municipalities can forecast the future condition of their assets by generating deterioration models to identify the pipes that require maintenance, rehabilitation, and replacement. The ultimate goal of many prediction models is to apply an appropriate mathematical technique to forecast the condition state of sewer pipes with the highest accuracy result. The current condition of sewer pipelines is often assessed through inspection techniques, however understanding the future condition of pipe systems needs a comprehensive deterioration model.

Deterioration models for sewer pipelines are classified into different categories. Morcouc and Lounis (2004) as well as Kley and Caradot (2013) divided the deterioration models into deterministic, probabilistic, and soft computing methods [14,15]. Yang (2004) used physical, artificial intelligence-based, and statistical categories for condition prediction [16]. Tran (2007) suggested deterministic and statistical models as a model-driven type and artificial intelligence-based models as a data-driven type [17]. Additionally, Altarabsheh (2016) classified the deterioration models into deterministic and probabilistic models [18]. It is obvious that accuracy of model prediction is highly dependent on selecting the proper modelling techniques for sewer deterioration [17]. In general, existing sewer deterioration models can be classified into two groups of statistical models and artificial intelligence models, as shown in Figure 1.

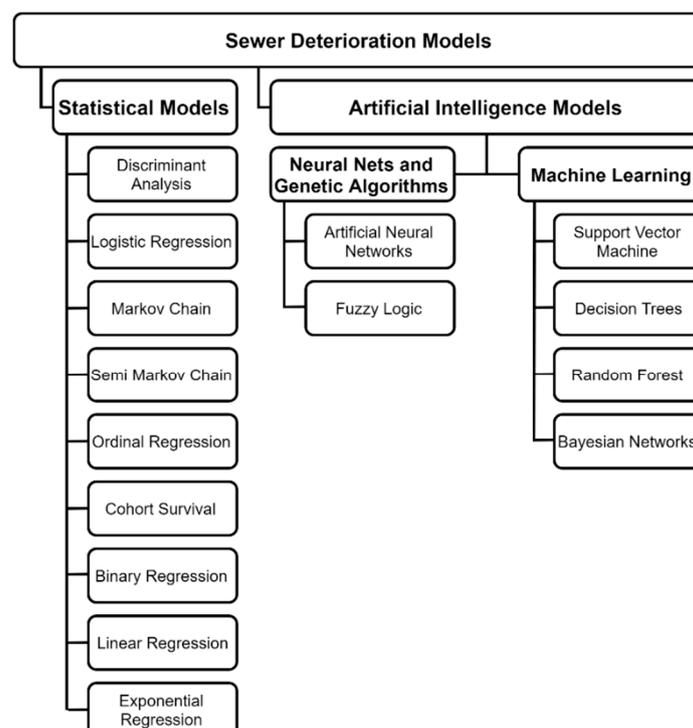


Figure 1. Classification of sewer deterioration models.

5. Statistical Models

The basic explanation of a statistical model is a random variable X , which represents a quantity whose outcome is uncertain. In statistical models, the probabilistic nature of historical data is used to describe the model output as a random variable. In any statistical analysis, estimates are “best guesses” based on the condition of given historical data [19]. Dasu and Johnson (2003) indicated that parametric density function is used in statistical models to measure the errors and identify probabilistic relationships between dependent and independent variables [20]. According to Tran (2007), predicting the ordinal data type and considering the probabilistic nature of the underlying deterioration process can be the advantages of statistical models [17], while the sensitivity of statistical models to noisy data and the methodologies to measure the errors are disadvantages of these models. Numerous statistical models, such as linear regression, exponential regression, logistic regression, Markov chain, Semi-Markov chain, ordinal regression, and cohort survival models were used to predict the condition of sewer pipes in previous studies. In particular, three different statistical models are discussed in this paper: (1) The linear regression model; (2) the Markov-chain-based models; and (3) and the logistic regression model.

5.1. Linear Regression Models

5.1.1. Model Description

The simplest linear regression model involves only one independent variable, and the dependent variable can be predicted based on their relationship. The regression model states that the true mean of the dependent variable changes at a constant rate as the value of independent variable increases or decreases. Therefore, the equation of a straight line shows the function relationship between the true mean of Y_i and X_i as shown in Equation (1) [21].

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (1)$$

where i is facility index, Y_i is dependent variable for facility i , β_0 , and β_1 are parameters to be estimated, X_i is independent variable, and ϵ_i is random error term. Multiple linear regression can be used to predict condition of sewer pipe with consideration of more than one independent variables. When deterioration of sewer pipes is modeled, condition state of the pipe is the dependent variable and independent variables contain pipe attributes such as pipe age, material, length, slope, and other environmental and operational factors. As the condition states of the sewer pipes are discrete values, the linear regression may have trouble predicting the categorical variables.

5.1.2. Previous Studies

Chughtai and Zayed (2007a, 2007b, and 2008) used the multiple regression technique to predict the deterioration mechanisms of sewer pipelines. Various factors, such as, pipe material, depth, length, age, diameter, bedding, road type, and slope were considered as independent variables to build the model. The best subset analysis was used to select important variables in this paper. The significance of the variables was investigated by different statistical test including F-test, t-test, and residual analysis, lack of fit test, and Durbin-Watson test. Four regression models were developed to predict the condition of concrete, asbestos, cement, and PVC pipes. The result showed 72% to 88% accuracy, and they suggested inspection priority should be given to the pipes with extremely steep bed slopes [22–24].

Bakry et al. (2016a, 2016b) used a regression analysis technique to develop a condition prediction model for sewer pipes, which had been rehabilitated before by CIPP method. The data were obtained from closed-circuit television (CCTV) inspection reports of Quebec CIPP rehabilitations. Various physical, operational, and environmental factors were used to generate the models. The regression models were validated using coefficient of multiple determinations and the result revealed range between 80% and 97%. In addition, the accuracy of the models was determined by calculating mean

absolute error and root mean square error. Linear deterioration curves were developed in this paper by examining the effect of increasing the age while changing the dependent variables [25,26].

5.1.3. Model Discussion

The major advantage of linear regression is the simplicity of the model. Furthermore, the relationship between dependent and independent variables can be easily translated. However, the linear regression model is too simplistic to display the probabilistic nature of pipe deterioration [17,27–29]. In addition, the condition states of sewer pipes are typically described as discrete values, and linear regression is not an appropriate model for classification and forecasting the categorical variables. In linear regression models, the result is obtained from the relationship between mean values of dependent and independent variables and sometimes they are not enough strong for models with multiple input variables. Moreover, this model is very sensitive to outliers. In general, application of linear regression is not suggested to develop a condition prediction model for sewer pipes.

5.2. Markov Chain Models

5.2.1. Model Description

The Markov chain was developed by Andrei Markov in 1906 as a discrete-time stochastic process. A Markov chain is a mathematical model of a random phenomenon over a unit of time to predict the future based on the present values and regardless of the past effects. The time can be discrete, continuous, or ordered set [30]. The Markov chain-based deterioration model assumes that conditional probability does not change over time and for all states i and j and all t , probability is independent of time as shown in Equation (2) [31].

$$P(X_{t+1} = j | X_t = i) = p_{ij} \tag{2}$$

where P_{ij} is the transition probability that, given the system in state i at time t , will be at state j at time $(t + 1)$. Generally, the transition probability matrix ($m \times m$ matrix) is used to calculate the transition probabilities. For example, consider a set of pipe state condition, $C = \{C1, C2, C3, C4, C5\}$. When a sewer pipe is in condition 1, a series of probabilities $P11, P12, P13, P14,$ and $P15$ determine the condition state of pipe in the next period. The deterioration process starts in one of the states and moves from one to another. If the sewer pipe is currently in condition $C3$, it moves to condition $C4$ in the next step with a probability of $P34$. This probability is called transition probability and only takes into account the current condition of pipe, without considering the historical data and previous conditions. The transition probability matrix is given in Equation (3).

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{bmatrix} \tag{3}$$

Then, the probability of being in different states at time $t + 1$ can be estimated by total probability theorem, as shown in Equation (4).

$$P_j^{t+1} = \sum_{i=1}^j P_{ij} \times P_i^t \tag{4}$$

where P_i^t is the probability of being in state i in year t [32]. Once the probability matrix is identified, the future condition of pipes can be easily obtained by Markov model.

5.2.2. Previous Studies

Extensive studies have been carried out to predict the deterioration of sewer pipes by developing Markov chain models. Wirahadikusumah et al. (2001) used Markov-chains-based models in combination with nonlinear optimization for generating infrastructure management modeling for sewer pipes. In this study, a frequency analysis technique was used to develop transition probabilities of Markov deterioration model for large combined sewers in Indianapolis. The sewer database was divided into 16 groups and simple linear regression was developed to identify relationship between time and condition of pipes. The transition matrix was generated by assuming that the condition of sewer pipe moves to poorer condition or stays at current condition. It means a pipe in condition 4 cannot improve and move to condition 2. Finally, a nonlinear optimization technique was used to minimize the sum of absolute difference between regression result and Markov chain estimations. The outcome of this study was the deterioration curve for sewer pipes to illustrate the changes in condition states while the pipe is aging [33].

Micevski et al. (2002) developed a Markov model for the structural deterioration of storm water pipes. The pipe dataset was randomly categorized into two separate datasets, and Bayesian techniques were used to identify the parameters of Markov model. The Metropolis–Hastings, which is a member of the family of Markov chain Monte Carlo (MCMC), was used to calibrate the model. The validation of the model was performed through hypothesis testing to determine if the Markov model is appropriate for storm water pipe deterioration. The result indicated that the Markov model was consistent (at the 5% significant level) and can be used for storm water pipe deterioration. In addition, pipe diameter, construction material, soil type, and exposure classification were found as significant variables that influence deterioration of pipes [32].

Jeong et al. (2005) used Markov chain to develop deterioration model for a wastewater infrastructure system. The model was generated by the inspection database obtained from the city of San Diego. The ordered probit model approach was used in this study to estimate transition period and transition probability matrix. The estimation result showed that pipe age, size, length, and slope are significant variables affecting deterioration of pipes. Additionally, the ordered probit approach was identified as an effective method to generate the model with less data groups and using categorical variables. However, the developed model could not be validated for condition states 3 and 4 in this study [31].

Le Gat (2008) developed a mixed multi-state deterioration process by a non-homogeneous Markov chains process to model the deterioration of urban drainage infrastructures. GompitZ analysis method was used to estimate the parameters of the time dependent transition probabilities through maximum marginal likelihood estimation. The GompertZ model considered a set of pipelines as a set of generic objects that are different based on their covariate values. The dataset was divided into different categories based on pipe diameter, sewer type, and installation period. Cross validation method was used to split the data randomly for test and validation process. The result of this study indicated that a statistical model like GompitZ cannot predict the exact condition of a given pipe and only condition probabilities can be estimated. Another problem in applying GompitZ methodology is that calibration of this method is very difficult and risk of misclassification is very high if population of pipes is not sufficient in database [34].

Scheidegger et al. (2011) developed a network condition simulator (NetCoS) to provide a synthetic population of sewer pipes based on historical inspection database. This model can be used to benchmark deterioration models and select an appropriate data management strategy. A semi-Markov chain technique was used to model deterioration of sewer pipes and transition probabilities. The deterioration of sewer pipes was defined by a set of survival function in this study. A survival function described condition states of sewer pipe based on age-dependent probabilities. Then, semi-Markov chain computed the probabilities of changing the condition of pipes. The strength of NetCoS is that it is not limited to certain type of distributions and it is very flexible to generate more complex data. However, the main problem of this model is that it is not possible to validate the model by real-life data [35].

Balekelayi and Tesfamariam (2019) applied a Bayesian geosadditive regression model to predict sewer pipe deterioration scores from a set of predictors categorized as physical, maintenance, and environmental data. Sewer data were collected from city of Calgary and three categories of covariates were included in the regression model: (1) Physical data—materials, length, diameter, age, depth, slope, and residential and commercial connections; (2) maintenance data—repairs, flushes, cleaning, degrease, backups, and root cuts; and (3) environmental data—the geographical location of pipes in a community. The results highlighted the importance of considering a semiparametric modeling approach, because some of the continuous covariates have nonlinear effects on the structural state of pipes. Furthermore, the Bayesian inference captures the uncertainty in the data [36].

5.2.3. Model Discussion

Compared to other statistical models, Markov models offer certain advantages and disadvantages. The main advantage of Markov models is its flexibility to predict the dependent variables. In addition, sequence dependencies can be modeled by Markov technique. Pipe deterioration is a complex and sequential event and the structure of Markov models offers a powerful algorithm to forecast the future condition of pipes. The result of Markov models can be used to manage a network or groups of pipes for future maintenance and inspection planning. The major challenge in the use of Markov models is lack of data on the past and present condition of pipes. Markov models require pipe grouping, and each group needs sufficient amount of data for development and validation of the model. Furthermore, development of the condition transition probabilities is a very complex and difficult process during implementation of the model.

5.3. Logistic Regression Model

5.3.1. Model Description

Logistic regressions are used to analyze the relationship between multiple independent variables and a categorical dependent variable. In logistic regression, the probability of occurrence of an event is estimated by fitting data to a logistic curve. Binary logistic regression and multinomial logistic regression are the most common types of logistic regression models [37]. Binary logistic regression is typically used when the response variable involves two categories (success or failure) and in the case of more than two response variable, multinomial logistic regression is applicable. Equation (5) presents the multiple logistic regression formula when multiple explanatory variables are used in the model [38].

$$\begin{aligned} \log\left[\frac{\pi}{1-\pi}\right] &= \log\left[\frac{P(Y = 1 | X_1, X_2, \dots, X_p)}{1 - P(Y = 1 | X_1, X_2, \dots, X_p)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \\ &= \alpha + \sum_{j=1}^p \beta_j X_j \end{aligned} \tag{5}$$

where X_1, X_2, \dots, X_p are independent variables, α is the intercept parameter for category i , and β is the regression coefficients. Multinomial logistic regression is used when multiple levels of categorical response variables are in the model. Equation (6) shows the multinomial logistic regression formula.

$$\begin{aligned} \log\left[\frac{\pi}{1-\pi}\right] &= \log\left[\frac{P(Y = i | X_1, X_2, \dots, X_p)}{1 - P(Y = k | X_1, X_2, \dots, X_p)}\right] = \alpha + \beta_{i1} X_1 + \beta_{i2} X_2 + \dots + \beta_{ip} X_p \\ &= \sum_{j=1}^p \beta_{ij} X_j \end{aligned} \tag{6}$$

where $i = 1, 2, \dots, K - 1$ correspond to categories of the dependent variable, X_1, X_2, \dots, X_p are independent variables, α is the intercept parameter for category i , and β is the regression coefficients associated with dependent category i . The probability that $Y = 1$ can be measured using an exponential transformation as shown in Equation (7).

$$P(Y = 1 | X_1, X_2, \dots, X_p) = \frac{e^{\alpha + \sum_{j=1}^p \beta_j X_j}}{1 + e^{\alpha + \sum_{j=1}^p \beta_j X_j}} \quad (7)$$

An important parameter in logistic regression is the odds ratio that measures the relationship between explanatory and response variables as shown in Equation (8).

$$\frac{\pi(X)}{1 - \pi(X)} = \exp(\alpha + \beta x) = e^{\alpha} (e^{\beta})^x \quad (8)$$

As condition states of sewer pipes are defined by discrete values (1, 2, . . . , n), the logistic regression is able to determine the probability of pipe being in each condition. In addition, odds ratio can show the effect of influence factors that degrade condition of sewer pipes.

5.3.2. Previous Studies

Logistic regression is widely used to model the deterioration of sewer pipes. Davies et al. (2001) developed a logistic regression model to predict the structural condition of rigid sewer pipes. The main objective of this study was to identify influenced factors affecting deterioration of sewer pipes. The condition of sewer pipes was divided into two categories of good and poor condition and the logistic transformation was used to estimate the probabilities. Stepwise forward and backward methods and binary logistic regression were employed in this study to select appropriate dependent variables. The result indicated that pipe material, diameter, length, sewer type, location, groundwater, and soil corrosivity are the influence factors that affect deterioration of sewer pipes. The main weakness of this study was that there is no information regarding validation and accuracy of the model. Additionally, only p -test was used to determine the significance of the dependent variables.

Ariaratnam et al. (2001) used logistic regression to predict condition states of sewer pipes. A linear regression variable selection method was used to specify the suitable independent variables in the model. Significance of the variables in this study was examined by Wald Test and likelihood-ratio test. The likelihood-ratio test revealed that pipe age, diameter, and sewer types are the significant variables in the model. A sensitivity analysis was performed to validate the logistic regression model. However, sensitivity analysis is not enough to determine the performance of logistic regression model [39].

Koo and Ariaratnam (2006) generated a logistic regression model to predict the deterioration of sewer infrastructure systems. The city of Phoenix, Arizona, wastewater collection database was used to develop binary logistic regression. Expert judgment was used to select pipe age, maximum velocity, and cumulative flow as dependent variables in the model. They divided the dependent variables into three separate groups with combination of 27 sub-classes. P-test, Wald Test, and likelihood-ratio test were used to assess the significance of the variables in the model. The result reflected that maximum velocity is not a significant factor in the model. The performance of logistic regression was not validated in this study [40].

Ana et al. (2009) investigated the influence of sewer physical properties on the structural deterioration of the sewer pipelines using logistic regression. This study used the backward stepwise regression method for selection the predictor variables. The significance of the dependent variables was assessed by carrying out Wald Test and likelihood-ratio test. They also investigated the interaction effects of independent variables. For example, length of sewer pipes may be found insignificant in the deterioration model but may become significant when combined with another independent variable. Sewer age, material, and length were found significant in this study and no validation method was used to validate the result of logistic regression [41].

Lubini and Fuamba (2011) developed a logistic regression model for the deterioration timeline of sewer systems. This model was applied to a case study in Quebec City, Canada, and pipe age, diameter, material, length, and slope were the contributing factors to generate the model. Several statistical tests such as overall model test, strength of association, likelihood-ratio test, and Wald Test were used to assess the significance of independent variables. A deterioration curve was developed in this study for maintenance and operational planning. However, the performance and accuracy of the logistic regression model was not validated [42].

Salman and Salem (2012), employed three statistical models including ordinal regression, multinomial logistic regression, and binary logistic regression to model the deterioration of wastewater collection lines. Five different ordinal regression were generated, and the likelihood-ratio test was used to determine the relation of dependent and independent variables. The result indicated that none of the ordinal regression models satisfied the odds assumptions. Also, developed multinomial logistic regression obtained just 52% accuracy. Binary logistic regression was the only model that could predict condition of sewer pipes with 66% accuracy. This study provided different deterioration curves and equations, which are useful to understand behavior of individual pipes in network. Moreover, logistic regression models were validated by confusion matrix and real data. The result of binary logistic regression revealed that pipe size, length, slope, age, material, and sewer type are the significant factors in the model [43].

Logistic regression model was used by Sousa et al. (2014) to assess structural deterioration of sewer pipelines. A complete model including all independent variables and a reduced model including only significant variables were developed in this study. Wald Test and likelihood-ratio test were used to identify significance variables in the model. The logistic regression model was validated by confusion matrix and the result indicated 65% accuracy [44].

Kabir et al. (2018) developed a Bayesian logistic regression model to predict the structural condition of sewer pipelines. In this study, Bayesian model averaging technique was used to identify significant variables and the condition of sewer pipes were predicted by logistic regression. P-test, Wald Test, likelihood-ratio test, and Durbin–Watson test were employed to determine the significance of the independent variables. The condition states of sewer pipes were divided into two categories including good and poor conditions. The performance of the model was validated through confusion matrix. The main weakness of this model is that the pipe data were grouped based on pipe material, and the model could not predict condition of pipe by considering all pipe material [45].

Malek Mohammadi et al. (2019) utilized a logistic regression model to predict condition of sanitary sewer pipes. The framework of this study was based on collected data from the City of Tampa, Florida. A variety of independent variables such as, pipe age, material, diameter, depth, length, slope, watertable, and soil type were used to run the model. P-test, Wald test, and likelihood-ratio test were used to determine significant variables. Multiple logistic regression was not able to predict all five condition states of the sewer pipes, however the binary logistic model predicted the condition rating of sanitary sewer pipes with 81% accuracy. The performance of the model was validated by confusion matrix. The result indicated that pipe age, material, diameter, length, and watertable are the significant factors affecting deterioration of sewer pipes [46].

Model Discussion. The simple concept and prediction power of logistic regression to forecast discrete values made it very popular to assess condition of sewer pipes. Logistic regression is the most frequently used regression model for the analysis dataset with two or more discrete outcome variables [38,47]. Furthermore, there is a simple relationship between the coefficients and the odds ratio in logistic regression, and the most important variables affecting deterioration of pipes can be identified in this model [47–49]. This feature provides a better understanding of sewer pipes deterioration process and also required data that need to be collected during inspection or data collection. In addition, the effect of important variables can be considered in the design and construction of new sewer systems.

The major weakness of logistic regression is the sufficient amount of data required to generate stable and meaningful model. Moreover, the linear nature of logistic regression is not flexible enough

to identify nonlinear decision boundaries and also more complex relationships. However, logistic regression is still one of the most popular models to predict condition of sewer pipes.

6. Discussion and Conclusions

Several statistical and artificial intelligence techniques are used to model deterioration of sewer pipes. In this paper, the most common statistical models for predicting deterioration and condition states of sewer pipes have been presented. Table 2 illustrates a summary of the models and the independent variables used to develop the models.

Table 2. Sewer condition prediction models.

Authors	Year	Model	Independent Variables
Davies et al.	2001	Logistic regression	Age, Material, Diameter, Depth, Length, Sewer Type, Location, Corrosivity, Road Type, Other Factors
Ariaratnam et al.	2001	Logistic regression	Age, Material, Diameter, Depth, Sewer Type
Wirahadikusumah et al.	2001	Markov chain	Material, Depth, Soil Type, Groundwater
Lubini and Fuamba	2001	Logistic regression	Age, Material, Diameter, Length, Slope
Micevski et al.	2002	Markov chain	Material, Diameter, Soil Type
Koo and Ariaratnam	2006	Logistic regression	Age, Flow, Other Factors
Chughtai and Zayed	2008	Linear regression	Age, Material, Diameter, Depth, Length, Slope, Bedding Type, Road Type
Gat	2008	Markov chain	Age, Diameter, Sewer Type
Ana et al.	2009	Logistic regression	Age, Material, Diameter, Depth, Length, Slope, Sewer Type, Location
Salman and Salem	2012	Ordinal regression Logistic regression Binary regression	Age, Material, Diameter, Depth, Length, Slope, Sewer Type
Sousa et al.	2014	Logistic regression	Age, Material, Diameter, Depth, Length, Slope
Bakry et al.	2016	Multiple regression	Age, Material, Diameter, Depth, Length, Sewer Type, Soil Type, Groundwater, Surface Type, Traffic
Gedam et al.	2016	Linear regression	Age, Material, Diameter, Depth
Kabir et al.	2018	Bayesian logistic regression	Age, Material, Diameter, Depth, Length, Slope, Up Invert, Down Invert, Other Factors
Malek Mohammadi et al.	2019	Logistic Regression	Age, Material, Diameter, Depth, Length, Slope, Groundwater, Soil Type
Balekelayi and Tesfamariam	2019	Bayesian Regression	Age, Material, Diameter, Depth, Length, Slope, Groundwater, Residential and Commercial Connections, Repairs, Flushes, Cleaning, Degrease, Backups, and Root Cuts

Availability of pipe inspection and soil data is the fundamental of developing condition prediction models. Unfortunately, most cities and agencies do not have an integrated database with all the required information, and the available databases typically involve uncertainty and missing values [43].

Employment of GIS based databases are suggested for data inventory and management. Several data layers can be joined together in GIS, and they can be updated and analyzed at the same time.

All the presented models in this paper were able to predict deterioration and future condition of sewer pipes. However, one of the main challenges was the validation of the developed models. Most of the studies just determined the significant variables in the models without considering any validation technique. Several validation methods, such as confusion matrix or receiver operating characteristic curve (ROC curve), can be used to validate classification models. The confusion matrix is used to identify the number of elements that have been correctly or incorrectly predicted for each class. In confusion matrix, for every test samples the actual class is compared to the class that was assigned by the trained classifier. ROC curve illustrates the exchange between true positive to false positive rates. In ROC curve, the X-axis illustrates a false positive rate (specificity) and the Y-axis presents a true positive rate (sensitivity) [50].

The pipe group models, such as Markov, with the ability of forecasting behavior of pipe networks, are useful to allocate budget for maintaining and rehabilitating groups of pipes. Meanwhile, pipe level models such as logistic regression are more appropriate for assessing the condition of individual pipes and prioritizing each pipe segment [6,51]. Additionally, probabilistic-based models such as Markov and logistic regression are useful to determine and analyze the risk of pipe failure in the network.

In general, based on the application of the presented models, linear regression models are not appropriate to forecast the discrete values, and classification algorithms are more valuable to estimate deterioration of sewer pipes. Markov models are very useful for predicting the condition of pipe networks, however development of this model was not simple and required sufficient amount of data. More research has been done with the employment of logistic regression due to simplicity of the model and the ability to identify influence factors that affect deterioration of sewer pipes. Identifying these factors is very vital to reduce the cost of data collection and risk of pipe failure by bearing in mind their importance during design and construction phases. Table 3 presents the performance and prediction power of sewer condition prediction models, reviewed in this study. Additionally, the main advantages and disadvantages of reviewed models are presented in Table 4.

Numerous deterioration models were presented in this paper. However, condition prediction models for individual sewer pipes have not been fully examined yet and the result of most studies reflected that it is possible to assess future condition and behavior of sewer pipes through novel artificial intelligence models. AI models such as neural network, support vector machine, decision trees, and new machine learning models not only consider the linear relationship between dependent and independent variables but are also capable of investigating their associations in several dimensions [47].

Table 3. Comparison summary of sewer condition prediction models.

Applicability	Logistic Regression	Markov Chain	Linear Regression
Predicting condition of pipe groups	Moderate	Good	Poor
Predicting condition of individual pipes	Good	Moderate	Good
Predicting categorical dependent variables	Good	Moderate	Poor
Conceptual and computational simplicity	Good	Poor	Good
Identifying relationship between dependent and independent factors	Good	Poor	Moderate
Calculation of condition probabilities	Good	Good	Poor
Flexible to deficiency of data	Moderate	Poor	Good

Table 4. Advantages and disadvantages of sewer condition prediction models.

Prediction Models	Advantages	Disadvantages
Logistic Regression	Does not require too many computational resources, incredibly easy to implement, highly interpretable, not required input features to be scaled, capable of predicting probabilities, does not require normal distribution of independent variables, capable of predicting influence variables	not very useful for non-linear and complex problems, can only predict a categorical outcome
Markov Chain	can predict categorical and continuous variables, strong statistical foundation, can be combined with other models, works for complicated distributions in high-dimensional spaces	difficult to implement and validate, requires large number of data, Markov assumptions may not be applicable for all datasets
Linear Regression	very easy to implement, capable of predicting influence variables, highly interpretable	can only identify linear relationships between variables, only looks at the mean of the dependent variable, sensitive to outliers

7. Future Research Needs

There is a need for more research to predict condition of sewer pipes with higher accuracy and confidence level. In addition, more investigation is required to identify the influence of physical and environmental factors that affect deterioration of sewer pipes. Few studies have considered the effect of independent variables on condition of sewer pipes. Moreover, there is a still lack of splitting trained and test data samples in statistical models. Some advanced techniques such as k-fold cross validation can be used to improve the validity of the models. Moreover, the validation of these models should be improved by development of more advanced models and validation techniques.

8. Acronyms

AI	Artificial Intelligence
ASCE	American Society of Civil Engineers
AWWA	American Water Works Association
C	Pipe State Condition
CCTV	Closed-circuit Television
CIPP	Cured-in-Place Pipe
EPA	Environmental Protection Agency
Equation	Equation
<i>i</i>	Facility Index
I/I	Infiltration/Inflow
m	Matrix
MCMC	Markov Chain Monte Carlo
NetCoS	Network Condition Simulator
<i>P</i>	Probability
<i>P_{ij}</i>	Transition Probability
PVC	Polyvinyl Chloride
<i>t</i>	Time
U.S.	United States
<i>X_i</i>	Independent Variable
<i>Y_i</i>	Dependent Variable for Facility
α	Intercept Parameter
ϵ_i	Random Error Term
β	Regression Coefficient

Author Contributions: Conceptualization, methodology, and investigation, M.M.M.; review and editing, M.N.; writing—Original draft preparation and formal analysis, V.K., R.S., N.S., T.A.

Funding: This research received no external funding

Acknowledgments: I would like to express my most sincere gratitude and appreciation to my academic advisor, and mentor Mohammad Najafi, P.E., F. ASCE, and Director of the Center for Underground Infrastructure Research and Education (CUIRE). I also would like to thank my friend at CUIRE who helped me to prepare this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaushal, V.; Serajiantehrani, R.; Najafi, M.; Hummel, M. Seismic Hazards Estimation for Buried Infrastructure Systems: Challenges and Solutions. In Proceedings of the 2nd International Conference on Natural Hazards & Infrastructure, Chania, Greece, 23–26 June 2019.
2. ASCE. *Report Card for America's Infrastructure*; American Society of Civil Engineers (ASCE): New York, NY, USA, 2017.
3. EPA. *Report to Congress: Impacts and Control of CSOs and SSOs*; EPA: Washington, DC, USA, 2004.
4. Kaushal, V. Comparison of Environmental and Social Costs of Trenchless Cured-in-Place Pipe Renewal Method with Open-Cut Pipeline Replacement for Sanitary Sewers. Ph.D. Thesis, University of Texas at Arlington, Arlington, TX, USA, 2019.
5. EPA. *Innovation and Research for Water Infrastructure for the 21st Century*; EPA: Arlington, VA, USA, 2007.
6. Ana, E.V.; Bauwens, W. Modeling the structural deterioration of urban drainage pipes: The state-of-the-art in statistical methods. *Urban Water J.* **2010**, *7*, 47–59. [[CrossRef](#)]
7. American Water Works Association (AWWA). Buried no Longer: Confronting America's Water Infrastructure Challenge, Boulder, CO. 2012. Available online: <https://www.awwa.org/Portals/0/files/legreg/documents/BuriedNoLonger.pdf> (accessed on 1 July 2018).
8. Opila, M.C. Structural Condition Scoring of Buried Sewer Pipes for Risk-Based Decision Making. Ph.D. Thesis, University of Delaware, Newark, DE, USA, 2011.
9. Kaushal, V.; Najafi, M.; Love, J.; Qasim, S.R. Microbiologically Induced Deterioration and Protection of Concrete in Municipal Sewerage System: A Technical Review. *J. Pipeline Syst. Eng. Pract.* **2019**, *11*, 03119002. [[CrossRef](#)]
10. Najafi, M.; Gokhale, S.B. *Trenchless Technology: Pipeline and Utility Design, Construction, and Renewal*; McGraw-Hill: New York, NY, USA, 2005.
11. EPA. *Condition Assessment of Wastewater Collection Systems*; EPA/600/R-09/049; EPA: Boston, MA, USA, 2009.
12. Davies, J.P.; Clarke, B.A.; Whiter, J.T.; Cunningham, R.J. Factors influencing the structural deterioration and collapse of rigid sewer pipes. *Urban Water* **2001**, *3*, 73–89. [[CrossRef](#)]
13. Al-Barqawi, H.; Zayed, T. Condition rating model for underground infrastructure sustainable water mains. *J. Perform. Constr. Facil.* **2006**, *20*, 126–135. [[CrossRef](#)]
14. Kley, G.; Caradot, N. *D1.2. Review of Sewer Deterioration Models*; Kompetenzzentrum Wasser Berlin gGmbH: Berlin, Germany, 2013.
15. Morcuos, G.; Lounis, Z. Maintenance optimization of infrastructure networks using genetic algorithms. *J. Autom. Constr.* **2004**, *14*, 129–142. [[CrossRef](#)]
16. Yang, J. Road Crack Condition Performance Modeling Using Recurrent Markov Chains and Artificial Neural Networks. Ph.D. Thesis, Department of Civil and Environmental Engineering, College of Engineering, University of South Florida, Tampa, FL, USA, 2004.
17. Tran, H.D. Investigation of Deterioration Models for Stormwater Pipe Systems. Ph.D. Thesis, Victoria University, Melbourne, Australia, 2007.
18. Altarabsheh, A.G. Managing Urban Wastewater System Using Complex Adaptive System Approach. Ph.D. Thesis, Purdue University, West Lafayette, IN, USA, 2016.
19. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: London, UK, 2011.
20. Dasu, T.; Johnson, T. *Exploratory Data Mining and Data Cleaning*; John Wiley & Son: Hoboken, NJ, USA, 2003.
21. Rawlings, J.O. *Applied Regression Analysis: A Research Tool*; Wadsworth & Brooks, Cole Advanced Books & Software: Pacific Grove, CA, USA, 1989.
22. Chughtai, F.; Zayed, T. Sewer pipeline operational condition prediction using multiple regression. In Proceedings of the Pipelines 2007: Advances and Experiences with Trenchless Pipeline Projects, Boston, MA, USA, 8–11 July 2007.

23. Chughtai, F.; Zayed, T. Structural condition models for sewer pipeline. In Proceedings of the Pipelines 2007: Advances and Experiences with Trenchless Pipeline Projects, Boston, MA, USA, 8–11 July 2007.
24. Chughtai, F.; Zayed, T. Infrastructure Condition Prediction Models for Sustainable Sewer Pipelines. *J. Constr. Facil.* **2008**, *22*, 333–341. [[CrossRef](#)]
25. Bakry, I.; Alzraiee, H.; Kaddoura, K.; Masry, M.E.; Zayed, T. Condition Prediction for Chemical Grouting Rehabilitation of Sewer Networks. *J. Perform. Constr. Facil.* **2016**, *30*, 04016042. [[CrossRef](#)]
26. Bakry, I.; Alzraiee, H.; Masry, M.E.; Kaddoura, K.; Zayed, T. Condition Prediction for Cured-in-Place Pipe Rehabilitation of Sewer Mains. *J. Perform. Constr. Facil.* **2016**, *30*, 04016016. [[CrossRef](#)]
27. Madanat, S.; Ibrahim, W.H.W. Poisson Regression Models of Infrastructure Transition Probabilities. *J. Transp. Eng.* **1995**, *121*, 267–272. [[CrossRef](#)]
28. Morcoux, G.; Rivard, H.; Hanna, A.M. Modeling Bridge Deterioration Using Case-based Reasoning. *J. Infrastruct. Syst.* **2002**, *8*, 86–95. [[CrossRef](#)]
29. Moteleb, M. Risk Based Decision Making Tools for Sewer Infrastructure Management. Ph.D. Thesis, University of Cincinnati, Cincinnati, OH, USA, 2010.
30. Konstantopoulos, T. *Markov Chains and Random Walks*; Springer: New York, NY, USA, 2009.
31. Jeong, H.S.; Baik, H.S.; Abraham, D.M. An ordered probability model approach for developing Markov chain-based deterioration model for wastewater infrastructure systems. In Proceedings of the Pipelines 2005: Optimizing Pipeline Design, Operations, and Maintenance in Today's Economy, Reston, VA, USA, 21–24 August 2005; pp. 649–661.
32. Micevski, T.; Kuczera, G.; Coombes, P. Markov Model for Storm Water Pipe Deterioration. *J. Infrastruct. Syst.* **2002**, *8*, 49–56. [[CrossRef](#)]
33. Wirahadikusumah, R.; Abraham, D.; Iseley, T. Challenging issues in modeling deterioration of combined sewers. *J. Infrastruct. Syst.* **2001**, *2*, 77–84. [[CrossRef](#)]
34. Gat, Y.L. Modelling the deterioration process of drainage pipelines. *Urban Water J.* **2008**, *5*, 97–106. [[CrossRef](#)]
35. Scheidegger, A.; Hug, T.; Rieckermann, J.; Maurer, M. Network condition simulator for benchmarking sewer deterioration models. *Water Res.* **2011**, *45*, 4983–4994. [[CrossRef](#)] [[PubMed](#)]
36. Balekelayi, N.; Tesfamariam, S. Statistical Inference of Sewer Pipe Deterioration Using Bayesian Geoadditive Regression Model. *J. Infrastruct. Syst.* **2019**, *25*, 04019021. [[CrossRef](#)]
37. Park, H.A. An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. *J. Korean Acad. Nurs.* **2013**, *43*, 154. [[CrossRef](#)]
38. Agresti, A. *An Introduction to Categorical Data Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
39. Ariaratnam, S.T.; El-Assaly, A.; Yang, Y. Assessment of Infrastructure Inspection Needs Using Logistic Models. *J. Infrastruct. Syst.* **2001**, *7*, 160–165. [[CrossRef](#)]
40. Koo, D.H.; Ariaratnam, S.T. Innovative method for assessment of underground sewer pipe condition. *Autom. Constr.* **2006**, *15*, 479–488. [[CrossRef](#)]
41. Ana, E.; Bauwens, W.; Pessemier, M.; Thoeye, C.; Smolders, S.; Boonen, I.; Gueldre, G.D. An investigation of the factors influencing sewer structural deterioration. *Urban Water J.* **2009**, *6*, 303–312. [[CrossRef](#)]
42. Lubini, A.T.; Fuamba, M. Modeling of The Deterioration Timeline of Sewer Systems. *Can. J. Civ. Eng.* **2011**, *38*, 1381–1390. [[CrossRef](#)]
43. Salman, B.; Salem, O. Modeling Failure of Wastewater Collection Lines Using Various Section-Level Regression Models. *J. Infrastruct. Syst.* **2012**, *18*, 146–154. [[CrossRef](#)]
44. Sousa, V.; Matos, J.P.; Matias, N. Evaluation of artificial intelligence tool performance and uncertainty for predicting sewer structural condition. *Autom. Constr.* **2014**, *44*, 84–91. [[CrossRef](#)]
45. Kabir, G.; Balek, N.B.C.; Tesfamariam, S. Sewer Structural Condition Prediction Integrating Bayesian Model Averaging with Logistic Regression. *J. Perform. Constr. Facil.* **2018**, *32*, 04018019. [[CrossRef](#)]
46. Malek Mohammadi, M.; Najafi, M.; Tabesh, A.; Riley, J.; Gruber, J. Condition Prediction of Sanitary Sewer Pipes. In Proceedings of the ASCE Pipeline Conference, Nashville, TN, USA, 21–24 July 2019.
47. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; Wiley: Hoboken, NJ, USA, 2013.
48. Atique, F. Analysis of Urban Pipe Deterioration Using Copula Method. Ph.D. Thesis, University of Delaware, Newark, DE, USA, 2016.

49. Salman, B. Infrastructure Management and Deterioration Risk Assessment of Wastewater Collection Systems. Ph.D. Thesis, University of Cincinnati, Cincinnati, OH, USA, 2010.
50. Malek Mohammadi, M. Development of Condition Prediction Models for Sanitary Sewer Pipes. Ph.D. Thesis, University of Texas at Arlington, Arlington, TX, USA, 2019.
51. Malek Mohammadi, M.; Homing, A.; Najafi, M.; Korky, S.J.; Serajiantehrani, R. Development of a Framework for Design and Installation of a New Lining Material for Water Pipeline Renewal. In Proceedings of the ASCE Pipeline Conference, Nashville, TN, USA, 21–24 July 2019.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).