infraPLAN Ask the Experts! #2

Are you still using Desktop Scoring to assess the Likelihood of Failure of pipes?

Machine Learning is a better alternative, more accurate, easier to use, affordable.

Assessing the Likelihood of Failure (LOF) of a pipe using desktop or GIS scoring, in view of prioritizing the replacement or inspection of water pipes, remains a popular approach. While it is perceived to be simple and scientific, it can actually be quite difficult to complete, rather inaccurate, and limited in scope.

Because of its predictive capacity and analytical flexibility, Machine Learning is more precise (even for small utilities), easier to develop, and further-reaching. It does require that breaks be associated to the ID of the pipes they occurred on which is no longer a problem at most utilities.

Thanks to our modeling expertise, the careful steps we take to bring data to its highest possible level of quality, and access to 1,000,000+ data points, our Machine Learning models predict up to 85% of the breaks occurring during the next 2 years (on 10% of the pipes; worse ranked first).

In this article we compare approaches leading to the determination of a LOF score for each pipe (desktop scoring vs. advanced analytics). We discuss limitations and capacities.

LOF scoring

The approach requires that the following be identified:

- variables that define a pipe and affect its potential degradation
- values those variables can take
- weight given to each variable
- score given to each value

info@infraplan-llc.com www.infraplan-llc.com

Variables taken into account typically pertain to the pipes physical and operational characteristics, and their environment. They include material, date of installation, date of abandonment, diameter, length, and, if available, soil, pressure, number of previous breaks, traffic, groundwater, etc.

Desktop or GIS scoring approach presents the following issues and limitations.

- **Weight**: The weight given to variable 1 may need to differ for different values of variable 2. For example, soil corrosivity is different for plastic versus Cast Iron pipes. Which calls for a different weight for soil for each material. Yet most scoring systems fail to adjust that weight, rather applying the same weight across the board.

- **Value**: The value a variable can take may call for a different score for different values of another variable. For example, the value “1950-1970” of the variable “period of installation” can yield a score of 1 for material 1 (good), and a 5 for material 2 (bad).

- **Equation/model**: The 2 previous points call for the creation of a different scoring equation/model for each material which makes it difficult to compare scores; a pipe with an overall score of “2” computed with one equation could actually be in worse condition than a “3” for a pipe computed with another equation.

- **Future LOF**: Scoring generates a LOF score for each pipe as of now, not in the future. Forecasting the future physical condition requires that the occurrences (breaks) be analyzed over a period of time in the past and be projected in the future. This cannot be achieved with scoring.

- **Number of breaks**: The LOF score cannot be translated into a number of breaks, or break rate using desktop or GIS scoring.

- **Abandoned pipes and their breaks**: The LOF of a pipe is drawn from the knowledge of the behavior of all pipes that have experienced the same degradation which, in addition to the pipes that are still in service, includes the pipes that may have been abandoned (which tend to have experienced more breaks). Properly incorporating their contribution to the LOF of the remaining pipes cannot be achieved with simple scoring because a scoring approach does not consider how long each pipe has been in service. An article will be dedicated to the importance of abandoned pipes.

- **Granularity**: Simple scoring often leads to many pipes having the same 1-5 score, making prioritization of pipe inspections or replacements rather difficult, and potentially leading to arbitrary decisions.

The only advantage of scoring is that it allows estimating a LOF score even when break data is not available at the pipe level. This also means that values and weights are therefore assigned based on expert knowledge and perception; which is the main drawback as far as accuracy goes: we are essentially guessing.

Advanced analytical approaches

If breaks are assigned to pipes, the LOF of a pipe can be estimated using advanced analytical approaches such as multi-variable regression or Machine Learning.

Those approaches present the following advantages compared to simple desktop or GIS scoring:

- They can "see" inferences between many variables and their values, looking at the millions of possible combinations of values the variables can take in a way human minds can't, which allows assigning LOF scores effortlessly. The data "decides" what the weight and scores ought to be, and how pipes compare, alleviating the difficulties and sources of error described in the previous section. For example, there is no need for several equations for various groups of pipes.

- The LOF can be translated in predicted number of breaks, and this, not only as of now but also in the future (on a year-by-year basis over an entire planning horizon), a precious commodity for long-term R&R planning simulations.

- When properly configured, advanced analytics allow taking abandoned pipes into account which greatly improves a model predictive capacity and accuracy, a contribution that will only increase as the percentage of abandoned pipes grows larger.

- LOF scores from advanced approaches come as continuous digital values offering much richer granularity than the 1-5 scores from simple desktop scoring.

- Finally, and perhaps most importantly, (and thanks to the above), advanced analytical predictions are far superior as shown in the validation results below.

Validation

Our validation approach relies on making predictions using breaks up to 2 years prior to the date of the analysis and using those last 2 years to validate the predictions. We rank the pipes based on their predicted LOF (worst first) and look at the percentage of actual breaks (for the 2 years of validation) a certain percentage of the number or length of pipes

info@infraplan-llc.com www.infraplan-llc.com

experienced. The higher the percentage of predicted breaks, the better the model. We show results for 3 systems with different size and condition.

System 1 - Large system - High break rate

2,457.7 miles of Cast Iron pipes and 2,511.7 miles of Ductile Iron pipes.

The study took place in 2021; breaks from 2005 to 2020; break data from 2005-2018 is used to predict breaks for 2019-2020, for which we also have actual breaks.

In the Table below, we show validation results obtained with:
1. Desktop scoring, 2. Multi-variable regression, and 3. Machine Learning.

Each cell contains 2 values, first the percentage of breaks identified for Ductile Iron, and then for Cast Iron pipes. To compare apple to apple, the desktop and Machine Learning results are broken down by material because there was a need to create 2 multi-variable regression models, one for each material.

DI/CI	% Breaks (2019-2020) avoided (worst pipes first)		
% Pipes Targeted	**Desktop scoring**	**Multi-variable regression**	**Machine Learning**
1	0.6/1.1	34.0/12.1	44.2/15.3
5	**12.2/5.9**	**54.5/35.7**	**63.5/38.1**
10	20.2/9.2	69.4/52.0	72.0/55.2

Results show that, for example, the top **5%** worst pipes as ranked by
1. Desktop scoring, 2. Multi-variable regression, and 3. Machine Learning, experienced **12.2%, 54.5%,** and **63.5%** of the 2019-2020 breaks (Ductile Iron pipes), and **5.9%, 35.7%,** and **38.1%** of the 2019-2020 breaks (Cast Iron pipes), respectively.

Machine Learning has a good predictive capacity, much better than with desktop scoring (the scale of the improvement depends on the percentage of pipes, the material and approach; from 3.5 to 14 times better). In addition to performing slightly worse than Machine Learning, multi-variable regression needs expert calibration. Machine Learning is therefore a better option.

System 2 - Large system - Medium break rate

1,490.0 miles of Cast Iron pipes and 713.6 miles of Ductile Iron pipes.

info@infraplan-llc.com www.infraplan-llc.com

The study took place in 2022; breaks from 2002 to 2021; break data from 2002-2019 is used to predict breaks for 2020-2021. We compare predicted breaks with the actual breaks for 2020-2021.

In the Table below, we show failure modeling validation results obtained with:

1. Desktop scoring, 2. Machine Learning.

No multi-variable regression model was run for that project.

ALL PIPES	% Breaks (2020-2021) avoided (worst pipes first)		
% Pipes Targeted	**Desktop scoring**	**Multi-variable regression**	**Machine Learning**
1	17.8	NA	30.6
5	**19.4**	NA	**79.3**
10	45.2	NA	84.9

Results show that, for example, the top **5%** worst pipes experienced **19.4%** of the 2020-2021 breaks when ranked with desktop scoring versus **79.3%** when ranked with Machine Learning.

The performance of the Machine Learning model is even better for system 2 than for system 1. Depending on the percentage of pipes considered, the predictive capacity is 1.8 to 4 times better than with desktop scoring.

System 3 – Small system – Erratic break trend

91.9 miles of Cast Iron pipes and 111.2 miles of Ductile Iron pipes.

The study took place in 2022; we had breaks from 2002 to 2021. However, the last 2 years showed an out of trend break number due to a change in operations (50% higher than for previous years). Therefore, we decided to create a validation model with breaks 2002-2017 (and validation on 2018-2019) as validation on 2020-2021 would be biased (better validation but higher overall break projections).

ALL PIPES	% Breaks (2018-2019) avoided (worst pipes first)		
% Pipes Targeted	**Desktop scoring**	**Multi-variable regression**	**Machine Learning**
1	NA	NA	8.2
5	NA	NA	**30.6**
10	NA	NA	**58.2**

Results show that, for example, the top **5%** worst pipes experienced **30.6%** of the 2018-2019 breaks; at **10%** it is **58.2%.**

Even for a small system (approximately 200 miles) with an erratic break rate, Machine Learning yields a good failure forecasting capacity. Furthermore, this project also shows that while calibration requirements are minimal when using Machine Learning, the user must first gain some understanding of the history of breakage of the system.

In an upcoming article, we will further discuss how, while data quality is paramount for any statistical approach (it is not more stringent because of increased forecasting capacity), Machine Learning does offer advantages when it comes to data availability but also because of its capacity to clean data.

info@infraplan-llc.com www.infraplan-llc.com