infraPLAN
Water Mains R&R & AI

Ask the Experts! #2
LOF Scores

Click below to read all ATE articles.
Article 1: 1pc Water Main R&R Rate
Article 2: LoF Scores
Article 3: LoF Case Studies
Article 4: Data Needed
Article 5: Abandoned Pipes
Article 6: Missing Data
Article 7: Structural Data Issues

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, desktop or GIS scoring can actually be quite difficult to complete, inaccurate, and limited in scope. Because of its predictive capacity and analytical flexibility, Machine Learning is more precise, easier to develop, and further-reaching. It does require that breaks be associated to the ID of the pipes they occurred on which, nowadays, is the case at most utilities.

In this article we compare approaches used to determine the LoF score of each pipe (desktop scoring vs. advanced analytics such as multi-variable regression and Machine Learning). We discuss capacity and limitations.

Desktop scoring

Desktop scoring requires that the following be identified:

- **Variables** that define a pipe or its environment, and contribute to the degradation of the pipe. They include material, date of installation, date of abandonment, break history (number and date), diameter, length, and, if available, soil, pressure, traffic, groundwater, etc.
- **Weight** given to each variable; proportional to the contribution (of the variable) to degradation
- **Values** those variables can take, and score given to each value; also proportional to the contribution (of the value taken by the variable) to degradation

For each pipe, the score that corresponds to the value of a variable is multiplied by the weight (w) that variable carries (w x score). This is done for each variable. The LoF score is the sum of the (w x score) entities for all the variables.

For example:

- The **variable** "material" may take the **value** Cast Iron (**score** of 2) or Ductile Iron (**score** of 1) and be given a **weight** of 2.
- The **variable** "soil" may take the **value** Good (1) or Bad (2) and be given a **weight** of 3.
- Pipe 1 (Cast Iron, Bad soil) LOF score is: 2 x 2 + 3 x 2 = 10.

Desktop Scoring is typically done in excel using data found in GIS, or directly in GIS. The approach presents analytical limitations; first and foremost, its subjectivity.

Desktop or GIS scoring approach presents the following 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.
- **Value**: The value a variable can take may call for a different score for different values of another variable. For example, the value "1960-1965" of the variable "period of installation" can yield a score of 1 for material 1 (good), and a 5 for material 2 (very bad).

The two previous points illustrate a frequent problem in statistics called a bias. Identifying the role of each variable (and their values) based on the other variables requires careful statistical techniques which cannot be done manually, and is therefore typically not undertaken when developing desktop scoring; this can lead to errors.

- **Equation/model**: The two previous points call for the creation of a different scoring equation/model for each value of a certain variable which makes it difficult to compare LoF 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 with another equation. Furthermore, which variables should be granted their own model and equation?

- **Future LoF**: Scoring generates a LoF score for each pipe as of now, not in the future. Forecasting the future physical condition may be necessary when working on long-term planning. It 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 desktop scoring.
- **Number of breaks**: The LoF score cannot be translated into a number of breaks, or Break Rate, using desktop or GIS scoring. Therefore, the aging curve of a group of pipes (future Break Rate by Age), valuable for long-term planning, cannot be drawn.
- **Abandoned pipes and their breaks**: The LoF of a pipe stems from the knowledge of the behavior of all pipes that have experienced degradation which, in addition to the pipes that are still in service, includes the pipes that have been abandoned (which tend to have experienced more breaks. Therefore, they are “data rich,” so of substantial modeling value. Properly incorporating their contribution to the LoF of the remaining pipes cannot be achieved with scoring because that approach does not consider how long each pipe has been in service. Article 5 is dedicated to the importance of abandoned pipes.
- **Granularity**: Simple scoring often leads to many pipes having the same 1-5 score, with many pipes ending up having a "3", making prioritization of pipe inspections or replacements rather difficult, and potentially leading to arbitrary or poor 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 the breaks are assigned to the pipes, the LoF of a pipe can be estimated using advanced analytical approaches such as multi-variable regression powered or not by Machine Learning. Those advanced analytical 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 many possible combinations of values the variables can take in a way human minds cannot, which allows assigning LoF scores effortlessly.

- Advanced analytical approaches have **built-in techniques** that identify and adjust statistical biases, assign the right weight and scores to each variable based on the weights and scores of the other variables, and determine 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 score can be translated in predicted number of breaks** on a year-by-year basis over an entire planning horizon, a precious commodity for long- term R&R planning simulations as this allows drawing the aging curves.
- When properly configured, **advanced analytics allow taking abandoned pipes into account** which greatly improves a model predictive capacity and accuracy. Not having taken abandoned pipes into account may not have been of much consequence so far if the percentage of abandoned pipes has remained low. But their contribution to break predictions will only increase as more abandonment takes place. Article 5 is dedicated to the importance of abandoned pipes.
- LoF scores from advanced approaches come as continuous digital values offering much **richer granularity** than desktop scoring integer scores.
- Finally, and perhaps most importantly, (and thanks to the features described in this article, and robust data cleaning), the **capacity to predict breaks with advanced analytics is far superior**.

The formidable predictive capacity of advanced analytics (versus desktop scoring) will be illustrated in Article #3. We compare the modeling performance for three systems of different size and condition using different modeling approaches. We will see that, thanks to our modeling expertise, and the careful steps we take to bring data to its highest possible level of quality, our Machine Learning models **predict up to 90% of the breaks occurring during the next two years** (on **10% of the pipes**; worst **ranked first)**, much better than desktop scoring, and slightly better than multi-variable regression not powered by Machine Learning.

Our Machine Learning models predict up to 90% of the breaks occurring during the next two years (on 10% of the pipes; worst ranked first), up to 6 times better than desktop scoring.

Click below to read all ATE articles.
Article 1: 1pc Water Main R&R Rate
Article 2: LoF Scores
Article 3: LoF Case Studies
Article 4: Data Needed
Article 5: Abandoned Pipes
Article 6: Missing Data
Article 7: Structural Data Issues

Contact us for a free discussion on using advanced analytics to maximixe your R&R plan!

infraPLAN-LLC helps water utilities, large and small, achieve savings on CIP expenses using our ground breaking platform, **infraSOFT.**

www.infraPLAN-llc.com
(917) 349-6386

Annie Vanrenterghem, PhD, CEO