# Water Utilities and Machine Learning

Using Analytics to Reduce Non-Revenue Water Leaks

Aqueduct Inc.

Table of Contents

## Introduction

Water utilities are a critical piece of infrastructure in the United States and worldwide, distributing potable drinking water to billions of people all over the planet. Modern society is entirely dependant on the proper functioning of these facilities. Though the feats and abilities of modern water infrastructure are quite stunning, the problems facing utilities over the next twenty years are enormous.

Worldwide water demand is expected to be 40% greater than supply by 2040[1], creating new pressure on an already strained network. In the United States, infrastructure is aging rapidly, and water utilities will require an investment of $697 billion dollars over the next twenty years in order to meet existing and new demand.[2] Furthermore, American water utilities currently lose as much as 10-30% of processed water due to non-revenue water leaks.

Water utilities are in a bind. Managers are aware of the enormous losses created by leaks; non-revenue water is tragic both from a business and an environmental standpoint. On the business side, non-revenue leaks cause enormous financial losses as water disappears from the system without being billed; from an environmental standpoint, water is a scarce resource that we cannot afford to waste.

On a global level, societies are entirely dependent on the proper functioning of water utilities. While most urban environments can operate with some dysfunctional infrastructure in the short term, running out of water is a real possibility in many places. Case in point: Cape Town, South Africa has dealt with a recent drought that led the water municipality to threaten to shut off the taps by Summer 2018 if nothing changed.[3] This crisis, nicknamed ‘Day Zero’, occurs amid a highly sensitive political environment that is worsened by resource constraints. Similarly, the Syrian state collapse and civil war in 2011 was triggered by a prolonged drought that destroyed the country's agricultural base and drove farmers into the cities to protest.[4]

---

[1] Water Demand May Be 40% Greater Than Supply by 2040 – Bloomberg

[2] https://www.huffingtonpost.com/entry/our-water-infrastructure-deficit-is-the-real-crisis_us_58d2e997e4b002482d6e6de6

[3] https://www.newyorker.com/tech/elements/counting-down-to-day-zero-in-cape-town

[4] https://www.smithsonianmag.com/innovation/is-a-lack-of-water-to-blame-for-the-conflict-in-syria-72513729/

The message is clear. The stability of societies around the world are dependent on a continuous supply of potable water to their residential, commercial, and agricultural zones. Without clean water, there is no economy. Without clean water, there is no life. In many places around the world, we cannot afford to waste even a drop.

Water infrastructure is a complicated subject, and there is no silver bullet that will solve our global challenges in the years to come. We will realistically solve the problem through a range of political, commercial, and agriculture solutions, topics which are beyond the scope of this paper. With that said, research indicates that water utilities can reduce non-revenue leak challenges in part by developing efficiency insights through their existing data sets.

Data analytics have dramatically improved over the past ten years, and algorithms today are better than ever at pattern detection. Analytics are yielding powerful benefits in Fortune 500 companies, government programs, nonprofits and NGOs, and far more.

This white paper is intended for water managers looking to improve their utilities, and will focus on how machine learning, a particularly powerful form of predictive modeling algorithm, can reduce costs and improve efficiency for modern water utilities. Water utilities can plug these analytics programs into their existing data - from real-time SCADA flow to historical accident rates - and learn more about where leaks are occurring, variables that are causated with leaks or asset failure, best practices for water pressure management, and far more.

Most importantly, machine learning can yield deep predictive insights about the future of assets, and can either delay investment into new assets or eliminate investment entirely. Conversely, water utilities can use this powerful tool to identify assets that have a higher likelihood of failure in the near future. The Benefits and Return on Investment section highlights the economic benefits of a machine learning program.

The modern water utility manager operates within a number of significant constraints. In an ideal world, managers would have an unlimited amount of money to work with, regularly investing tens of millions of dollars into upgrading their infrastructure. Unfortunately, reality dictates that funds are limited. Managers must identify savings and efficiencies whenever possible in order to fully supply their community with high quality potable water.

We believe that the answer to this question lies in data and analytics technologies. Machine learning software is far cheaper than infrastructure investment, and, as we will demonstrate, has been found to produce results with a higher level of accuracy, even as the datasets themselves become more expansive and complicated. To put it simply, machine learning can significantly improve operating efficiency without the need to invest in new physical infrastructure.

With machine learning and better analytics, water utilities can thrive over the next twenty years and meet the enormous challenges that they face. Even amid population growth, water utilities can continue to supply the global population with the fresh water that they need to live healthy and productive lives.

## Challenges Facing Water Utilities

For water utilities, leakage and non-revenue water loss is simply a fact. Most customers are unaware of massive amounts of water lost, but the reality is that on average, American water utilities lose 10% of their water due to leakage, with some losing as much as 20-30%.[5] Even more so, the problem is getting worse, with overall pipe breaks up 27% over the last 6 years.

As experts have noted, breaks and leaks are far more serious than just mere inconveniences. With water leaking out of the system, “economic impacts include loss of treated water, increased maintenance budgets, overtime hours for service personnel, traffic and business disruptions, and damage to private property.” Antiquated infrastructure has a negative effect on pressure, safety and quality of potable water, and flow. The Army Corp of Engineers issued a report in 2017 that gave US water infrastructure a “D” in terms of the quality of distribution systems.[6] In short, this is a systemic problem that weakens individual utilities and damages society as whole.

In this white paper, we will focus on **non-revenue water loss** as our champion key performance indicator (KPI). Non-Revenue water is defined as all of the water in the system that is produced but not sold. This supply can exit the system early through leaks, bursts, and overflows, as well as unauthorized consumption, illegal taps, and meter errors. Non-revenue water represents pure downside, so reducing it is a critical metric of success for the water utility.[7]

---

[5] https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1173&context=mae_facpub
[6] https://www.infrastructurereportcard.org/
[7] http://www.waterworld.com/articles/print/volume-26/issue-80/editorial-features/water-audit-helps.html

Smart water managers are fully aware of this, and recognize the importance of maintaining the health of their infrastructure. Furthermore, it is preferable to improve the efficiency of the existing system instead of building new infrastructure or securing new sources, which can be very expensive and politically imprudent in many locations.

Public Water Systems (PWS's) have four main tools available to reduce their non-revenue water loss[8]:

1. *Pressure Management*

   Research indicates that data-driven water pressure management techniques can be an effective tool in preventing unnecessary pipe bursts while still meeting customer demand. Top utilities optimize their pressure throughout the system to reduce risks.

2. *Speed and Quality of Repairs*

   Infrastructure that is detected and repaired faster reduces the total loss to the system. Furthermore, high quality repairs can prevent recurring bursts and expensive duplicate repairs

3. *Active Leakage Control*

   A large percentage of leaks go unreported. A leak detection program, using sensors or other technology, can go a long way in identifying sources of loss.

4. *Asset Upgrade*

   Broken infrastructure paves the way for total system overhaul, and presents an opportunity to install brand new, high-tech infrastructure.

But before PWS's can begin to address their infrastructure problems, they must conduct a study to fully understand the scope of their challenges. They do this by completing a *water audit*.

## The Water Audit

Water utilities conduct a **comprehensive water audit** in order to determine the total water demand as well as the extent of their losses. Managers will typically take the outcomes of these studies and benchmark them against statistics from comparable water utilities in order to determine the health of their systems.

---

8 http://www.pacificwater.org/pages.cfm/water-services/water-demand-management/what-water-demand-management/four-methods-of-real-loss-management/

Furthermore, the water audit will map out the existing system and attempt to determine the most vulnerable parts of the system in order to predict the return on investment of various potential upgrades.[9] Contrary to conventional thought, older infrastructure is not necessarily more dysfunctional, and managers must gather data from a variety of sources in order to secure an accurate reading.

The water audit is almost entirely a data gathering process, and is designed to give water utility managers a complete picture of the strengths and weaknesses of their infrastructure. A completed audit is followed by two more steps, intervention and evaluation. Intervention is the action step, where utilities send in teams to make repairs, or make operational and administrative changes. Evaluation is a more reflective period, understanding the positive changes that have been made and identify future improvements.

A robust water audit and control problem is often one of the most effective systems that a water utility can implement. As the EPA noted in 2013, “Water loss control programs are often the most economical solution [to address] increasing demand.” It is a process that is geared towards efficiency - doing more with less - and allows the water utility to become a better organization without any abnormal expenditures.[10]

Specifically, better analytics and the introduction of machine learning algorithms into the water audit and control process can dramatically strengthen the system and improve the intelligence of water utility managers.

## Data and Water

The water audit process was defined as a data gathering period. For water utilities, the quality, quantity, and variety of the collected data is a major determinant of the effectiveness of the following intervention period. Furthermore, the analytics software used can reveal a massive number of hidden insights that could never have been discovered solely by a human analyst.

Machine learning is an advanced form of predictive modeling and data analytics that is far more powerful than human operators in detecting patterns and hidden insights. Though very new to water utilities, machine learning is increasingly being adopted in a

[9] http://www.awwa.org/portals/0/files/legreg/documents/waterlosscontrol508.pdf
[10] https://www.epa.gov/sites/production/files/2015-04/documents/epa816f13002.pdf

variety of industries, from retail to self-driving cars to electric utilities.[11] Executives and business managers are aware of the major business improvements available when artificial intelligence and advanced algorithms are brought into the fold.

As the blog Smart & Resilient Cities notes, “Machine learning is the only way to mathematically optimize a large set of variables, whether it is identifying visual patterns using computer vision for image recognition, working to understand speech or identifying pipelines that are likely to fail.”

Machine learning takes your historical data - input variables and output results - and processes this data to identify patterns and statistically estimate complicated functions. Because they “train” themselves utilizing historical data, machine learning algorithms are then able to take these processed patterns and “learn” to perform some sort of task (i.e., correctly detecting areas most susceptible to non-revenue water loss).

Once this training process has led to a sufficient level of accuracy, the machine learning algorithms test out their results by comparing real time data with its predictive model. When it detects anomalies, the algorithm records and notifies the system operator who can act accordingly.

When it comes to water utilities, there are an enormous number of variables that are positively causated with leaks or likelihood of leaks. Leakage prediction variables include pipe material age, and design, or delivery pressure and total demand. Machine learning algorithms can analyze all of the available data about your water utility, from your internal SCADA system or from geospatial satellite imagery. In short, there is an enormous amount of valuable data that can yield insights only under the condition that it has been matched with a machine learning algorithm.

### Machine Learning Process

As previously mentioned, machine learning can be incorporated into your existing water audit processes in order to dramatically improve the quantity and quality of actionable insights. The process of machine learning works in the following way:

Step 1: Collection and Assessment

[11] https://www.smartresilient.com/using-machine-learning-assess-infrastructure-replacement-needs

Many water utilities have a lot of data, but the sets are siloed in different places throughout the organization. Furthermore, the data may be choppy, low quality, or in completely different formats (e.g., some data may be quantitative and numerical, whereas other data may be more qualitative and descriptive). The first step in the machine learning process is to compile all of your data so as to cleanse and normalize it in a way that is ready for processing.

It is meaningful for water utilities to include as much data as possible, including sets that are externally acquired. This specifically means that satellite or drone imagery should be used if possible, due to the valuable insights that can be obtained from the analysis of the physical environment of water utilities.

Collection and aggregation can be a tricky and manual process, often identified as the more challenging process of the overall development of a machine learning algorithm. However, once it is in place, the machine learning algorithm is ready to work.

Step 2: Analysis

With a coherent data set, the machine learning algorithm “trains” on the information that is provided and learns to predict outcomes based on existing sets of variables.

For example, a machine learning algorithm can use satellite vectors to analyze soil corrosion data,and compare this against historical data containing previous locations of water main breaks or previously identified areas where non-revenue water leakage was detected. Furthermore, soils can be cross referenced against maps of pipe material. The process will uncover potential patterns or correlations between these levels of soil corrosion and the identified leaks. This would be incredibly tedious and prone to error and bias if conducted by solely human operators, and therefore can yield much faster insights about the quality of your infrastructure.

Likelihood of failure (LOF) is a particularly important metric. The machine learning algorithm estimates the likelihood that any given water main or stretch of pipeline is likely to fail over specified timeline. This is an actionable metric that directs your team how and where to operate.

Phase 3: Action Plan

The algorithm creates a viewable and downloadable report that is customized to the needs and priorities of your management team. This is the equivalent of shining a spotlight on your reporting process; the massive reduction in ambiguity can dramatically improve savings, delay or expedite investments, and strengthen the efficiency of your entire organization.

**Benefits and Return on Investment**

Ideally, the previous pages have demonstrated the enormous potential of machine learning for leak prevention and reduction for the modern water utility. Though the demand challenges facing municipal water organizations are extremely complex, machine learning can go a long way in improving the efficiency of systems and saving huge amounts of money along the way.

Water utilities are constrained by the bottom line and must make skilled decisions about their operations and investments. In general, water utility managers are burdened by the enormous financial commitment necessary for investments in new assets. The American Water Works Association estimates that from 2011 to 2050, water utilities will spend $1.7 trillion on repairing and expanding their physical infrastructure.[12]

New investment is unavoidable. But with the likelihood of failure metric, managers can gain a much clearer picture on the infrastructure that does or doesn't need to be replaced. When it comes to water utility asset management, the worst case scenario is to spend millions of dollars replacing a facility that can be fixed with much cheaper repairs.

A report of 155 water utilities in Pennsylvania demonstrated that 327 million gallons of water are lost per day as non-revenue. This is the equivalent of $60 million dollars in lost revenue leaking out of the group of utilities every year. Furthermore, the study estimated that ⅓ of this total could be saved in a cost effective manner, or $20 million a year. For some of the larger utilities within this sample size, an aggressive water loss prevention program could save hundreds of thousands of dollars, if not millions of dollars per year.[13]

This study does not take into consideration advanced analytics programs, so it does not provide an exact answer on the ability of machine learning to improve water utilities - and in fact, machine learning is likely to decrease the cost of identifying leaks while improving the overall scope of detection - likely more than ⅓ of leaks can be addressed

---

[12] https://medium.com/s/ai-for-good/how-ai-could-smarten-up-our-water-system-f965b87f355a
[13] https://www.nrdc.org/sites/default/files/pa-utilities-water-audit-data-evaluation-20170215.pdf

under a machine learning program. From any perspective, it is clear that improvements on the non-revenue metrics yield immediate business benefits to a water utility.

Furthermore, delaying the investment in a multi-million dollar facility is an enormous positive for a financially strained organization. With a clearer idea of LOF rates, machine learning is ideal for producing insights in this sphere.

**Implementation and Reporting**

Now that you possess both a solid understanding of benefits machine learning offers and an outline of what the machine learning process looks like, the only remaining item is the “how:” implementing a machine learning program within your organization.

Machine learning is a highly technical process. With that said, ML produces business insights that can be interpreted and reviewed by anyone on your team. In order to maximize the benefits of machine learning within the context of your decision making, on the input side you need to identify all possible variables that you would like the algorithm to analyze. On the output side, you can customize your reports so that the algorithm is delivering value to your process.

Setting up a machine learning program may seem like a daunting task, but implementation is as simple as scheduling a meeting. During this process, we discuss both your most important and highest-level goals and drill down into specifics on the data your organization maintains and has available, the processes you follow, historical performance in water audits, and more. By the end of our meeting, we will have a clear understanding of the data available and the steps required to begin building and training a machine learning algorithm specific and tailored to your data, processes, and overarching goals, and you will have the confidence and comprehension of what to expect from a reporting perspective and how you will be able to use the results of our machine learning algorithms within the context of your current operations.

Our team sets up automated monthly reports for you that specifically address your present business and infrastructure concerns, customized to your needs. Our reporting is incredibly robust, and we can quickly adjust the reporting format to the demands of your organization. The report is delivered as a dashboard, listing trends, KPIs, points of progress, warning signs and anomalistic behavior, and far more.

The machine learning insights and reporting will quickly become a regular contribution to your oversight meetings, and will improve your overall decision-making process and allocation of resources.