

Operational Analytics: Complex Calculations, Simple Execution

How to Prioritise Response Resources and Deploy Accompanying Action Plans

P Bonk, J Klaric

ABSTRACT

The water industry has entered an era where vast amounts of data are being produced from numerous sensors monitoring the network. This "Big Data" provides potential for far greater integration between operations, engineering and management.

This data, if accessed, configured and analysed in real time will allow utilities, councils and supporting consulting companies create new operational insights in real time. Using this additional information with proactive system management can enable the discovery and solution of the complex challenges confronting the industry with a degree of execution and response that has previously been out of reach.

The paper will discuss a step-by-step approach for operational analytics grounded in the discussion of successful case studies.

Keywords: Analytics, Operational, Big Data, Digital Twin, Artificial Intelligence

INTRODUCTION

Data-Driven Decisions

Numerous data sets from multiple sources provide potential for a higher degree of integration between operations, response teams, planning and management. Rapid growth in data collection is a progressive step for organisations however, the shortcoming to this success lies with the inability to articulate and interpret the data in real time for decision making purposes. By the time live datasets and small data analysis (spreadsheets, etc.) are manually combined and analysed, their value is greatly diminished. In the real world it's the difference between a system alert providing early identification by the utility of an

issue rather than relying on a customer or public complaint reporting a problem often well after the event has occurred.

When staff from all segments of the utility can access live data, water, sewerage and drainage networks will in effect be monitored by many more staff whose different skill sets can be used by way of stored logic, sampled and derived data feeds, and templated and scalable calculations to greatly improve operations and system management.

Design. Automate. Offload. Action

Operational analytics is the process of continuing to improve business operations using information from data sets. In pursuit of proactive decision making, operational analytics allows for prioritisation of response resources & deployment of accompanying action plans. The operational analytics and artificial intelligence web based applications, Info360 & Emagin by Innovyze, to be discussed in this paper, contain workspaces designed to be utilised on a daily basis. These personalised workspaces include mapping real time or historical data from geospatially located sensors and alert mechanisms, visualisation of data through charts and built-in analytic displays, pre-designed visuals and metrics to monitor sensor health and data quality. These provide deeper insights on the operations and performance of water, sewerage and drainage networks.

Informing the Info360 workspaces, shown in Figure 1 for network insights & subsequent action, is a data modelling application which receives, in this example, multiple categories of live data (automated meter reading/automatic meter infrastructure, SCADA, water quality, demand, flow, pressure, level, pump data & hydraulic model historical/predictive runs). As well as network information, individual customer consumption from digital meters can be included to calculate live zonal demands, for comparison against fundamental aspects of networks from reservoir levels, pump operation and source flows.

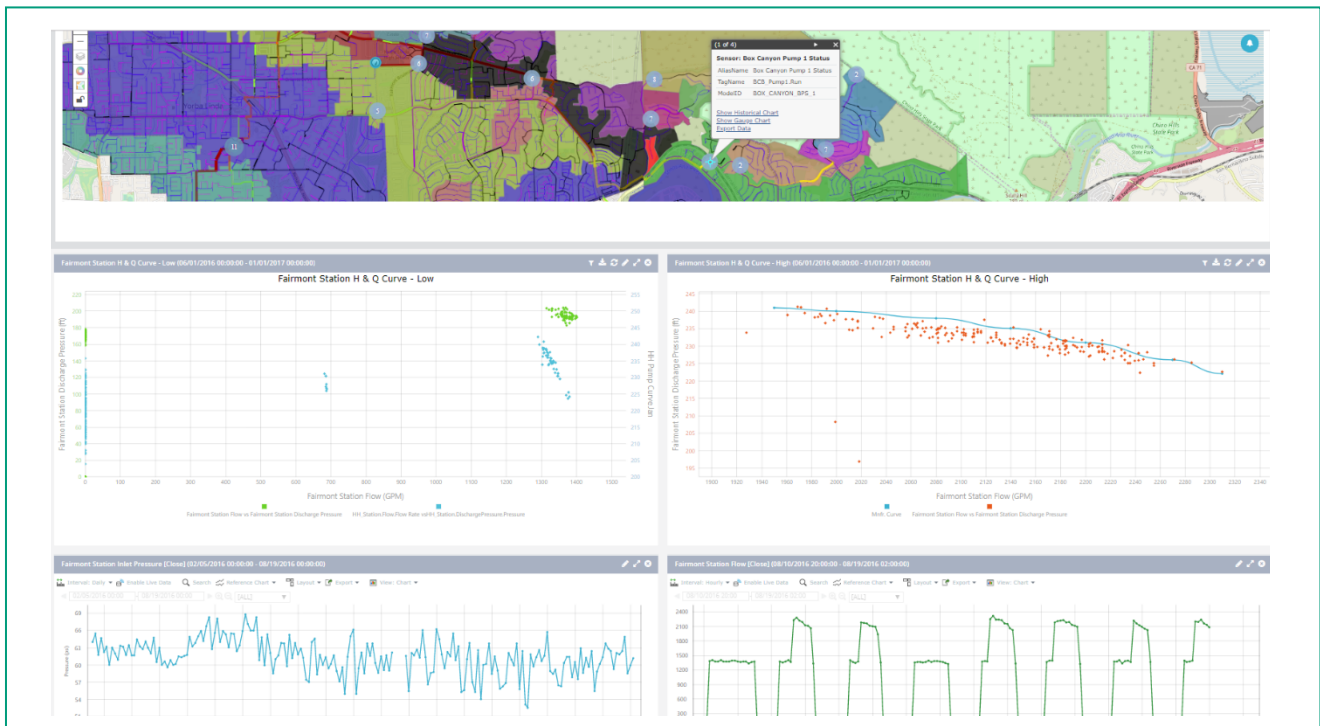


Figure 1: Info360 by Innovyze, workspaces showing multiple geospatially located assets & data tags built on a data modelling application which consumes multiple categories of live data (AMR/AMI, SCADA, water quality, demand, flow, pressure, level, pump data and hydraulic model historical/predictive runs) for water, sewerage and drainage network insights and subsequent action.

The automated data modelling and analytics engine constantly samples across the available data record to create a “Time Series model” for each sensor to capture the properties and behaviour of the measured data stream. The application is source agnostic, designed for operational analytics to be a fundamental aspect of proactive decision making which provides an immediate response to a system event. When action is necessary, operational analytics provides the newfound ability to decide where, when, why and how an operations, field and engineering team react and respond to system issues as they arise.

Utilise Internal and External Expertise

More data and dashboards are not enough for behavior change and adoption by users in their every day practices. The intent of this paper is to methodically demonstrate ways in which operational analytics are within every organisation's grasp utilising their existing skillsets, data sets, IT infrastructure and software tools. The paper will

discuss a step by step approach for operational analytics grounded in the discussion of successful case studies.

OVERVIEW OF SYSTEM BENEFITS

An overview of the system benefits found thus far include:

- Sensor health; reporting on data quality, total uptime and notification of stuck/failed sensors.
- Demonstrated cost savings relative to baseline operations with return on investment realised within months.
- Event detection and management; the early detection of system or component failures using simple analysis with basic statistical functions (variance equations) to more advanced analysis techniques (Fourier transforms).
- Proactive identification of leaks and bursts; by applying industry best practice analysis metrics such as real loss metrics and Infrastructure Leakage Indexes (ILI).

- Active System Operation; detection of apparent losses such as unauthorised consumption, data transfer errors or data analysis errors can be discovered and addressed.
- Better customer service; improving response time to bursts, pressure fluctuations and changes in dosing regimens.
- Anomaly detection; simple statistics and/or pattern recognition can be built in to identify system anomalies. The application allows the user to build calculations to successfully access data and analyse it in either real-time or historically.
- Increased reliability of data; templated approaches and methodologies can be created so that data is utilised for decision making as opposed to efforts solely directed on data cleansing.
- Moving beyond solely human judgement, utilising artificial intelligence to make predictive recommendations to operations and management on complex operational decisions.

Technology Drivers

Overall, operational analytics are being driven by:

- Organisations collecting large amounts of timeseries data yet often failing to realise the information contained within the collected data.
- Increased collaboration and communication between operations, response teams, engineers, management and board level decision makers.
- Cost savings from increased energy efficiency and effectiveness of daily system operations.
- Automation and daylighting of routine calculations and analysis measures often managed “out of sight” by a single engineer.
- Streamlining and standardising internal processes for teams with multiple stakeholders and needs.
- Understanding how planned changes in the system will affect customers.
- Identifying events that may be occurring in the system unnoticed.
- Quantifying how the system can be operated more effectively.
- Determining what impact each system is having on another.
- Prioritising current and future data collection.
- Improved resiliency of network ensuring key system constraints such as service reservoir volumes are maintained continuously over daily operational cycles.

Overcome Organisational Siloing

Often internal groups operate in isolation with their own data sets, shown figuratively as a ‘siloing effect’ in Figure 2. Whether one works within operations, response teams, engineering, planning, management through to board level decision makers, workspaces can be created to visualise, calculate, and report network performance based on a stakeholder’s unique data needs, timelines and responsibilities.



Figure 2: Stakeholders within an organisation experience separation based on lack of access to data and their respective unique data needs.

Unique data needs of key members of a water authority can be accommodated within an operational analytics platform. This could include:

Operators & Response Teams

- Event detection and response
- Understand loss of system integrity
- Asset performance and live levels of service
- Optimised controls
- Updating operating manuals.

Engineering

- System design and behaviour
- Event forecasting
- Understanding demand
- Leakage and non-revenue water

Management

- Non-revenue water
- Developing key performance indicators
- Collaboration amongst the team
- Knowledge management

Directors/Board Level

- Risk management
- Regulatory compliance
- Cost/benefits analysis

METHODOLOGY - TECHNOLOGY DISCUSSION

Source and Configure Real-Time Data Capture

Smart meters, Automated Meter Reading and SCADA are utilised to acquire large amounts of data which provides consumption metrics (authorised and unauthorised) of water supplied, variability in pressures across multiple levels of zones, asset performance in real-time and over periods of time to name a few.

As with all real-time data analytics, data streams must be assessed for quality, reliability, and suitability for appropriate metrics. As part of this, all SCADA managers, data analysts, and engineers will need to determine the availability of their data, frequency of data-dropouts, repeats or other systematic anomalies present in all real-world data. Data availability and reliability issues are an intrinsic part of any 'live', real-world data collection system.

With these inherent issues, Sensor Health, also known as a way of measuring data quality, can provide feedback and a better understanding of the data quality itself to ensure designers, operators and maintainers of SCADA systems may have this added level of confidence in the data they are collecting and providing for decision making processes within their organisations.

Figure 3 provides a simple example of how a user could demonstrate the uptime and downtime of sensors in a network. The summary chart was created via the software's built-in visualisation functionality and added to a user's workspace.

Dashlets are customisable visualisation tiles that utilise built-in and configurable calculations on the live and historical data. The dashlet shown in Figure 3 calculates the approximate amount of hours of downtime for a group of sensors. The calculation contained in this dashlet looks at each 30-minute block (over the past 31 days) for either no received signal or a constantly repeated signal and presents the number of half-hour blocks as a measure of downtime.

Uptime and Repeat Summary - All Pressures		
ID	Result	
DMA081.CPP.Press	740.5	
DMA009.CPP.Press	740.5	
DMA108.CPP.Press	740.5	
DMA116.PR.V.Press.In	740.5	
DMA116.PR.V.Press.Out	740.5	
DMA12.CPP.Press	740.5	
DMA007.CPP.Press	740	
DMA029.PR.V.2.Press.Out	737.5	
DMA081.PR.V.Press.In	724	
South DMA 81 PRV Pressure Analog	720	
DMA10.PR.V.2.Press.Out	671.5	
DMA097.PR.V.Press.Out	635	
DMA10.PR.V.2.Press.In	610.5	

« < ... 1 2 3 4 5 ... > »

1 of 29

Figure 3: A dashlet displaying a sortable summary list of sensor uptime for selected sensors within a network.

Dashlets, added by the user to their respective personal and or sharable workspaces, is a customisable and configurable view that can display network information in a range of formats such as charts, tables, mapping, alert summaries, scatter graphs, external website content, images, etc.

Sortable lists provide users with better information to prioritise maintenance schedules, targeting the longest-outage or most critical sensors first for repair or replacement.

A high level of analysis is readily available to a first-time user. Built-in mathematical functions within each dashlet, such as those shown in Figure 4, can be applied to timeseries data to provide a broad range of information such

as volume through a zone, rates of change over different periods, or even simply moving averages to smooth out spikes in the data.

The screenshot shows the 'New/Edit Dashlet' configuration window. The fields are as follows:

- Dashlet Name(*):** Uptime and Repeat Summary - All Pressures
- Description:** (empty)
- Group:** [ALL]
- Dashlet Type(*):** Table
- Data Source(*):** Pressure
- Sensor Type:** Pressure
- Expression(*):** $0.5 * IIF(Uptime(1) < 50 \text{ or } CountRepeat(48,6) > 24, 1, 0)$
- Data Interval(*):** 30 minute
- Group By Sensor ID:**
- Group Data By:** None
- Apply Function to Each Group:** Sum
- Time Range(*):** 01/01/2016 00:00:09 - 20/11/2018 00:00:09
- Days Offset:** 31
- Display Unit:** (empty)

Buttons at the bottom right: Finish, Close.

Figure 4: Configuration of dashlets allows users to add simple and/or complex mathematical expressions for deeper levels of analysis.

Dashlets actively query the raw data stream at pre-set intervals (e.g. half-hour intervals) and sum the total number of intervals during which no data was received from individual sensors within a group.

Dashlets, as shown in Figure 3, sort and display the data's availability summations within a sortable table for ease of establishing the largest contribution to overall downtime within a SCADA system, and thus can help prioritise maintenance and replacement decisions.

Once the quality of the overall data has been established, it is then up to the user to decide on what data sources are to be used – for example, raw sensor data such as a flow or pressure sensor on a particular pipe, or derived data, such as net usage through a network or district metered area or averaged data from multiple sources to construct a pump efficiency curve.

Beyond solely summary charts, dashlets can provide colourful and comprehensive images such as pie charts,

shown in Figure 5, informing the total uptimes of all the flow sensors for a network over a given historical time period (day, week, month, year, etc.). Graphical representations of uptime provide a quick ‘eye-test’ based on any unusual variability between chart segments relative to remaining sections of the pie-chart. Users would immediately be alerted of data quality issues over the preceding day or week prompting them to investigate the root cause before basing critical decisions on measured data.

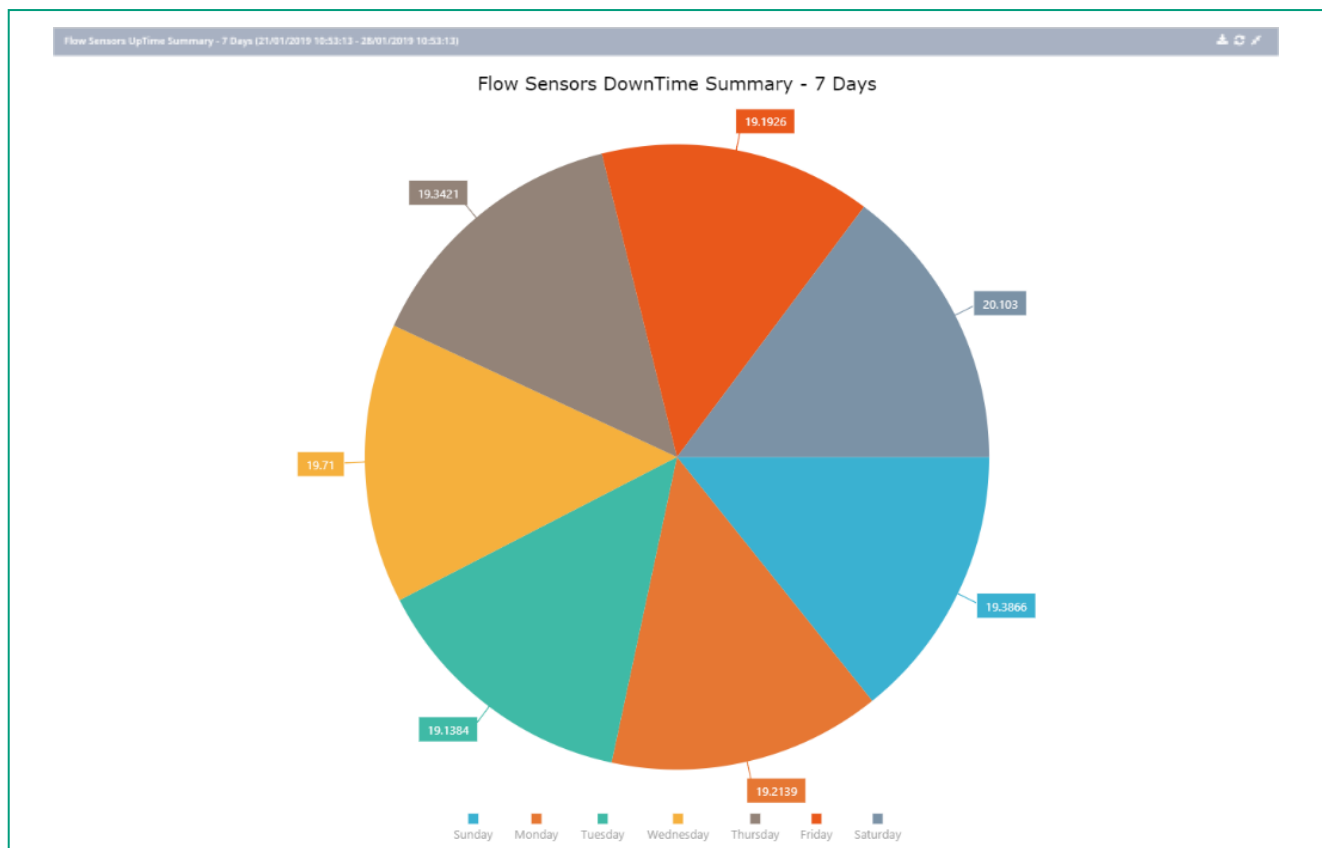


Figure 5: Pie chart dashlet of flow sensor total uptime for previous seven days.

Being able to drill down into which particular sensors are contributing most to total uptime/downtime can be useful on a day-to-day operational perspective.

Figure 6 shows a bar chart dashlet for a data drill down on a single sensor with daily uptime (X-axis is day of the month) on an hourly interval (Y-axis average daily uptime).

For a greater overview of sensor performance and data quality beyond individual dashlets and single metrics, pre-built and designed sensor health workspaces are available to users; accessible from the software’s home page, once logged in to the application.

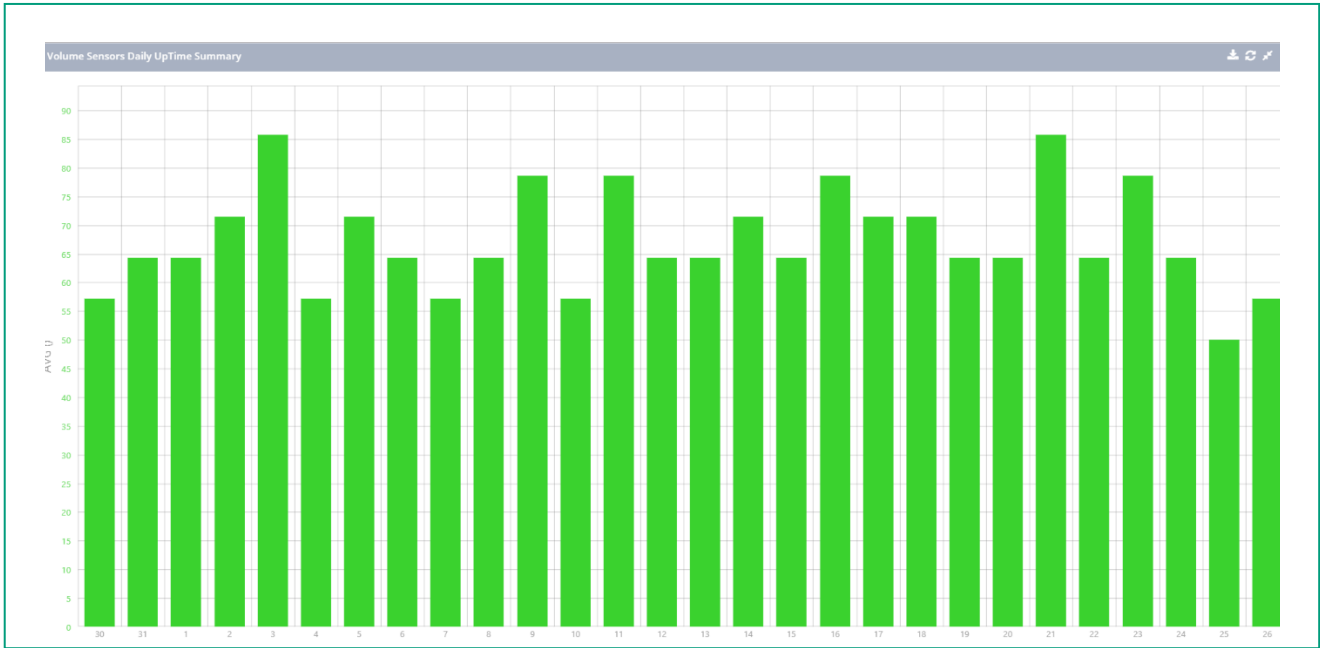


Figure 6: Bar chart dashlet viewing a single sensor’s average daily uptime on an hourly interval (X-axis; day of the month, Y-axis; Average daily uptime – hourly interval).

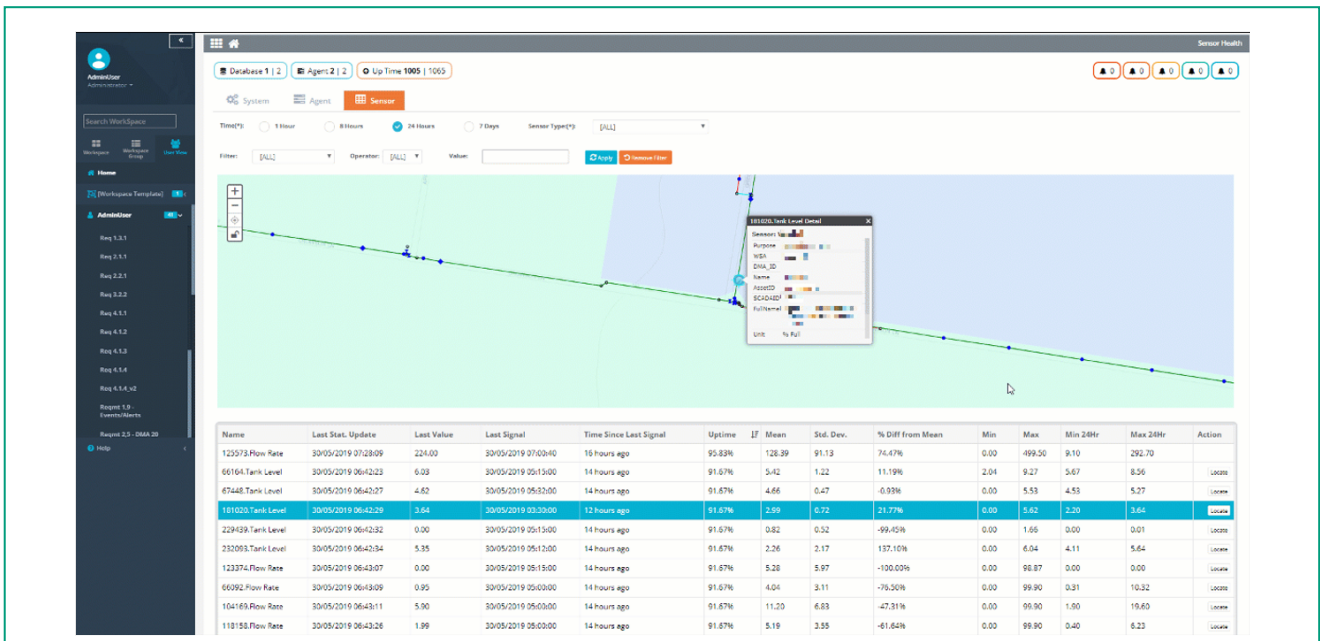


Figure 7: Sensor Health and Data Quality Workspaces pre-designed within the software and available to users on log-in brings to light gaps in live and historical data, up-time of sensors and informs the prioritization of maintenance schedules of sensors; all geospatially located, viewable on a workspace.

Figure 7 shows a screenshot of the information available to users upon login and connection to data sources.

In short, a sensor health analysis quickly and effectively brings to light gaps in both live and historical data, uptime of sensors, informs prioritisation of maintenance schedules of sensors as well as assisting the identification of 'where' and 'when' for future data collection locations.

Sensor health, the quality of an organisation's data, is now available for the user to actually understand the quality and dependability of its decisions with a known degree of confidence.

Identifying and Applying Key Performance Indicators (KPIs)

Deploying key metrics and insights beyond engineering services involves an application with workspaces open, configurable and viewable by varying levels of analysis and technical expertise specific to the needs of the respective stakeholder – operators, engineers, field crews, management, and key decision makers.

The operational analytics application allows the user to build complex calculations within built-in analytics tools. Scripting and SQL (Structured Query Language) skills are not required for the user to successfully access data in real-time and build workspaces based on their unique needs.

Configurable calculations, or automated workflows, are essential for complex calculations that have traditionally been built within spreadsheets and remained unavailable to stakeholders beyond teams with highly technically oriented skillsets such as engineering services.

Outlining a few of the examples of key configurable calculations and KPIs utilised within the application thus far include:

- Sensor health; reporting on total uptime and notification of stuck sensors.
- Sensor uptime; providing a sortable list of problematic sensors based on pre-defined criteria.
- Sensor drift; identifying variations from normal system behaviour.
- Sensitivity analysis; identifying false negatives and false positives, such as detecting a burst that did not occur (i.e. assuming a swimming pool being filled when a burst has occurred).
- Burst detection and reporting; using anomalous flow detection against live/historical usage profiles, highest usage instances, against flushing occurrences, misplaced versus lost volumes.
- Identifying closed valves; using anomalous pressure values to identify valves left in a closed position after maintenance
- Pressure Reduction Valve (PRV) drift - % spread for measure of control.
- Pump efficiency; calculations outputted and displayed with scatterplots to observe pump performance.
- Identifying high/low pressures; via data enveloping.
- District Metered Areas (DMAs) usage; display of usage verses time against all DMAs.
- Boundary breach versus sensor issues.
- Identifying data faults; highlighting data that is inadequate for network analysis.
- Mass balances; using billing data to conduct mass balances.
- Critical pressure point performance; comparison of Critical Point Pressure (CPP) to Decision Support System (DSS) to determine drift from normal system behaviour.
- Targeting system losses; quantifying monthly trunk main losses.
- Current Annual Real Losses (CARL); Unavoidable Annual Real Losses (UARL) and International Leakage Index (ILI) calculations for whole of network and DMAs.

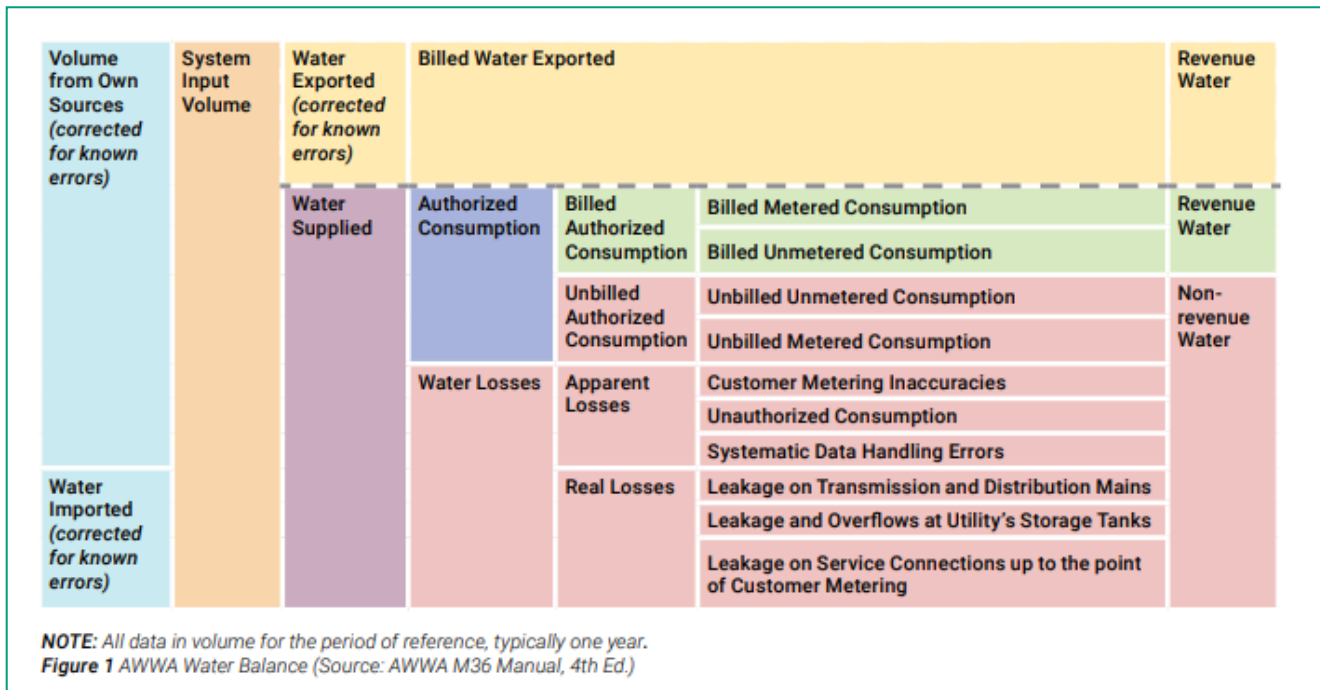


Figure 8: American Water Works Association’s (AWWA) Water Balance for authorised consumption, sources of revenue water, non-revenue water and categorizes the various sources of water loss (Source: AWWA M36 Manual, 4th Ed).

Figure 8 shows the American Water Works Association’s water balance for authorised consumption, sources of revenue water and non-revenue water and categorises the various sources of water loss. Metrics such as Infrastructure Leakage Index (ILI) is an increasingly accepted method for water loss audits globally. ILI is calculated using Current Annual Real Losses (CARL) and Unavoidable Annual Real Losses (UARL), with both metrics being commonly used for water loss audit reporting.

Demonstrating how the configurable calculations and automated workflows are represented within the operational analytics software, Figure 9 shows the workflow for CARL and Figure 10 progresses further through the respective workflow to demonstrate how UARL is represented within the software. Key water balance metrics, calculated and completed for one section of the network, may be templated, mimicked and scaled across to other pressure zones and DMAs and utilised at the level of whole of network, providing a streamlined and standardised approach to building and applying key network performance metrics.

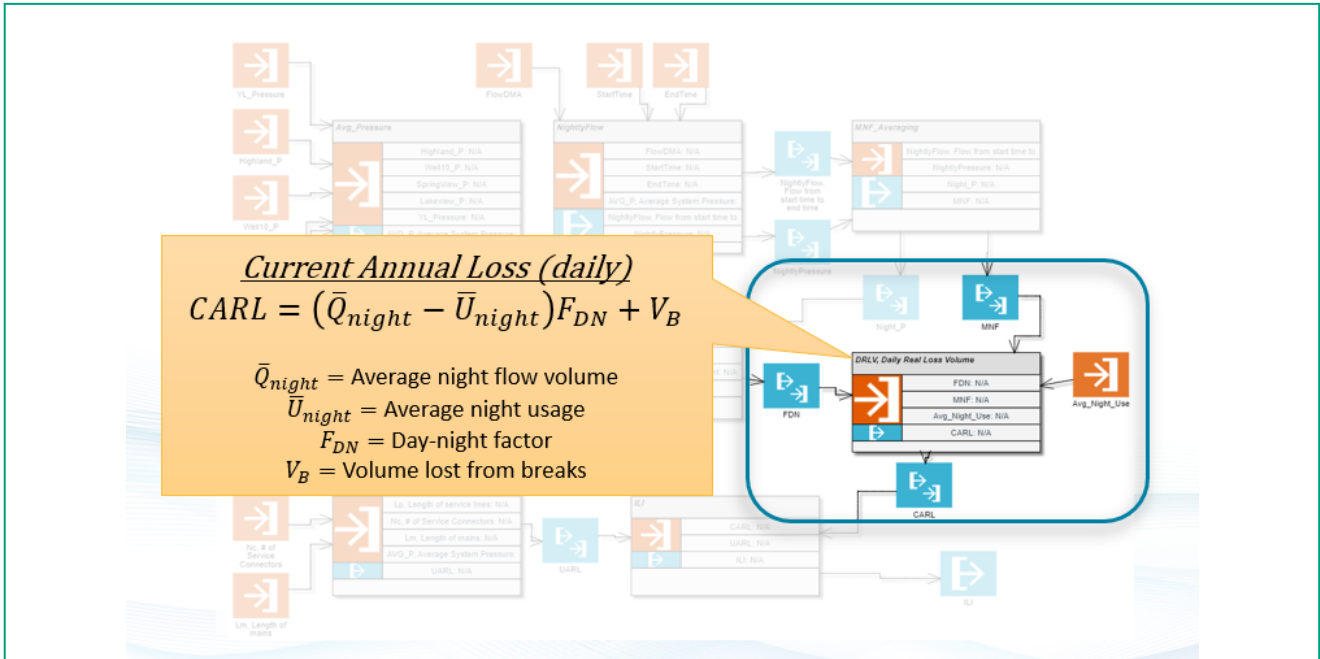


Figure 9: Calculations such as Current Annual Loss (daily) embedded within the Operational Analytics data modelling tool (source for CARL calculation: C. Lenzi et al. Procedia Engineering 70 (2014) 1017-1026) for streamlined, standardised and daily trending for real-time and historical analysis.

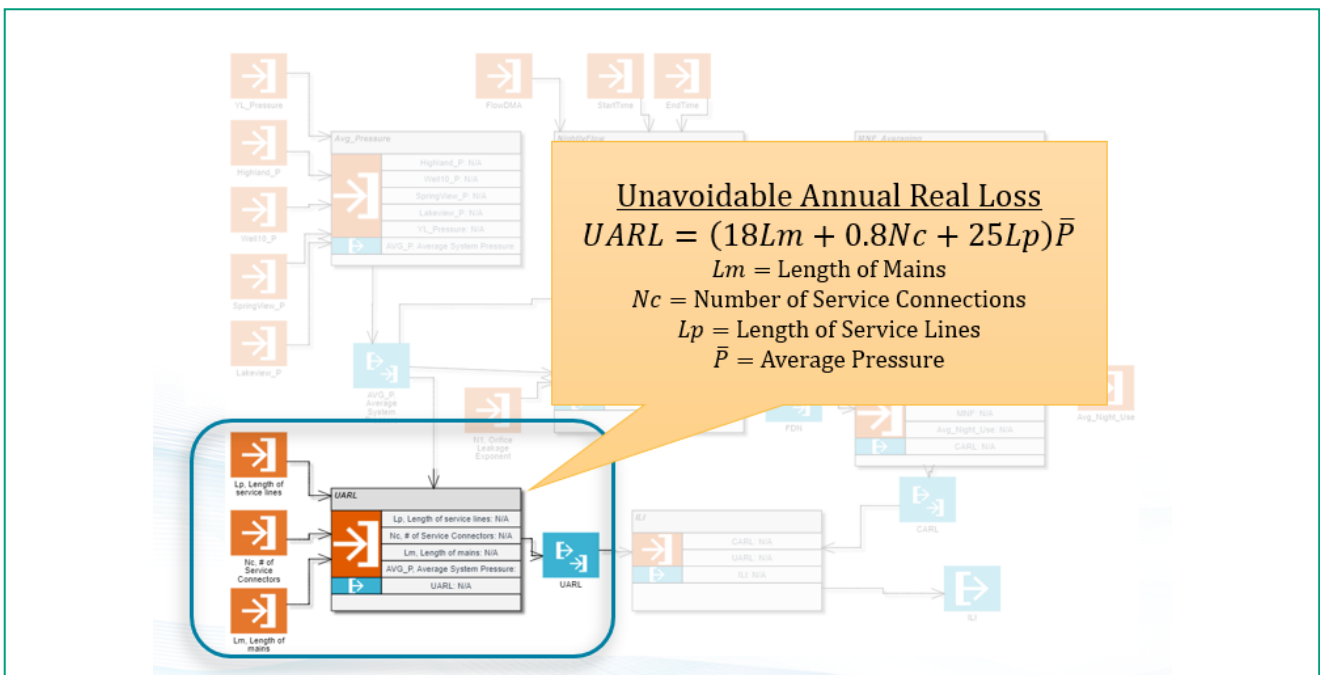


Figure 10: Progressing further through the calculation schematic, Unavoidable Annual Real Loss (daily) embedded within the Operational Analytics data modelling tool (source for UARL calculation: C. Lenzi et al. Procedia Engineering 70 (2014) 1017-1026) for streamlined, standardised and daily trending for real-time and historical analysis.

Analysis for Operational Insights

Individual workspaces are built to visualise, calculate, inform network operations and report network performance based on the unique needs across the organisation from operations, response teams, engineering, planning, management through to board level decision makers.

Once key KPIs have been identified and applied, the quality of the operational insights will be directly related to the quality of the data itself. Relating results to data quality is paramount for actual insights.

"We identified a sensor was down for 6 months, yet we've been relying on it for pressure values to make decisions on our networks." (Large Council, Australia)

With data quality analysis in place and the generation of key KPIs, metrics relating to leakage, water balance and loss, pipe breaks and storage once typically calculated on an annual or quarterly basis can now be calculated on a sub-daily basis; allowing for trending analysis of key metrics that have not been available without time consuming and labour-intensive efforts. When analysing operational insights, the KPIs may now be applied on a sub-daily basis, to larger scale, whole of network analysis for event detection and system performance.

Leakage performance may now be calculated for every hour of the day and each zone in the system. The sub-daily data may now be plotted in real-time for insights into how water loss is being addressed and to further detect system anomalies that may indicate pipe breaks, water theft or malfunctioning valves.

To progress operational insights, event management analysis may be applied to the individual sensors or at the DMA level. On an individual sensor, simple statistical functions such as variance or rate of change may be applied to detect anomalies and events at that specific location. Yet as analysis scales up to the higher network level such as a respective metered zone or to a given DMA, network anomalies are harder to discern and differentiate. At the zonal and DMA level, a network sensitivity analysis must be conducted and considered to account for the wide variety of occurrences in a system that regularly contribute to deviations from typical system behaviour. Key considerations include:

- A pipe burst and a swimming pool filling can both be identified with a metric such as variance. How can one differentiate between regular occurrences and events requiring rapid triaging of operations and response crews?
- A pressure value is drifting, a flow value spiked, or a tank level dropped, why and what does this mean?
- What's the sensitivity of the statistical functions applied to each event?

- How do I determine a false/negative versus a false/positive?
- How does one account for the trade-off between "detecting every occurrence" and "detecting essential events?"
- How does one optimise while considering key system constraints such as service reservoir levels and volumes?

Developing event detection analytics for large areas can seem complex at first thought, and certainly highly complicated mathematical and statistical models can be built, yet improved understanding of system behaviour can be achieved with simple analysis techniques.

One relatively simple method for identifying very large anomalous usages is to determine a 'high usage threshold' for each hour of the day, and set alerts to detect when the usage within a particular hour exceeds such a threshold.

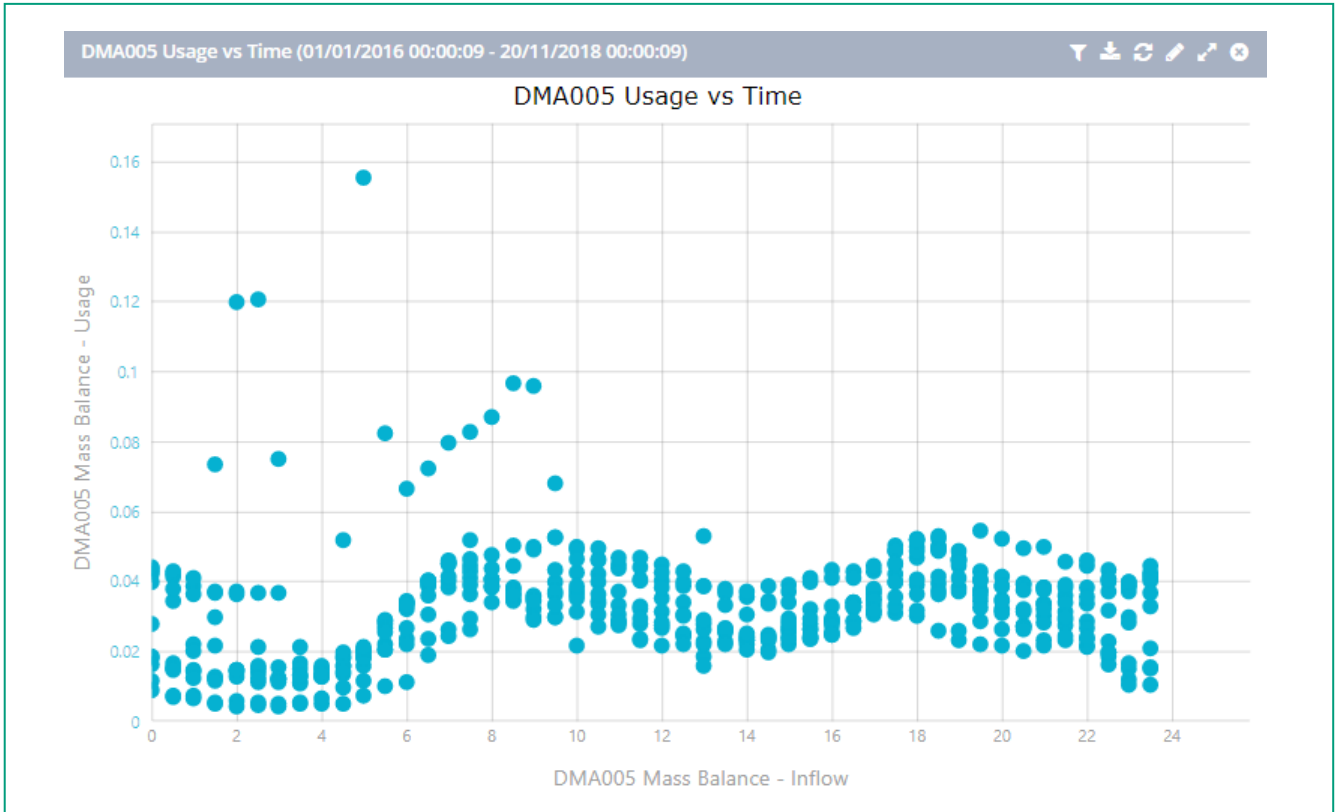


Figure 11: Scatterplot dashlet of DMA usage versus time for a 24-hour period used to identify large anomalous usages.

Figure 11 shows a scatterplot of usage versus inflow for a given DMA averaged over a 24-hour period.

The respective DMA 24-hour diurnal usage profile derived by plotting inflow against usage provided the standard system performance to apply additional statistical functions with the aim of extracting events of significance.

Figure 12 shows a summary table where a statistical function, within the dashlet functionality of the software analysed where the 95th-percentile and two-standard deviations above the mean usage for each hour of the day was calculated for a particular DMA. Furthermore, the applied statistical function identified usages above set thresholds for longer than two consecutive hours for a deeper differentiation of classification as an 'anomalous event'.

Statistical analysis can incorporate rules to separate out anomalous events deserving the attention of operations and response teams. The usage during these anomalous events was subtracted from the average usage for the respective hours for the time spent above the thresholds and summed to give a list of approximate volumes lost during these 'events', shown in Figure 12.

The list was then compared to known historical events that had occurred within the DMA over the last few years and the volumes generally agreed (sometimes even to within 10% of the estimated volume of the known event) with actual system events such as a pipe burst incidents and routine pipe flushing occurrences.

The usage pattern applied in this case was used to filter out the sensitivity of a given analysis in alerting potential false negatives and false positives.

Figure 13 shows a dashlet with built-in metrics to detect and determine typical usage behaviour versus an actual event such as a burst for a given pressure zone.

In short, progress can be achieved with a simple variance function for analysing anomalies across a single meter or increasing the complexity of analysis by applying Fourier transform functions and advanced statistical or data science techniques at the DMA level to separate out deviations in system behaviour. In relation to complexity of approach, a 'spectrum of analysis' can be applied for better system understanding.

DMA20 Event Detection Summary (01/01/2018 00:00:30 - 05/12/2018 21:53:30)			
Date	Sensor ID	Value	
24/03/2018	DMA20 Cumulative Loss.Volume Misplaced	2.7027	
12/06/2018	DMA20 Cumulative Loss.Volume Misplaced	0.6777	
11/06/2018	DMA20 Cumulative Loss.Volume Misplaced	0.6338	
10/06/2018	DMA20 Cumulative Loss.Volume Misplaced	0.5092	
09/06/2018	DMA20 Cumulative Loss.Volume Misplaced	0.4258	
23/10/2018	DMA20 Cumulative Loss.Volume Misplaced	0.3309	
30/08/2018	DMA20 Cumulative Loss.Volume Misplaced	0.3026	
27/07/2018	DMA20 Cumulative Loss.Volume Misplaced	0.3015	
23/03/2018	DMA20 Cumulative Loss.Volume Misplaced	0.1903	
26/07/2018	DMA20 Cumulative Loss.Volume Misplaced	0.1744	
25/01/2018	DMA20 Cumulative Loss.Volume Misplaced	0.053	
08/06/2018	DMA20 Cumulative Loss.Volume Misplaced	0.0277	

Figure 12: Tabular dashlet showing a sorted summary from greatest cumulative volume lost for events beyond mean usage over a 24-hour period for a respective DMA.

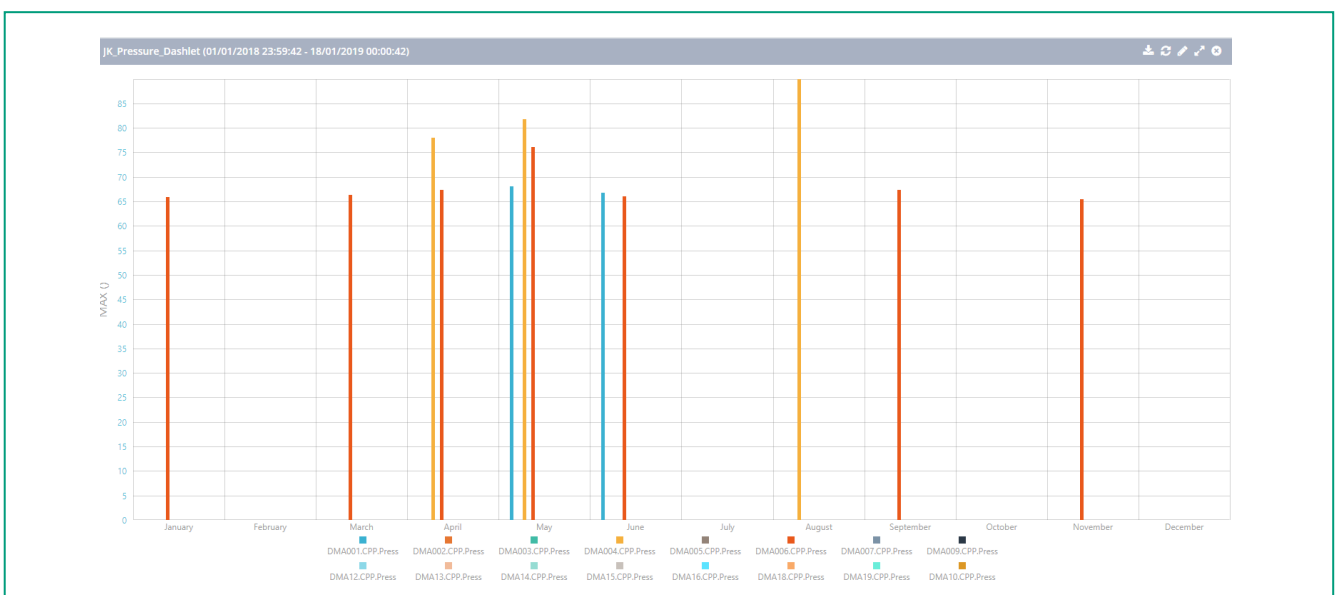


Figure 13: Dashlets with built-in metrics to detect and determine typical usage behaviour vs an actual event such as a burst for a given pressure zone.

A User Scenario: Utilising External Expertise

A local Council has aspirations of gaining a deeper understanding of its network by completing an advanced event management analysis of its network utilising the latest data science and statistical methods available to the water industry.

Lacking the internal expertise and experience to take on the complexity of advanced event management analysis, an external data science consulting company is sourced to collaborate with the local knowledge of Council's engineering services team.

Working alongside the consulting data scientists, Council engineers, utilising their intimate understanding of their respective network, provide their prior years flow and pressure data along with known locations of historical incidences. Collaborating with their in-depth knowledge of historical network operation practices, the engineering team work with the data scientists to relate network asset locations to the timeseries data sets handed over to the consulting team.

Statistical analysis metrics; successfully applied to create differentiation between standard system behaviour, insignificant network anomalies and events of consequence such as actual bursts are added to the third party operational analytics platform. The analytics platform enables data to be made visually available and geospatially represented to Council's engineering team via the application's dashlet and mapping features within its respective web-based workspaces. The analysis developed by the consultants are handed over within templated workspaces, saved and applied to multiple live disparate data feeds for future event detection purposes. For scalability, the saved analysis is available to be replicated and applied to remaining sections of the respective network.

Traditionally, data once available solely to Council's engineering services has now been deployed with the latest event detection analysis techniques to operations with web-based workspaces viewable within their control centre for the application of pattern recognition on events as they happen, to field crews on their handheld tablets to geolocate an incident shortly after it happens. In parallel, the potential exists for network incidences to now be escalated to management for visualisation of event happenings and prioritisation of resourcing, with pre-configured KPIs available for post event diagnosis and assessment.

In pursuit of operational insights, whether through internal expertise or external consultants for supplementing engineering and data science capacity, the respective operational analytics platform allows for greater depth and detailed analysis to be applied, saved, standardised,

reused (mimicked), streamlined and scaled across to other parts of the network.

An application open to the industry and market forces ensures a continual pursuit of innovation and ability to integrate the latest data science and artificial intelligence techniques for system event management and leakage analysis.

Insights Utilised for Action

New insight and understanding becomes available once the vast array of data is accessed for action-oriented system management.

From reactive to active system operation; asset management and leakage management programs, apparent losses such as unauthorised consumption, data transfer errors or data analysis errors can now be discovered, addressed and solved.

Insights achieved from configurable calculations are now available in automated workflows beyond traditional spreadsheet methods and siloed engineering services team members expanding the system operations and performance insights to operations, SCADA managers, response teams and management observing KPIs on network performance.

Automated workflows available to the business include:

- Sensor health analysis brings to light gaps in live and historical data and uptime of sensors, informs prioritisation of maintenance schedules of sensors as well as the prioritisation of 'where' and 'when' for future data collection locations.
- Calculations available once a year are now available on a daily or sub-daily basis such as Current Annual Real Losses (CARL), Unavoidable Annual Real Losses (UARL) and daily Infrastructure Leakage Index (ILI) for performance metrics on water losses across a network at various supply level fidelities.
- Geospatially located trending analysis to relate measured system phenomena to known network events with notification of a system event for deployment of crews to the right place and prioritisation of available resources.
- Provide reliable levels of service while balancing the costs of pump and tank operations.
- Observe longer timescales where action over time is required i.e. "month on month, DMA pressure performance is getting worse."
- Pump health analysis – operate pumps in a uniform way to reduce deterioration.
- Work towards differentiating the thousands of SCADA alarms generated on a yearly basis with the aim to avoid the significant false positives from being buried amongst the mass of alerts.

- Minimise the cost of operations; predictive pump scheduling utilising artificial intelligence to make predictive recommendations to operations and management on complex operational decisions.

A User Scenario: Design, Automate, Offload, Action

A utility renowned for its industry leading example of pursuing more intelligent water networks prides itself in its vast live data collection efforts. Internally, this prominent water enterprise has started to realise the growing amount of collected information does not necessarily lead to strategic and actionable decisions.

Shifting its focus from data collection to event management analysis, the respective utility employs a small team of engineers to work with its resident data scientist internally to conduct a retroactive analysis of historical events that have occurred within their network.

Extending the retroactive event detection analysis, the respective utility implements an operational analytics platform with geospatially located sensors, alert mechanisms and incident visualisation through built-in analytics dashlets to discern where, when, why and how an operations, field and engineering team interpret, react and respond to system issues in real-time as they arise.

The retroactive analysis revealed a major event which occurred last autumn; a catastrophic pipe burst. The event occurred in the early evening and went unnoticed for 14 hours eventually being detected by a ratepayer who contacted customer service and asked; “why is my street under water?”

With a major burst going unnoticed an entire street became flooded over the course of the event. The said retroactive analysis by the engineering team revealed a final volume lost from the network of approximately 3.0 ML of water.

Utilising a relatively simple statistics technique, the retroactive review picked up the burst in the first 2 hours which would have saved 80-90% of the burst volume water lost. Had operations been notified; preventative measures could have been taken to mitigate asset costs from the damages associated with flooding an entire residential street.

With the operational analytics platform deployed in a live environment for future burst occurrences, the web-based workspaces with live network overviews and incident summaries within the software may now be used by operations, response teams and customer service to mitigate, prevent damages and reach out to customers prior to the immense frustrations that are felt due to damages to their neighborhood and homes.

More data and dashboards do not always lead to more confident decisions. Actual meaning derived from available data will inform action from a live decision support tool. Workflows must design, automate and offload on behalf of the user for behaviour change and user adoption to occur.

SUCCESS STORIES

The outlined methodology was applied to the following successful case studies.

Pipe Burst Detection & Tank Level Variability – Small Council, Northern New South Wales, Australia

Data was provided from a small northern NSW Council for an early Saturday afternoon pipe burst in a residential area that went undetected by council staff for a period of 4 hours until called in by the public. Data provided to Innovyze included 5 days of tank outflow data leading up to a historic burst, (residential zone supplied only from flow monitored tank). Figure 14 shows a set of the flow data provided for the historical analysis of the pipe burst detection study.

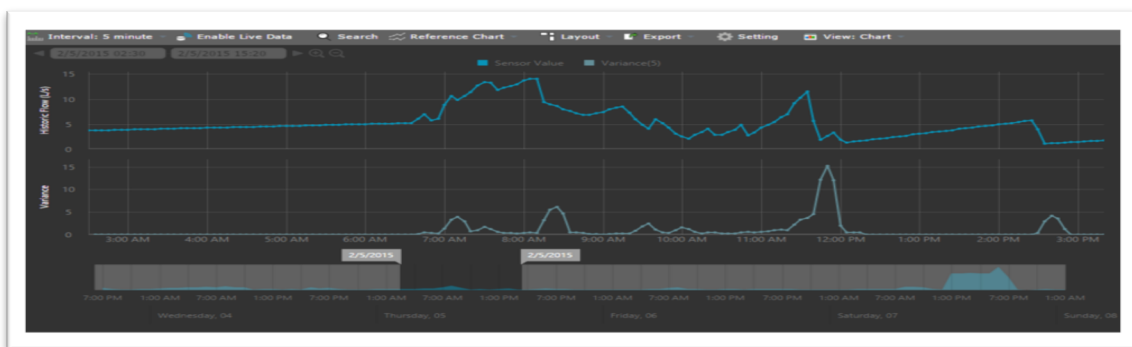


Figure 14: Residential flow data provided for historical analysis on the pipe burst detection study.

Simply using a dollar per litre lost approach, \$2,000 would have been saved as the alert detection reduced response times to 10 minutes.

What was “seen” by operations? When Council forensically investigated the burst, operators recalled:

- Tank level dramatically dropped; no SCADA alarm generated as it was within normal operating bands.
- When tank level SCADA alarm was triggered, the cause was thought to be failure of the pump supplying the tank rather than a pipe burst in the zone supplied by the tank.



Figure 15: Burst detection using changes in the variance of SCADA data received from sensors was an effective method of detecting status changes within a water distribution network, including the detection of both leaks and bursts even within a sampled data environment.

Figure 15 shows the flow data provided with a derived time series (refer to Equation ‘1’ below) utilising the variance function for event detection. A brief review of the 5 days leading up to the burst with the variance timeseries shown in Figure 15 revealed that for normal system behaviour the variance did not exceed a value of 15.

Whereas when utilising the same variance equation (Var, 5) through to the pipe burst incident, the derived variance time series drastically spiked to a value upwards of 600 indicating a burst had occurred.

The following variance function defined in the equation below, where “μ” is the average of the dataset and xi the individual readings, was applied to a total 7 days of flow data with 5 days of flow patterning leading to the burst assumed to represent typical zone demand.

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2,$$

Concepts demonstrated by this study include:

- The Council had recently implemented SCADA but the data historian pulled data from SCADA on a daily schedule at midnight. This meant that all data analytics performed by the historian were useless to the control room operators.
- Existing IT infrastructure and skillsets were used to build the search, track and scheduling mechanism that successfully detected the burst from a historic time series.
- Built in queries and search functionality; therefore, scripting or SQL syntax skills were not required to analyse historical data and/or build alerts.
- No specialist SCADA implementers required to setup alerts.
- Timeseries data accessed by all (secure – read only access).

Live Data Saves Money – Yorba Linda, California USA

Yorba Linda Water District (YLWD) has 25,000 potable water connections, serving residents in the hilly terrain of Yorba Linda and parts of Placentia, Brea, Anaheim, and Orange County in California, USA. The District imports

about a third of its water, supplementing its groundwater supply. Yorba Linda is utilising operational analytics to monitor its water usage and asset efficiency. YLWD is starting the process of calculating non-revenue water (NRW) and leakage management calculations such as International Leakage Index (ILI), validating pump curves, monitoring tank levels and pump usage and continuously calibrating their hydraulic models with insights from both the engineers and operators.

The initial operational analytics works at YLWD have resulted in improved maintenance of pump stations (a “lead” pump unexpectedly had much more wear and tear from higher usage and a discrepancy in performance) and clearer insights into tank levels and pump runtimes.

“But this time, with this analytics application, we were able to bring in our operations team together with engineering staff for just two days. So, not only did it expedite the calibration effort, it also impressed our operators as they had direct input during the meetings. It wasn’t even a full eight hours – with their assistance, it was just part of the day.” (Anthony Manzano, Senior Project Manager, Yorba Linda Water District)

Water Management System (WMS) Implementation – Large Council, South East Queensland, Australia

The South East Queensland Council went to market for a real-time water monitoring solution and selected Innovyze to install and implement a Water Management System (WMS) for NRW & leakage management, event management and water network performance management for supply fidelities such as whole of network, pressure managed areas and demand metered areas.

Web-based, real-time workspaces were built to daylight and extend existing works and practices and build initial workspaces for key performance indicators such as ILI.

A key consideration and focus of this particular implementation was around burst reporting. Specific analysis was completed on anomalous flow detection against live/historical usage profiles, highest usage instances, against flushing occurrences, misplaced vs lost volumes while filtering out the sensitivity of a given analysis in alerting on potential false negatives and false positives.

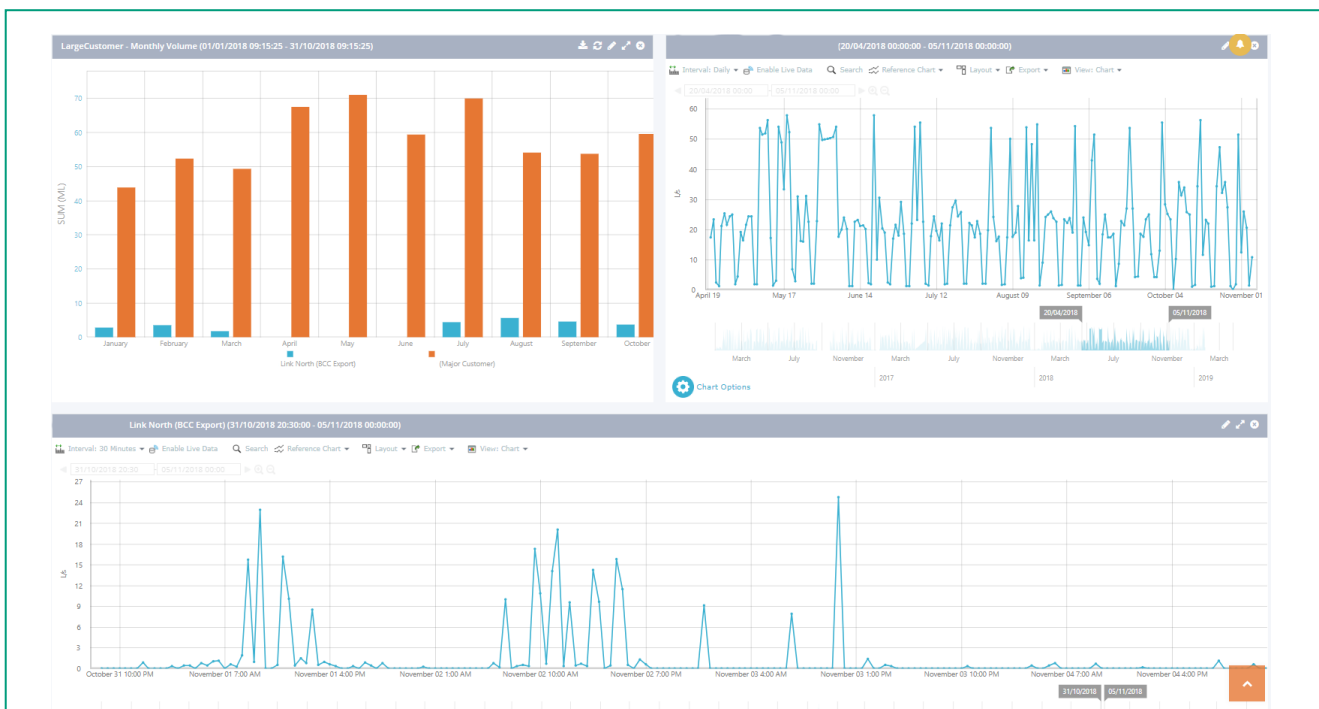


Figure 16: Large customer report differentiating significant demand events from the network such as tanker truck fillings, usage from industry, to an actual system incident such as a pipe burst.

Initial workspaces built for the implementation were to be mimicked and scaled across remaining sections of the network by Council engineers and consultants following the implementation. Figure 16 shows a large customer report differentiating significant demand events from the network such as tanker truck fillings, industrial use to an actual system incident such as a pipe burst.

Workspaces built for the operations team, engineers and management as part of this implementation include:

- Water balance metrics and quarterly usage reports.
- Identifying data that was not to the standard required for adequate network analysis.
- Using billing data to conduct mass balances with bulk provider.
- Sensor uptime with a sortable list of problematic sensors based on pre-defined criteria.
- Displaying problematic sensors from a pressure standpoint.
- Current Annual Real Losses (CARL), Unavoidable Annual Real Losses and International Leakage Index (ILI) calculations for whole of network and DMAs.
- Burst reporting: anomalous flow detection against live/historical usage profiles, highest usage instances, against flushing occurrences, misplaced versus lost volumes.
- Pump efficiency calculations with scatterplots to observe pump performance.
- Critical pressure, point performance, comparison of CPP to DSS.
- Monthly trunk main losses.

Application of Artificial Intelligence for Water and Wastewater Systems – United Utilities, United Kingdom

Innovyze’s artificial intelligence platform and algorithms were applied at United Utilities, United Kingdom (Abdul Gaffoor 2019). Confronted with complex operational decisions dependent on human judgement and experience, United Utilities opted to deploy AI-driven real-time optimisation selecting the Oldham District metered zone due to its remote-control capabilities and high degree of implementation at its sites.

Background on the United Utilities Oldham District metered zone includes (Abdul Gaffoor 2019):

- Oldham is the 5th most populous area in the Greater Manchester Region.
- Supplies 55 MLD (or 20,000 ML annually).
- Services 19 DMAs and 3 large industrial users.
- 5 out of 10 pump stations remotely controlled.

- 4 out of 10 service reservoirs monitored.

The AI platform generated real-time pump schedules, minimising the cost of operations within United Utilities’ compliance and maintenance requirements. The technology demonstrated (Abdul Gaffoor 2019):

- Machine learning was leveraged to predict expected demand and dynamically adapt operations with least cost trajectory.
- Machine learning models are an ideal tool for predicting the future states of complex and dynamic systems for their ability to self-adapt and adjust to new patterns that emerge from ongoing network behaviour.

After a 12-week program, summarised below in Table 1, Innovyze’s AI platform, in collaboration with United Utilities, demonstrated the following return on investment, outcomes and benefits (Abdul Gaffoor 2019):

- Generated 22% cost savings (approximately 3 £/ML) relative to baseline operations, corresponding to a payback period of 5 months.
- Improved resiliency of network by imposing terminal constraints on service reservoirs to ensure volumes were continuously maintained over daily operational cycles.
- Enhanced visibility of network by providing staff with impact of operational decisions on key performance indicators.
- Potential to save 4,000 staff hours in terms of alarm management and response time.
- CO2 emission reduction equivalent of 300 homes.

	MIN	AVG.	MAX.
Baseline Annual Operational Cost (£)		230,367	
Optimized Annual Operational Cost (£)	192,326	180,148	125,823
Savings (%)	17%	22%	45%
Savings per Annum (£)	38,041	50,219	104,544
Normalized Savings (£/ML)	2.1	2.8	5.7
Payback Period	7 months	5 months	2 months

Table 1: Summary (minimum, maximum and average) of cost savings and investment return periods achieved after 12-week program between United Utilities (Oldham District DMA) and Innovyze’s AI platform (Abdul Gaffoor 2019)

CONCLUSION

Smart meters, AMR, SCADA and other data collection processes are generating bigger, better results but the data generated is at present under-utilised. The "Big Data" era provides potential for far greater integration between operations, response teams, engineering systems, management and board level decision makers. In pursuit of proactive decision making, operational analytics allows for prioritisation of response resources and deployment of accompanying action plans often in real time.

The operational analytics and artificial intelligence web-based applications, Info360 and Emagin by Innovyze, discussed in this paper, contain workspaces designed to be utilised on a daily basis include mapping, geospatially located sensors and alert mechanisms, visualisation of data through charts and built-in analytics dashlets, pre-designed visuals and metrics to monitor sensor health and data quality to provide deeper insights on the operations and performance of water and sewer networks.

Overcome the Inertia of Existing Systems

The approach of sourcing and configuring real-time data capture, identifying KPIs for application, running an analysis to create operational insights and utilising such insights for proactive system management will allow utilities, councils and supporting consulting companies to discover, address and solve the complex challenges confronting the industry with a higher degree of execution that has previously been elusive and out of reach for many to take on.

Design, Automate, Offload and Action

More data and dashboards do not always lead to more confident decisions.

Insights achieved from configurable calculations are now available in automated workflows beyond traditional spreadsheet methods and siloed engineering services team members, expanding the system operations and performance insights to operations, SCADA managers, response teams and management observing KPIs on network performance.

Actual meaning derived from available data will inform action from a trusted decision support tool. Workflows must design, automate, offload and action on behalf of the user for behaviour change and user adoption to occur.

Moving beyond solely human judgement for action, artificial intelligence applied to operational analytics may now be utilised to make predictive recommendations to

operations and management on complex operational decisions.

Utilising Internal and External Expertise

In pursuit of operational insights and directed action, whether through internal expertise or an external consultant for supplementing engineering and data science capacity, greater depth and detailed analysis may be applied, saved, standardised, reused (mimicked), streamlined and applied to other parts of water, sewerage and drainage networks with the use of the operational analytics and artificial intelligence software by Innovyze.

Operational analytics, artificial intelligence and the accompanying spectrum of analysis techniques from simple statistical methods through to the application of machine learning algorithms may now be part of every utility and council's digital journey.

An application open to the industry and market forces ensures a continual pursuit of innovation and ability to integrate the latest data science and artificial intelligence methods available to the water industry.

REFERENCES

American Water Works Association "AWWA Water Balance" AWWA M36 Manual, 4th E.

CARL Calculation: Lenzi, C 2014, Procedia Engineering 70, 1017-1026

UARL Calculation: Lenzi, C 2014, Procedia Engineering 70, 1017-1026

Abdul Gaffoor, T 2019, Application of Artificial Intelligence for Water and Wastewater Systems

THE AUTHORS



Patrick Bonk

Patrick Bonk is a part of the Innovyze Asia-Pacific team as the resident Sewer Hydraulics Engineer. Patrick, a professional Environmental Engineer has over 8 years of experience with Hydrologic/Hydraulic Water/Wastewater modelling including experience in human centred design & user scenario development to transform user needs into product design concepts. He joined Innovyze in 2014 and is a software solutions lead for Operational Analytics decision support tools and software advancements within the context of smart water systems and digital twins.



Jonathan Klaric

After completing undergraduate degrees in theoretical physics and mathematics, Jonathan was a PhD candidate in applied mathematics at The University of Queensland before starting at XP Solutions as a mathematician and numerical engine developer. While in academia, Jonathan also worked on a variety of research projects such as computer imaging, temporal wormholes, mathematical epidemiology and a project on Galactic dust extinction with NASA. Jonathan also has a role as the Resident Scientist at UQ's Wonder of Science program.