OPERATIONAL ANALYTICS – THE CALCULATIONS CAN BE COMPLEX, BUT THE EXECUTION IS SO SIMPLE

Patrick Bonk[1], Jonathan Klaric[2]
1. Innovyze, Gold Coast, QLD, Australia
2. Innovyze, Brisbane, QLD, Australia

KEYWORDS

Analytics, Operational, Big Data, Real-Time, Decision-Support

ABSTRACT

The Big Data era provides potential for far greater integration between operations, engineering and management.

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.

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

INTRODUCTION

Data-Driven Decisions

The Big Data era provides potential for a higher degree of integration between operations, response teams, planning and management. However, the current live data bottleneck caused by securing these live datasets combined with the use of 'small data' analysis software (spreadsheets, etc.) when the stale time series data is finally accessed, stifles innovation. When the skills sets of technical and non technical 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 by way of stored logic, sampled and derived data feeds, templated and scalable calculations and associated alert generation.

Design. Automate. Offload. Action

In pursuit of proactive decision making, operational analytics allows for prioritisation of response resources & deployment of accompanying action plans. The operational analytics web based application, Info360 by Innovyze to be discussed in this paper contains 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, sewerage and drainage networks.

Informing Info360's workspaces, shown in Figure 1 for network insights and subsequent action is 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). Individual customer consumption from digital meters can be combined to calculate live zonal demands, for comparison against fundamental aspects of networks from reservoir levels, pump operation and source flows.

The automated data modelling and analytics engine constantly samples across the available data historian or store to create a "TS model" for each sensor to capture the properties and behaviour of the underneath data stream. The application is source agnostic, designed for operational analytics to be a fundamental aspect of proactive decision making and quicker to action when reacting to a system event. When action is necessary, operational analytics provides the newfound ability to discern 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 utility and council'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, including data quality, total uptime and notification of stuck sensors.
- Event detection and management. "Spectrum of Analysis," from simple analysis with basic statistical functions (variance equations) to more advanced analysis techniques (Fourier transforms).
- Improve identification of leaks and bursts by applying industry best practice analysis metrics such as Real Loss metrics and Infrastructure Leakage Indexes (ILI).
- From reactive to active system operation; apparent losses such as unauthorised consumption, data transfer errors or data analysis errors can now be discovered and addressed.
- Better customer service including reduction in response time to bursts, deviations in system pressure affecting customers etc.
- Simple statistics and/or pattern recognition can be used for system anomaly detection. The application allows the user to build complex calculations within built-in analytics tools where scripting and SQL skills are not required for the user to successfully pull data and analyse it in real-time (& historical).
- Configurable calculations, or automated workflows for scalability of complex calculations that have traditionally been built within static spreadsheets.
- Templated approaches and methodologies in place so that data is utilised for decision making as opposed to efforts solely directed on data cleansing.

Technology Drivers:

Overall, Operational Analytics are being driven by:

- Utilities and Councils collect large amount of timeseries data yet often underutilise collected data within their decision-making processes.
- Increase collaboration and communication between operations, response teams, engineers, management and board level decision makers.
- Capitalise on 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.
- Streamline and standardise internal processes for teams with multiple stakeholders and needs.
- Understand how planned changes in the system will affect customers.
- Identify what events may be occurring in the system unnoticed.
- Quantify how the system can be operated more effectively.
- Determine what impact each system is having on another.
- Optimally prioritise current and future data collection.

Overcome Organisational Siloing

Whether one works within Operations, Response Teams, Engineering, Planning, Management through to Board level decision makers workspaces are 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.

Establishing the needs of stakeholders wtihin a given water authority, replicable approaches based on actual use cases and system behaviour may now be created. Unique data needs of key members of a water authority now accomodated within an operational analytics platform include:

Operators & Response Teams

- Event Detection and Response
- Understand loss of system integrity
- Asset Performance & live levels of service
- Optimised controls
- Updating operating manuals.

Engineers

- System Design and Behaviour
- Event Forecasting
- Understanding Demand
- Leakage and Non-Revenue Water

Management

- Non-Revenue Water
- Developing Key Performance Indicators
- Collaboration amongst their team
- Knowledge Management

Directors/Board Level

- Risk Management
- Regulatory Compliance
- Cost/Benefits Analysis

METHODOLOGY - TECHNOLOGY DISCUSSION

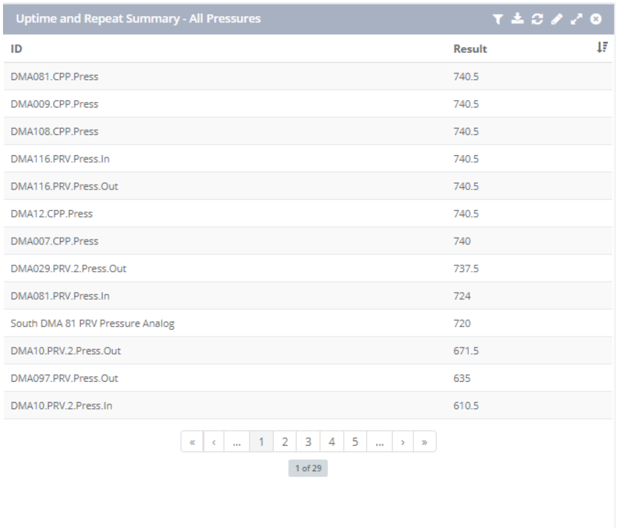
Source and Configure Real-Time Data Capture

Smart meters, AMR and SCADA are utilised to acquire large amounts of data which is used for 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-drop outs, repeats or other systematic anomalies present in all real-world data. A natural part of any data collection and analysis lifecycle, data availability issues are an intrinsic part of any 'live', real-world data collection system. With these inherent issues, Sensor Health analysis can provide feedback and richer understandings on the data quality itself to ensure designers, operators and maintainers of SCADA systems may have this added level of understanding on 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 dashlet functionality within the software and added to a user's workspace.

The built-in and configurable calculations on the backend of each visualisation tool (dashlets, placed on web hosted workspaces) 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

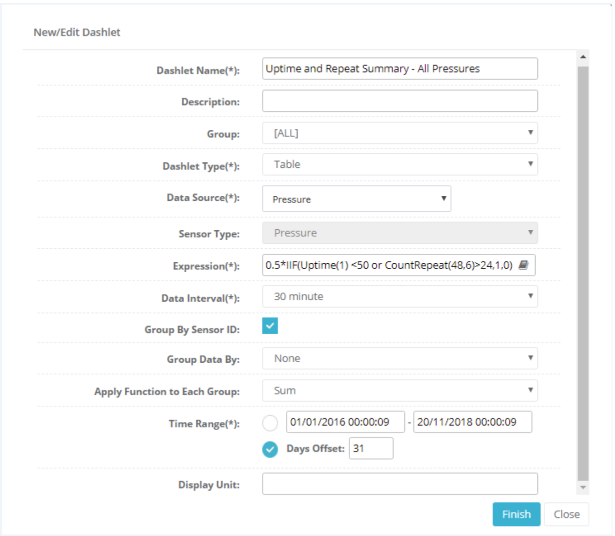
ID	Result
DMA081.CPP.Press	740.5
DMA009.CPP.Press	740.5
DMA108.CPP.Press	740.5
DMA116.PRV.Press.In	740.5
DMA116.PRV.Press.Out	740.5
DMA12.CPP.Press	740.5
DMA007.CPP.Press	740
DMA029.PRV.2.Press.Out	737.5
DMA081.PRV.Press.In	724
South DMA 81 PRV Pressure Analog	720
DMA10.PRV.2.Press.Out	671.5
DMA097.PRV.Press.Out	635
DMA10.PRV.2.Press.In	610.5

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 business intelligence to prioritise maintenance schedules to target the longest-outage sensors first for repair or replacement.

Deepening the level of analysis readily available to a first time user, built-in mathematical functions, shown in Figure 4 within each dashlet can be applied to the timeseries data to display a variety of output data based on derived calculations such as volume through a zone, rates of change over different periods, or even simply moving averages to smooth out spikes in the data.



New/Edit Dashlet

Dashlet Name(*):	Uptime and Repeat Summary - All Pressures
Description:	
Group:	[ALL]
Dashlet Type(*):	Table
Data Source(*):	Pressure
Sensor Type:	Pressure
Expression(*):	0.5*IIF(Uptime(1) <50 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:	

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 half-hour intervals and sums the total number of half-hour 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 guide maintenance and replacement decisions to prioritise longest-standing issues.

Once the quality of the overall data has been established, it is then up to the user to decide on what sorts of 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 DMA 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 septants relative to remaining sections of the pie-chart. Users would immediately be alerted of data quality issues over the preceding day or week prompting users to deploy an investigation for a deeper look into its root cause before basing critical decisions on measured data.

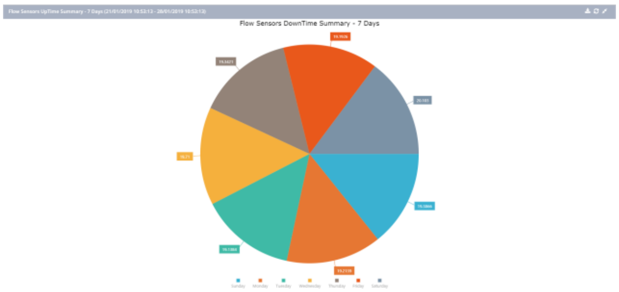


Figure 5: Pie chart dashlet of flow sensor total uptime for previous seven days.

Further to the necessity of high-level overviews of data for monitoring and reporting purposes, in lieu of a network-wide outage, being able to drill-down into which particular sensors are contributing most to total uptime/downtime can be more 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 up-time on an hourly interval.

Figure 6: Bar chart dashlet viewing a single sensor's average daily up-time on an hourly interval.

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 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 live and historical data, up-time of sensors, informs prioritisation of maintenance schedules of sensors as well as informs the prioritisation 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.

Identify the Key Performance Indicators (KPIs) to Apply

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 where scripting and SQL skills are not required for the user to successfully pull data in real-time and build workspaces based on their unique data 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, including total uptime and notification of stuck sensors.
- Sensor uptime with a sortable list of problematic sensors based on pre-defined criteria.
- Determine sensor drift from normal system behaviour.
- Sensitivity analysis to avoid false negatives (detecting a burst that did not occur) or false positives (observing a swimming pool being filled when in actuality a burst occurred).
- Burst detection and reporting: Anomalous flow detection against live/historical usage profiles, highest usage instances, against flushing occurrences, misplaced vs lost volumes.
- Boundary breach versus sensor issues.
- 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.
- DMA usage versus time against all DMAs.
- Daylighting data that was not to standard required for adequate network analysis.
- Using billing data to conduct mass balances with bulk provider.
- Critical pressure, point performance, comparison of CPP to DSS to determine drift from normal system behaviour.
- Monthly trunk main losses.
- Current Annual Real Losses (CARL), Unavoidable Annual Real Losses and International Leakage Index (ILI) calculations for whole of network and DMAs.

Figure 8 shows the American Water Works Association's water balance for authorised consumption, sources of revenue water, 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 Current Annual Real Loss and Figure 10 progresses further through the respective workflow to demonstrate how Unavoidable Annual Real Losses (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, DMAs and utilised at the level of whole of network providing a streamlined and standardised approach to building and applying key network performance metrics.

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 KPI's, metrics relating to leakage, water balance and loss, pipe breaks and storage 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 aforementioned 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 my 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 a respective metered zone or at the level of the DMA, a network sensitivity analysis must be conducted and considered to account for the wide variety of occurrences in a given 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 has 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?"

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 progressing understandings of system behaviour can be achieved with accessible 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 shows a scatterplot of usage versus inflow for a given DMA averaged over a 24 hour period.

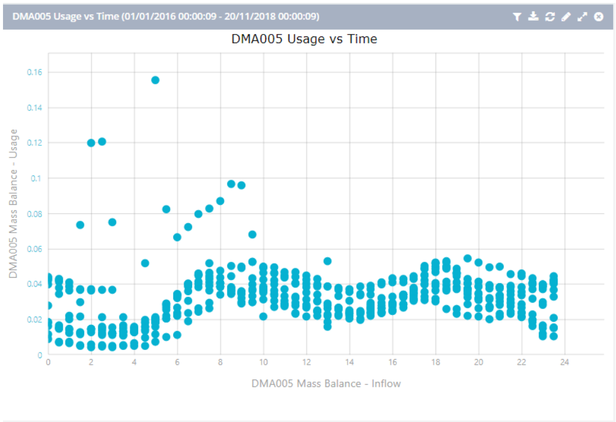
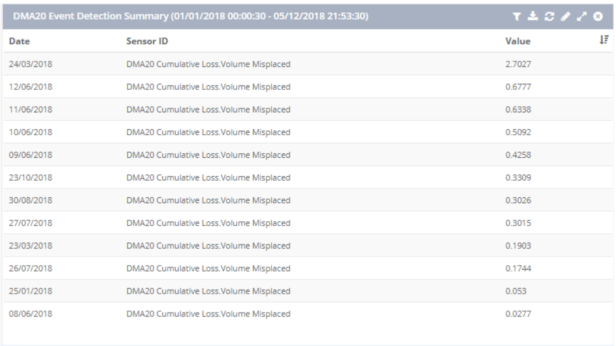


Figure 11: Scatterplot Dashlet of DMA Usage vs Time for a 24 hour period used to identify large anomalous usages.

The respective DMA 24-hr 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 incorporating robust rules to truly separate out anomalous events deserving the attention of Operations and Response teams. The usages during these events of statistical significance were then subtracted from the average usage for their respective hours for the duration of 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) to actual system events such as a pipe burst incidents and routine pipe flushing occurrences.



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.

The usage pattern applied in this case was used to filter out the sensitivity of a given analysis in alerting potential false negatives (detecting a burst that did not occur) and false positives (observing standard system behaviour when in fact a burst occurred). Figure 13 shows a dashlet with built-in metrics to detect and determine typical usage behaviour vs 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.

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 by 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 anamolies 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 are available to be replicated and applied to remainining sections of the respective network.

Traditionally, data available solely to Council’s Engineering Services has now been deployed with the latest event detection analysis technicques 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 and 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, re-used (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 statistical analysis techniques for system event management and leakage analysis.

Insights are Utilised for Action

The insights and understanding, once a vast array of data reads from the network are now at the level 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 KPI's on network performance.

Automated workflows available to the business include:

- Sensor Health analysis brings to light gaps in live and historical data, up-time of sensors, informs prioritisation of maintenance schedules of sensors as well as informs 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 resource prioritisation.
- Provide reliable levels of service while balancing the costs on 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.

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 in real-time discern where, when, why and how an operations, field and engineering team interpret, react and respond to system issues as they arise.

The retroactive analysis revealed a major event in particular, 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 to 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.

Simply using only a $ 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 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 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 mu is the average of the dataset, was applied to a total 7 days of flow data with five days of flow patterning leading to the burst assumed to represent typical zone demand.

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

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 was 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 U.S.A

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, US. 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.

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.
- Daylighting data that was not to 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 vs lost volumes.
- Pump efficiency calculations with scatterplots to observe pump performance.
- Critical pressure, point performance, comparison of CPP to DSS.
- Monthly trunk main losses

CONCLUSION

Smart meters, AMR, SCADA and other data collection processes are generating bigger/better but underutilised data sets. 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 & deployment of accompanying action plans.

The operational analytics web based application, Info360 by Innovyze discussed in this paper contains 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 KPI's 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.

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, re-used (mimicked), streamlined and applied to other parts of water, sewerage and drainage networks with the use of the operational analytics software Info360 by Innovyze.

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

REFERENCES

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

CARL Calculation: C. Lenzi et al. Procedia Engineering 70 (2014) 1017-1026

UARL Calculation: C. Lenzi et al. Procedia Engineering 70 (2014) 1017-1026)

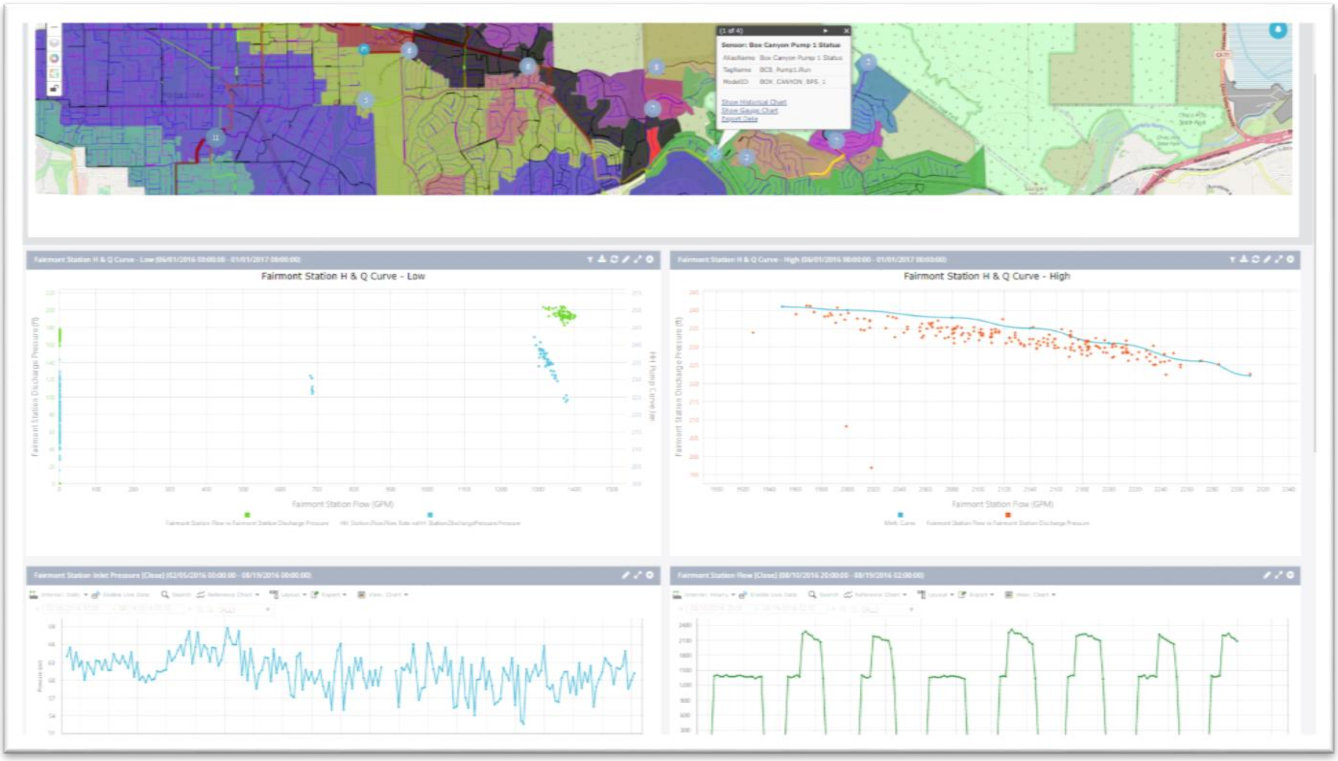
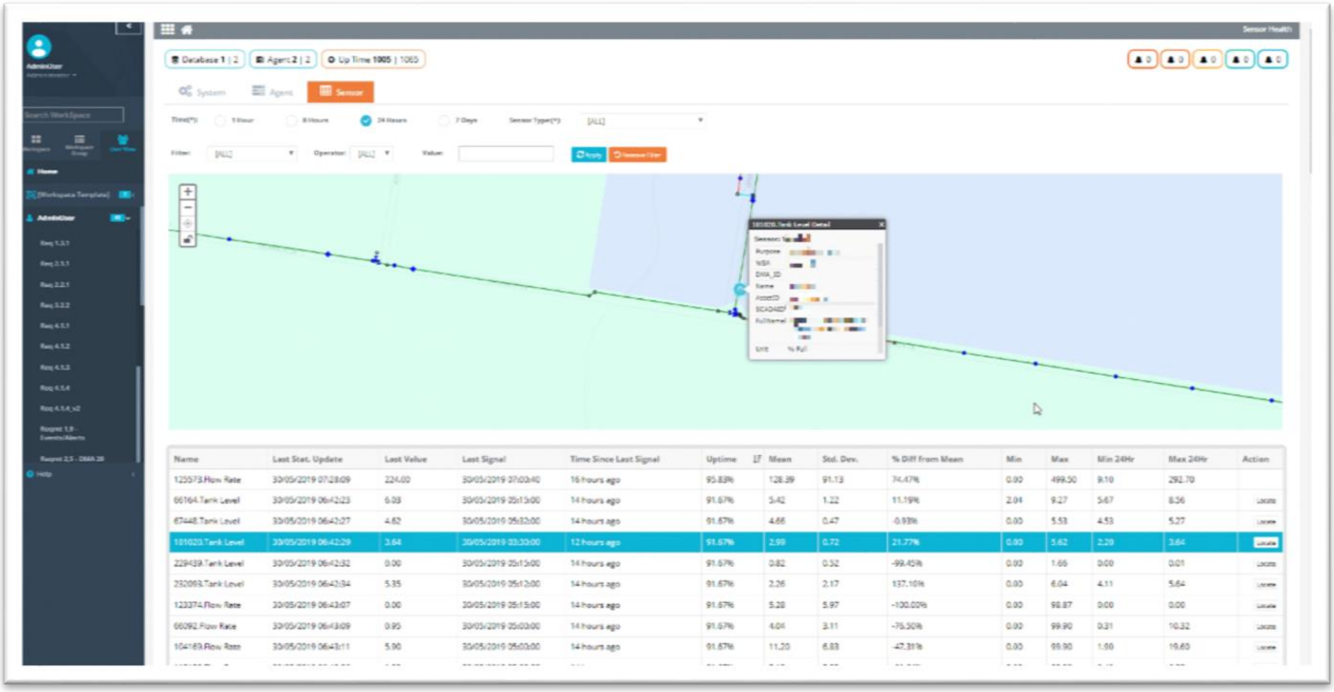
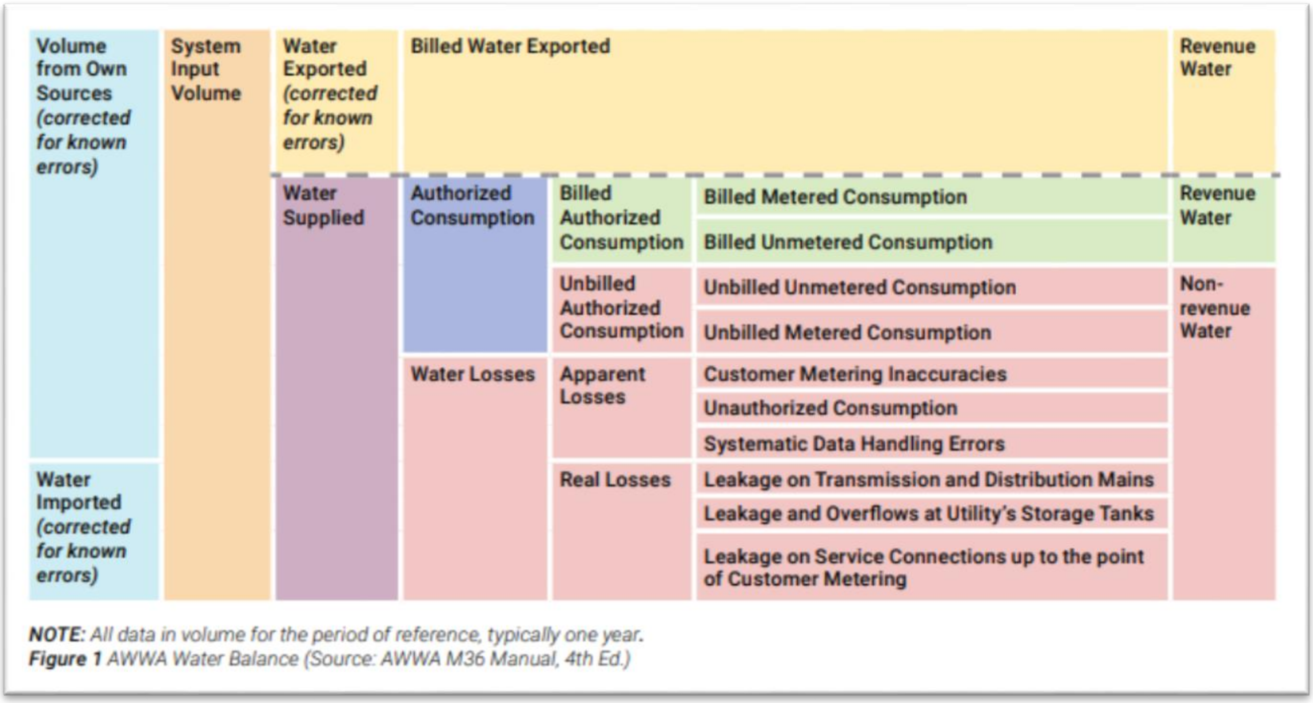


Figure 1: Info360 by Innovyze, workspaces showing multiple geospatially located assets & data tags built on a data modelling application which consumes multiple categories of live data (AMR/AMI, SCADA, water quality, demand, flow, pressure, level, pump data and hydraulic model historical/predictive runs)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



Name	Last Stat. Update	Last Value	Last Signal	Time Since Last Signal	Uptime	Mean	Std. Dev.	% Diff from Mean	Min	Max	Min 24Hr	Max 24Hr	Action
125573.Flow Rate	30/05/2019 07:28:09	224.00	30/05/2019 07:03:40	16 hours ago	95.83%	128.39	91.13	74.47%	0.00	499.50	9.10	292.70	
66164.Tank Level	30/05/2019 06:42:23	6.03	30/05/2019 05:15:00	14 hours ago	91.67%	5.42	1.22	11.19%	2.04	9.27	5.67	8.56	Locate
67448.Tank Level	30/05/2019 06:42:27	4.62	30/05/2019 05:32:00	14 hours ago	91.67%	4.66	0.47	-0.93%	0.00	5.53	4.53	5.27	Locate
101020.Tank Level	30/05/2019 06:42:29	3.64	30/05/2019 03:33:00	12 hours ago	91.67%	2.99	0.72	21.77%	0.00	5.62	2.20	3.64	Locate
229439.Tank Level	30/05/2019 06:42:32	0.00	30/05/2019 05:15:00	14 hours ago	91.67%	0.82	0.52	-99.45%	0.00	1.66	0.00	0.01	Locate
232093.Tank Level	30/05/2019 06:42:34	5.35	30/05/2019 05:12:00	14 hours ago	91.67%	2.26	2.17	137.10%	0.00	6.04	4.11	5.64	Locate
123374.Flow Rate	30/05/2019 06:43:07	0.00	30/05/2019 05:15:00	14 hours ago	91.67%	5.28	5.97	-100.00%	0.00	98.87	0.00	0.00	Locate
66092.Flow Rate	30/05/2019 06:43:09	0.95	30/05/2019 05:03:00	14 hours ago	91.67%	4.04	3.11	-76.50%	0.00	99.90	0.31	10.32	Locate
104169.Flow Rate	30/05/2019 06:43:11	5.90	30/05/2019 05:03:00	14 hours ago	91.67%	11.20	6.83	-47.31%	0.00	99.90	1.90	19.60	Locate

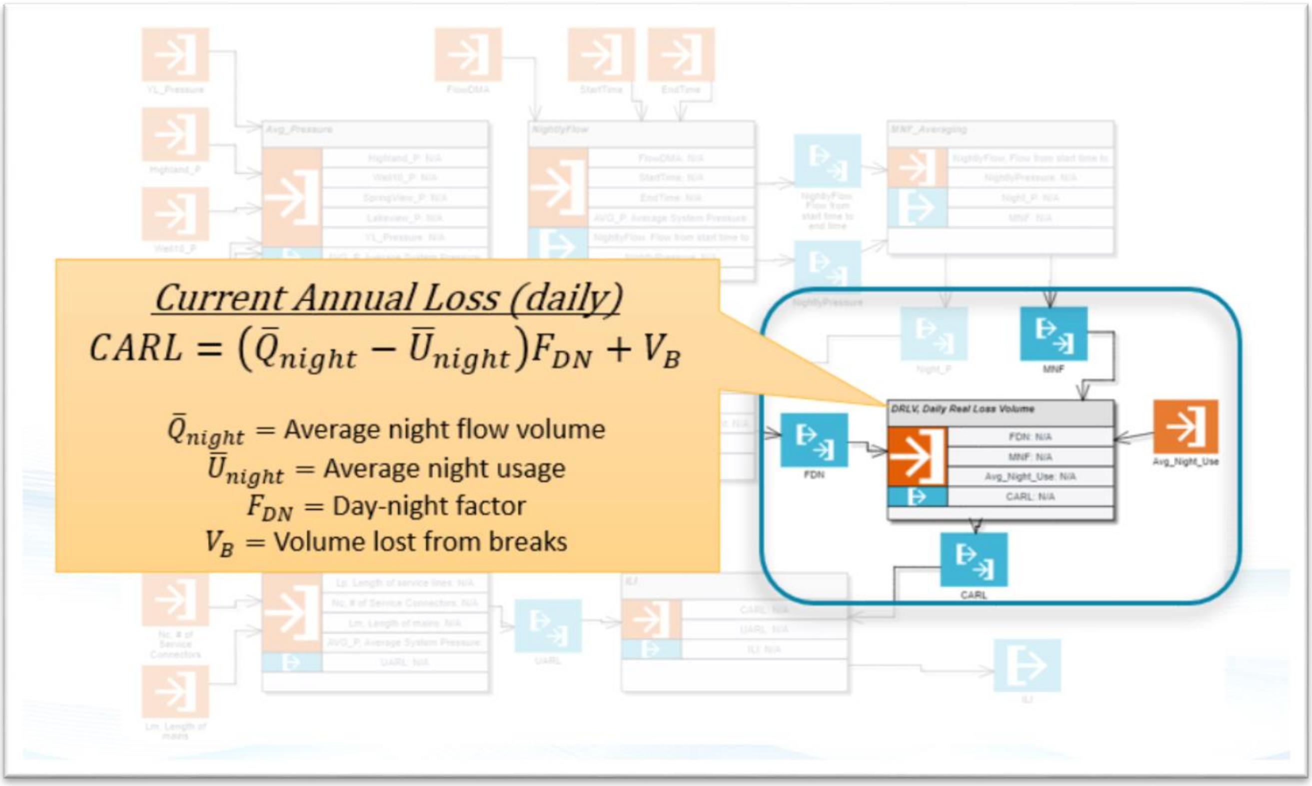
Figure 7: Sensor Health and Data Quality Workspaces pre-designed within the software and available to users on log-in brings to light gaps in live and historical data, up-time of sensors and informs the prioritization of maintenance schedules of sensors; all geospatially located, viewable on a workspace.



Volume from Own Sources *(corrected for known errors)*	System Input Volume	Water Exported *(corrected for known errors)*	Billed Water Exported			Revenue Water
		Water Supplied	Authorized Consumption	Billed Authorized Consumption	Billed Metered Consumption	Revenue Water
					Billed Unmetered Consumption	
				Unbilled Authorized Consumption	Unbilled Unmetered Consumption	Non-revenue Water
					Unbilled Metered Consumption	
			Water Losses	Apparent Losses	Customer Metering Inaccuracies	
					Unauthorized Consumption	
					Systematic Data Handling Errors	
Water Imported *(corrected for known errors)*				Real Losses	Leakage on Transmission and Distribution Mains	
					Leakage and Overflows at Utility's Storage Tanks	
					Leakage on Service Connections up to the point of Customer Metering	

NOTE: *All data in volume for the period of reference, typically one year.*
Figure 1 *AWWA Water Balance (Source: AWWA M36 Manual, 4th Ed.)*

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).

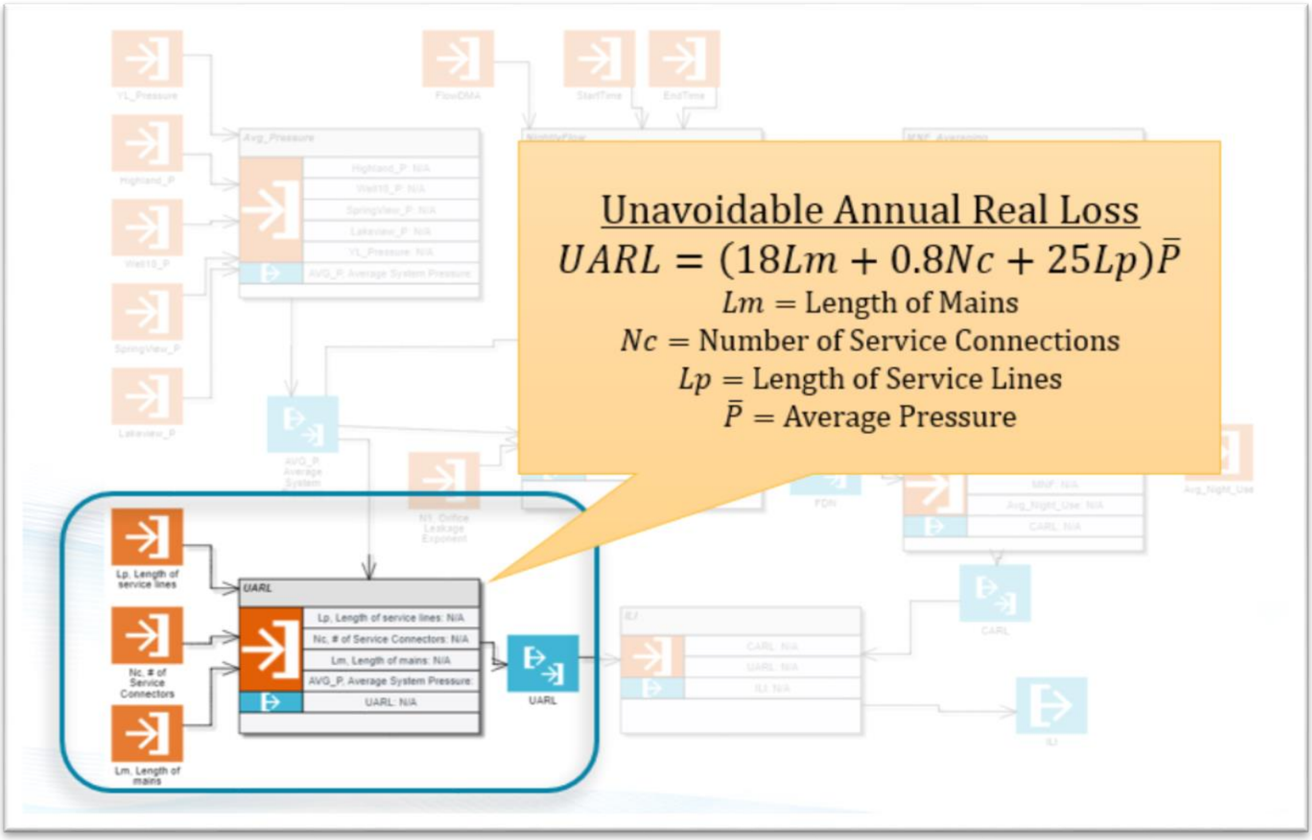


Current Annual Loss (daily)

$$CARL = (\bar{Q}_{night} - \bar{U}_{night})F_{DN} + V_B$$

$\bar{Q}_{night}$ = Average night flow volume
$\bar{U}_{night}$ = Average night usage
F_{DN} = Day-night factor
V_B = Volume lost from breaks

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.



Unavoidable Annual Real Loss

$$UARL = (18Lm + 0.8Nc + 25Lp)\bar{P}$$

Lm = Length of Mains

Nc = Number of Service Connections

Lp = Length of Service Lines

$\bar{P}$ = Average Pressure

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.

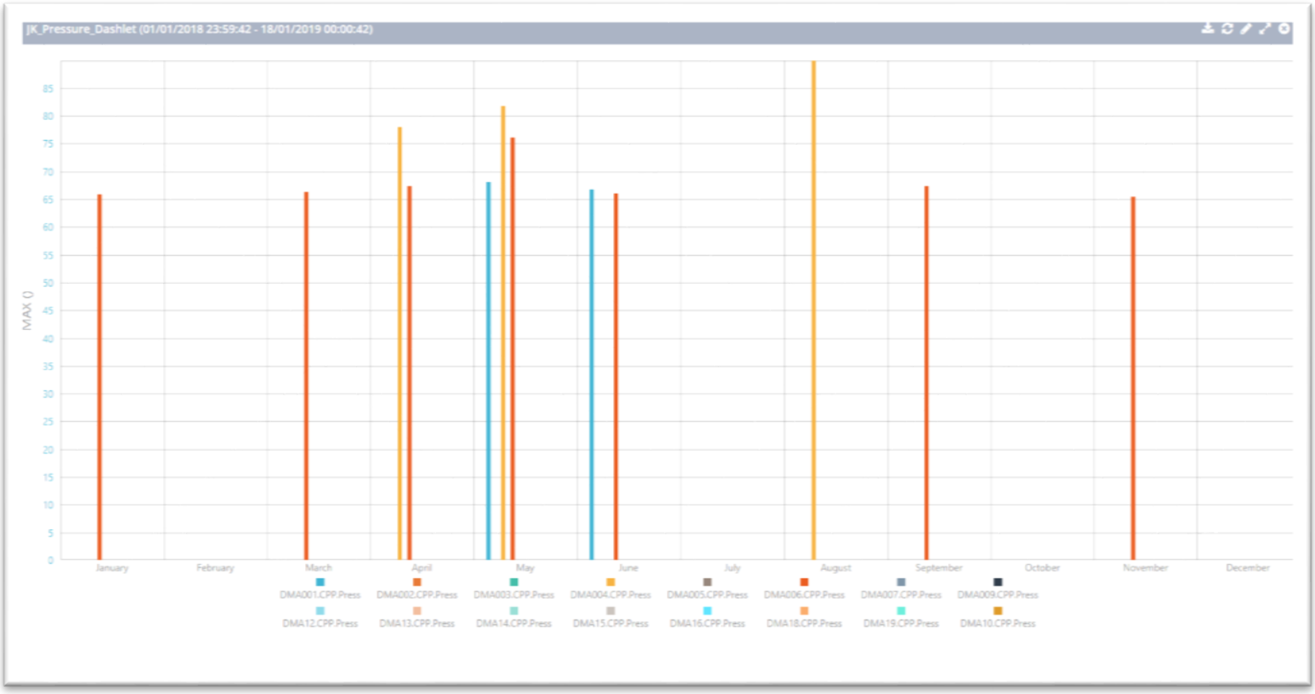


Figure 13: Dashlets with built-in metrics to detect and determine typical usage behaviour vs an actual event such as a burst for a given pressure zone.

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

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.

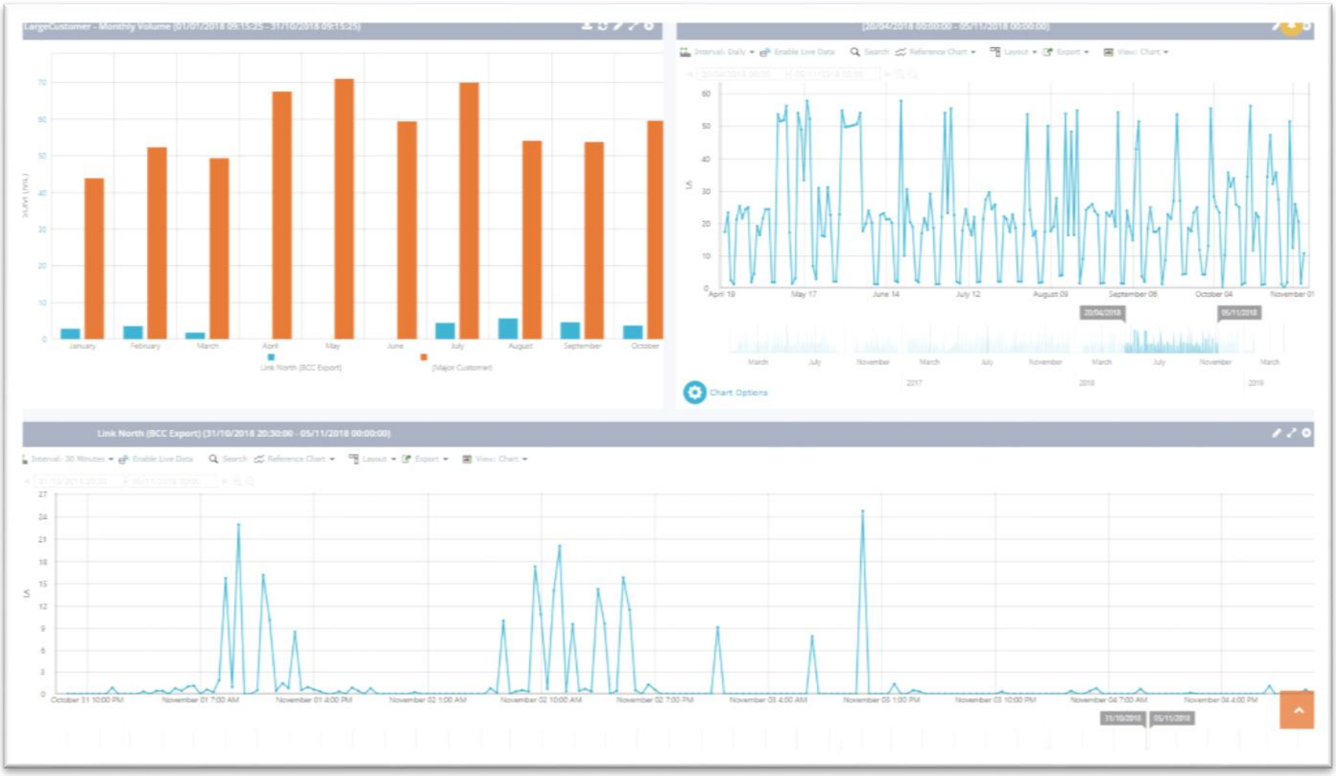


Figure 16: Large customer report differentiating significant demand events from the network such as tanker truck fillings, usage from industry, to an actual system incident such as a pipe burst.