Specification for Proof of Concept trials to utilise machine-learning algorithms in order to provide early warning of sewer capacity issues

Background

Our longer-term vision is to develop and maintain intelligent wastewater networks that will provide real-time and forecast performance information.

This will enable the prediction and resolution of surcharging sewers before they become service failures (pollution or flooding incidents). Failing that it will deliver a dramatic improvement in response time to service failures.

With the increased number of monitors in the wastewater network (mostly at combined sewer overflows; a result of the Event Duration Monitoring programme), during wet weather events Wessex Water's Control Room is receiving an increasing number of alarms. These alarms are currently being generated when depths in sewers reach defined levels – either approaching a spill or at spill level (indicating that sewage is escaping from the network).

During rainfall events, these alarms are mostly just verifying high sewer levels and permitted spills at overflows. However, there is a growing risk that some overflows may spill prematurely due to insufficient capacity caused by issues such as blockages or pumping incapacity. During wet weather it is currently difficult to separate expected spills which occur during heavy rainfall from premature spills that are caused by restrictions in the network, usually caused by (partial or total) blockages.

Smart algorithm-based tools are now available to provide a level of decision support by applying machine-learning techniques to the data from wastewater depth monitors. These tools are able to predict when performance at a particular monitored point on the network is within the expected range (given the prevailing conditions) or outside the expected range - which might give early warning of a problem.

With better predictive warnings, we will be able to respond to issues before they escalate into service failures.

Wessex Water wishes to engage suppliers to carry out proof of concept (POC) trials to demonstrate the value of applying machine-learning algorithms to such data. This document outlines the scope and specification for such trials.

Proof of Concept trial

Suppliers are invited to submit proposals to provide a machine learning tool within the proof of concept. The POC will collate sewer depth information and compare this to time series data as well as upstream rainfall intensity data. The POC solution will produce an acceptable operating depth band for the prevalent conditions.

Over time as the volume and quality of the data collated reaches significant maturity, we expect to see the machine learning algorithms predict an acceptable operating depth band. Where depths are subsequently recorded outside (above or below) the acceptable operating depth band, generate an alarm to initiate a response.

The proposed POC trial will be carried out in our Bath catchment. This area has known hydraulic issues and a high density of monitoring. Details of the catchment can be found in *appendix 1*.

Output required from POC trial

We are committed to reducing alarm load in the control room and ensuring that potential pollution and flooding incidents are avoided. We will apply a number of measures and metrics for any POC trial and will carry out a primary assessment before offering the POC. Full details of the measures and metrics are provided in *appendix 2*.

Trial metrics will be based around the following:

- Minimisation of control room alarms
    - During regional rainfall events the system will mute alarms/alerts when sufficient rainfall causes the high level to trigger
    - Only a set percentage of the alarms/alerts should be false or ghosts
        - As the system learns we will expect to see this reduce
- Identifying network issues
    - We will look to simulate this at a random locations and timings within the POC
- Alert latency and response time
    - Any POC will have to run in real time
- Demonstrate compatibility with existing systems
    - Alarm management system
    - Work management
    - Business reporting

We are looking to improve operational performance and provide an improved risk management approach. We will be measuring against our current practices and approach.

Information available for successful suppliers

For a proof of concept Wessex Water will provide the following data depending on the location of the data management solution (full specification in *appendix 3*):

- Data that can be provided for on-premise solutions:
    - Historical rainfall – near real-time
    - Predicted rainfall – near real-time
    - Network sensor outputs – near real-time
    - Network configuration – at implementation and as updated
    - Site sensor outputs – near real-time
    - Site configurations – at implementation and as updated
- Data that can be provided near real time for off-premise or cloud-based solutions:
    - Network sensor outputs – near real-time (note that the definition of real time is as received by the control room. Some sites feed back data every 24 hours unless alarming)
    - Network configuration – at implementation and as updated
    - Site sensor outputs – near real-time (note that the definition of real time is as received by the control room. Some sites feed back data every 24 hours unless alarming)
    - Site configurations – at implementation and as update
- Catchment information
    - We will provide the location of the catchment to be monitored. Details of the catchment/s are provided in *appendix 1*
    - We will provide the locations of any monitoring devices including hydraulic relationships (if requested) via a skeletonised network view

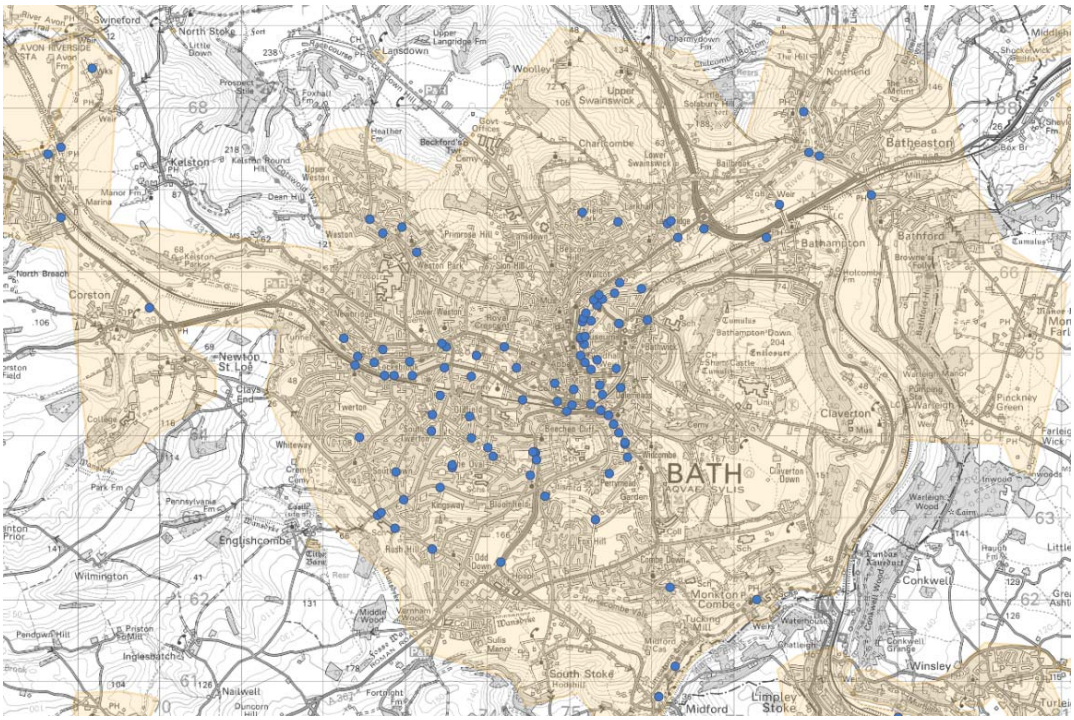
- We will provide the site configuration at each monitoring point
    - Existing alarm levels
    - Site measurements

Our preference would be an off-premise solution in the first instance, to avoid internal costs.

Appendix 1

The Bath catchment has been chosen to be the most suitable location for a machine learning POC. Details of the Bath catchment for the trial purposes are below:

- Bath STW Catchment
- STW Saltford Mead Lane
- Properties
    - 42000
- Domestic population
    - 93500
- Total population (inc. workers/hotels etc)
    - 109543
- Sewer length
    - 540742m
- Catchment area
    - 62902423m^2



- In sewer depth monitors
    - Total active monitors
        - 89
    - Of which at least 1 year of good-quality data
        - 51
    - Of which at least 1 year of poor-quality data
        - 16
    - Of which less than 1 year of good-quality data
        - 20
    - Of which less than 1 year of poor-quality data
        - 2

Appendix 2

Initially a desktop exercise will be carried out to identify suppliers from the proof of concept proposals to take forward to a POC trial. Wessex Water will be looking at the feasibility of deployment, existing case studies, data and information security, and commercial viability.

During the trial the following measures will be used:

Alarm volumes

Given the likely scale of devices in the network we need to think about what is acceptable for false/ghost alerts/alarms. The South West has between 10 and 15 rainfall days per month on average. This indicates that, with what will be approaching 1500 monitoring devices, a significant number of alerts/alarms per month could be avoided.

We suggest there is a pre-defined learning period, this should range between 3-6 months as a maximum for new devices plus the historic data provided on Marketplace.

Metric/Measure:

- At go-live of the operational trial
    - 90% of alarms will be genuine issues
- After 6 months of the operational trial
    - 95% of alarms will be genuine issues

The solution will be able to filter transitory events from the measuring device that are not caused/triggered by flow or rainfall.

Enhancement

Any proposed analytical system will need to demonstrate improvement over and above systems and processes currently employed in the business. We will need to evidence the following:

- The system identifies sites that are meeting their permit during a rainfall event in the catchment and does not create an alert/alarm
- The system identifies sites that are not meeting their permit during a rainfall event in the catchment and does create an alert/alarm

The solution will be measured against our current alarm volumes.

Connectivity

Any proposed analytical system will have to demonstrate seamless connectivity to Wessex systems such as our Alarm Management System and our Business Reporting Tools.

- We operate a Servelec Scope telemetry system and alarm management tool. The proposed solution will need to demonstrate how alarms will be visualised.
- We use QlikView for business reporting. The proposed solution will need to demonstrate how it will integrate.

Initially this will be a robust paper-based evidential exercise. Beyond the trial connectivity will need to be demonstrated before any work is awarded.

Appendix 3

Details of data available

- Rainfall data
    - If on-premise the rainfall data that can be provided are as follows
        - Met-office Nimrod data
            - In binary data file format
            - 1km resolution
            - 5-minute time resolution
            - Updated every 5-15 minutes
        - Met-office Nowcast Data
            - In Binary format GRIB 2 (Gridded Binary)
            - 2km resolution
            - 15-minute time resolution
            - Forecasts 6-hours ahead
            - Updated every 15-minutes
    - Rainfall data cannot be provided to any off-premise hosted solution
        - Any off-premise hosted solution will need to provide rainfall data for the analytics that is of equivalent quality to that we are currently receiving
- Network and site sensor outputs
    - Analogue and digital signals
        - Depth/level monitors
            - 2-minute readings
            - Transferred once per day if operating in normal range
            - Also transferred when a predetermined threshold is breached at the site
            - The data transfer period may be adjusted for the trial to provide more frequent data
        - Pump run/stop signals
            - Variable time stamps
            - Transferred once per day if operating in normal range
            - Also transferred when a predetermined threshold is preached at the site
            - The data transfer period may be adjusted for the trial to provide more frequent data
        - Flow measurement
            - If available, more information supplied by request
    - If on-premise
        - Delivered via a direct push into a SQL database
        - A publisher/subscriber method
        - If off-premise – to be confirmed