HEALTH-BASED TARGETS PERFORMANCE REPORTING - USING VIRTUAL SCADA TAGS TO FACILITATE DATA ANALYSIS

Peter Prevos [1]

1. Coliban Water, Bendigo, VIC

ABSTRACT

Analysing SCADA output is challenging due to the relatively large volumes of data and issues with data quality. Coliban Water has developed software to convert SCADA data to facilitate analysis, called Virtual tags, and to report on this data.

This paper outlines the development of a reporting system to analyse SCADA data in accordance with the Microbial Health-Based Target guidelines published by WSAA (2015).

INTRODUCTION

Safe drinking water is a necessary condition to allow communities to live, grow and enjoy. While in many parts of the world water supplies are the direct or indirect cause of about ten percent of the burden of disease (Hrudey and Hrudey, 2004), Australian communities enjoy water that poses almost no risk to human health. The question remains, however, exactly how low the level of risk actually is. To define microbial safety the World Health Organisation uses the metric of Disability Adjusted Life Years (DALYs), which represents the period a person is burdened with an illness (i.e. loss of time in good health) and years lost through premature death.

Safe drinking water is a necessary condition to allow communities to live, grow and enjoy. While in many parts of the world water supplies are the direct or indirect cause of about ten percent of the burden of disease (Hrudey and Hrudey, 2004), Australian communities enjoy water that poses almost no risk to human health. The question remains, however, exactly how low the level of risk actually is. To define microbial safety the World Health Organisation uses the metric of Disability Adjusted Life Years (DALYs), which represents the period a person is burdened with an illness (i.e. loss of time in good health) and years lost through premature death.

The microbial Health-Based Targets (HBT) approach analyses water treatment plant performance to assure that the level of public health risk is less than 1 micro DALY per person per year.

The Manual for the Application of Health-Based Treatment Targets, published by the Water Services Association of Australia (2015) provides guidance to estimate the potential burden of disease caused by reticulated drinking water. The HBT Manual defines methods for determining the minimum treatment requirements for drinking water supplies by assessing the relative microbiological risk in the source water for the supply and provides a suite of decision rules to estimate the effectiveness of treatment barriers. The effectiveness of the treatment barriers is expressed in Log Reduction Values. A Log Reduction Value (LRV) is the logarithm of the proposition of pathogens removed from the water. The HBT Manual defines the Water Safety Continuum which relates LRV to public health risk in micro DALY.

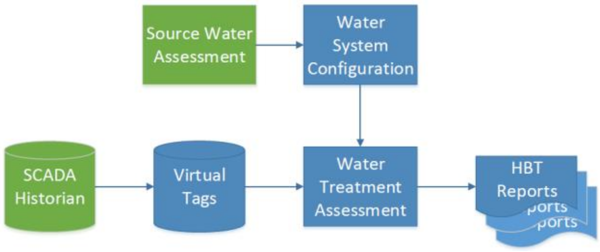

Figure 1: System Modules.

Several water utilities around Australia have successfully implemented the process defined in the HBT Manual. Coliban Water's experience with this process and feedback obtained from other utilities indicates that the HBT Manual is a useful analytical tool that greatly enhances a utility's knowledge about the performance of its pathogen barriers. Using the guidelines in the Manual assists with making informed decisions about improving processes or justifying investments.

Another shared observation is that the decision rules in the Manual are difficult to review on a regular basis because of the large amounts of data to be analysed. Spreadsheets collapse under the pressure of crunching through the amount of data required to assess barrier performance. The assessments are typically undertaken using data at one-minute intervals over the period of one month, which means that for an average treatment plant about half-a-million data points need to be analysed. Given the nature of SCADA data, the process also involves a large amount of manual data cleaning. These issues make the process of implementing the HBT Manual time consuming and cumbersome.

Using the HBT Manual to assess pathogen barrier performance on a regular basis requires a lot of resources for medium-sized water utilities that typically manage a large number of drinking water supply systems. Undertaking these assessments regularly is valuable because it adds value to otherwise amorphous SCADA data locked away in databases. To fully realise the benefits of the HBT approach, Coliban Water has developed software to automate the decision rules in the HBT Manual and produce a monthly report of pathogen barrier performance.

It should be noted that the reports produced by this software can be used for post-hoc analysis and governance, they are not a replacement for operational monitoring.

METHOD

The software consists of four modules (Figure 1). The first module extracts and transforms relevant data from SCADA, using the Virtual Tag approach. The second module stores the configuration of the water treatment processes. The third module analyses the data accordance with the HBT Manual. The final module presents a performance report for each water treatment plant in Coliban Water's service region.

Water System Configuration

Before analysis can be undertaken the configuration of each drinking water supply system is entered into a control screen, using an MS Access front-end to the database. The configuration file contains the list of treatment sites, the barriers used at each treatment site and the relevant Virtual Tags required for analysis (Figure 2).

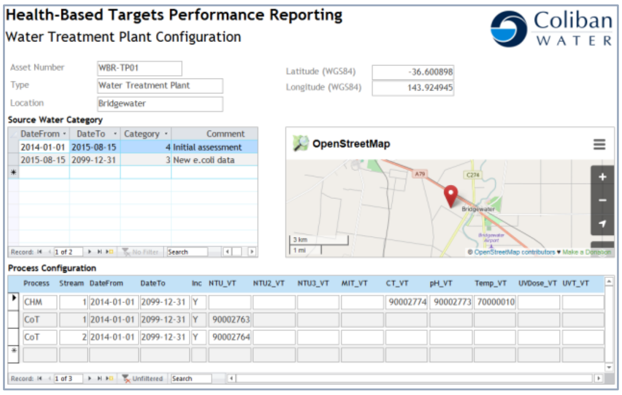

Figure 2: Configuration screen.

The screen contains the location of each plant, its source water category and the types of treatment barriers. The location of the plant is used for spatially oriented reporting of the outcomes.

The source water category (category 1–4) is determined for each water treatment site in accordance with the HBT Manual (Tier 1 assessment). These categories are used to determine the recommended minimum pathogen reduction requirements as listed in Table 3 in the HBT Manual.

The software is not able to accommodate minimum Log Reduction Values determined using a Tier 2 Quantitative Microbial Risk Assessment (QMRA). For systems where a QMRA has been undertaken the relevant control table in the database needs to be amended.

Source water categories can periodically change and the software records the period during which a system is assigned to a particular source category. The comment field is used to refer to the relevant Source Water Category Assessment.

The pathogen barrier configuration lists all process streams associated with each plant. All process types described in the HBT Manual are incorporated:

- Filtration
  - Coagulation, Flocculation & Dissolved Air Flotation
  - Coagulation, Flocculation & Sedimentation
  - Direct filtration
  - Conventional Treatment
  - Second Stage Filtration
  - Micro-Filtration
  - Ultra-Filtration
- Disinfection
  - Chlorination
  - Chloramination
  - Ultraviolet Disinfection
  - Ozone

Processes can consist of multiple streams. For example, a filter with two cells is represented as two separate filter processes. Where a process consists of multiple streams, the worst performing stream is used to report performance.

All menus are date-driven so that the analysis is undertaken in accordance with the most recent information. This can be used to reflect changes in treatment configuration over time, without changing the analysis of historic data. This feature is useful when, for example, systems temporarily converting a system from chloramination to chlorination.

Each process at a water treatment site is associated with a series of Virtual Tags. These are data sets created from SCADA data to facilitate analysis.

Virtual Tags

A common problem in data analysis is that information is created and collected to manage operational issues and not to solve future analytical problems. SCADA is an operational system that allows operators to make the best decisions. The system is not necessarily configured for post-hoc analysis of the data.

SCADA historian systems store the data for future retrieval. However, the primary function of archiving is to make the data accessible for future use, but not necessarily in a format suitable to solve future problems. A more precise term would be SCADA archivist instead of SCADA historian. An archivist stores what is available, while a historian is somebody who interprets archives to create new knowledge.

To analyse SCADA data in accordance with the HBT Manual, a series of queries has been developed to extract data from SCADA and create a separate database of the tags required to undertake the analysis. This database is called Virtual Tags. A Virtual Tag is a stream of SCADA data converted to a format suitable for analysis.

The concept of Virtual Tags is based on the principles of Data Science series (Leek, 2015; Peng, 2015). The core principle of Data Science is reproducibility of analysis. This requires that the data is stored in an appropriate way and that all outcomes can be reproduced by another analyst. Reproducibility is difficult to achieve using spreadsheets because the code and the data are merged. Using SQL or any other coding language (R, Python, MathLab) to analyse data provides greater transparency and reproducibility.

Second requirement is that the structure of the data is fully documented. The HBT software documents all SCADA tags used for analysis and the logical relationships between them.

Additionally, the data is stored as a regular time series, which implies that a measurement is available for each single minute of the analysis period.

The Virtual Tag approach alleviates two problems. Firstly, raw SCADA data is generally stored as an irregular time series, which complicates analytics. An irregular time series is a set of data points stored over time where the period between measurements is irregular. Secondly, the logical relationship between individual SCADA tags is not recorded within the system, for example, filtered water turbidity is measured regardless of whether the filter was running or not. These relationships can be extracted from the PLC and differ greatly between treatment sites. The Virtual Tag approach solves these issues by creating regular time series for each tag and, where required, combining several SCADA tags into one Virtual Tag.

The software uses three types of Virtual Tags.

Dummy Tags

Dummy Tags are Virtual Tags for which no physical measurement is available, but the data is required for analysis. For example, disinfection effectiveness is dependent on water temperature.

Where this data is not available, a constant value is assigned to a Dummy Tag as an estimate of temperature.

Source Tags

A Source Tag is a copy of the raw SCADA data for a particular tag, converted to a regular time series. There are two types of source tags:

- Measurements: Numerical assessment of the state of the system (e.g. turbidity, flow, tank level)
- Observations: Status of a system (e.g. valve position, alarms, status indicators)

The irregular time series presented by SCADA are converted to a regular time series with an interval of one minute. The Virtual Tag process stores the measured values (rounded to the nearest minute). Values between the measured points are derived using nearest-neighbour interpolation. To be able to distinguish between measured and interpolated values, measured values are marked as “actual”.

This approach is visualised in Figure 3. The orange squares are those available in the SCADA system. The Source Tag contains all blue dots. Whilst this approach increases the amount of data, it significantly simplifies analysis. The accuracy of this approach depends on the rules applied to the storing of SCADA data and the intrinsic accuracy of the instrument.

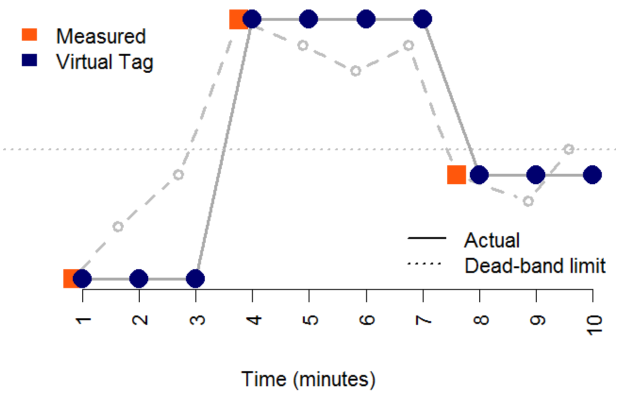

Figure 3: Virtual Tag creation.

Derived Tag

The last type of Virtual tag is the Derived Tag, which is a combination of one or more Virtual Tags. Examples of a derived tag are the conversion of tank volume from percentage to kilolitres, disinfection contact time, conversion of equipment status signal to a measurement, and so on.

The software contains twelve transformation rules to create Derived Tags, most importantly:

Ignore Under or Over Threshold

Only uses values for a given tag where the value of a related tag is below or above a given value. These rules are most commonly used to exclude filtered water turbidity values when the flow rate from the filters is less than a certain value. The value is most commonly set as 1 L/s. This ensures that analysed data only relates to times when the filter was active.

Cone and Cylinder Volume

This rule calculates the volume of a tank that with a conical section at the bottom based on tank level.

Disinfection contact time

Calculates disinfection contact time (C.t) based on disinfectant concentration, tank volume and flow, multiplied by a baffle factor. The C.t value is calculated in mg/L.min, over a period of one minute.

Multiply and Divide by Constants

Multiplies the values of a Virtual Tag with one constant and divides it by a second constant. An example of this type of transformation rule is: multiply tank percentage by maximum volume and divide by 100 to get tank volume.

Pass of Fail: Under or Over Threshold

Assigns a value of 1 or 0 (pass or fail) to the Virtual Tag when another Virtual tag is greater than or lesser than a given value. This rule is used to assess status signals provided by membrane filters.

On or Off Within Range

Sets the value of the Derived Tag at 1 (on) when the input is between two values. Sets the value at 0 (off) when this condition is not met. This can be used to assess the status flag of, for example, a UV system and determine when the system was running.

Sum of Samples

Uses the sum of a range of similar measurement to determine a new value. Used for flow meters to determine total flow.

Weighted Average

Provides a weighted average of two tags, based on two other tags. This method is used to create a weighted turbidity for a two-cell filter unit based on individual flow rates, to be used to assess the effectiveness of a disinfection downstream process.

Other rules are available and additional rules can be easily created to allow for greater flexibility.

Some Virtual Tags require the input of other Virtual Tags and the software cycles through all tags until all transformations have been implemented. The most complex Virtual Tags require three iterations.

During the Virtual tag creation process no raw data is deleted. Measured values and derived values are stored separately. Each data point also includes an indicator whether to use this data point or not.

Virtual Tag Example

The following example illustrates the Virtual Tags approach, using the Pyramid Hill water treatment plant. This plant consists of two processes relevant to HBT: conventional treatment and chlorination. To assess these processes, eight Source Tags are extracted from SCADA Historian.

- Treated Water Free Chlorine
- TWS Level
- Filtered Water pH
- Filtered Water Flow
- Filtered Water Turbidity 1
- Filtered Water Turbidity 2
- Pre TWS Turbidity
- Treated Water Flow

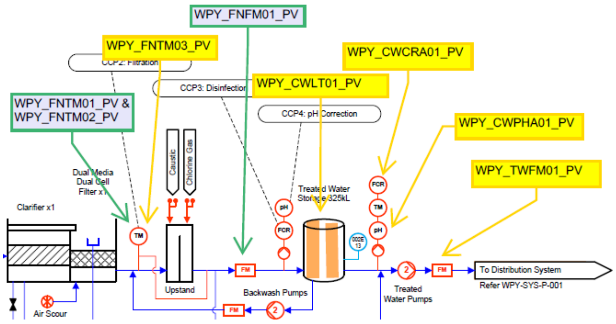

Figure 4: Pyramid Hill WTP.

These Source tags are transformed to Source Tags and Derived Tags using a series of processes:

- Treated Water Free Chlorine and pH data is only used where Treated Water Flow is less than 1 L/s.
- Turbidity measurements are only used where Filtered Water Flow is more than 1 L/s.
- The Treated Water Storage (TWS) Level is converted from a percentage to kilolitres by multiplying the percentage by the total volume (325 kL).
- Contact Time is calculated by combining the Derived Tags for Treated Water Free Chlorine, TWS Volume and Treated Water Flow, multiplied by a baffle factor (0.30).

The value of 1 L/s is used instead of 0 L/s because very low flows can be observed when the plant is not in operation.

This process results in five Virtual Tags stored as a regular time series with one-minute intervals:

- Filtered Water Turbidity 1
- Filtered Water Turbidity 2
- Contact Time
- Pre TWS Turbidity
- Filtered Water pH

These five Virtual Tags are then used by the water treatment assessment module to estimate the Log Reduction Values.

The Virtual Tag approach can be easily expanded to incorporate other data sources beyond those required for HBT assessment. The system is modular and new transformation rules can be easily introduced to further expand this approach.

Water Treatment Assessment

The HBT performance reporting system estimates LRV credits on the first day of each calendar month for the period of the whole of the previous calendar month, using the relevant system configuration and Virtual Tags.

Each of Coliban Water's water treatment plants consists of two or more pathogen barriers (treatment processes). For each type of process, the HBT Manual provides one or more decision rules to estimate the LRV credits. For example, if the 95th percentile of filtered water turbidity of a conventional treatment process is less than 0.15 NTU and no spikes over 0.30 NTU for more than 15 minutes have been detected during a whole month, then that process is considered to have the following LRV credits: Protozoa: 4.0, Bacteria: 2.0, Viruses: 2.0. (Table 7 of the HBT manual).

Where a process consists of more than one stream, such as multiple independent filter cells, both streams are assessed separately and the lowest performing stream is used to report Log Reduction Values.

To allow for the automation some interpretation has been applied to the rules defined in the HBT Manual. The decision rules provided by the HBT Manual can be divided in filtration and disinfection rules. The rules for filtration are all based on assessing data over the period of one month. The rules for filtration are based on an immediate assessment. One month of data is used for disinfection in order to equalise the reporting periods.

Another interpretation of the rules in the manual is to use the 99^{th} percentile in lieu of the maximum value and the 1^{st} percentile in lieu of the minimum value. This approach was taken to be able to ignore spurious low or high values.

The output of this module is a table that holds the outcomes of the calculations (95^{th} percentiles and so on) and the Log reduction Value Credits.

HBT Reports

Reporting of the results occurs on three levels:

- Region
- System
- Barrier

The three levels of reporting are linked which allows analysts to query the outcomes in great detail.

Region

At the regional level, the system reports on the Log Reduction Value shortfalls for the three pathogen types. To report on these values the 'cheesecake diagram' was developed. This is a pie chart where information is reported through the colour of the slices instead of their size (Figure 5).

The colours follow a modified traffic light arrangement, with colour adjusted to be colour blind safe. Green indicates that the plant for that specific pathogen type has achieved an LRV credit. Blue indicates that the plant operates in the safe zone for that pathogen type and yellow and caramel indicate that the system is not performing optimally.

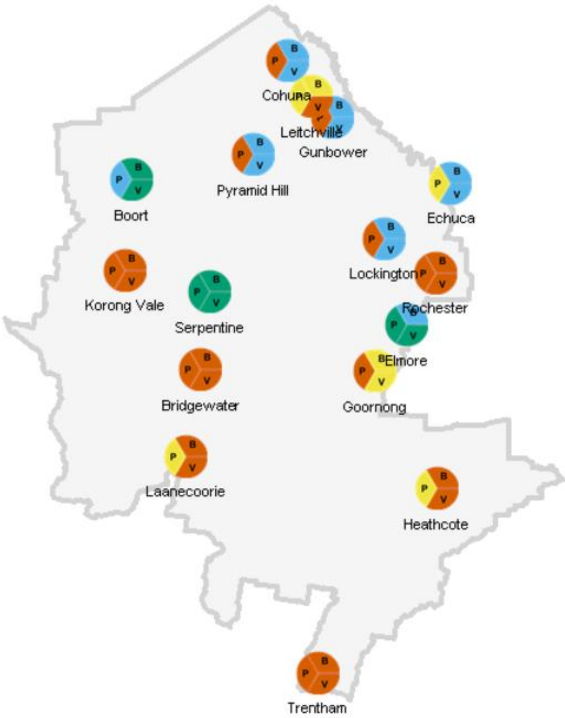

Figure 5: Regional performance report (simulated data).

System

Clicking on the relevant cheesecake provides further analytical detail at the system level. The bar chart in Figure 6 shows the required LRV for each pathogen type (horizontal line), the achieved LRV per month over the past year and the maximum achievable LRV in shaded colours.

Figure 6: System performance report.

Barrier

At the deepest level, the performance of each pathogen barrier can be investigated. Graphs are available to view the daily maximum and minimum values and to view daily trends of each of the Virtual Tags used in the assessment.

Software Operation and Maintenance

All software has been written in Microsoft SQL Server 2012 (Service Pack 1) and reports were developed in Reporting Services (SSRS).

The software runs a daily service to update the Virtual Tags and HBT assessments are undertaken monthly.

The software has reached maturity but is undergoing continual improvement. Further development opportunities screen to assist users create Virtual Tags and the additional of additional assessment rules for water recycling plants.

Coliban Water shares this software with other water utilities under an open source arrangement. The source code and documentation are freely shared with other water utilities under a no obligation and no warranty arrangement.

The objective of this approach is to collaboratively improve the software to come to a common understanding of analysing water treatment plant data in accordance with HBT principles. Interested parties can contact the author to discuss sharing arrangements.

CONCLUSION

The HBT Reporting software has been running for several months and has provided valuable intelligence on the performance of our water treatment plants.

Implementing this system has also improved our understanding of the information provided by the SCADA system. The construction of the Virtual Tags has required us to investigate this data at a level of detail not previously undertaken.

The Virtual Tag approach has provided a valuable resource that not only can be used for assessing Log Reduction Values, but they can also be used for other business intelligence processes, such as chemical consumption and water balances.

ACKNOWLEDGMENT

I would like to acknowledge the hard work of Brad Dole from Information Management Services, who converted our vision into code. This project would also not have been possible without the tireless work by Gary Schurr. Finally like to acknowledge my manager David Sheehan who has provided me with the freedom to experiment and innovate.

REFERENCES

Hrudey SE & Hrudey EJ (2004) *Safe Drinking Water: Lessons From Recent Outbreaks in Affluent Countries*. IWA Publishing, London.

McBride G (2005) *Using Statistical Methods for Water Quality Management: Issues, Problems and Solutions*. John Wiley & Sons.

NHMRC (2014) Health Based Targets for Microbial Safety of Drinking Water. Discussion Paper.

Leek, J. (2015) *The Elements of Data Analytic Style*. LeanPub.

Peng, R. (2015) *R Programming for Data Science*. LeanPub.

Prevos, P. and Sheehan, D. (2015), Health-Based Targets performance reporting. *Water* 42(7):64-66.

Water Services Association of Australia (WSAA) (2015): *Drinking Water Source Assessment and Treatment Requirements. Manual for the Application of Health-Based Treatment Targets*, WSA 202—2015-1.