

# IMPROVING DECISION MAKING IN WATER PLANT OPERABILITY THROUGH BAYESIAN BELIEF NETWORKS

How to get more from existing instrumentation data to achieve better process performance

T Trinh, C Pelekani, G Leslie, P Le-Clech

## ABSTRACT

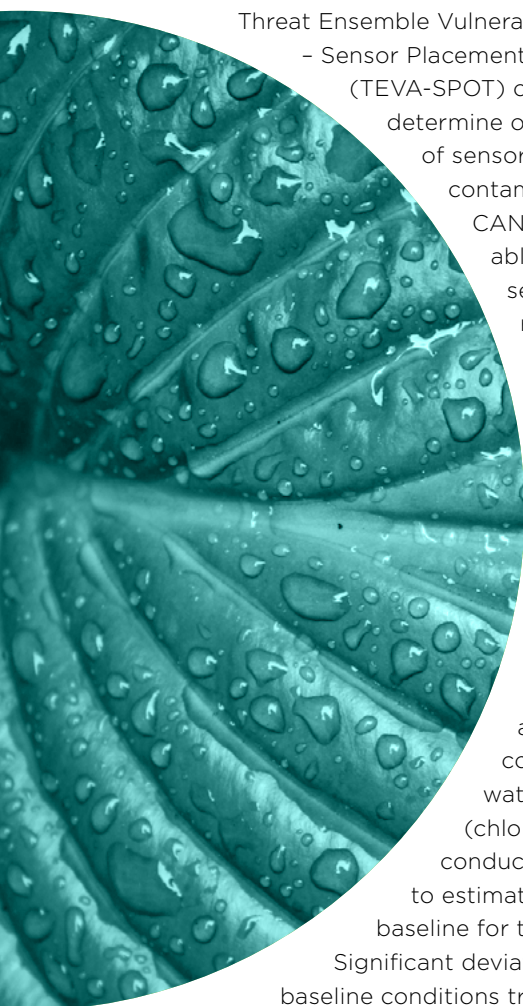
Real-time process and water quality monitoring has improved compliance and risk reduction in water treatment plants; however, it is a challenge to manage and extract maximum value from the terabytes of data generated by on-line instruments. This study used Bayesian Belief Network (BBN) to expand the use of historical data for improving decision making in water treatment plant operations. BBNs were developed and validated using on-line turbidity data and related operational conditions at the Mount Pleasant Filtration plant in South Australia. Data was converted to probability functions for possible causes and corresponding corrective actions conditions of high turbidity at the filter outlet. This quantitative statistical information can be used to develop appropriate response to “out of normal” operation events, e.g. events that cause turbidity excursions and other non-compliant conditions during operation.

## INTRODUCTION

The increase in stringency of water quality requirements in Australia has driven the need for improved data collection and process monitoring practices at water treatment plants (WTPs).

On-line monitoring has significant advantages over traditional monitoring techniques in providing real time water quality and process information. It is widely used across process industries, including Australian water utilities. Through supervisory control and data acquisition (SCADA) systems, data can be visualised in real time, and be sent to an archival data storage system (Mussared et al., 2015). Although SCADA systems provide for alarming and call-outs to operational staff when parameters are outside prescribed ranges, they cannot automatically correlate alarms with likely initiating factor(s). Operators then have to trend multiple parameters to assist in identifying the likely failure cause. Commercial software packages are available for data analysis. For example, OSIsoft PI allows data extraction and transformation for further retrospective analysis through a Microsoft Excel platform (Mussared et al., 2015). Microsoft Business Intelligence (Microsoft BI) allows data processing and visualisation to produce automated reports, and send alerts to operational staff when parameters fall outside prescribed ranges (Mussared et al., 2015).

Other data processing software have been developed by the US Environmental Protection Agency for specific event detection (Mussared et al., 2015, Hall and Szabo, 2009, Hart and McKenna, 2012, Storey et al., 2011).



Threat Ensemble Vulnerability Assessment – Sensor Placement Optimisation Tool (TEVA-SPOT) can be used to determine optimum placement of sensors for detection of contamination events. The CANARY algorithm tool is able to detect unusual sensor responses in real-time to analyse deviations in water quality from the set baseline for identification of contamination events (Mussared et al., 2015, Hart and McKenna, 2012). The Hach Event Moni-tor™ Trigger System analyses five commonly measured water quality parameters (chlorine, turbidity, pH, conductivity and TOC) to estimate a water quality baseline for the monitored system. Significant deviations from these baseline conditions trigger an alarm sent

to operators in real-time, and an auto sample is taken at the designated location. The system subsequently compares the computed algorithmic values to the archived fingerprints containing a wide range of threat contaminants specific to the system being monitored to estimate event type (e.g. water main burst, change in water source etc.) (Mussared et al., 2015, Hart and McKenna, 2012). Despite the clear benefit of online monitoring for risk reduction and improved compliance, obtaining the full effective value of large volumes of data created by online instruments is still an ongoing challenge.

Over the last decade, Bayesian Belief Network (BBN) has increasingly been used for modelling complex systems such as ecosystems and environmental management systems (Uusitalo, 2007, Aguilera et al., 2011). For example, BBN has been used to analyse factors influencing wildfire occurrence (Dlamini, 2010),

and to assess the public health risk associated with wet weather sewer overflows discharging into waterways (Goulding et al., 2012). BBN is a graphical model that represents a set of variables and their probabilistic dependencies. In BBN, variables are represented by nodes, and the relationships between variables are represented by directed arcs. Quantitatively, these relationships are expressed in conditional probability tables (CPTs) (Korb and Nicholson, 2011, Pearl, 2000). The graphical nature of BBN makes it an effective tool for modelling and communicating complex systems where there are multiple variables influencing each other (Sahely and Bagley, 2001). BBN is well-known for its ability to deal with uncertainty as the content of each variable is presented as probability distribution so BBN not only gives the result but also its expected frequency. In addition, BBN has a capability to combine different sources of knowledge, e.g. expert knowledge and real data (Uusitalo, 2007), and it is relatively easy modified and updated with new data and knowledge (Sahely and Bagley, 2001). Algorithms in BBN can also handle situations with missing observations which are often the case in environmental data. BBN is bidirectional so the same network can be used without modification to diagnose causes to specific problems given information about the output variables or to predict increases in operational efficiency given information about the input variables (Sahely and Bagley, 2001).

Although BBNs offer numerous benefits for modelling complex systems, their applications in water treatment systems are still very limited. A few studies have investigated the application of BBNs in diagnosing upsets in lab-scale wastewater treatment systems (Sahely and Bagley, 2001, Cheon et al., 2008). However, application of BBNs based diagnosis systems for full-scale water treatment processes has not been reported. Based on available on-line turbidity data and related operational inputs from a filtration process at the Mount Pleasant WTP, this case study aims to develop BBNs, which could allow the determination of probability of possible causes and corresponding corrective actions for given high filter outlet turbidity readings. Such quantitative statistical information from the models may assist operators to decide appropriate course of action when facing 'out of normal' operational events and thus improve effectiveness of decision making.

## MATERIALS AND METHODS

### Mount Pleasant WTP

Mount Pleasant WTP sources its water from the River Murray via an off-take from the Mannum-Adelaide pipeline. Mount Pleasant WTP has two independent treatment process trains, each with a design production capacity of 1.25 ML/d (2.5 ML/d total). This case study focuses on the filtration process associated with Stream 1 (Conventional Treatment). TStream 1 consists of MIEX pre-treatment (for enhanced removal of dissolved organic carbon), powdered activated carbon contact tank (normally employed for algal taste and odour challenge events), chemical coagulation, two-stage flocculation, high rate clarification (tube settlers) and dual media gravity filtration (Figure 1). There are two dual-media filters. Under normal operation only one of the filters is on-line, whilst the other is off-line. Operation switches when the on-line filter is backwashed to remove accumulated solids and restore filtration capacity. The outlet valve position is the primary indicator of the filter state. Digital state tags for the filters do not currently exist for data trending.

### Software Used

In this study, model development, evaluation and validation were performed in BayesiaLab 5.3. BayesiaLab is a powerful software package that provides an integrated workspace to handle BBNs. Apart from common algorithms (e.g. inference and parameter learning) included in conventional BBNs computer applications, this software includes useful features such as seven discretization methods, missing value processing, structure learning, model averaging,

cross-validation and numerous types of plots and reports. After developed and validated, the model was transferred to Netica 5.18 for easier communication. Netica was used because its interface is easier to follow for people who are not very familiar with BBNs. In addition, Netica is widely accessible as a free trial version can be used for networks having equal or less than 15 nodes, so broader community can view and use the model developed from this study.

### Model Development and Validation

BBN development is an iterative process that often requires several iterations before a final valid model is achieved. The major steps in developing a BBN comprise: (1) define model objective and scope; (2) collect and format data; (3) define model structure including selecting variables (nodes), deciding states of variables and connections between them; (4) parameterize the model; (5) evaluate and validate the model.

### Variable Selection and Available Data

From discussion with SA Water staff, as well as from information provided in the Mt Pleasant WTP Schematic Flow Diagram (SA Water, 2014), the Operations Manuals (SA Water, 2004), and the Water Quality Operating Plan (SA Water, 2013), the following 9 nodes were identified for the turbidity sensor BBN:

1. Raw water turbidity (NTU): describes the turbidity of raw water;
2. Raw water flow (l/s): describes the inlet flow of stream 1;
3. Outlet turbidity (NTU): describes the online turbidity readings of individual filter outlet;

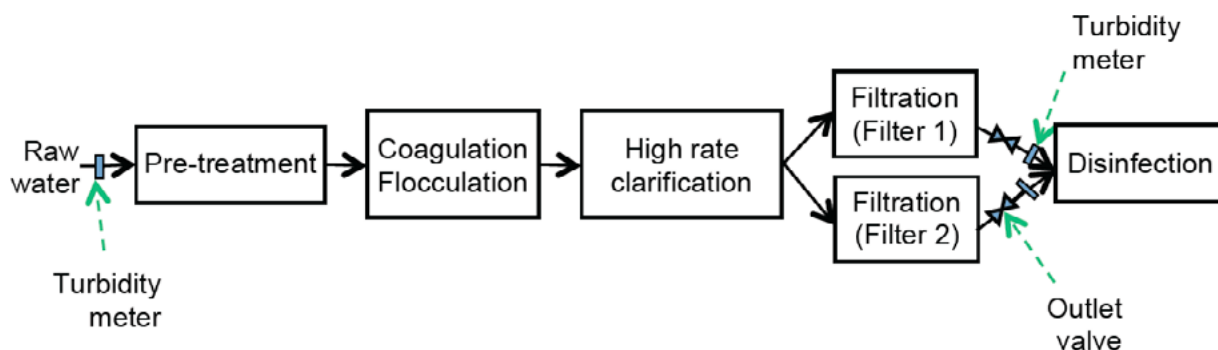


Figure 1 Process schematic of stream 1 of Mt Pleasant WTP

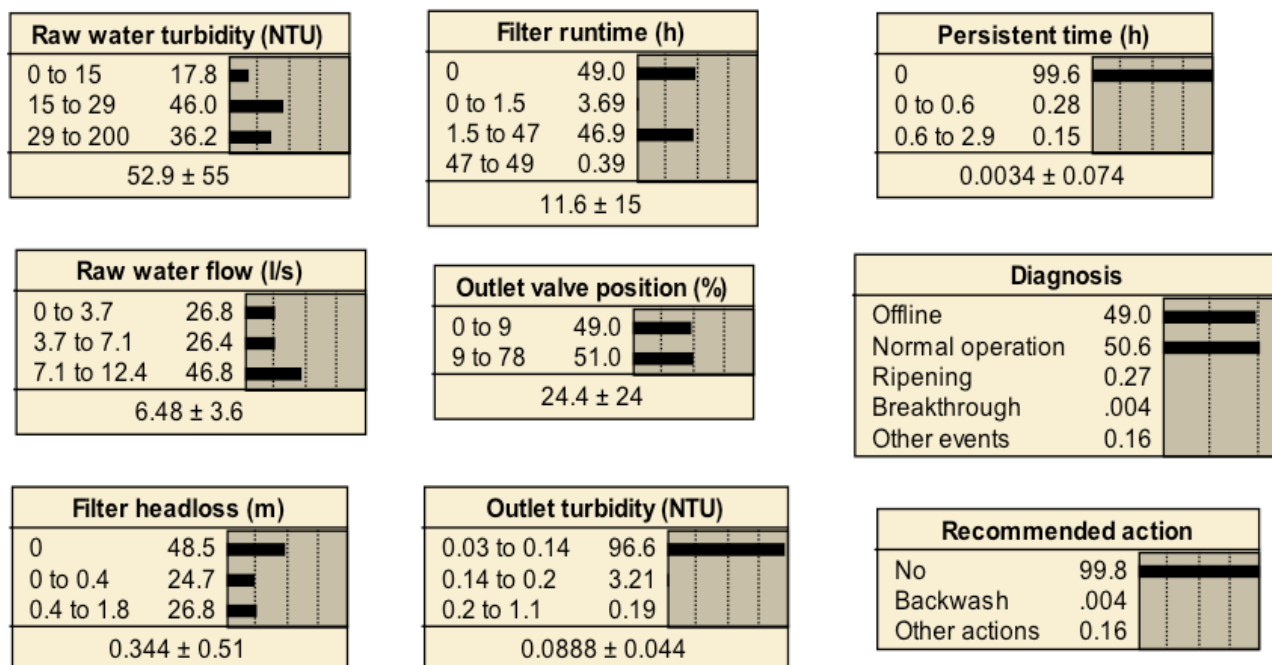


Figure 2 States and intervals for selected variables

4. Outlet valve position (%): indicates whether the filter is online (>9%) or offline (<9%);
5. Filter head loss (m): presents head loss of individual filters during the filtration cycle;
6. Filter runtime (h): presents actual runtime of individual filters during the filtration cycle; runtime is an important operational parameter for the filtration process;
7. Persistent time (h): describes how long outlet turbidity readings  $\geq$  Hi alarm set point of 0.14 NTU persist; information from Table 1A in Appendix shows that the persistent time is important for identifying causes and defining corresponding actions; During January-December 2015, the outlet turbidity rarely reached the HiHi alarm set point with delay, so the persistent time for HiHi alarm was not included to simplify the model;
8. Diagnosis: describes diagnosis for high turbidity readings, e.g offline, normal operation, ripening, breakthrough, turbidity meter failure, influent flow changes, influent water quality changes, changes in coagulation doses etc.;
9. Recommended action: describes recommended actions in response to high turbidity readings. e.g

changing coagulation dose, calibrate/maintain turbidity meters, backwash filters, shutdown filters, re-start filters etc.;

Among them, raw water turbidity, inlet flow, outlet turbidity, outlet valve position, and filter head loss data were obtained directly from the SCADA files featuring 90470 entries over January to December 2015. Filter runtime was calculated for each filtration cycle, starting when the outlet valve position changed from < 9% to > 9%, and ending when the valve changed vice versa. Persistent time was also calculated based on the duration of high turbidity readings. Diagnosis was identified based on information about identified events during the high turbidity periods from abstract files, or indicated by SCADA data. Recommended actions were identified based on event diagnosis and guidance in the Water Quality Operating Plan.

### States and Intervals for Variables

The states and intervals of the variables are presented in Figure 2. These states and intervals were defined based on insight from discussion with SA Water staff, as well as information from the Operations Manual (SA Water, 2004), and the Water Quality Operating Plan (SA Water, 2013) of Mt Pleasant WTP.



Raw water turbidity (NTU) node and Inlet flow rate (l/s) were given 3 states each by auto-discretisation with density approximation method. This discretisation method detects changes in the sign of the derivative of the density function in order to identify local optima. An interval boundary is subsequently added between each local optimum.

Filter head loss (m) were also given 3 states by auto-discretisation with density approximation method.

Outlet turbidity (NTU) node was given 3 states with breakpoints of Hi alarm value of 0.14, and HiHi alarm value of 0.20.

Outlet valve position (%) node was given 2 states as the filter is online when the outlet valve position > 9% and it is offline when the outlet valve position < 9%.

Diagnosis node was given 5 states including:

- ▶ Offline: when the filter is offline
- ▶ Normal operation: when the filter is online and outlet turbidity reading is below the Hi alarm set point of 0.14 NTU;

- ▶ Ripening: if the outlet turbidity readings  $\geq 0.14$  NTU occur during the first 1.5 h of the filtration cycles, and if the persistent time is less than 0.6 h, and if there is no information indicating that may due to other possible causes. 0.6 h was arbitrary defined as a sum of 10 min to trigger the alarm, plus 15 min allow compromised time. An additional 11 min data interval or variation in estimating the runtime was also considered in this study as a conservative measure.

- ▶ Breakthrough: if the high turbidity readings occur at the end of the filtration cycles, and if there is no information indicating that may due to other possible causes;

- ▶ Other events: if the high turbidity readings do not fall into the above windows. Other events could include turbidity meter failure, influent flow changes, influent water quality changes or changes in coagulation doses etc. These events are grouped together as there was no evidence for identifying each specific cause in this case study;

Filter runtime (h) node was given four states, including: 0 h - when the filter is off-line; 0-1.5 h - as it is one of criteria to define ripening cause; 1.5-47 h - when high turbidity readings are likely due to other causes; and > 47 h - when high turbidity readings are likely due to breakthrough.

Persistent time (h) node was given three states, including: 0 h - when turbidity readings are below the alarm set point; 0-0.6 h - as it is one of criteria to define ripening cause; and > 0.6 h - when high turbidity readings are likely due to other causes. Within this study, the 0.6 h limit was set up as a conservative value. In practice, alarm lasting for more than 0.25 h would indicate compromised data.

Recommended action node was given three states, including: No - when the filter is off-line, in normal operation or ripening with persistent time  $\leq 0.6$ h; Backwash - when the filter reaches breakthrough; and 'other actions'. As discussed, in the case herein, there is no evidence for identifying each specific event in this category, so the recommended action is to go through a checklist of what events previously occurred to identify specific causes and related actions.



The actions could include change in coagulant dose, calibrate/maintain turbidity meters, shutdown filters, initiate manual backwash, re-start filters, etc. This is labeled as 'other actions' in the model.

## Expert Knowledge Input and Automatic Structure Learning

The connections between the nodes were determined based on a combination of expert knowledge and automatic structure learning. Expert knowledge was introduced through two fixed arcs (Diagnosis → Outlet turbidity, and Diagnosis → Recommended action) and a range of forbidden arcs (Table 2A in Appendix) before performing structure learning. The fixed arcs have to be present in the network learned, while the forbidden arcs are prohibited. Automatic structure learning was conducted using Taboo search, unsupervised structural learning in BayesiaLab 5.3.

## Model Evaluation and Validation

Evaluation was conducted through model walkthrough where each of the model components including structure, variables, states and their probability connection is given a closer look to check if it makes sense and if any modification is necessary. For model validation, the data file was randomly split into training file with 80% of data, and testing file with 20% of data (Pollino et al., 2007). During the model development

process, only the training dataset was used in order to avoid overfitting. Any modifications to the models were conducted prior to the final evaluation with the testing dataset. Prediction accuracy was used to evaluate and validate the model. High prediction accuracy (%) indicates good prediction of target node by the model and vice versa.

## RESULTS AND DISCUSSION

### BBNs for Filter 1 and Filter 2

The networks learned from Filter 1 data are presented in Figure 3. By applying the same procedure for structure learning, a consistent network structure was also obtained for Filter 2. During the structure development process, the raw water turbidity and inlet flow nodes were observed to be insensitive to other nodes. For example, sensitivity of these nodes to outlet turbidity node and diagnosis node was <0.4%. As a result, these nodes were removed from the two models.

In general, the initial probability distribution for most variables including outlet valve position, filter run time, persistent time, diagnosis, recommended action is similar between Filter 1 and Filter 2. There are some differences in the initial probability distribution for outlet turbidity and head loss between the two filters. The probability of head loss being in 0-0.4 m was 25% for Filter 1, while 59% for Filter 2.

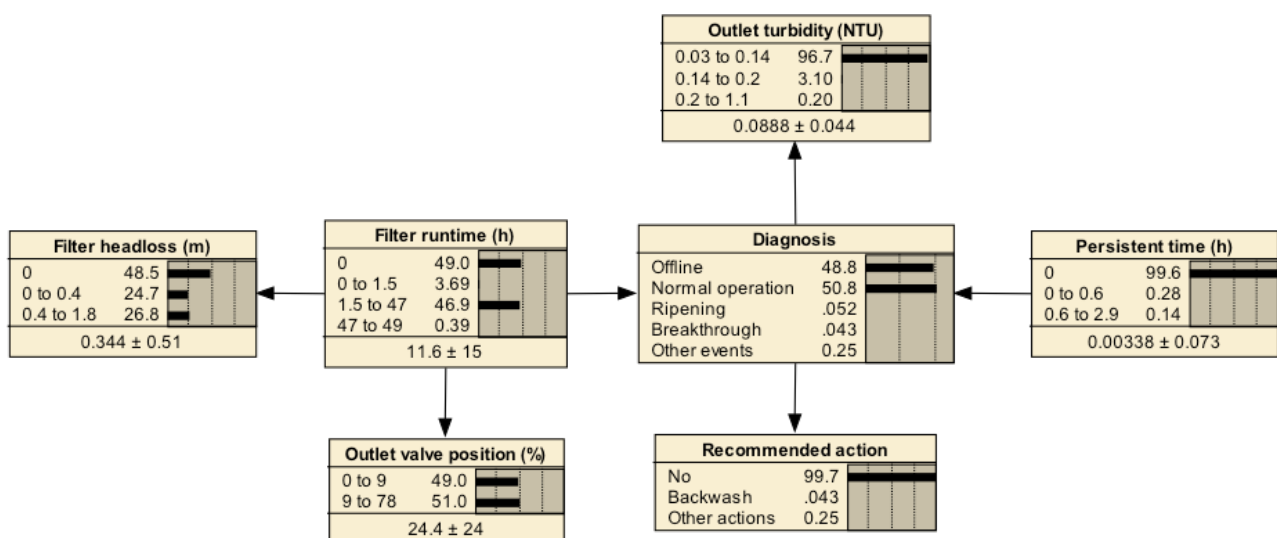


Figure 3 An initial BBN for filter 1

The probability of filter outlet turbidity being  $\leq 0.14$  NTU was 97% for Filter 1 and 99% for Filter 2. An assessment of filter performance conducted by SA Water staff identified a difference in head loss profile between the two filters. It was hypothesised that the quantity of filter media (sand and filter coal) may not be the same in the two filters. In addition, the outlet turbidity meter for Filter 2 is newer than for Filter 1 and also has a different light source (880 nm versus white light), so there is an offset of 0.02 NTU in Filter 2 outlet turbidity reading compared to Filter 1. These may be the reasons for the differences in the initial probability distribution for outlet turbidity and head loss between the two filters.

Some examples of useful information that the BBN can provide under given scenarios are presented in Figure 4.

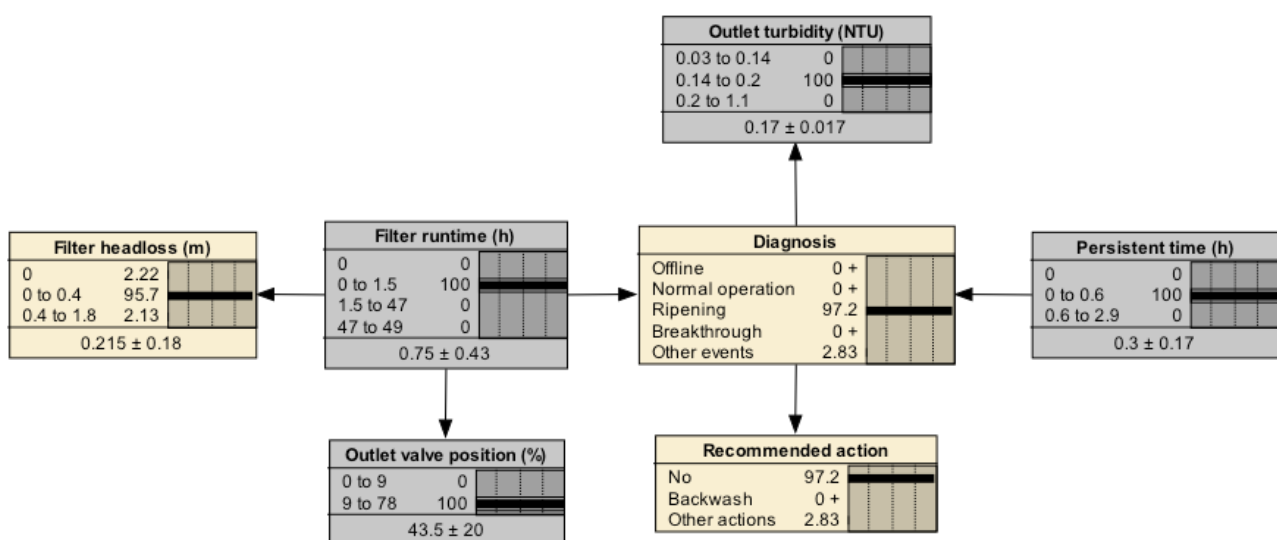
Figure 4a presents a scenario where high outlet turbidity of 0.14-0.2 NTU occurs within a runtime of 0-1.5 h and persists within 0.6 h. From the historical data, the BBN can diagnose that ripening is the most likely cause for this event with a probability of 97.2%, and because the event recovers within 0.6 h, no corresponding action is needed.

Figure 4b presents a scenario where high outlet

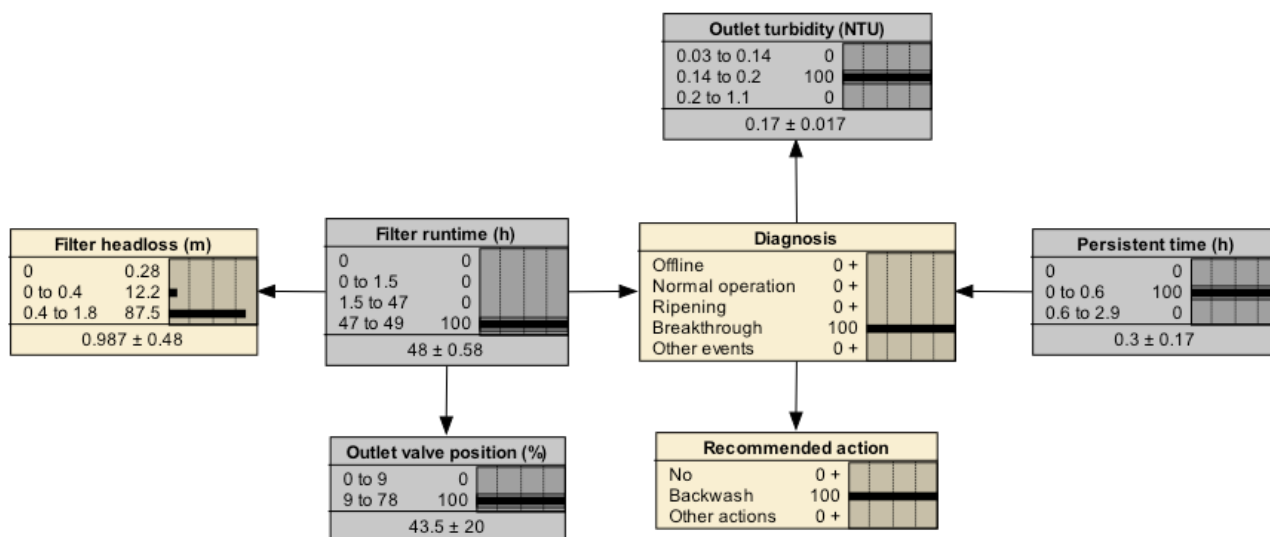
turbidity of 0.14-0.2 NTU occurs at the end of the filtration cycle and persists within 0.6 h. From the historical data, the BBN can diagnose that breakthrough is the most likely cause for this event with a probability of nearly 100%, and recommended corresponding action is to backwash the filter.

Figure 4c presents a scenario where high outlet turbidity of 0.14-0.2 NTU occurs in the middle of the filtration cycle and persists more than 0.6 h. From the historical data, the BBN can diagnose that other events most likely occurred with a probability of nearly 100%. The recommended action is to go through a checklist of previous events to identify specific cause and related action. This is labeled as 'other actions' in the model.

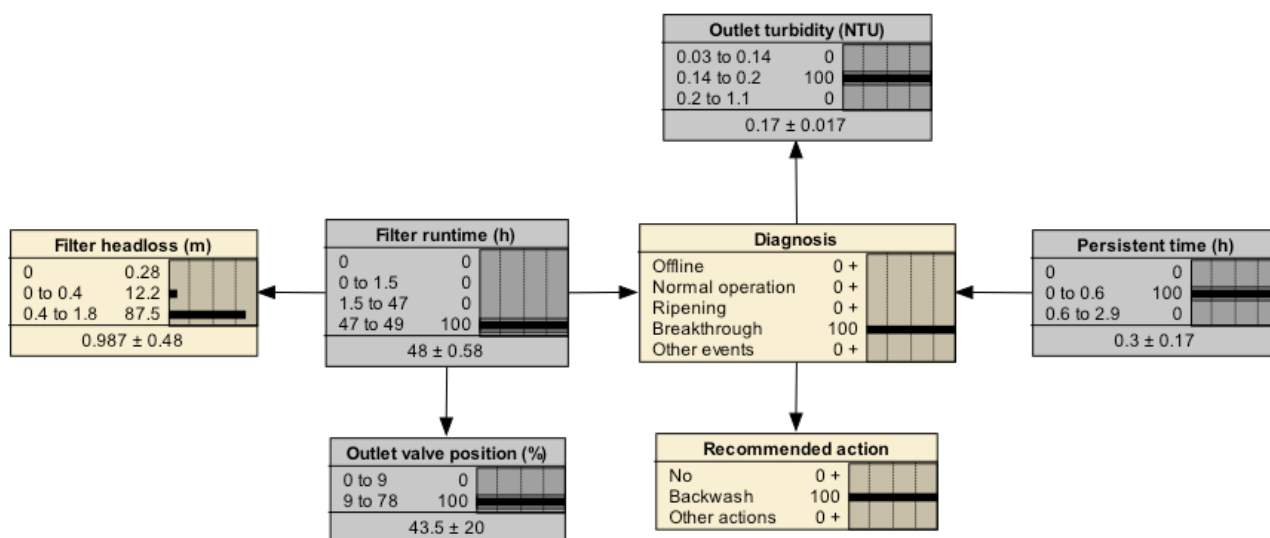
Similar information can also be obtained from the Filter 2 model. This quantitative statistical information provided by the BBNs is potentially useful for the plant operators in identifying and confirming possible causes and corresponding actions in given scenarios. When more information and data regarding other possible causes and corresponding actions are available. The BBNs can be updated and more states can be added into the diagnosis and recommended actions nodes to make other events and actions more specific.



a. A scenario indicating high probability of ripening



b. A scenario indicating high probability of breakthrough



c. A scenario indicating high probability of other events

Figure 4 Scenario analysis of the filter 1 turbidity sensor BBN (a. A scenario indicating high probability of ripening; b. A scenario indicating high probability of breakthrough; c. A scenario indicating high probability of other events)



### Model Evaluation and Validation

Model evaluation was conducted through model walkthrough where each of the components, including structure, variables, states and their probability connection were analysed. For model validation, the data file was randomly split into training file (80% of data), and testing file (20% of data) (Pollino et al., 2007). During the model development process, only the training dataset was used in order to avoid overfitting. Any modifications to the models were conducted prior to the final evaluation with the testing data set. Considering Diagnosis as a target node, the BBNs of both filters show prediction accuracy  $\geq 99\%$ .

### CONCLUSIONS

The quantitative statistical information provided by BBNs is potentially useful to plant operations staff in identifying and confirming possible causes and corresponding actions for a range of scenarios.

Applying BBNs in this environment presents a number of challenges, in particular obtaining appropriate data set and information for the model development and validation. Treatment plants with robust operational data records and sound management strategies usually have few “out of normal operation” incidents. Although full data sets could be obtained from these plants, data variation is low (few relevant incidents), which limits the full development and validation of comprehensive BBNs. On the other hand, treatment facilities featuring a wide range of “out of normal operation” incidents, are usually not well monitored and lack comprehensive data set. Therefore, for the purpose of developing and validating BBNs, treatment facilities with a balance of ‘out-of-normal’ and normal operation data records and operational responses, are preferred.

As a result of this study, it has been demonstrated that better decision tools can be developed from historical data. In addition, BBNs can be considered as a potential training tool for new operators as the models can stimulate better understanding of the links between different operating parameters in treatment processes. Furthermore, BBNs can also serve as a complementary strategy to existing management strategies to improve plant process reliability.

### ACKNOWLEDGEMENTS

This project was funded by WaterRA (Project 1075) with financial support from SA Water. Particular acknowledgements are given for Tahlia Sklifoff at SA Water, and Guido Carvajal Ortega and Keng Han Tng at UNSW.

### THE AUTHORS



**Trang Trinh** received her PhD in Environmental Engineering from The University of New South Wales, Australia. She is currently a postdoctoral research fellow in the UNESCO Centre for Membrane Science and Technology at the University of New South Wales.

email: [trang.trinh@unsw.edu.au](mailto:trang.trinh@unsw.edu.au)



**Prof Greg Leslie** is the Director of the UNESCO Centre for Membrane Science and Technology at UNSW Australia. Prior to joining UNSW, he worked in the public and private sectors on water treatment, reuse and desalination projects, including the Singapore NEWater and the Orange

County Water District in California. He has served on the Water Advisory Committee for the Prime Ministers Science Engineering and Innovation Council, among others, and currently serves on the Independent Advisory Panel for the Orange County Groundwater Replenishment Project. email: [g.leslie@unsw.edu.au](mailto:g.leslie@unsw.edu.au)



**A/Prof Pierre Le-Clech** has been working in the School of Chemical Engineering at UNSW Australia for the last 12 years, after completing his PhD on membrane bioreactors in Cranfield University, UK.

Over the years, he has studied many aspects of the water and wastewater treatment by membrane processes, focusing on membrane bioreactors and other hybrid membrane systems.

email: [p.le-clech@unsw.edu.au](mailto:p.le-clech@unsw.edu.au)



**Dr Con Pelekani** is Manager Water Treatment Performance & Optimisation at SA Water, Adelaide.

[con.pelekani@sawater.com.au](mailto:con.pelekani@sawater.com.au)

### REFERENCES

- AGUILERA, P. A., FERNÁNDEZ, A., FERNÁNDEZ, R., RUMÍ, R. & SALMERÓN, A. 2011. Bayesian networks in environmental modelling. *Environmental Modelling and Software*, 26, 1376-1388.
- CHEON, S.-P., KIM, S., KIM, J. & KIM, C. 2008. Learning Bayesian networks based diagnosis system for wastewater treatment process with sensor data. *Water Science and Technology*, 58, 2381-2393.

DLAMINI, W. M. 2010. A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modelling and Software*, 25, 199-208.

GOULDING, R., JAYASURIYA, N. & HORAN, E. 2012. A Bayesian network model to assess the public health risk associated with wet weather sewer overflows discharging into waterways. *Water Research*, 46, 4933-4940.

HALL, J. S. & SZABO, J. G. 2009. Distribution system water quality monitoring: Sensor technology evaluation methodology and results. USA: U.S Environmental Protection Agency.

HART, D. B. & MCKENNA, S. A. 2012. User manual for Canary version 4.3.2. EPA 600/R-08040B. USA: US Environmental Protection Agency.

KORB, K. & NICHOLSON, A. E. 2011. Bayesian artificial intelligence, USA, Chapman Hall/CRC Press.

MUSSARED, A., FABRIS, R. & CHOW, C. 2015. Optimisation of existing instrumentation to achieve better process performance (Waterra 1075)-Literature review.

PEARL, J. 2000. Causality: models, reasoning, and inference, US, Cambridge University Press.

POLLINO, C. A., WOODBERRY, O., NICHOLSON, A., KORB, K. & HART, B. T. 2007. Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling and Software*, 22, 1140-1152.

SA WATER 2004. Operations manual for Mt Pleasant water treatment plant.

SA WATER 2013. Water quality operating plan for Mount Pleasant water treatment plant.

SA WATER 2014. A schematic flow diagram for Mount Pleasant water treatment plant.

SAHELY, B. S. G. E. & BAGLEY, D. M. 2001. Diagnosing upsets in anaerobic wastewater treatment using bayesian belief networks. *Journal of Environmental Engineering*, 127, 302-310.

STOREY, M. V., VAN DER GAAG, B. & BURNS, B. P. 2011. Advances in on-line drinking water quality monitoring and early warning systems. *Water Research*, 45, 741-747.

UUSITALO, L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, 203, 312-318.

