Australian Water Recycling
Centre of Excellence

Project Report

National Validation Guidelines for Water Recycling: Comprehensive Bayesian Recycled Water Validation

A report of a study funded by the
Australian Water Recycling Centre of Excellence

D. Roser, G. Carvajal, B. van den Akker, A. Keegan, R. Regel and S. Khan, December 2015

National Validation Guidelines for Water Recycling: Comprehensive Bayesian Recycled Water Validation

Project Leader

Dr Stuart Khan
School of Civil Environmental Engineering
University of New South Wales
Sydney NSW 2052 AUSTRALIA

Telephone: +61 2 93855070

Contact: Stuart Khan s.khan@unsw.edu.au

Partners

Melbourne Water
Water Corporation (WA)
South East Water Ltd
SA Water

About the Australian Water Recycling Centre of Excellence

The mission of the Australian Water Recycling Centre of Excellence is to enhance management and use of water recycling through industry partnerships, build capacity and capability within the recycled water industry, and promote water recycling as a socially, environmentally and economically sustainable option for future water security.

The Australian Government has provided $20 million to the Centre through its National Urban Water and Desalination Plan to support applied research and development projects which meet water recycling challenges for Australia's irrigation, urban development, food processing, heavy industry and water utility sectors. This funding has levered an additional $40 million investment from more than 80 private and public organisations, in Australia and overseas.

ISBN: 978-1-922202-70-3

Citation:

D. Roser, G. Carvajal, B. van den Akker, A. Keegan, R. Regel and S. Khan (2015). *National Validation Guidelines for Water Recycling: Comprehensive Bayesian Recycled Water Validation*, Australian Water Recycling Centre of Excellence, Brisbane, Australia.

Date of publication: December 2015

Publisher:
Australian Water Recycling Centre of Excellence
Level 5, 200 Creek Street, Brisbane, Queensland 4000
www.australianwaterrecycling.com.au

This report was funded by the Australian Water Recycling Centre of Excellence through the Australian Government's National Urban Water and Desalination Plan.

Disclaimer

Use of information contained in this report is at the user's risk. While every effort has been made to ensure the accuracy of that information, the Australian Water Recycling Centre of Excellence does not make any claim, express or implied, regarding it.

National Validation Guidelines for Water Recycling: Comprehensive Bayesian Recycled Water Validation.

Never Stand Still | Faculty of Engineering | School of Civil & Environmental Engineering

Date: 24 December, 2015

Authors: David Roser, Guido Carvajal, Ben van den Akker, Alex Keegan, Rudi Regel and Stuart Khan
Address: School of Civil Environmental Engineering, University of New South Wales.
Phone: (02) 93855070
Email: s.khan@unsw.edu.au

Table of Contents

List of Tables in Document

List of Figures in Document

Executive Summary

In the context of a water treatment process or a recycling scheme, validation is taken to mean *"the confirmation that the treatment technology meets the specified performance targets."* Current Australian water recycling guidelines describe the concept of and need for validation but do not specify how the validation should be done.

The Australian Water Recycling Centre of Excellence (AWCoE) established a major collaborative research project to develop a National Validation Framework. This project, known as 'NatVal', was composed of a series of subprojects undertaken by various research groups. Some of these projects focused on validation protocols for specific treatment processes (membrane bioreactors, reverse osmosis membranes, biological systems). The project described in this report was tasked with identifying a framework, which could provide some consistency in the approach taken for various treatment processes and a means of validating an overall "system" in addition to the validation of its individual components.

After careful consideration of a wide variety of risk assessment and risk management tools, the use of Bayesian Nets (BNs) were identified as a means of collating information describing system performance, as well as producing validation conclusions through the formalised description of cause-effect relationships that define treatment process mechanisms and observational data. There are a large number of software packages available for constructing and analysing BNs. We reviewed a number of them and all had various advantages and disadvantages. For the sake of simplicity and consistency, the vast majority of work presented in this report was developed using Netica™ by Norsys Software.

Chapter 1 reiterates the project aims and objectives. It also provides a brief description of the initial review and planning that led to refining the project approach. The adopted approach, based on the use of BNs, was identified as a framework, which could:

1. Incorporate or allow for the use of all or the large majority of the 31 ISO 31010 risk management tools.
2. Include or clarify the relationship between primary contaminants, indicators and surrogates so that recycled water validation testing assessments could be cost effective and the data underpinning this either credible or at the least auditable.
3. Ensure quantification of risk, and treatment train and individual process effectiveness is transparent, standardised, readily auditable and straightforward to understand for regulators as well as technical specialists.

Chapter 2 provides a generalised introduction to Bayes theorem and BNs. It describes the key characteristics of BNs and their capabilities for modelling cause-effect systems. Furthermore, the chapter provides an overview of what the project team have learnt regarding the ways in which BNs could be used to support an operational recycled water validation framework.

Chapter 3 builds upon the introduction given in Chapter 2 to provide more specific descriptions of how BNs can be used operationally for recycled water system validation. It describes how the proposed framework is consistent with many existing practices and how relatively minor developments can offer the opportunity to exploit Bayesian inference related concepts, methods and tools. From the outcomes of the work presented in this report, this chapter presents a list of principles should be applied to recycled water system validation.

Chapter 4 presents a case study example for "Whole of System Validation". This example is based on data, provided by SA Water, on the performance of the Bolivar water recycling plant.

The overall system performance is characterised in terms of log reduction values (LRVs) for the multiple barrier system. This example shows how a BN can be constructed based on an existing understand of cause-and-effect relationships. It directly reflects the way the system was designed and is assumed to operate. Individual water treatment barriers are constructed with a particular design performance, which is subsequently cross-matched with validation data. The design and validation values can be assessed in combination or separately. Various scenarios (including combinations of scenarios) can be quickly tested to determine whether validation objectives can be assumed to have been met under each one.

The work summarised in Chapter 5 provides an introduction to the concepts of “naïve” and “semi naïve” BNs. This approach proved to be of significant value for identifying the (combined) predictive capability of various operational and monitoring parameters. Complex systems (in this case, activated sludge) are difficult to model since knowledge of cause-effect relationships between what can be measured (e.g. monitoring parameters) and what is desired to be known (e.g. pathogen LRVs) are often not well defined. Naïve and semi naïve BNs differ from ‘causal’ BNs in that they do not begin with a fully defined understanding of the system. Instead, procedures are used to ‘learn’ predictive relationships among the available data. We have developed a clear stepwise procedure for doing this. Only a brief summary is presented in this report chapter since the full details have now been published in a peer reviewed journal article (Carvajal et al., 2015), which is attached as Appendix A. We then compared the outcome using a causal BN, based on our expert understanding of system cause-effect. The development of Naïve and semi naïve BNs was aided by the use the general data mining software WEKA.

The first of two case studies undertaken in collaboration with Melbourne Water is presented in Chapter 6. This work was undertaken using data previously collected and provided by Melbourne Water for re-analysis. It provides an example of how operating parameters might be related to microbial concentrations or LRVs. The outcomes presented demonstrate the use of these data for:

- Initially constructing causal BNs for the three central analytes which characterise treatment effectiveness as LRVs;
- Comparing the result of calculating LRVs using a BN vs. Learning from primary data tables;
- Assessing model accuracy and hence prediction reliability;
- The use of semi-naïve BNs compared to causal BNs;
- Gathering summary statistics pertinent to LRV calculation and crediting;
- Improving understanding of system structure and function using Netica’s Sensitivity (to findings) analysis tool;
- Bayesian Validation;
- The use of WEKA in data mining especially large on line data sets.

The second case study undertaken with Melbourne Water is presented in Chapter 7. This work was an experimental field study designed to address uncertainties remaining from the work outlined in Chapter 6. The experimental campaign confirmed the insight sand estimated obtained during the initial validation assessment and indicated that the establishment of LRV credits for the treatment process (preozonation and biofiltration) was feasible. This work demonstration that variance in LRVs over the short-term were comparable to those which had previously been observed over a longer term. The performance of preozonation alone for disinfection was observed to be highly variable, but much more consistent results were achieved by considering the preozonation and biofiltration units as a single treatment step.

Chemical analysis undertaken for this work revealed only minor (3-13%) conversion of bromide to bromate. Online monitoring for turbidity and UV_{254} absorbance were shown to be potentially more useful than had previously been assessed.

Chapter 8 presents a case study application for pathogen LRV estimation from data collected from full scale activated sludge systems by CSIRO Land and Water (in support of another NatVal subproject). The data provided included influent/effluent pathogen concentrations for four activated sludge sewage treatment plants. Due to the complex biological nature of these systems, it was apparent that a naïve/semi naïve BN approach would be more suitable than a causal model, which would require explicit assumptions of cause-effect relationships between operational/monitoring parameters and achieved LRVs. This work demonstrated that significant bacterial and viral LRVs could be achieved, but a high variation in performance was also observed. Temperature and nitrogen concentrations were determined to be effective predictive parameters for LRV performance.

Chapter 9 provides a detailed case study of an ultrafiltration (UF) membrane treatment process undertaken in collaboration with SA Water. For comparison, both causal and semi-naïve models were developed, which yielded similar predictions for pathogen LRVs. The causal model was constructed based on the assumption that the different experiments, units, time-step and replicate measurements could independently influence inlet and outlet bacteriophage concentrations, and hence LRVs, in different ways and extents. Many of the statistics and observations generated from this study could have been alternatively generated using conventional means. However the BNs allowed the whole LRV picture to be captured in one platform and in a clear graphic format. This was helpful for communication, discussion and decision support in respect to achieving concurrence on LRV credits.

As a means of scoping the incorporation of ‘hazardous events’ in the validation process, an assessment of the use of BNs for fault tree analysis is presented in Chapter 10. Netica software was shown to be suitable for constructing basic fault trees with classical AND and OR gates. Furthermore, the BN constructed in this example allowed extensive exploration of which factors were dominant and trivial when it came to exploring the overall system reliability.

Chapter 11 provides an assessment of alternative means of summing LRV credits from multiple barriers. Where LRVs are known to be variable or uncertain, current techniques tend to adopt lower-range values such as 5th percentile values. If LRVs attributed to multiple barriers are treated in this way and then summed, the final ‘multiple barrier’ LRV is increasingly conservative, depending on the variability of each barrier and the number of barriers summed. This chapter explores the effect of these compounding conservative assumptions and demonstrates how probabilistic analysis (which may be achieved by conventional Monte Carlo assessment or by the use of BNs) can avoid provide an alternative means of assessment. The potential advantage is that the same level of conservatism (e.g., the use of a 5th percentile LRV value) can be maintained regardless of the number of barriers summed.

Chapter 12 describes a number of additional Bayesian applications, which were not explored in detail during this work, but are potentially valuable recycled water validation. These include applications of ‘sequential learning’ and ‘adaptation’. Furthermore, this chapter explores how the familiar likelihood/consequences risk matrix approach can be incorporated in BN models. Techniques are also described for more general recycled water system definition (prior to high resolution monte carlo analysis) and for identifying appropriate sample sizes required to achieve a specified level of confidence in validation monitoring.

In conclusion (Chapter 13), we present a generalised approach to water recycling process and scheme validation. It is proposed that causal models, guided by accepted understanding of

cause-effect relationships, will be suitable for some water treatment processes. These include processes for which validation protocols have already been developed and accepted, based on accepted relationships between operational conditions, monitoring observations and treatment performance (LRVs). However, in more complex, less well understood systems the use of semi-naïve BN techniques may greatly aid the identification of appropriate operational and monitoring parameters within which to define a system as having been validated for satisfactory treatment performance.

A glossary of technical and less-familiar terms used in this report is provided in Chapter 14.

Acknowledgements

Many organisations and individuals made contributions and provided invaluable in-kind resources to this work. These include industry partners who provided data, assistance and other resources; notably SA Water, Melbourne Water, Water Corporation (WA), CSIRO Land & Water, and The National Measurements Institute (NMI).

Other major individual contributing included Paul Monis (SA Water), Petra Reeve (SA Water), Santosh Ramesh (UNSW), James McDonald (UNSW), Fred Leusch (Griffith University), Simon Toze (CSIRO), Eduardo Cosa (Water Corporation), Stacey Hamilton (Water Corporation), John Mieog (Melbourne Water), Cheryl Lim (NMI), Daniel Deere (Water Futures), Cedric Robillot (AWRCoE), Mark O'Donohue (AWRCoE) and David Cunliffe (SA Health).

This research was supported by the Australian Water Recycling Centre of Excellence (NatVal2.2. WaterRA Project Number 3018-12) through the Australian Government's National Urban Water and Desalination Plan. Industry partners were Melbourne Water, Water Corporation (WA) and South East Water Ltd.

1. Introduction

1.1. Project aims and objectives

The specific aims and objectives adopted for this work were as follows:

1. Review available risk assessment methods for implementation in the validation of water recycling processes and projects.
2. Provide recommendations for the collection and incorporation of suitable chemical, microbial, or surrogate data in performance assessment/validation activities. This will specifically address factors such as the appropriate duration of validation testing.
3. Identify key characteristics of a framework to apply to the validation of water recycling unit processes to ensure consistency of data collection, statistical evaluation, and performance assessment.
4. Develop a rigorous basis for the incorporation of potential hazardous events (i.e., non-ideal operational conditions) and performance failures in the validation process.
5. Provide case studies of appropriate risk assessment methods for the validation of a specific water recycling process
6. Provide recommendation of practical approaches for combining the individual validation of unit processes to achieve the overall validation of multi-barrier water recycling projects.

1.2. Initial and interim planning, reporting and refining the project's approach

Performance validation is a key step in managing risks associated with water recycling projects. As a component of risk management, formalised validation guidance should be consistent with current best practices for risk management. Current Risk Management standards and guidelines provide an array of at least 31 diverse risk assessment and risk management tools (IEC/ISO, 2009, ISO, 2009, Standards Australia and Standards New Zealand, 2009, Standards Australia and Standards New Zealand, 2013). These tools are listed below in Table 1.

Review of the literature indicated all of the tools presented in Table 1 were applicable to water recycling, though a number had only seen moderate application in the water supply and treatment industry so far.

Further this list was itself somewhat incomplete and there were many emerging best practices which would be difficult to ignore. For example Quantitative Microbial Risk Assessment which is central to the Australian Recycled water guidelines and is being rolled out as best practice in USA water management as well (U.S. Environmental Protection Agency and U.S. Department of Agriculture, 2012) barely rated a mention in the ISO based standards (IEC/ISO, 2009).

Table 1. ISO 31010 Risk Management Tools

Tool Class	Tool Code	Tools and Techniques
Supporting	B01	Brainstorming
	B02	Structured or semi-structured interviews
	B03	Delphi
	B09	Structure « What if? » (SWIFT)
	B20	Human reliability analysis

Tool Class	Tool Code	Tools and Techniques
Look up	B04	Check-lists
	B05	Primary hazard analysis
Function Analysis	B06	Hazard and operability studies (HAZOP)
	B07	Hazard Analysis and Critical Control Points (HACCP)
	B13	Failure mode effect analysis
	B22	Reliability centred maintenance
	B23	Sneak circuit analysis
	B27	FN curves
Scenario analyses	B08	Environmental risk assessment
	B10	Scenario analysis
	B11	Business impact analysis
	B12	Root cause analysis
	B14	Fault tree analysis
	B15	Event tree analysis
	B16	Cause and consequence analysis
	B17	Cause-and-effect analysis
	B19	Decision tree
	B28	Risk indices
	B29	Consequence/probability matrix
	B31	Multi-criteria decision analysis (MCDA)
Controls assessment	B18	Layer protection analysis (LOPA)
	B21	Bow tie analysis
Statistical Methods	B24	Markov analysis
	B25	Monte Carlo simulation
	B26	Bayesian statistics and Bayes Nets
	B30	Cost/benefit analysis

[a]Some tools were not allocated to classes in ISO 31010. So the allocations shown include our suggestions based on what the tools are used for.

We concluded that the framework and tools we proposed must or should:

1. Incorporate or allow for the use of all or the large majority of the 31 ISO 31010 risk management tools.
2. Include or clarify the relationship between primary contaminants, indicators and surrogates so that recycled water validation testing assessments could be cost effective and the data underpinning this either credible or at the least auditable.
3. Ensure quantification of risk, and treatment train and individual process effectiveness is transparent, standardised, readily auditable and straightforward to understand for regulators as well as technical specialists.

Provisionally we formed the opinion that Monte Carlo risk assessment methods (Haas et al., 1999, Haas and Eisenberg, 2001) and Bayesian Networks (BNs) (Korb and Nicholson, 2011, Kragt, 2009) might significantly contribute to the solution to these broad problems as well as providing operational methodologies. This view was based on earlier experience of risk assessment application to recycled water (Roser et al., 2006, Khan and Roser, 2007, Khan et al., 2007) and an earlier effort to apply BNs to recycled water (Donald et al., 2009, Donald et al., 2010, Cook et al., 2013 (accessed)).

Accordingly, the methodology and case studies presented in this report seek to demonstrate how the use of BNs (incorporating Monte Carlo-type probabilistic assessment) can provide considerable power for the application of water recycling system validation practice.

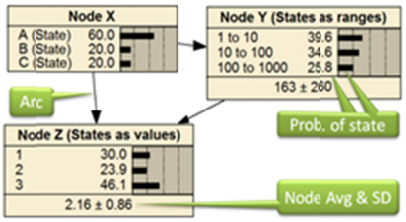
2. Bayes theorem: A framework for recycled water treatment validation

Bayesian Nets (BNs) are acyclic Cause=>Effect based computer models where variables ('Nodes') are linked to one another by unidirectional 'Arcs'. Typical Directed Acyclic Graph (DAG) format is a boxes+arrows flow diagram or network e.g. as used in HACCP. Unlike DAGs created for HACCP analysis by graphic programs (e.g. PowerPoint), links between Nodes must be mathematically defined, logically coherent and possess data integrity.

BN Nodes can represent most discrete concepts and variables that might interest a water engineer or scientist e.g. water quality, treatment options, expert opinions, true/false. Each Node takes on one or more different 'States' e.g. categories, values, value ranges; with probabilities summing to 1.0. BN software can display Node data in the form of probabilities and miniature bar graphs. When a Node represents a quantitative variable, its average and standard deviation is also displayed. BNs exploit Bayes Theorem [Equation 1]; the set theory based rule that '*a priori*' data (A_i /historical probability estimates) can be combined with *a posteriori* (B_j /new) data to improve '*a priori*' estimate accuracy. e.g.

A water manager has data on filter failure from (new evidence) local tests and the manufacturer (old priors). Bayes Theorem describes how to combine the two data sets and obtain a better (new posterior) failure likelihood estimate.

$$P(A_i|B_j) = \frac{P(A_i)\,P(B_j|A_i)}{\sum_i P(A_i)\,P(B_j|A_i)} \quad \text{[Equation 1]}$$



Where $P(A_i|B_j)$ = the conditional (posterior) probability of each *A i*th event given each *B j*th event; $P(A_i)\,P(B_j|A_i)$ = the product of that (prior) *i*th *A* probability and the conditional probability of each *B j*th event given the *A i*th event; $\sum_i P(A_i)\,P(B_j|A_i)$ = the sum of the products of each (prior) *A i*th probability and the conditional probability of each *B j*th event given each *A i*th event.

BNs make possible model conceptualization, definition, probability calculations and exploration in a single platform. Bayes theorem, together with graphical control elements in modern BN software, allows the impacts and implications of changing Nodes or Node states to be rapidly and interactively assessed e.g. exploration of diverse What-If? risk exposure and management scenarios. BNs are somewhat analogous to spreadsheets. They are a general computer modelling platform which can be used for many similar or diverse tasks. Nodes, like spreadsheet cells, can reference one another, and Node states can depend conditionally on other linked cells. BN data may be inputted as single values, be calculated using algorithms, transformed using formulae or imported as large data sets e.g. CSV files.

However BN programming structure sets them apart. Instead of traditional linear computer code and functions, the Node heart involves defining BN relationships (arcs) between independent variables ('parent' Nodes) and dependent variables ('child' Nodes) in the form of probability value matrices known as contingency tables. These tables define how the probability of a 'child' Node taking on a particular state, value or range, depends on the state/value/range of each 'parent'. BN software ensures that Arcs are logically consistent in a manner analogous to relational databases enforcing referential integrity. The probability table based structure allows BNs to be a more flexible and intuitive modelling tools than spreadsheets.

2.1. BN support for an operational recycled water framework

Central to practical application of causality theory, algebra and symbols has been the development of Directed Acyclic Graph (DAG) type BNs. BN software is a class of Graphical User Interface (GUI) computer languages, designed to model networks, variables and relationships between variables, and facilitate causal reasoning. Based on our experience it appears they can provide a common platform for most recycled water validation activities[1],[2] including:

1. Defining and constructing quantitative relationship networks reflecting physical recycled water processes and systems using either machine optimization techniques or expert beliefs of how system components and variables (management options, treatment processes, risk, manufacturer specifications) are inter-related;
2. Defining the relationships between variables (e.g. CCPs) clearly, unambiguously and auditably;
3. 'Learning' of relationships from historical/supporting water treatment databases tables (e.g. .CSV files from manufacturer measurement data from comparable systems), and new validation test data (both qualitative and quantitative) and combining these data with one another and expert opinion into contingency table values;
4. Performing diverse Monte Carlo equivalent QRA style calculations and estimate risk with sufficient precision for decision making;
5. Directly linking ***a prior*** data and new ***evidence*** to generate new ***posterior*** risk estimates for comparison with tolerable risk via Bayesian Inference - we suggest this be termed 'Bayesian Validation';
6. Backcasting i.e. defining set goals such as tolerable risk targets and assessing what treatment performance is required to achieve these.
7. Concisely capturing and communicating beliefs about recycled water system design in a form which can be rapidly audited and modified and used to explore a range of alternative risk scenarios including in a workshop setting;
8. Estimating the predictive accuracy of process and system models which underpin decisions;
9. Exploring different risk exposure scenarios including those associated with hazardous events;

---

[1] Programmatically BNs are more comparable to spreadsheets in that they provide a general tool for systems thinking, design and exploration i.e. an operational risk assessment and management framework for analysis and interpretation of Validation data. Also as with the early development of spreadsheets there are still a range of competing packages. Consider for example MS Excel with rarely-used spreadsheets such as Visicalc, Lotus 123, Quattro and Supercalc.

[2] The proposed central place of BNs in operational validation arises from their ability to undertake calculations involved in defining cause => effect networks and chains and capture the results concisely. Prior to BN development the calculation of tables defining the relationships between variables (nodes) in large networks turned out to be challenging. Part of BN technology development was to find algorithms for rapidly calculating the (Bayesian) probabilities.

10. Generating Bayes model variants supporting conclusions regarding validation;
11. Generating summary statistics for key variable uncertainty and how this changes with changing scenarios;
12. Linking treatment system performance models to wider considerations (downstream exposure risks, influence of external barriers such as access control, management options).

2.1.1. What Bayes appears to offer operationally

1. BNs appear capable of achieving the equivalent of virtually all ISO 31010 risk tools within a single data management and software platform. So for example BNs appear able to both replace/substitute for conventional consequence X likelihood matrices and act as a platform for developing mathematically sound fault tree and event tree analysis.
2. In the case of single treatment processes BNs appeared capable of:
    a. Quantifying pathogen LRVs in detail and sufficiently precisely for validation purposes and providing measures of model precision and credibility;
    b. Quantifying the relationship between pathogen LRVs and operational parameters and physicochemical surrogates and microbial indicator LRVs and their predictive capacity;
    c. Conceptually generating multifactorial 'known unknown' (latent) variables able to capture process groups (e.g. the complex of processes controlling particle aggregation which are indirectly reflected in monitoring parameters such as turbidity and suspended solids) cf. (Hincks et al., 2014, Bollen, 2002, Clark, 2005).
3. In the case of whole treatment trains, BNs appeared capable of generating whole of system (combined treatment process and supplementary buffers) validation and risk estimates comparable to those obtained using conventional QMRA programs (e.g. @Risk) with the additional benefits of:
    a. Readily allowing backcasting (given model and operating conditions identifying and documenting potential/likely causes higher up the treatment train);
    b. Readily allowing modification of model design/exploration of alternative system models;
    c. Allowing system behaviour under modified risk scenarios to be explored in real time making their use in workshop or meeting settings viable;
    d. Easy scoping of hazardous event impacts;
    e. Ready incorporation/exploration of the effects of management options or removal of treatment steps.
4. Initial data mining of input information for possible relationships can be done using complementary software such as WEKA (Witten et al., 2011 , Twardy et al., 2006).
5. 'Learning' the arcs/CPTs of pre-designed BNs is possible by various means e.g. by:
    a. Populating node contingency probability table values simply by importing standard database (record X field) files into the BN software. The software will automatically count the frequencies with which particular CPT combinations occur as percentages;

b. Generating node CPT tables using interpolation methods (similar to curve fitting) routines such as 'Expectation maximisation' and 'Gradient' methods.

6. Separately BN structure may be 'machine learnt' (and hence be relatively free of human bias) in the case of both semi-naïve e.g. TAN, BAN(Carvajal et al., 2015) and causal e.g. CaMML (Korb and Nicholson, 2011) BNs.
7. Diverse assessment of model prediction reliability and precision (Marcot, 2012, Marcot et al., 2006) is possible including:
    a. Receiver Operating Characteristic (ROC) curves;
    b. Estimation of true positive, true negative, false positive and false negative rates (The problems arising from assessing a test or result's value purely from its true positive rate alone are described by Dowie (Dowie, 2006).

This potential does not of course negate the value of other specialised software and approaches. But BNs seem to provide a general quantitative operational tool for first cut risk management of recycled water systems and processes.

2.2. Operational and conceptual BN application

Beyond the operational needs of water managers, Bayes also provides a conceptual/theoretical framework, 'Bayesian inference' (Ellison, 1996). When we validate a water treatment system or process, we collect a range of data sets and combine them to **infer** how the system will operate in the future to reduce risks to a predictable degree provided standard operational conditions are maintained. A system or process is then licensed based on the LRVs we infer it can achieve.

Historically such inference has been done using commonsense and expert opinion with different quantitative data sets used one by one to varying degrees for decision support. The application of Consequence X Likelihood matrices to draw conclusions about risk also illustrates the inference process.

Bayes Theorem offers a great advance on this approach. It provides mathematical rules by which most data and expert opinions collected for, or supporting, water recycling can be combined (e.g. scientific literature, validation testing results). The combined data can then be used to infer and ask questions about system behaviour as a whole based (or large parts thereof) to an extent not possible otherwise. The process is termed 'Bayesian Inference' and lies at the heart of the 'frequentist' v. Bayesian controversy which many statistics users have encountered but are often mystified by. An explanation of Bayesian inference is provided by Ellison (1996) using ecosystems for illustration.

Bayesian inference is computationally challenging. However, BN software make the process simple in much the same way as conventional statistics packages make ANOVA and T-tests elementary. In fact when we use BNs to draw conclusions, explore scenarios, and analysis recycled water system interactions Bayesian inference is precisely what we are doing. This is unsurprising as BN software were developed to aid inferences based on methods for framing causal relationships by researchers such as Judea Pearl and his colleagues (Pearl, 1996, 1999, 2000).

Further, risk assessment itself can be thought of a Bayesian process as evidenced by Fenton and Neil (2012). We use our prior knowledge of risks e.g. pathogen levels in sewage, in combination with new data e.g. proposed complex treatment plant design to infer the likely output levels of pathogens which pose a health hazard. While this may seem nothing special on

initial consideration, if accepted as given, it means that the rich tool kit of Bayes methods and concepts becomes available for validation.

This in turn explains why, from our own assessment of BNs and Bayesian inference it appears together they addressed our original four project aims:

1. develop an overview of validation risk assessment/management methods;
2. identify characteristics of a ‘risk based’ framework to apply when validating water recycling processes that ensure consistency of data collection, statistical evaluation, and performance assessment ;
3. recommend procedures for collecting and incorporating chemical, microbial, or surrogate data into validation activities;
4. outline strategies for the incorporation of hazardous events, suboptimal operational conditions and performance failures within validation.

3. Operational validation using a BN framework

3.1. Recycled water Validation concept

In this Chapter, we propose primary activities required to achieve validation. Other activities are also identified, which may precede or complement a central validation study. These proposals draw upon an understanding of Bayesian Inference and BNs. The proposals are not necessarily new, but their inclusion here is intended to systematise existing best practice based on risk assessment and management, such that they may be framed as Bayesian inference activities employing BNs.

New approaches to 'best practice' can take time to implement, so the following should be noted about the recommendations below:

- Much of what is proposed should already be undertaken guided by Australian Guidelines for Water Recycling (AGWR), HACCP and risk management principles.
- The proposed framework is designed not to replace risk-based management, but reframe it slightly in respect to terminology and how water managers think about validation so as to be able to exploit Bayesian inference related concepts, methods and tools especially BNs e.g.:
    - the implications of Bayesian inference for how to relate contaminant concentration and removal to levels of indicators and surrogates;
    - the use of BNs to facilitate communication between project proponents, regulators and auditors;
    - clear definition of how ***prior*** data on treatment systems and concepts and ***new evidence*** collected for validation relate to one another and generate ***posterior*** estimates of risk, treatment effectiveness etc.;
    - clarifying the role and use of quantitative data in combination with more qualitative information such as expert opinion;
    - scoping the viability of a proposed scheme up front and whether it is likely to be fit for purpose and cost effective;
    - introducing methods for validating our beliefs about recycled water system structure and function generally;
    - clarifying how all available data can be integrated into the validation process.

3.2. Principles for validation using BNs and Bayesian inference

From the outcomes of the work presented in this report, we propose that the following principles should be applied to recycled water system validation. Subsequently, we have worked to demonstrate the application of these principles in the diverse case studies presented in this report.

3.2.1. Recycled water system validation principles

- Validation should be based on/employ Bayes Theorem, Bayesian Inference and the use of BN wherever appropriate.
- The goals of validation experiments, whether descriptive or manipulative should be framed in terms of parameter estimation and hypothesis testing and developed using BN development best practice.
- When we make an assertion about system validation results we do not claim the interpretation with certainty but assert a belief having a (high) degree of probability reflecting the ***prior*** and ***new evidence*** based knowledge available to us. We also recognise these probabilities may be modified in the future by additional knowledge (e.g. during revalidation).
- Bayesian inference provides a framework and mechanism based on probability calculus, to quantify the uncertainty in parameter estimates, and to determine the probability that an explicit validation hypothesis is true given, and "conditional on", a set of ***prior*** data.
- BN design should not be over-complicated to ensure communication is not compromised and beliefs are clear. Whole system BNs need not include detailed sub-units describing individual processes in the same BN.
- The basis for defined node characteristics should be documented. Where causal BNs are constructed the rationale for arcs should also be documented unless they are self-evident e.g. two water treatment processes following one another in sequence as part of the scheme's design.

3.2.2. Individual process conceptualisation and definition

- Individual process definition should use BNs and ***prior*** data to predict ***posterior*** system performance as a first step.
- It is unlikely that commercial validation testing will provide sufficient data for a quantitative BN model to be constructed through learning. However, validation data should be sufficient to be combined with ***priors*** to assess the credibility of the primary recycled water treatment system (model).

3.2.3. Recycled water system structure

- Validation process managers should require recycled water system designers to provide them with full quantitative details of assumptions, design specifications and performance expectations and other information which can be used to develop ***prior*** knowledge of how the recycled water system is likely to perform and operate including:
    - Manufacturer data and claims;
    - Relevant scientific literature which is either representative or comprehensive;
    - Data from comparable systems in use elsewhere;
    - Expert opinion;
    - Designer beliefs about risk levels, LRV reductions, hydraulics, operating parameters.
- Validation designers should provide such information in a format suitable for constructing BN nodes and arcs such as database tables and probability density functions.

3.2.4. Elicitation of BN *prior* data

- Validation teams should design their testing and assessment methodology conceptually based on Bayes Theorem of combining prior data and new evidence to develop posterior estimates of the likelihood that a recycled water system meets its aims.
- Validation teams should develop or adapt BN descriptions provided by designers or develop their own based on the available data.
- Validation teams should employ best practice data elicitation for evaluating any previously constructed BNs or constructing new BNs describing their system and its components.

3.2.5. Design specifications and designer claims

- These should be quantitative as far as possible and include statistics for nodes (=variables).

3.2.6. Expert opinion

- Expert opinion should be captured in a numerical or categorical format suitable for entry into BNs.
- Expert opinion should be supported by appropriate documentation or scientific literature.

3.2.7. Manufacturer data

- Where available, primary manufacturer validation data sets should be obtained.
- Alternatively, manufacturers should provide quantitative predictions of system performance and details of the operating conditions under which they were recorded.
- Ideally they should provide information on how system components interact which can be tested in validation studies.

3.2.8. Benchmarks (risk, treatment LRVs, process specifications)

- The initial BNs constructed should include benchmarks. If not these should be added by the validation team.
- The primary benchmarks should take the form of risk measures and be probabilistic (e.g. illness, infection DALY, Hazard Quotient, cancer risk, toxicity).
- Other benchmarks may include ecological, environmental and monetary ones in the interest of water recycling being ‘cost effective’ and holistic.
- Specialised utility and decision nodes may be included if seen as useful or appropriate.

3.2.9. Revalidation

- At times, process units or processes will require revalidation.
- Revalidation studies should be undertaken in a comparable fashion to initial validation studies with the exception that earlier validation data should now be included as ***priors*** and revalidation data should take on the role of initial validation data.

3.2.10. Contaminants, indicators and surrogates

- There is no conceptual restriction on the use of novel or existing indicators or surrogates in place of primary contaminants provided credible relationships between them can be established applicable to the system under consideration based on refereed scientific literature or high quality technical reports and studies including ones by manufacturers.
- Claims about relationships need to be defined quantitatively and supported with relevant data. This should include operating conditions where the data supporting the relationships was obtained.

3.2.11. Regulators and auditors

- Regulators (and other stakeholders) should become familiar with the concept of Bayes Theorem and Bayesian inference and the basics of BNs, their structure and their operation so there is a common language of communication between all stakeholders. Communication can then be clear and two-way.
- Acceptance or rejection of validation data and findings (***priors, new evidence, posteriors***), on the basis of its quality or inferences should be based on or take the form of acceptance or rejection of Bayesian inferences captured in design and validation BN nodes and arcs which in turn are based on clear evidence, established policy, national or internationally credible guidelines.
- Acceptance or rejection or validation data and findings should be undertaken cooperatively, identifying areas where beliefs differ and identifying how these might be addressed if there is disagreement.
- All beliefs should be captured in the form of BNs which are accessible to all stakeholders.

3.2.12. Subsequent verification and monitoring data

- Verification and monitoring of recycled water systems should also involve the application of Bayesian inference and BNs so that it can be incorporated into revalidation activities and revisions of the initial BNs.
- Verification and monitoring should be framed and justified using Bayesian inference and BNs e.g. Programs for monitoring data collection should identify what is going to be inferred, about how it is believed the system operates, and how hazardous events are to be addressed.

3.2.13. Gathering '*New Evidence'* (the historical concept of validation studies)

Overall system conceptualisation

- BNs are to be used to conceptualise:
    - Stakeholder beliefs about the recycled water system validation process of interest.
    - Individual processes and beliefs about their structure and function.

On the needs and aims of new evidence collection

- The aims of validation data collection should be defined in terms of Bayesian inference and captured by specifying how the primary design BN will be modified.
- The place of new validation data and how it is to be combined with ***prior*** data should be defined.

Validation study analytical and SCADA data

- Validation study data should be collected with a view to it being used in the fashion of ***new evidence*** in Bayesian inference and BNs.
- Validation study data should be used to assess whether a system or process are fit for purpose as follows - either:
    - Modify appropriate BN design settings to obtain new predictions of performance and risk estimates for comparison with Benchmarks.
    - Bayesian Validation – this involves adding new validation data (new findings) and nodes to the primary BN describing ***prior*** system structure function and performance and examining the change to posterior probability i.e. entering findings one by one into a net and following how the posterior probability changes in for example nodes defining risk and the likelihood that the system overall complies with regulatory requirements.

Validation study BNs

- Validation BNs should be constructed for the whole system - This net will most likely take a causal chain format i.e. source to tap though more complex nets are possible e.g. Page et al. (2010).
- Validation of individual processes may include construction of naïve or semi naïve BNs to quantify transformations e.g. LRVs and their basis, as an alternative to causal nets. The LRV summary data can then be used in evaluating revised (***posterior***) system performance.
- Validation BNs should be confirmed as indicating a scheme is satisfactory/fit for purpose before significant construction and analytical resources are expended to validate concept soundness.

Numbers of validation data samples

- For confidently validating a treatment process a minimum of 10 to 15 data points is recommended per process.
- For small systems fewer validation study measurements (5-10 per process) may be sufficient if the ***posterior*** probability indicates there is a high margin of safety (at least 1 $\log_{10}$) so that there is a low likelihood benchmark risks will be exceeded and this is sufficient for risk under normal operating conditions to be tolerable.

Hazardous events

- BNs can be used to model many alternative impact and management scenarios using counterfactual reasoning. Hazardous events are simply and extreme version.

- It is impractical to experimentally validate a scheme's ability to tolerate all possible events. And some event combinations are likely guaranteed to create unacceptable risks.
- To cope with this the following approach is proposed:
    - Scope possible hazardous events of concern.
    - Obtain qualitative and, if possible, quantitative data on their likelihood and consequences;
    - Identify events of most concern.
    - Reformat data on these events in a form suitable for entry as '***new evidence***' in the design of BN. Add new nodes if necessary;
    - Evaluate potential impacts.
    - Consider the impacts of multiple plausible concurrent hazardous events e.g. high rainfall leads to poor pre-treatment and disinfection failure.
    - Ensure contingencies are in place for dealing with simple and complex hazardous events e.g. timely warning methods, pre application water storage.
    - Incorporate these in the design BN.
    - Develop a BN based summary of how hazardous events are to be managed.
- Another practical approach to hazardous event characterization and management is as follows. Hazardous Events can be first identified, characterized and provisionally prioritized using the established risk matrix approach e.g. (Nadebaum et al., 2004). Subsequently priority events can be further characterized using appropriate Risk assessment scenario tools such as fault tree analysis and event tree analysis (e.g. Fenton and Neil, 2012).

Below detection contaminant presence

- Use BNs to evaluate the impact of slightly below detection limit levels in all no- detect samples (worst case).
- Include nodes which add contingency factors.
- Use BNs to evaluate the impact of slightly below detection limit levels.

3.2.14. Miscellany

Cost effectiveness

- Defining system structure and function in BN format should ensure cost effectiveness as far as practicable. For this reason it is desirable that validation start before system construction or development of an experimental program or support for validation tasks be allowed for early on in a scheme's lifetime.

Data management

- Copies of all significant BNs and input data tables and the source of the data should be maintained electronically.
- Hardcopies of BN details should also be kept for future reference.

4. Whole of System Validation: Bolivar water recycling plant

The example presented in this chapter describes simulated validation of SA Water's Bolivar STP, and supply of feed water to a recycled water plant. The primary barriers to microbial risks are the activated sludge plant (ASP), the STP lagoon system, a coagulation + filtration system and chlorination.

To validate the recycling system SA Water quantified removal of *Cryptosporidium* and Adenovirus from the STP Activated Sludge, lagoon system and a coagulation + filtration step. Other considerations employed in the original fitness for purposes assessment were the end use options for which different inactivation credits were assigned (commercial crops, woodlots and municipal watering).

4.1. Employing Best Practice BN development steps

We developed a whole of system BN (Figure 1 - Figure 8) suitable for undertaking a range of validation tasks and which could be adapted to other recycled water circumstances.

Compared to many BNs used as examples by Kragt (2009) our model appeared straightforward to develop. This was largely because water treatment systems are inherently well defined in terms of purpose and structure as illustrated by the results of applying their 5 recommended BN use and development steps.

The results of applying these steps were as follows:

Step 1 Define model objectives system and scale (Kragt, 2009)

Our primary objective was to develop a system to validate treatment effectiveness in terms of risk and implement this. The system boundaries were the treatment plant and downstream recycled water uses and exposed populations of normal individuals.

Step 2 Conceptual model of the system

The basis of the BN was the changes in pathogen concentration between raw sewage entering the STP and the final human risk/exposure calculation (Figure 1).

The main challenge was defining appropriate and informative independent nodes. It would have been possible conceptually to have a simple linear chain of variable as in a standard QMRA, as Netica^{tm} allows complex conditional and probabilistic expression based algorithms to create its contingency tables. But this would have made scenario analysis more difficult so most distinct concepts were assigned their own nodes.

The causal BN structure settled on (Figure 1) was moderately complex and comprised:

- A central catchment to consumer backbone of treatment nodes (Group 1.) and exposure nodes (Group 6.) which calculated the decreasing pathogen concentrations. Each of the post treatment node (Group 1) contingency tables summarized calculations for this node/variable which combined the prior concentration and the primary treatment LRVs (Group 2.) (cf. Figure 2);
- The treatment and other barrier LRV nodes (Group 2.);

- Intermediate design LRVs (Group 3.) and Validation LRVs (Group 7.) which provided alternative inputs for the primary LRVs (Group 2.);
- Environmental barrier credit point value LRVs (Group 4.);
- A range of option selection nodes (Group 5.) whose setting could be used to controlled selection of pathogen, design v validation LRVs, ingestion volumes, exposure frequencies and other settings.

*Step 3 Parameterize the model with (**prior**) data*

Parameterization was undertaken in four different ways in the current model:

- Some node contingency tables were directly populated by manually entering probabilities (sum = 1.0) for each parent state combination (An example of parent state values are the two different pathogens for which there were different possible input concentration ranges);
- Some node tables were calculated using probability density functions (e.g. design Yeast LRV concentrations);
- Some node tables were calculated using simple algorithms which related these (child) nodes to parent nodes e.g. post lagoon concentration involved the subtraction of the lagoon LRV from the post Activated Sludge Process (ASP) treatment concentration;
- Some node tables were left empty so they could be freely varied in scenario analysis between the available states (e.g. option nodes such as pathogen type or the different reuse options which determined multiple environmental barriers).

To emulate Monte Carlo style modelling, parametric 'continuous' nodes commonly had large numbers of bins (i.e. ranges) – the typical number of ranges (states) was from 10 to 20. These nodes are identifiable from their normal distribution-like bin percentage plots. Discretisation (binning into states of ranges) was undertaken using tools in Netica[3].

Alternatively, it would have been possible to learn the validation data node contingency tables (Group 7) from a .csv file or spreadsheet of the original treatment data instead or using that data to estimate LRV probability density functions.

In the model shows though the probability density functions describing the treatment LRVs were entered into node dialogue boxes as algorithms and used to calculate the contingency tables using Netica's 'Equation to Table'[4] wizard.

Construction of 'design LRV' tables used a similar method. Indicative process performance (LRV) ranges presented in the Australian Guidelines for Water Recycling were treated as the boundaries of log transformed 'Uniform' distribution PDFs.

[3] Netica[tm] has tools which can generate discretization ranges in equal value ranges on either an arithmetic or (in this instance) logarithmic scale. Alternatively where discretization reflects actual data each bin size can be defined to cover a given proportions of the total number of record. The number of bins can also be defined.

[4] Other BN software uses different methods.

To reproduce a BN similar to the one we constructed, it would also be necessary for a user to understand:

- Nature node design especially;
    - How nature nodes can be discrete or 'continuous'
    - How categories are treated differently from numbers;
- The idea of states and state titles;
- That every time a new node is added to a net or node discretization is changed new contingency tables need to be recalculated. This involves BN compilation and recalculation for which short cut buttons are available;
- Equation functions and their programming syntax[5] ;
- Frequent saving of a net is desirable to avoid data losses in case of software crashes;
- Once a BN is built it needs to be recompiled before introducing ***new evidence*** and other manipulations.
- New evidence can be introduced in many ways e.g.:
    - It can be 'learnt' from data files.
    - Belief bars can be manipulated manually using the mouse cursor.
    - New findings can be manually entered into contingency tables as probabilities or raw data.
    - A new node might be added.
- Where new evidence is entered the appearance of a node will change if it is a child of this node directly or indirectly.

The best way to learn the above without direct assistance from someone familiar with BN design and concepts would be to undertake the tutorial exercises provided by Netica^{tm}. Further help comes with the Netica^{tm} software in detailed help files.

Step 4 Evaluate the model

For comparison we constructed an analogous model in MS Excel using @Risk v.5.7. This comparison showed that other than the extreme (e.g. 99th) percentiles the model (log scale) statistics and output were the same to 2-3 significant figures. And the BN estimates for average and standard deviation were at worst conservative. This highlighted how BNs can provide an initial easily audited first cut quantitative risk analysis which may be revised if desired.

Another alternative would be for the validation team to do a reality check e.g. evaluate algorithm correctness and undertaken trial checks.

[5] Most similar to Java, C or C++ but not unlike Visual Basic or spreadsheet formulae either. Early in BN construction we found this one of the more difficult tasks to get used to. While Netica^{tm} help was very informative the dialogue box and editor were very basic and using a text editor such as Wordpad is advised.

*Step 5 Scenario analysis (by generating different **posteriors**)*

The primary scenario analysis we undertook was to compare final risk of infection given design vs validation input data for different pathogens, end uses, exposure volumes and exposure frequencies.

In the final model some nodes were added to allow contingency factors to be included and total system LRV to be calculated.

The scenario analysis demonstrated, for example, how given our conservative validation assumptions of 50 exposures per year and 10 mL per exposure:

- For *Cryptosporidium* and water recycled into woodlots:
    - the total water treatment barrier credit averaged 6.08±1.5 logs
    - the environmental barrier credit totalled 7 logs.
    - The median risk of infection was 2.2×10^{-7} per person per year.
- For Adenovirus and water recycled into woodlots:
    - the total water treatment barrier credit averaged 7.47±1.5 logs
    - the environmental barrier credit totalled only 0.5 logs.
    - The median risk of infection was 1.7×10^{-7} per person per year.

The range of statistics which could be generated was very great so discussions between stakeholders are desirable to identify what primary statistics might be identified and what might be generated in a workshop situation to aid decision making.

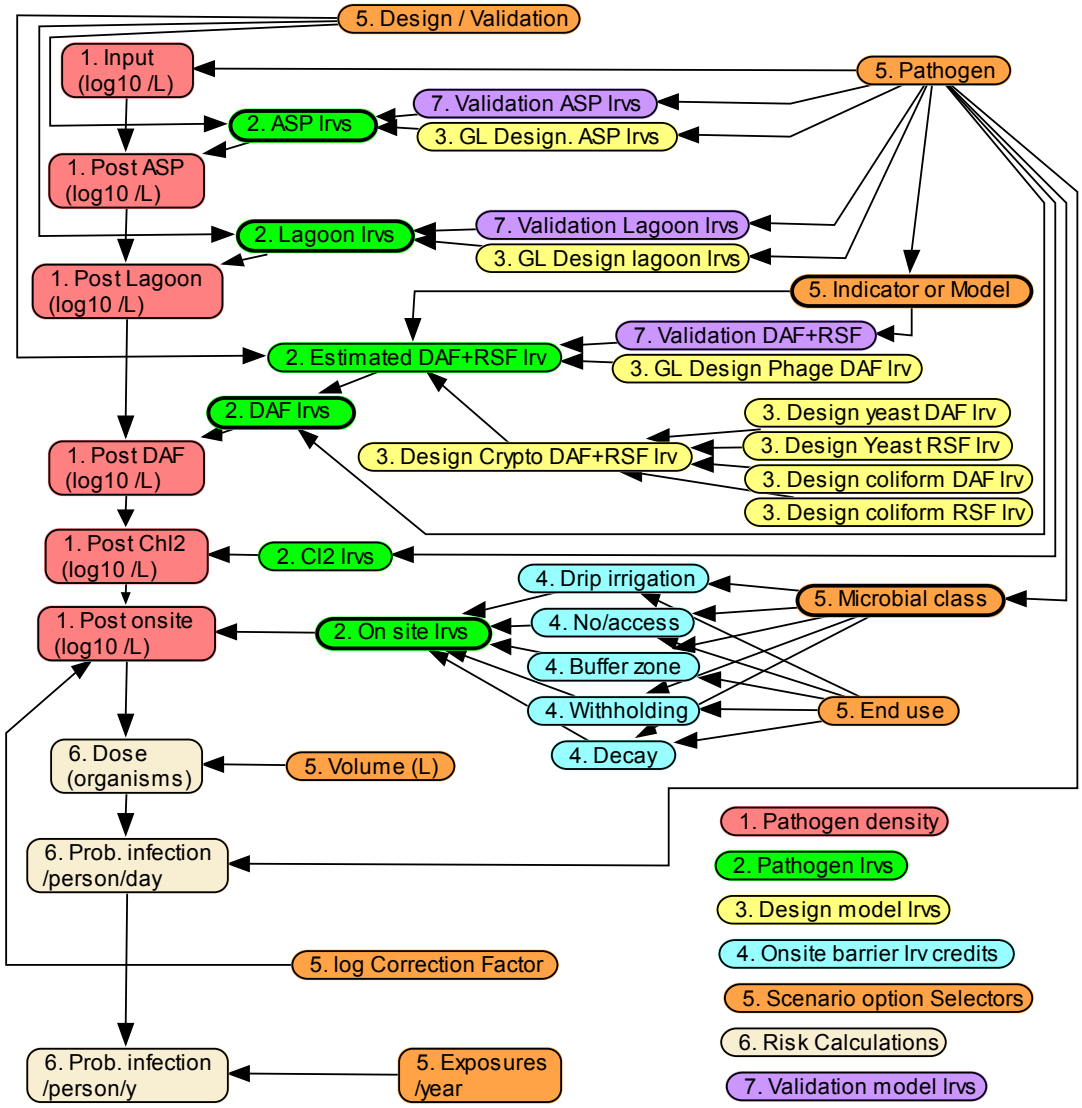
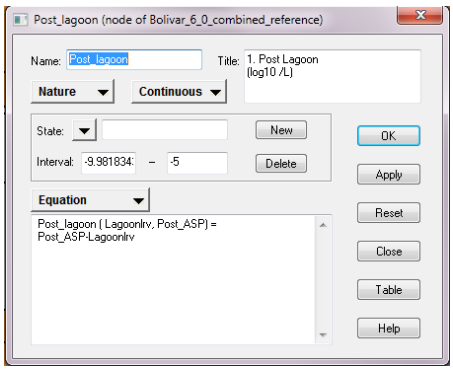


Figure 1. Labelled Box BN of Bolivar Recycled Water System



2. Lagoon Irvs	1. Post ASP (lo...	-9.98183 to -5	-5 to -4	-4 to -3	-3 to -2	-2 to -1	-1 to 0
0 to 0.5	-3.99624 to -1	0	8	34	33	25	0
0 to 0.5	-1 to -0.5	0	0	0	0	51	49
0 to 0.5	-0.5 to 0	0	0	0	0	0	100
0 to 0.5	0 to 0.5	0	0	0	0	0	45.5
0 to 0.5	0.5 to 1	0	0	0	0	0	0
0 to 0.5	1 to 1.5	0	0	0	0	0	0
0 to 0.5	1.5 to 2	0	0	0	0	0	0
0 to 0.5	2 to 2.5	0	0	0	0	0	0
0 to 0.5	2.5 to 3	0	0	0	0	0	0
0 to 0.5	3 to 3.99277	0	0	0	0	0	0
0.5 to 1	-3.99624 to -1	0	26.5	32.5	36.5	4.5	0
0.5 to 1	-1 to -0.5	0	0	0	0	100	0
0.5 to 1	-0.5 to 0	0	0	0	0	50.5	49.5
0.5 to 1	0 to 0.5	0	0	0	0	0	100
0.5 to 1	0.5 to 1	0	0	0	0	0	51
0.5 to 1	1 to 1.5	0	0	0	0	0	0
0.5 to 1	1.5 to 2	0	0	0	0	0	0
0.5 to 1	2 to 2.5	0	0	0	0	0	0

Figure 2. Calculation of post lagoon concentrations a. algorithm used for calculation b. contingency table section.

4.2. Whole of system BN features

The system summarised in Figure 1 performs the role of a HACCP diagram in that it identifies barriers where controls can be placed. But the model can do much more by virtue of being able to view the information contained in increasing complexity, and alter scenarios to explore what can be inferred given those scenarios, especially:

- The primary BN model structure (Figure 1)
- The algorithms and node interrelationships (Figure 2);
- The expanded belief bar format (Figure 3) which details each node graphically and in summary statistics (Figure 4) for any given valid set of ***prior*** beliefs and ***new evidence***; and which can be made more intelligible by selective expansion of nodes (Figure 5);
- Analyses of the BN as a whole, and for individual nodes, via various tools and wizards in the software e.g. 'Sensitivity to findings[6]' (of a selected node) (Figure 6).

Accordingly, BNs capture most of the central interests of validation stakeholders in a format which is intuitive, concise and standardised. The detail can be readily be drilled into, in real time as far or as little as the user desires within the one software platform. To exploit their potential, users need to understand, what is in effect, a language of "Cause=>Effect" and inference. Some important features which are illustrated by this specific BN are:

1. Nodes are essentially variables which capture data already known. So Figure 3, Figure 4 and Figure 5 show what is already known i.e. current ***priors*** data (beliefs on contaminant concentrations and treatment data).
2. Some node tables can be left blank to make scenario variation easy. In Figure 5, Group 5 performs this role.
3. '***New evidence'*** in the present instance is entered by altering a node probability setting. So for looking at 'Design' scenarios Design probability is adjusted to 100% from the initial state of "50/50" where no 'belief' has been entered. So where a BN has been set up, validation simply involves flipping the Design/Validation node setting for a desired scenario.
4. LRV choices can be as simple as a single value reduction credit (e.g. on site barrier LRV credits which may be selected individually or collectively (end-use node)) or be the outcome of a complex choice e.g. the *Cryptosporidium* DAF/filter reductions.
5. All nodes will change to reflect new/specific scenario settings once the choice settings are entered (such as 100% probability that the pathogen is *Cryptosporidium)*. The probability values at this point describe the new ***posterior*** state of variables e.g. the actual probability of infection for that scenario.
6. In the present instance 'microbial class' and 'indicator or model' node selectors are unexpanded because their choices were automatically determined based on which pathogen was selected.

[6] Approximately analogous to correlations.

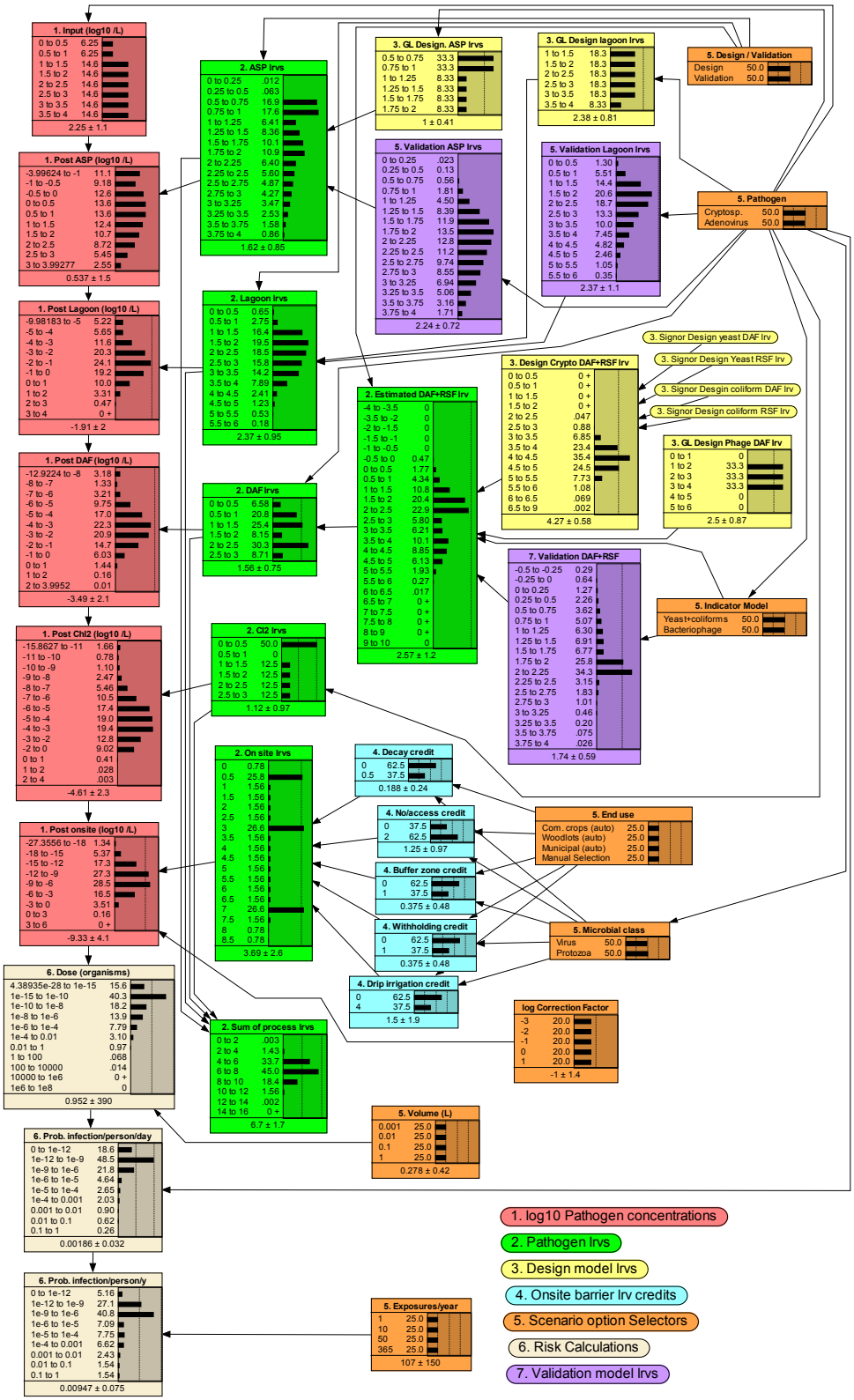


Figure 3. Fully expanded BN for Bolivar recycled wastewater plant

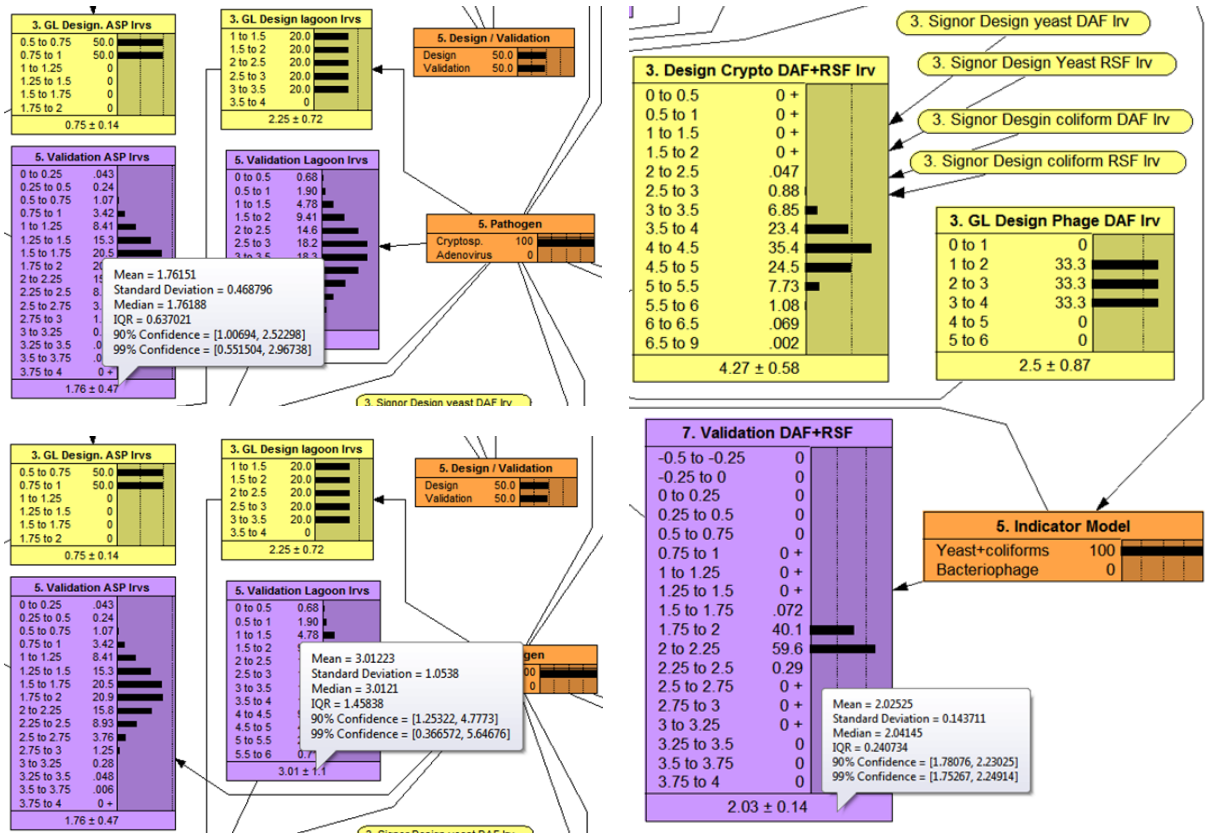


Figure 4. Main LRV processes validated at Bolivar showing the summary statistics that Netica[tm] displays for each node.

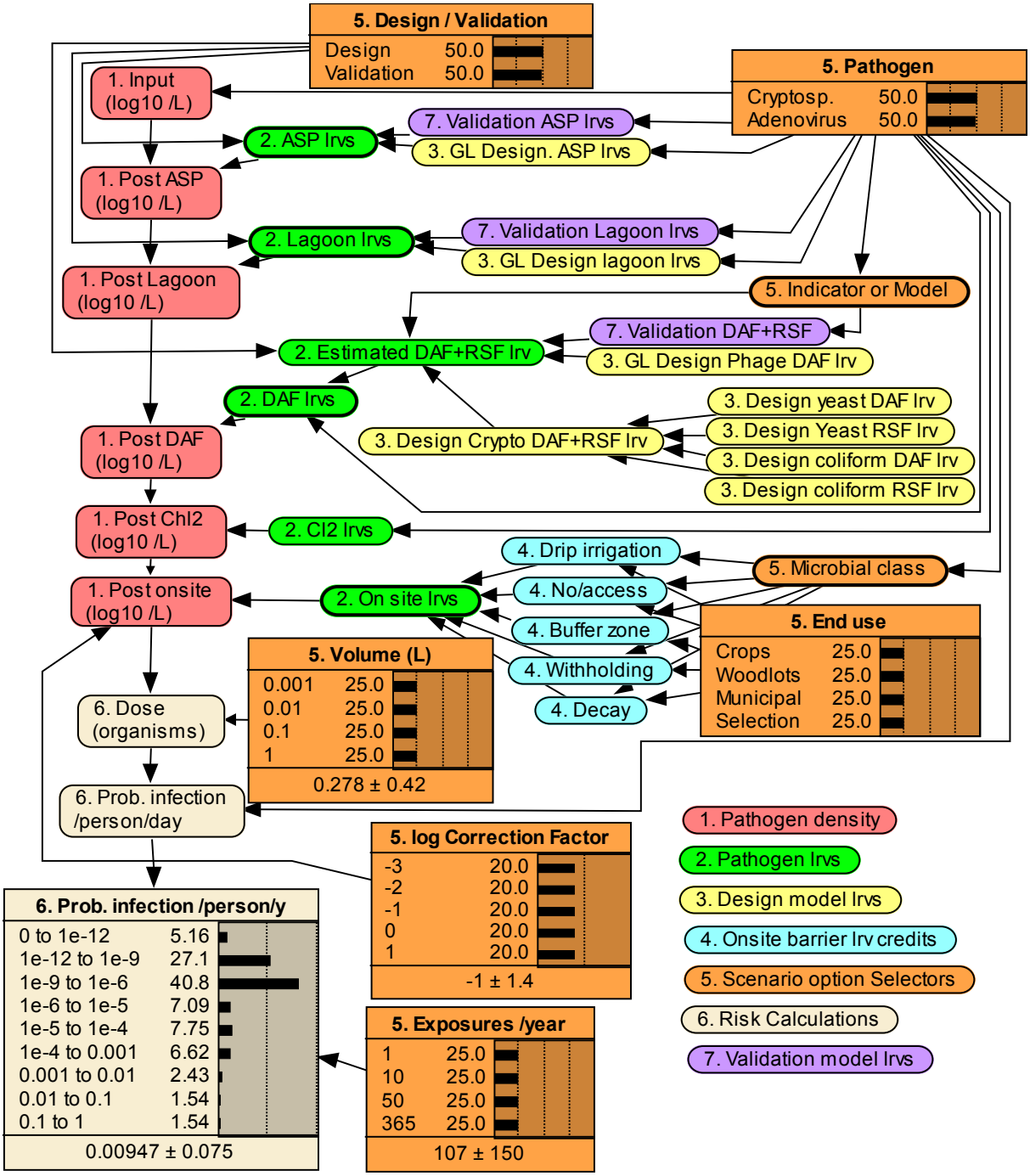


Figure 5. Partially Expanded Bolivar Recycled Water System BN Model

Sensitivity of '6. Dose (organisms)' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
6. Dose (organisms)	532.9	100	2.33356	100	0.6045466
1. Post onsite (log10 /L	269.7	50.6	1.49024	63.9	0.1834335
1. Post Chl2 (log10 /L)	73.46	13.8	1.20939	51.8	0.1287903
1. Post DAF (log10 /L)	66.67	12.5	1.02583	44	0.0862112
1. Post Lagoon (log10 /L	21.52	4.04	0.84422	36.2	0.0581024
6. Prob. infection/perso	15.82	2.97	1.87938	80.5	0.3721685
6. Prob. infection/perso	0.992	0.186	1.01561	43.5	0.0775354

Figure 6. Extract of 'Sensitivity to Analysis' of Dose node

4.3. Understanding how the BN works

Validation in practice is illustrated in Figure 7a and b. In Figure 7a design validation is shown for *Cryptosporidium* assuming exposure volume is 10 mL, the end-use is "woodlots" and 10 exposures per year. In practice an assessor or auditor might want to check further assumptions. LRVs from the guidelines are not shown but these can be viewed using the expanded BN (Figure 3).

The median probability of infection for the scenario settings shown can be seen to lie in the range 10^{-6}-10^{-9}/y/person. This is a ***posterior probability*** which has been inferred from a combination of (guideline) design treatment performance estimates, the assumption of environmental barriers being very effective, and other givens such as the dose response algorithm which is located in Node "6. Prob infection / person/day/"

Figure 7b shows another different revised infection posterior probability given the selection of Validation = 100% which selects the LRVs in group 7. There is a slight increase in infection risk.

In both scenarios the ***posterior*** probability of infection for *Cryptosporidium* is <1% that infection probability will exceed 10^{-6} per year. So if the benchmark is 10^{-4} per year at the 95th percentile, the system has been validated as fit for purpose.

If a regulator is not convinced by the assumptions, beliefs or other priors, a BN allows all these to be easily altered and results automatically recalculated.

The same result could obtained by using spreadsheet based QMRA modelling using Palisade @Risk but summarising and presenting the prior assumptions new evidence and posterior findings would be far more difficult.

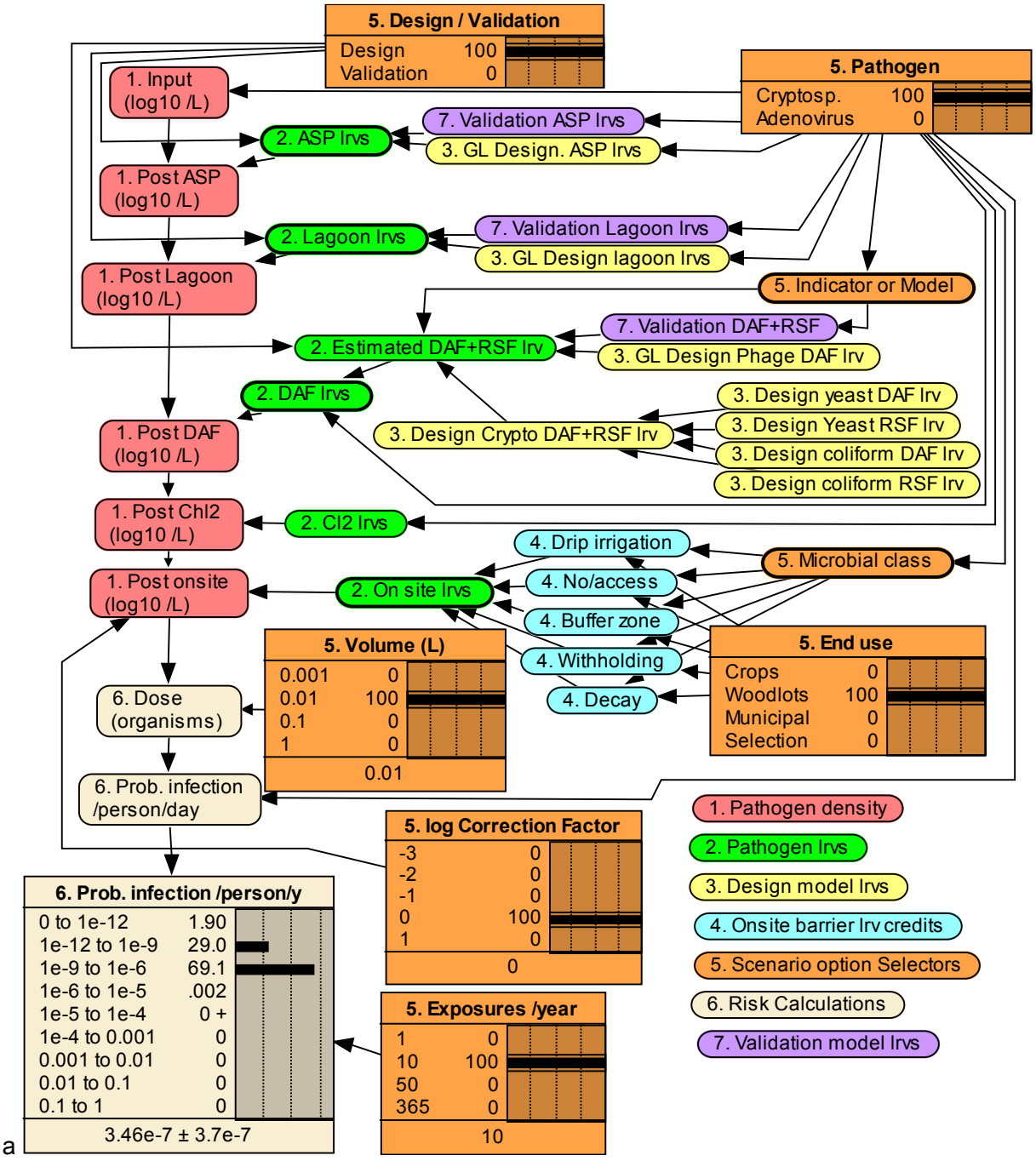
Consideration of potential hazardous event scenarios can also be achieved within the BN. Figure 8 shows a 'Hazardous Event' counterfactual scenario with the following changes:

- Adenovirus is the contaminant of concern;
- Crop irrigation is the end use.
- The average consumption volume has been reduced to 5 mL, the amount of water that might be retained by a lettuce.
- Exposures per year have been increased to an average of 30 per year.

It can be seen that yearly risk probability now exceeds the 10^{-4} per person per year with probability estimated to be 14%. Examination of the detailed BN by selective expanding nodes to their belief bar format will show this result which reflects mostly the reduced protection

associated with cropping as an end use barrier. These are only some of the ***posteriors*** that can be estimated the systematisation of inference which BNs allow.

A feature of BNs which Figure 8 illustrates is that '***new evidence'*** scenarios are not confined to single states or ranges. So for in the case of 'exposures per year' and 'volume of use', the effect of two alternative states occurring during the same year (e.g.10 and 50 exposures per year = average 30 per year) is shown.



a

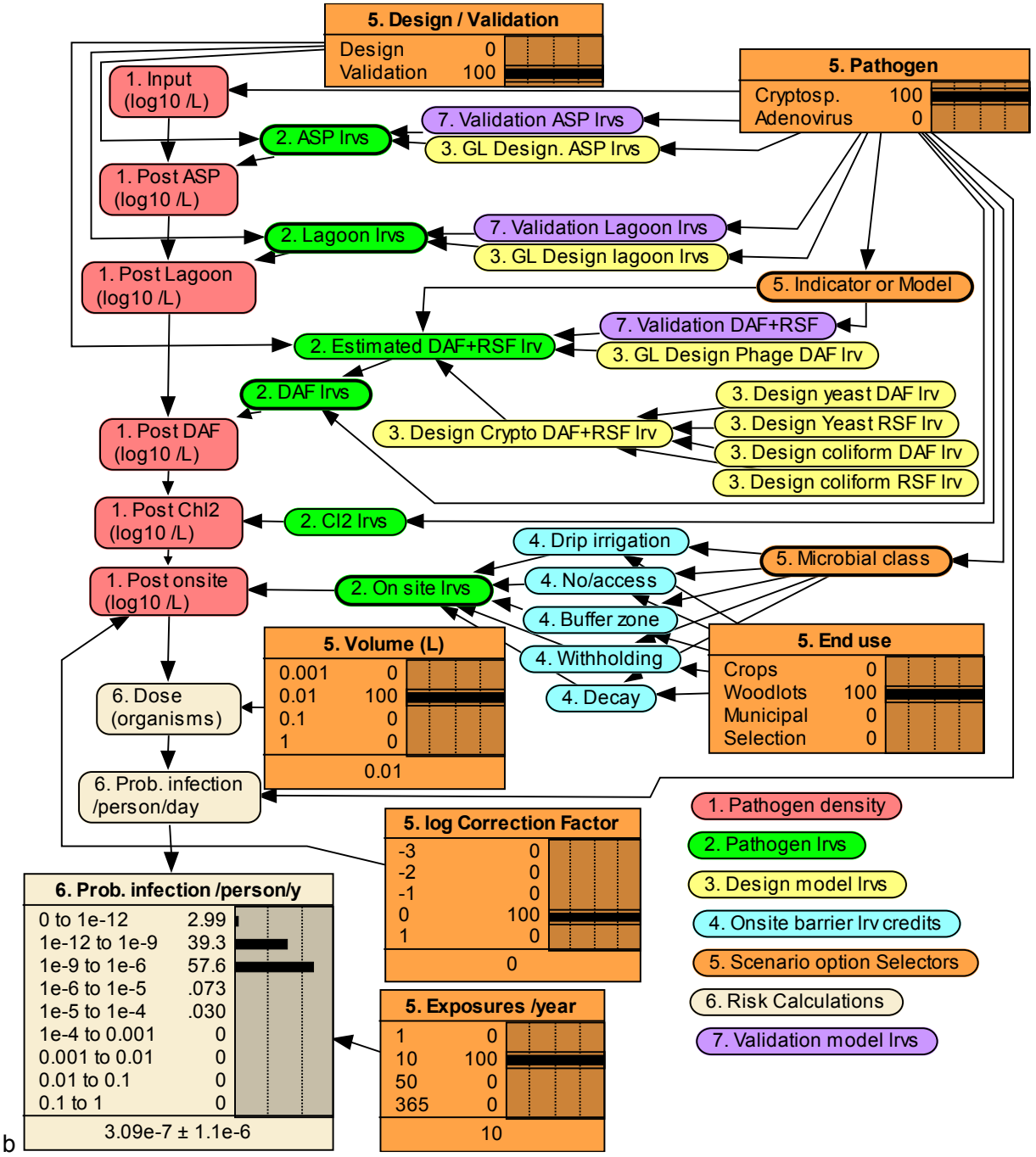


Figure 7. Proposed Bolivar design specifications a. compared with b. Validation **study based predictions**

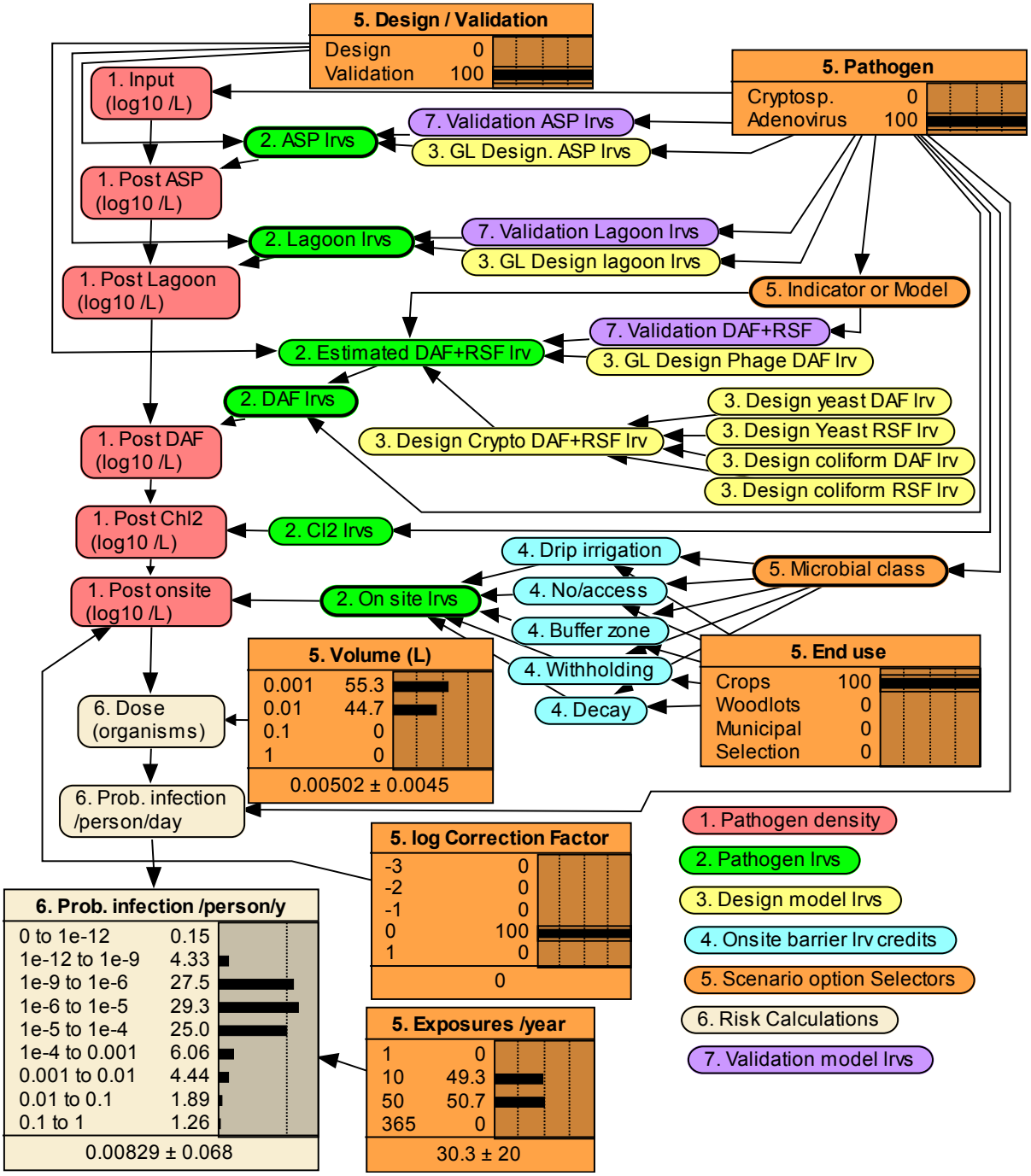


Figure 8. Hazardous Bolivar operation scenario

4.4. Development of monitoring recommendations

One of the very useful by-products of our exploration of semi-naïve (Carvajal et al., 2015) and causal BNs is that it showed how indicators:

- could be identified, evaluated and compared for their accuracy and predictive power;

- less expensive indicators and surrogates could be validated

Specifically this case study found that for the activated sludge process, two microbial indicators (total coliform and enterococci), and three water quality parameters (turbidity, SS and ammonia) could be used to predict the LRVs for *Cryptosporidium*. The prediction of the removal efficiency through water quality or operational parameters will depend on the process in study and the reliability of the data. The same procedure could be applied to determine the predictability of any pathogen or microbial indicator LRV.

An alternative which we have yet to fully explore is validating turbidity as a measure of bacterial removal. This said Figure 9 from the causal network developed also using the Smartwater project data (Flapper et al., 2012) and introduced above shows the influence of Particle Density (State Values 0 = low turbidity and SS) and somewhat reduced organic matter processing (State value 1) on bacterial populations.

Figure 9 shows how there was a very high likelihood of a high Cryptosporidium LRV when SS and Turbidity (combined in the form of a Latent node named Particle Density) and bacterial indicator densities especially *E. coli* and enterococci, were in their lower probability bins.

Thus it appears plausible that routine microbial monitoring might be periodically foregone in place of undertaking other useful monitoring.

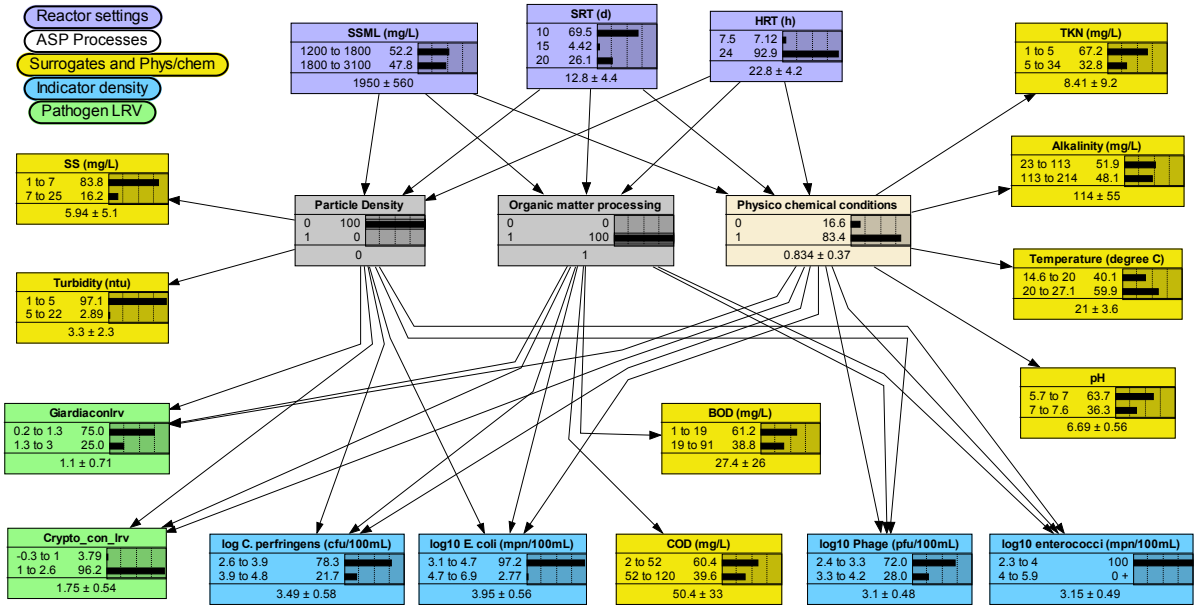


Figure 9. Influence of Particle Density and Organic matter processing on bacterial reduction

4.5. Bayesian Validation

A particular recycled water system validation challenge BNs can address well is that of integrating multiple disparate information sources provided they can be framed quantitatively. The Queensland Guidelines - section 3.8.7 (Queensland Water Supply Regulator et al., 2013) propose consideration of all sources of information during the validation process but the integration process advanced is a relatively crude semi-quantitative one. Bayes provides a more

quantitative approach which we suggest be call 'Bayesian Validation' and involves the incorporation and integration of one or multiple priors.

Several approaches are available. Below we outline the 'enter findings' method for validation of a treatment train with the incorporation of priors. This approach is illustrated using the Bolivar water recycling system and a hypothetical UF system as models. The concentration reduction BN backbone is at the top of each Figure in yellow. Subsequently we found that more interactive and detailed validation was possible. A second alternative distribution integration method is described below. Finally we found that it was possible to do such integration interactively. The method is illustrated in the Melbourne Water Eastern Treatment Plant and particularly SA Water Glenelg case studies.

4.5.1. 'Enter Findings' method

The first approach illustrated in Figure 10 and Figure 11, assumes a ***prior*** parametric distribution for the LRVs. ***Priors*** in the first approach are specified for the parameters of the distribution e.g. mean and standard deviation for normal distribution in the example shown. These are shown in blue.

The nodes for the ***new evidence*** validation test data to be entered are shown in green. One datum is entered at a time which accounts for the large number of validation nodes.

The data entry process is illustrated in using the activated sludge process as the example. The first expanded validation entry node has initially no data but in the example b. an LRV estimate of 2.25 to 2.5 has been estimated in the usual way. The BN node colour changes to orange to show this change. Netica[tm] also includes an 'enter finding' (essentially enter new evidence) facility. So it is possible to add similar values into nodes formatted in the more concise labelled box style too. These settings may be removed at any time.

The overall result of this analysis is a group of distributions for each subset of parameters, which gives a higher estimated LRV variability. So seven validation data points have been added to generate validation ***posterior*** estimates of the mean and standard deviation. The revise BN in turn has been used to assess whether there has been an improvement on the original design expectation.

4.5.2. Distribution integration method

The second method (Figure 12) uses a nonparametric approach. It incorporates the ***prior*** data directly into the distribution or interest. This requires assigning relative weightings to the earlier ***prior*** and the ***new evidence*** validation data set. The outcome in this case is a single distribution. This process is automated by software such as Netica[tm]. In this methodology the BN "learns" the probabilities directly from the validation dataset so there are no single samples explicitly in the network as nodes. This approach makes use of other functionalities of BNs such as learning and 'fading'.

The way this is done in Netica[tm] is as follows:

- The validation test data is saved in a .CSV file or equivalent
- The distribution node which may be generated in any way e.g. the UF LRV node generated using an algorithm, is selected.
- Netica[tm]'s 'Process Case' option is selected from the Cases (learning from data) menu.

- Then for each case in the file, Netica[tm] reads the case and enters it as findings into the BN. Finally Netica[tm] does belief updating to find probabilities for all the nodes that didn't have new findings.
- Operationally what happens during this process is that Netica[tm] provides an option allowing the user to say what proportion of the new distribution will come from the old distribution data and how much from the new validation evidence. The revised combined posterior distribution can then be assessed for example against a benchmark.

This process may sound complicated but in practice it takes a few seconds to implement.

4.5.3. Interactive Distribution integration

The final method builds on the previous one. It includes nodes which allow the user to interactively change the weighting given to the different priors and new evidence using another belief bar node. The method is most clearly illustrated in the Glenelg case study.

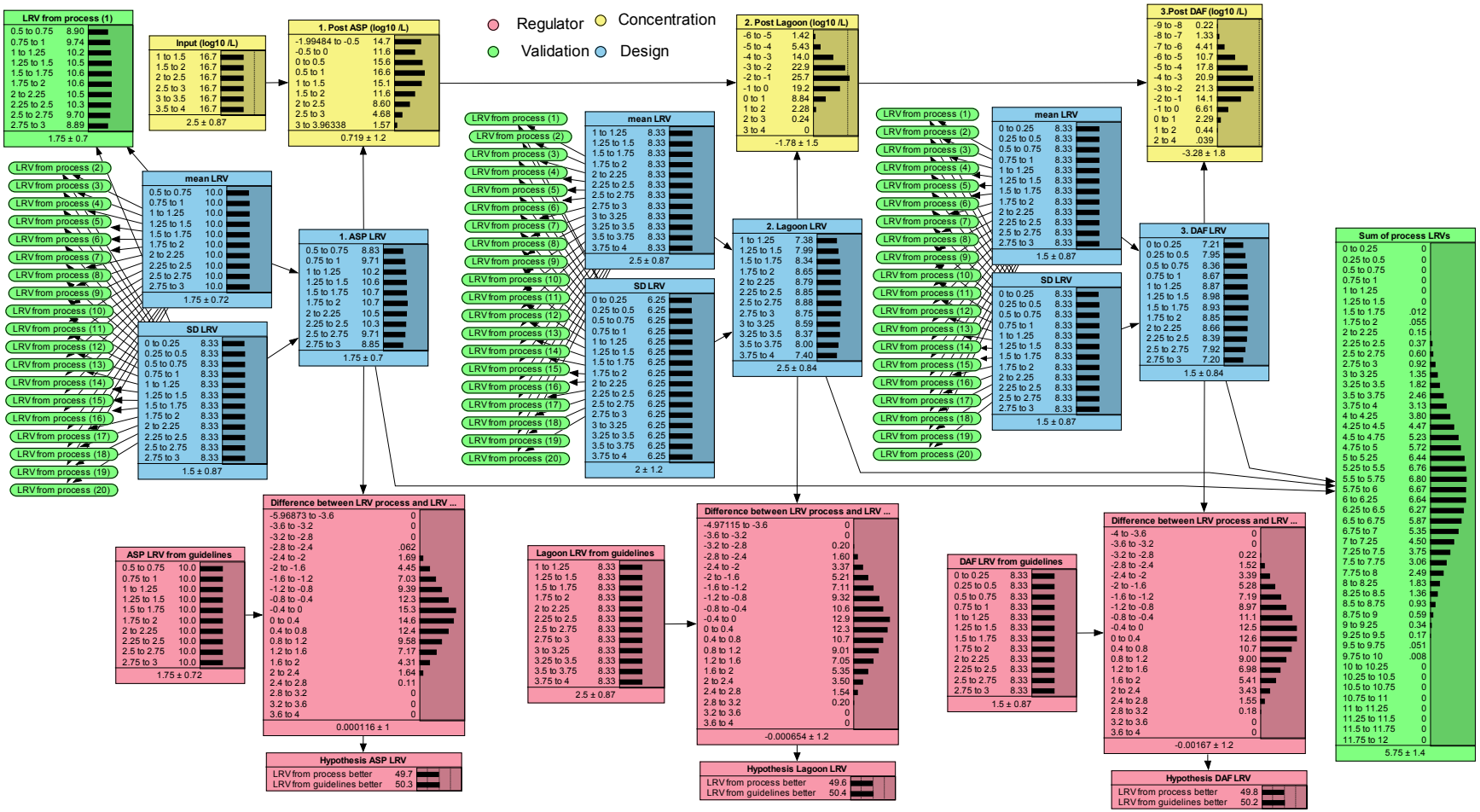


Figure 10. Bayesian Validation **of Bolivar STP treatment processes approach 1 "Enter Findings" method**

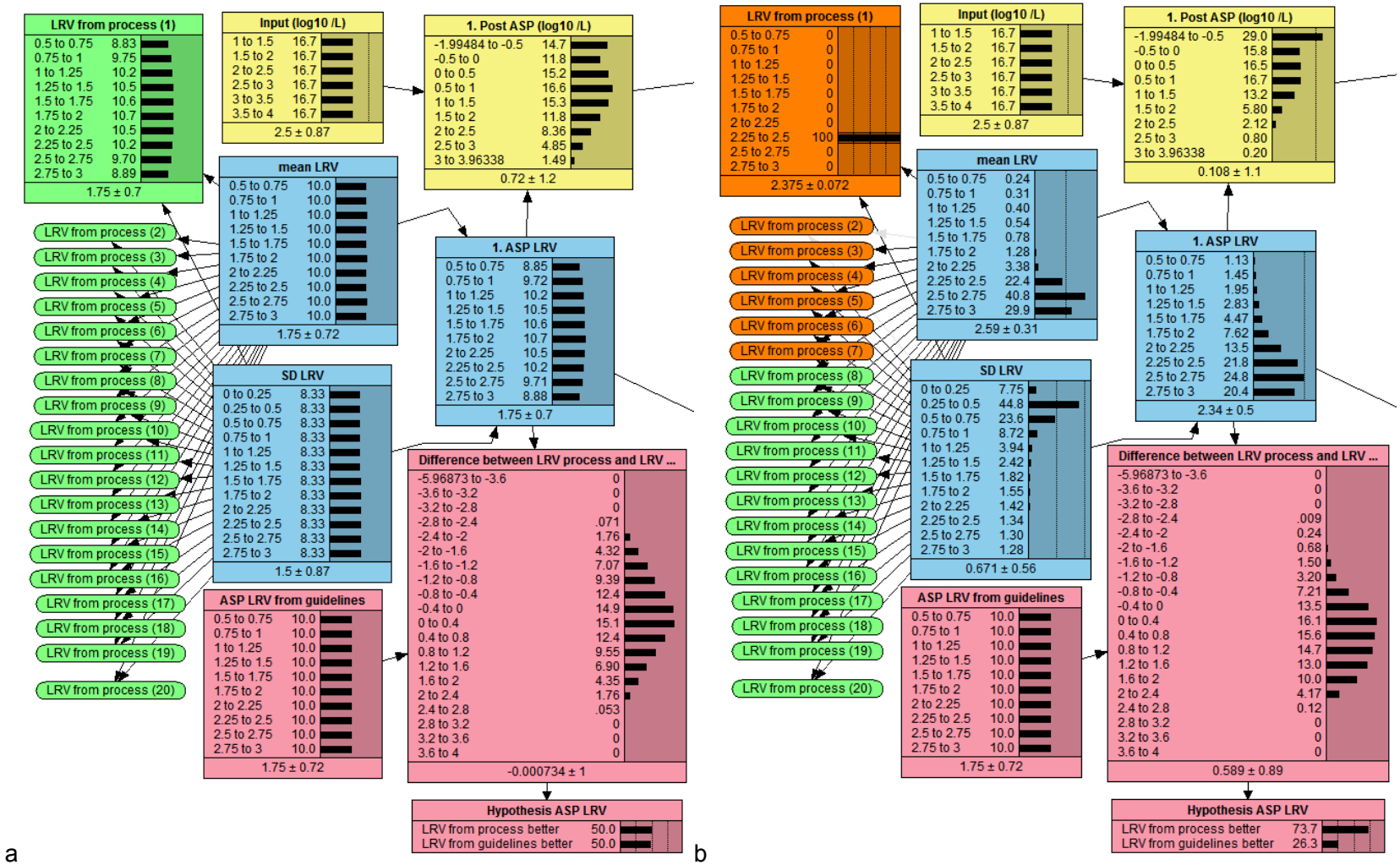


Figure 11. Illustration of a. before and b. after applying "Enter Findings method to Bolivar STP system

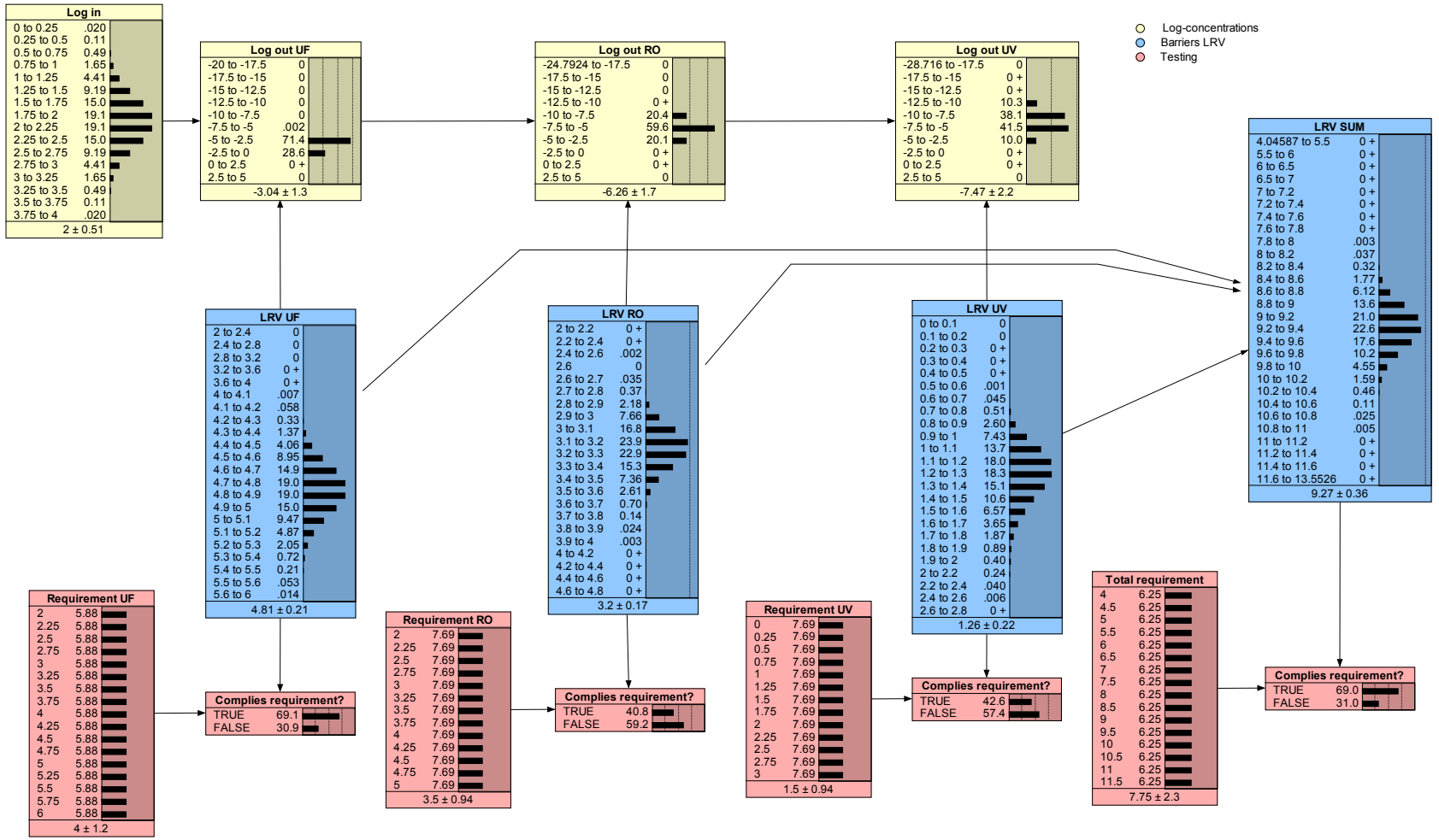


Figure 12. Bayesian Validation of UF STP treatment processes approach 2 “Distribution Integration” method

5. Process validation with large data sets: Naïve & semi-naïve v. Causal BNs

To explore how BNs could be used to estimate LRVs for input into whole of system models where there are large data sets, such as result of substantial research projects, we explored two avenues, the development of intuitive causal nets similar to those described in the previous section and an alternative – naïve and semi-naïve BNs (NBs and SNBs respectively).

In this section we outline firstly the second of these options as this appeared to provide the simpler way for estimating and validating LRVs in an unbiased fashion.

5.1. Naïve and semi-naïve BNs

Naïve Bayes models (NB) are non-causal BN models commonly used for classification problems (Kjræulff and Madsen 2008). They often provide good accuracy, while offering simplicity and efficiency. Their construction employs a range of objective rules and tests, which address modelling traps with causal BNs including the use of inappropriate variables, modeller bias and over-fitting. By definition, the structure of a NB model always employs a "class node" which is the only parent of each other node (attribute nodes), all of which are conditionally independent given the class node. LRVs are a 'logical class' node.

In the case of the related SNBs, the independence assumption is relaxed by allowing some arcs between the attribute nodes as well as the class node using link selection rules not necessarily involving a choice by the investigator. Examples of SNBs include Tree Augmented Naïve (TAN) Bayes models which Netica^{tm} is able to construct in which the nodes depend on the class node and at most one other node (Korb and Nicholson 2011), and BN Augmented Naïve Bayes (BAN) models, where two or more arcs are allowed between nodes additional to the class node (Cheng and Greiner 1999).

Work on this involved analysis of the real world Activated Sludge pathogen reduction data set developed and published by Flapper et al. (2012). The latter looked at factors controlling *Cryptosporidium* and *Giardia* reduction. Approximately 98 measurements were taken under varying operating conditions and a range of water quality measures and microbial indicators were concurrently measured.

The results of this have been used to develop a full paper, which has now been published in *Water Research*. The title and abstract are presented below. In addition we have reproduced here:

- Figure 13 which summarizes the procedure for developing naïve and semi Naïve nets for LRVs;
- Figure 14 which illustrates the naïve BNs developed[7]; and
- Figure 15 which illustrates the optimum semi-naïve BNs.

[7] One proposed use for NBs in medical diagnosis. Experience has shown that NBs can usefully link diseases to symptoms even though this may seem causally counter intuitive. The problem is framed as disease => symptoms and signs. But in the case of poisoning for example clearly the true causal link is toxin => disease. This illustrates how defining the causal order in Bayes Nets can be slippery.

The conclusions we drew from this work included the following. Naïve and semi-naive Bayes models can be used to predict and manage pathogen reductions. The methodology we developed is objective, systematic and applicable to analysing water treatment processes more generally. Though the study identified operational parameters potentially useful for the prediction of *C. parvum* removal efficiency, modelling *G. lamblia* suggested that its removal by AS is not sufficiently understood and cannot yet be quantified based on removal of microbial indicators, even though assignment of average reduction credits of ≥1 $\log_{10}$ is still reasonable judging by the raw LRV probability density function.

Our non-causal models also provided a reference and starting point for BN modelling by identifying those variables most likely to be useful when constructing causal models with the minimum of nodes. The SNB models provide an objective way of estimating the maximum accuracy that is possible with a causal Bayes model. The models were relatively easy to understand which should assist uptake by non-experts in BNs interested in and other non-standard treatment approaches. Finally, the method here can reduce disagreements between model developers about what form BNs should take.

Abstract

Modelling pathogen $\log_{10}$ reduction values achieved by activated sludge treatment using naïve and semi naïve Bayes network models

Authors: Guido Carvajal, David J. Roser, Scott A. Sisson, Alexandra Keegan, Stuart J. Khan

Abstract: Risk management for wastewater treatment and reuse have led to growing interest in understanding and optimising pathogen reduction during biological treatment processes. However, modelling pathogen reduction is often limited by poor characterization of the relationships between variables and incomplete knowledge of removal mechanisms.

The aim of this paper was to assess the applicability of Bayesian belief network models to represent associations between pathogen reduction, and operating conditions and monitoring parameters and predict AS performance. Naïve Bayes and semi-naïve Bayes networks were constructed from an activated sludge dataset including operating and monitoring parameters, and removal efficiencies for two pathogens (native *Giardia lamblia* and seeded *Cryptosporidium parvum*) and five native microbial indicators (F-RNA bacteriophage, *Clostridium perfringens*, *Escherichia coli*, coliforms and enterococci).

First we defined the BN structures for the two pathogen $\log_{10}$ reduction values (LRVs) class nodes discretised into two states (<and ≥ 1 LRV) using two different learning algorithms. Nine metrics, such as Prediction Accuracy (PA) and Area Under the receiver operating Curve (AUC), provided a comparison of model prediction performance, certainty and goodness of fit. This comparison was used to select the optimum models.

The optimum Tree Augmented naïve models predicted removal efficiency with high AUC when all system parameters were used simultaneously (AUCs for *C. parvum* and *G. lamblia* LRVs of 0.95 and 0.87 respectively). However, metrics for individual system parameters showed only the *C. parvum* model was reliable. By contrast individual parameters for *G. lamblia* LRV prediction typically obtained low AUC scores (AUC <0.81). Useful predictors for *C. parvum* LRV included solids retention time, turbidity and total coliform LRV. The methodology developed appears applicable for predicting pathogen removal efficiency in water treatment systems generally.

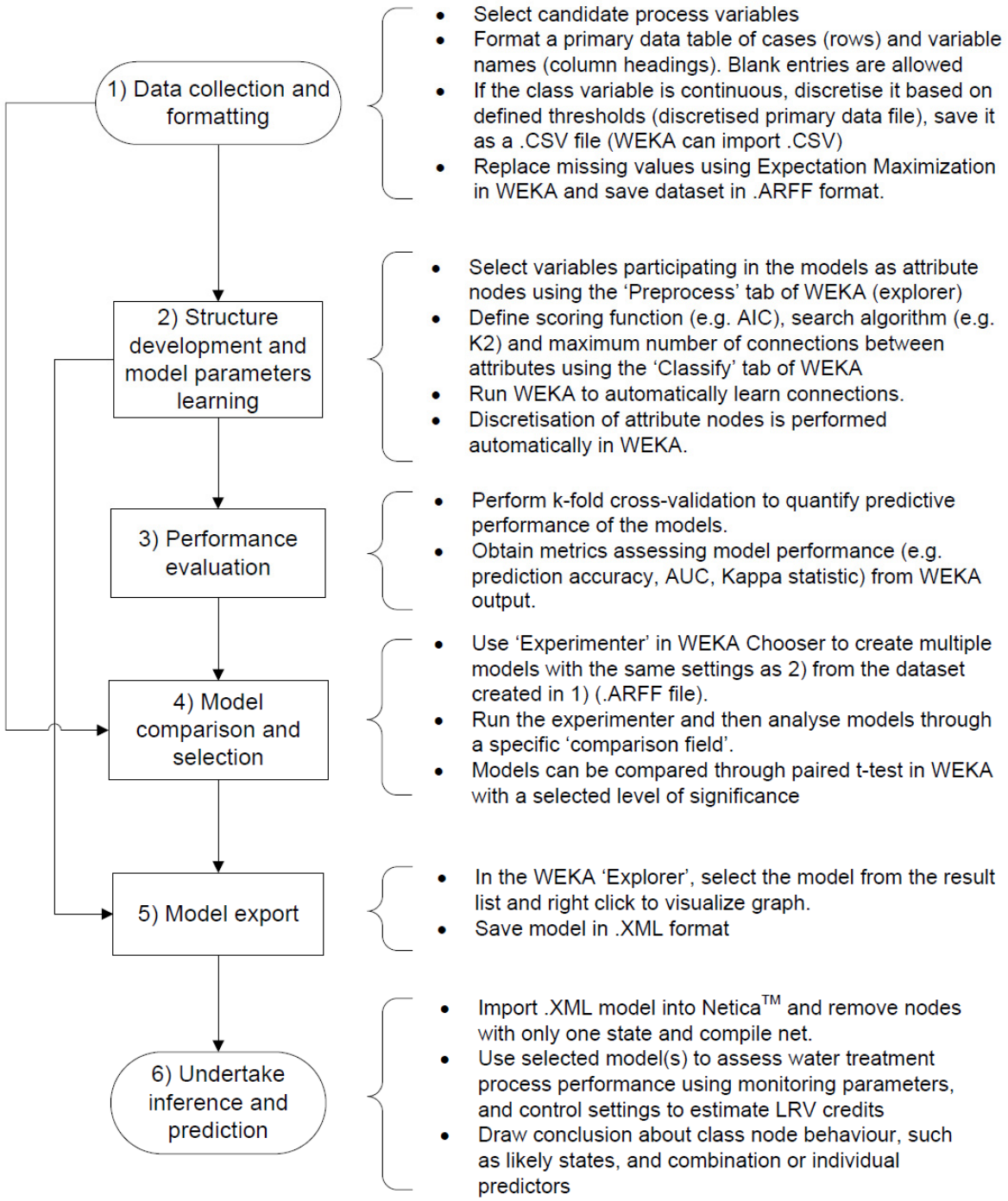


Figure 13. Procedure for developing a semi-Naïve BN

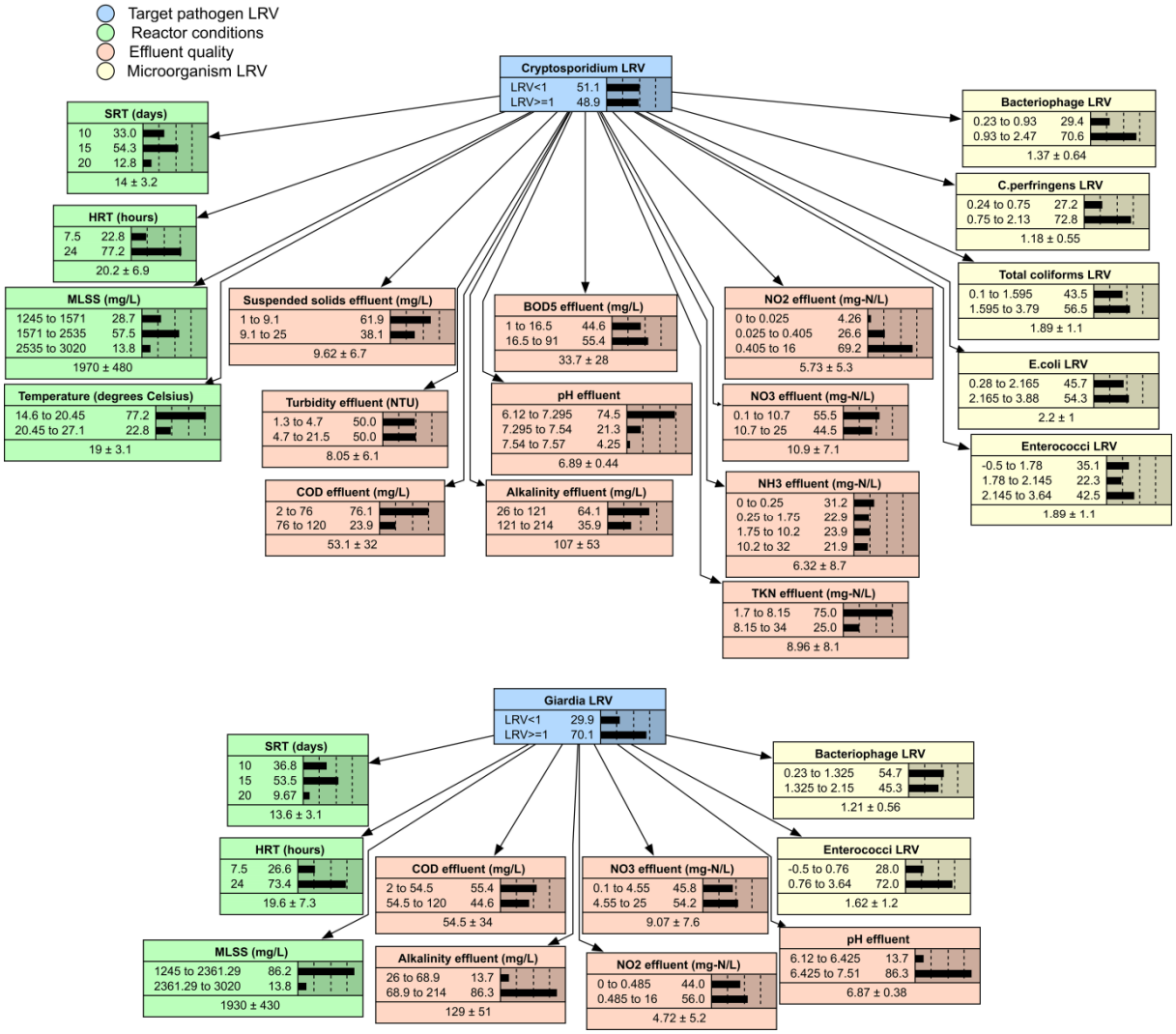


Figure 14. Naïve Bayes models for (a) *C. parvum* LRV and (b) *G. lamblia* LRV showing discretization ranges.

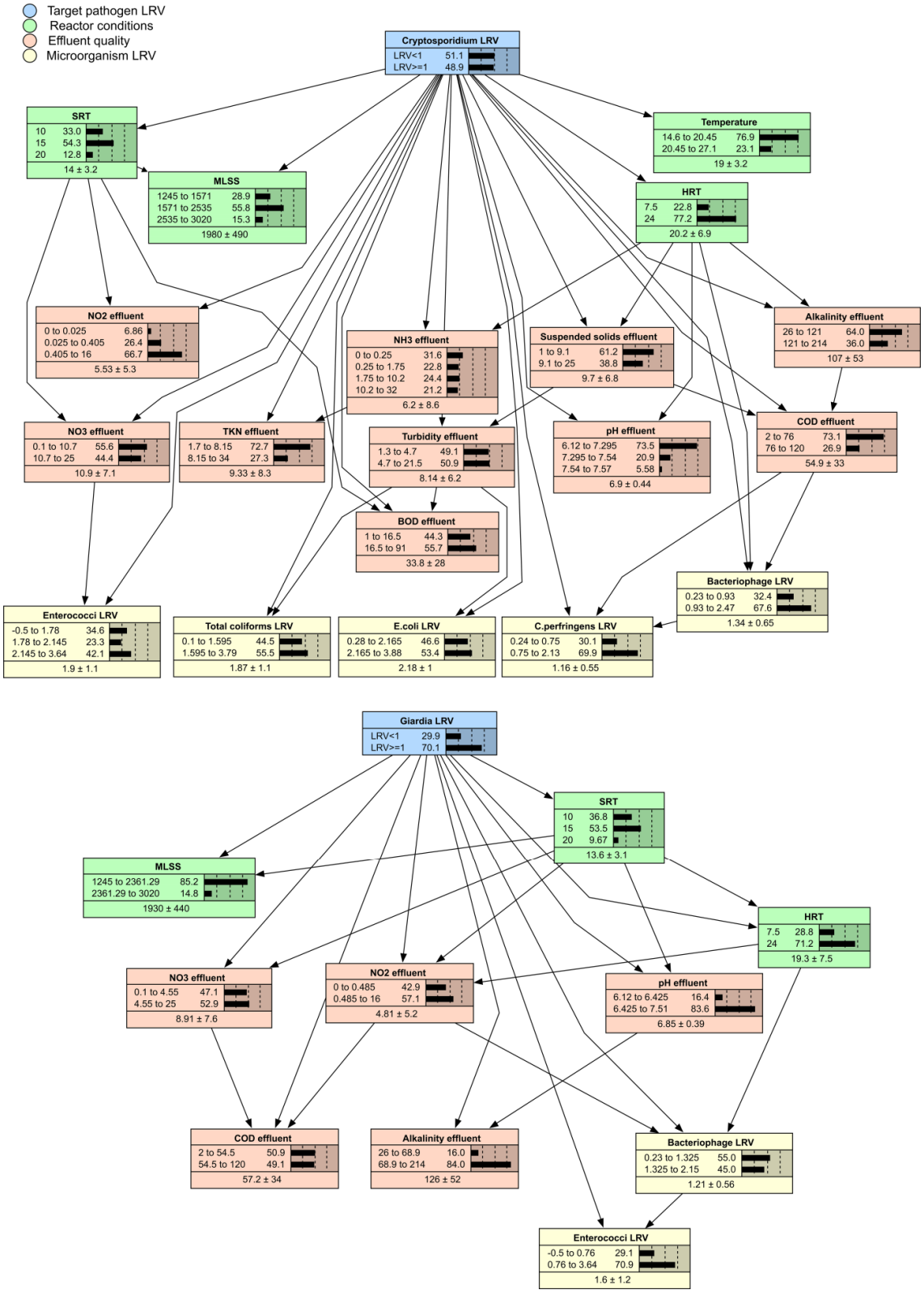


Figure 15. Optimum semi-Naïve BN for C. parvum LRV (BAN) and G. lamblia (BAN).

5.2. Causal process BNs

Prior to exploring the semi/naïve approach we also trialled the development of causal nets for quantifying LRVs. It was immediately evident that depending on our logical inferences a diversity of plausible BNs was possible to generate. Furthermore, the number of states possible for any given node was also extensive and varied.

Figure 16 illustrates one of the final causal networks resulting, which we use here to illustrate a range of issues arising when using causal nets.

Reassuringly for BN validation though the results of different 'common sense' conceptualizations all pointed to similar findings to one another and to those from the semi/naïve Bayes analyses described in the previous section.

This situation contrasts with the whole of system analysis. In the latter instance it was straightforward on first principles e.g. physical design of Bolivar, to assign causal arc directions. In the AS process considered here though the causal directions and dependencies were less clear as we do not fully understand how AS works at a microbiological level even though we do have a great deal of conceptual information. The complexity is comparable to that faced by clinicians using tools of varying power to diagnose diseases. This does not reflect water research deficiencies but rather the complex nature of microbial consortia and how they interact with their environment.

Rather than assume no beliefs we adopted the view that the system likely functioned as follows:

- The system operating conditions were controlled by the Activated sludge solids biomass, the sludge retention time and the hydraulic retention time.
- These controlled parameters and factors whose classes included particle density, organic matter processing and physicochemical conditions. Control occurred in complex ways and probably included many biological feedback cycles. Though we had no collective parameter to describe these three process classes, latent nodes provided a means for beginning to define these complex variables.
- The pathogen LRVs and the microbial indicator concentrations also reflected particle density, organic matter processing and physicochemical conditions but not the measures related to these directly. But relating indicators to pathogens was not a problem for the BN to simulate because of its backcasting capacity.
- 'Particle Density' was viewed as reflecting (causing) the varying observed levels of SS and turbidity, 'Organic Matter' processing was reflected in the liquid BOD and COD measurements and 'Physicochemical Conditions' were major drivers of pH, temperature, Alkalinity and TKN[8].

The parameters shown were selected prior to our incorporating more formal data mining using WEKA. Instead they were derived by selecting variables which showed statistically significant correlations to LRVs (as R^2) exceeding 0.05 (5%) as measured using MS Excel. The R^2 values also showed there was a high relationship between indicator concentrations and LRVs. The latter were used in the model shown instead of LRVs because it would be easier to measure concentrations rather than LRVs in the field and concentration determined ultimately dosage and infection risk.

[8] Logically the reverse of this last point seems more plausible but taking this perspective would have led to having to the physicochemical node having too many parents.

Importantly the list of preferred parameters was comparable to that developed later for semi/naïve Bayes modelling indicating the Bayes approach gives comparable output data to more traditional approaches[9]. This illustrates how 'expert opinion' is a useful starting point if not always optimal.

The primary BN developed is shown in detail in Figure 17a. In this instance we first trialled discretisation of each node into five states or value ranges[10] as an intuitive compromise between overfitting and poor resolution of node variability.

As pathogen LRVs were our primary interest we variously entered 'New evidence'/counterfactual reasoning into the BN by setting the *Cryptosporidium* LRV to its different ranges (Figure 17b). The result was striking. In line with 'common sense' and Flapper et al.'s review (2010) when the highest removal (range 1.4 to 2.6 logs average 2.0) was set to 100%, concurrently the three posterior bacterial concentrations altered to their lowest values. SS and turbidity also tended to their lowest values. The latent nodes also changed markedly especially that for particle density. Sensitivity to findings (Figure 18a) for the *Cryptosporidium* node confirmed bacterial indicators and particulates were most closely associated with *Cryptosporidium* LRVs consistent with the previous section.

Other provisional conclusions for the study system were as follows:

- Counterintuitively the *Giardia* LRV did not relate well to *Cryptosporidium* LRV nor to bacterial indicators or particulate concentration.
- SRT appeared to be the preferred control parameter for optimising *Cryptosporidium* reduction.
- Particle size parameters were confirmed as key surrogates of *Cryptosporidium* removal and organic matter levels and other physico-chemical factors seemed to play a lesser role.

A caveat on these conclusions regarding *Cryptosporidium* emerged when we looked at the predictive accuracy of the model. This was important because we saw the latter as needed for high for confidence in assigning LRV credits to activated sludge for *Cryptosporidium*. 'Testing with cases' was done by repeatedly randomly splitting the data set into a primary model defining data sets and 'test with cases' data sets (78/20) using random number selection. This yielded an error prediction rate of ≈40% where 5 nodes were used (Table 2) and suggested significant overfitting.

Accordingly we repeated the procedure but with only 2 states or ranges per node (Figure 19a and b). This reduced the error rate to a much more satisfactory 11% and other test measures were even more improved (Table 2).

The superficial downside was that the estimated LRV (Figure 19b) was reduced to 1.8 and its standard deviation increased to 0.46. The 'Sensitivity to findings' (Figure 18b) also decreased, although the order of importance of the variables was much the same. We concluded that though the 1.8±0.46 LRV was smaller and less precise than the initial 2.0±0.35 LRV value it was much more credible. Also it proved closer to the >1 LRV cut-off indicated by the semi-naïve BN subsequently.

[9] We undertook stepwise regression analysis as well.

[10] The exceptions SRT and HRT only had 2 and 3 states respectively.

These results highlighted how when estimating LRV credits there is a need to check the accuracy of the underlying prediction model to ensure excess credits are not assigned. Put another way, BNs+data mining methods+accuracy tests provided a means of credibly calculating LRVs.

Separately our analysis supported the use of surrogates and latent nodes for characterising and predicting reductions other reduction processes. It is proposed that novel indicators be introduced for assessing recycled water fitness for purpose but an unanswered question has been how best to relate them to more important parameters such as the primary contaminants. The approach here shows how BNs can provide a logic for this which does not require that contaminants be dependent on surrogate and indicator variables when logically they or at least measurements of them are not.

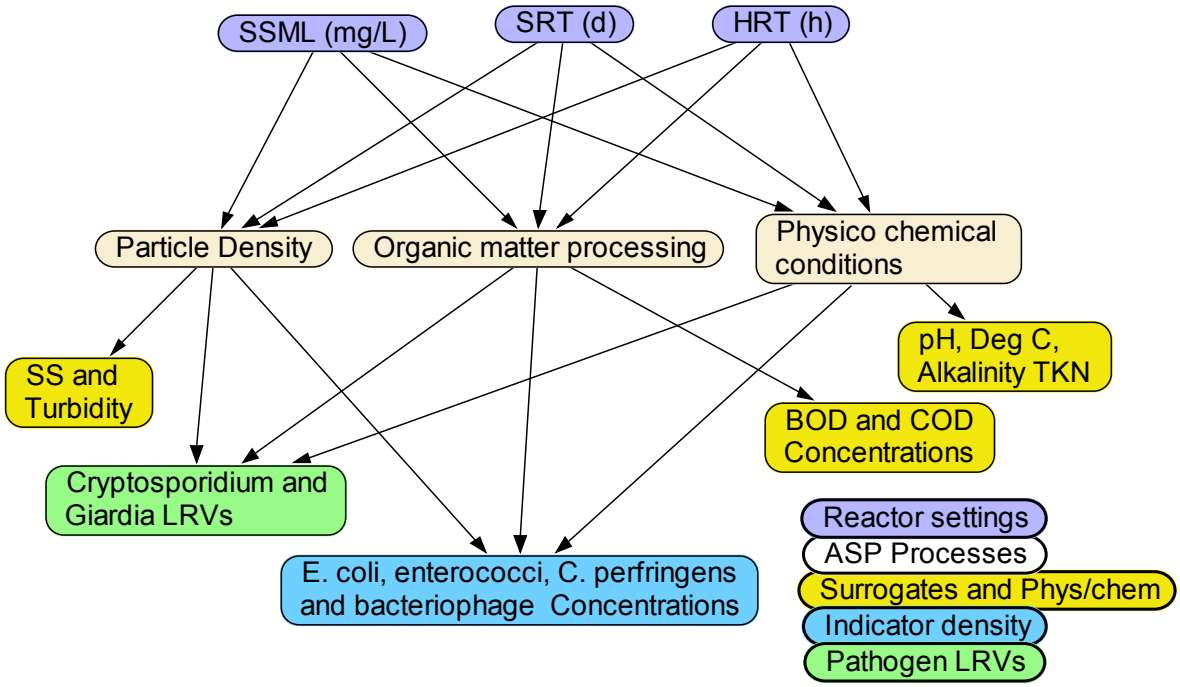


Figure 16. Concept map of the causal network

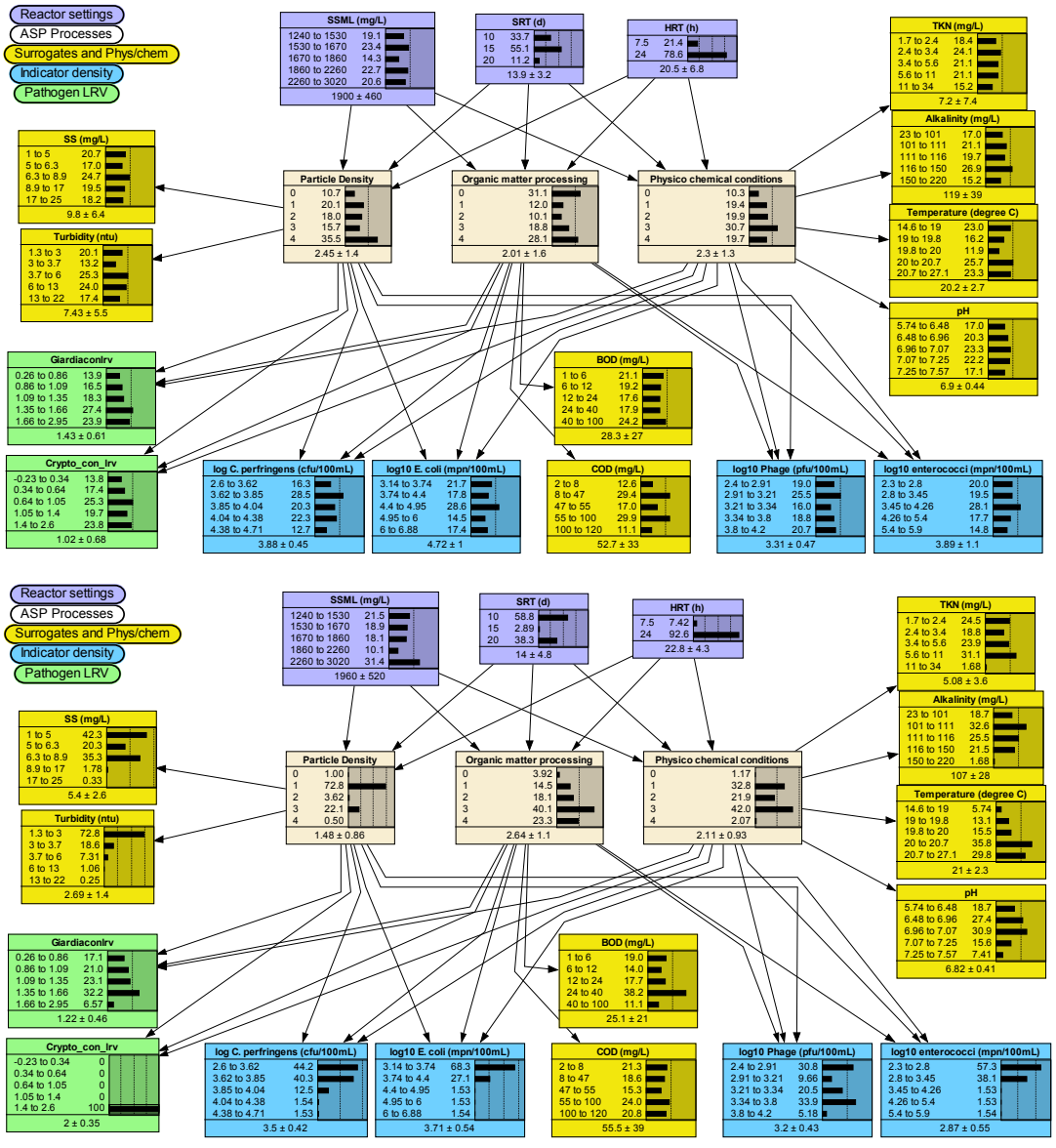


Figure 17. Activated sludge causal network constructed using EM learning – 5 ranges per node a. Primary BN, b. Posterior probabilities when *Cryptosporidium* LRV is set to maximum range

Sensitivity of 'Crypto_con_lrv' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
Crypto_con_lrv	0.461	100	2.28969	100	0.6277174
Particle Density	0.2433	52.8	0.72476	31.7	0.0887026
log10 E. coli (mpn/100mL	0.2362	51.2	0.61001	26.6	0.0609421
Turbidity (ntu)	0.2309	50.1	0.60999	26.6	0.0825560
log10 enterococci (mpn/1	0.2115	45.9	0.53538	23.4	0.0440742
SRT (d)	0.1799	39	0.40780	17.8	0.0397074
log C. perfringens (cfu/	0.146	31.7	0.35685	15.6	0.0329309
Physico chemical conditi	0.1431	31	0.40479	17.7	0.0219026
SS (mg/L)	0.1245	27	0.29578	12.9	0.0150902
Alkalinity (mg/L)	0.08813	19.1	0.24820	10.8	0.0128955
Temperature (degree C)	0.08373	18.2	0.20807	9.09	0.0104894
TKN (mg/L)	0.08092	17.6	0.20208	8.83	0.0097836
pH	0.05442	11.8	0.15344	6.7	0.0077738
log10 Phage (pfu/100mL)	0.04075	8.84	0.33011	14.4	0.0234254
Organic matter processin	0.03627	7.87	0.27008	11.8	0.0138519
SSML (mg/L)	0.03542	7.68	0.11974	5.23	0.0053620
HRT (h)	0.03342	7.25	0.07173	3.13	0.0036534
Giardiaconlrv	0.02687	5.83	0.18491	8.08	0.0082991
BOD (mg/L)	0.01166	2.53	0.14296	6.24	0.0100430
COD (mg/L)	0.004078	0.885	0.10160	4.44	0.0065104

a.

Sensitivity of 'Crypto_con_lrv' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
Crypto_con_lrv	0.6975	100	0.99638	100	0.2487482
Particle Density	0.3002	43	0.46600	46.8	0.1427851
log10 E. coli (mpn/100mL	0.243	34.8	0.36654	36.8	0.1155916
SRT (d)	0.2099	30.1	0.31209	31.3	0.0998295
Turbidity (ntu)	0.1994	28.6	0.29718	29.8	0.0948331
log10 enterococci (mpn/1	0.1847	26.5	0.27163	27.3	0.0878335
Physico chemical conditi	0.1769	25.4	0.26469	26.6	0.0841378
Organic matter processin	0.1383	19.8	0.20957	21	0.0657890
SS (mg/L)	0.1313	18.8	0.18911	19	0.0624392
log C. perfringens (cfu/	0.08633	12.4	0.12389	12.4	0.0410593
pH	0.07009	10	0.09878	9.91	0.0333368
Alkalinity (mg/L)	0.05693	8.16	0.07976	8	0.0270755
HRT (h)	0.05278	7.57	0.07797	7.82	0.0251036
TKN (mg/L)	0.03845	5.51	0.05388	5.41	0.0182887
Giardiaconlrv	0.02914	4.18	0.04048	4.06	0.0138593
COD (mg/L)	0.01944	2.79	0.02696	2.71	0.0092479
Temperature (degree C)	0.01928	2.76	0.02677	2.69	0.0091714
BOD (mg/L)	0.01672	2.4	0.02318	2.33	0.0079526
log10 Phage (pfu/100mL)	0.01056	1.51	0.01467	1.47	0.0050220
SSML (mg/L)	0.002442	0.35	0.00337	0.338	0.0011616

b .

Figure 18. Sensitivity to findings analysis of Cryptosporidium LRVs for Activated sludge causal network constructed using EM learning – a. 5 ranges/states per node b. 2 ranges/states per node

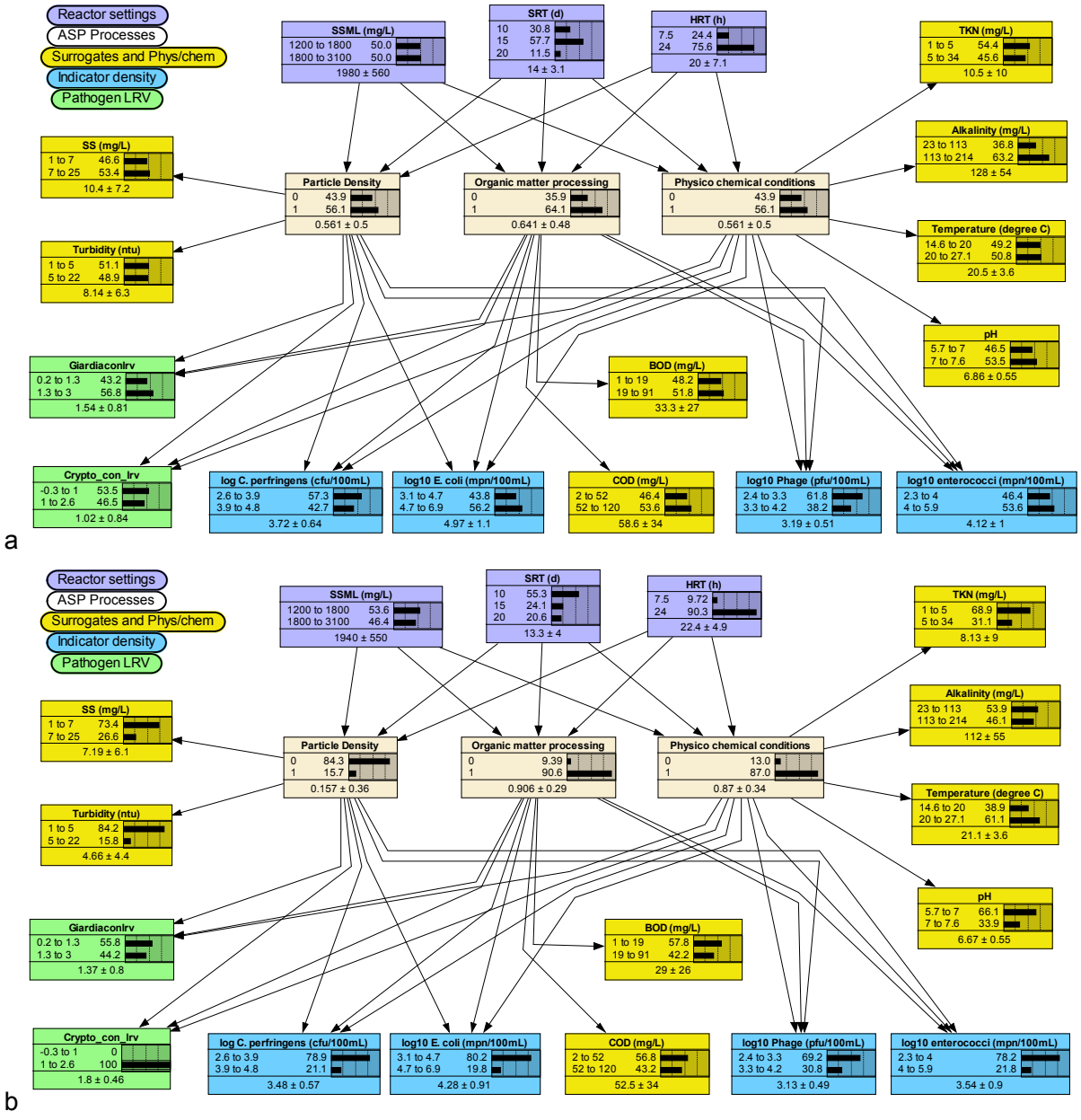


Figure 19. Activated sludge causal network constructed using EM learning – 2 ranges per node a. Primary BN, b. Posterior probabilities when *Cryptosporidium* LRV is set to maximum range

Table 2. Comparison of 'Test with Data' scores for *Cryptosporidium* LRV Node (Crypto_con_lrv)

5 ranges per node	.....Predicted..... [Confusion matrix] -0.23 0.34 t 0.64 t 1.05 t 1.4 to Actual ------ ------ ------ ------ ------ ------ 0 2 0 0 0 -0.23 to 0.34 0 4 0 0 0 0.34 to 0.64 0 1 1 1 1 0.64 to 1.05 0 0 0 0 2 1.05 to 1.4 0 0 0 1 5 1.4 to 2.6 Error rate = 44.44% :Scoring Rule Results: Logarithmic loss = 2.488 Quadratic loss = 0.7756 Spherical payoff = 0.5344
2 ranges per node	...Predicted.. [Confusion matrix] -0.3 t 1 to 2.6 Actual ------ ------ ------ 7 2 -0.3 to 1 0 9 1 to 2.6 Error rate = 11.11%: Scoring Rule Results: Logarithmic loss = 0.3626 Quadratic loss = 0.2041 Spherical payoff = 0.8937
Ideal	...Predicted.. [Confusion matrix should follow diagonal] -0.3 t 1 to 2.6 Actual ------ ------ ------ 9 0 -0.3 to 1 0 9 1 to 2.6 best Error rate = 0% Scoring Rule Results: best Logarithmic loss = 0 (possible range 0-∞) best Quadratic loss = 0 (possible range 0-2) best Spherical payoff = 1 (possible range =0-1)

Notes

1. The test results shown were generated by Netica. A diverse range of other metrics is possible to generate for example using WEKA or other BN software e.g. the ROC scores which are discussed and illustrated further in our semi-Naïve BN paper.
2. The confusion matrix is particularly useful as it shows not only inaccuracies but how great each inaccuracy is. It effectively identifies false positives and negatives and true positives and negatives.
3. Marcot (2012) provides an excellent review of the topic.

6. Melbourne Water ETP 1: LRV estimation & validation using historical data

6.1. Introduction

This case study was undertaken as a test of the proposed BN validation process using real collected data from Melbourne Water's Eastern Treatment plant (ETP).

Among other things Melbourne Water has acquired approximately 1757 microbial measurement records for *E. coli*, somatic coliphage and *C. perfringens.* The microorganisms were measured at 6 stations along the water recycling part of the ETP system between October 2013 and September 2015. Furthermore, Melbourne Water has concurrently collected physicochemical data and high frequency on-line parameter measurements for the same system.

A description provided by John Mieog from Melbourne Water is reproduced in Figure 20. Microbial as well as physico-chemical measurements were collected at the 6 points indicated by the upward arrows.

EASTERN TREATMENT PLANT

Data file description:	This file contains pathogen indicator data across the tertiary treatment process at the Eastern Treatment Plant.
	Samples are taken routinely independently of plant conditions at the time.
	The data includes periods when the process was both compliant and non-compliant with all Class A recycled water HACCP CCPs.
	No recycled water was supplied to Class A customers when HACCP CCPs were non-compliant.
Treatment process description:	Secondary effluent is fed to the tertiary treatment step either via the Effluent Holding Basins (wet weather storage) or directly via a bypass.
	Secondary effluent feedwater is lifted by the Tertiary Supply Pump Station (TSPS) and flow throug the remainder of the process is by gravity.
	Pre-Ozone treatment is not validated for pathogen log reduction credits.
	Biological Media Filtration (BMF) is not validated for pathogen log reduction credits.
	Post-Ozone is validated for pathogen log reduction credits based on measured ozone Ct (≥0.6-log *Crypto*, 4-log Virus & Bacteria).
	UV is validated for pathogen log reduction credits based on calculated log inactivation(4-log *Crypto*, 4-log Bacteria).
	Chlorine is validated for pathogen log reduction credits based on measured chlorine Ct (≥0.8-log *Virus* & Bacteria).
	The chlorine process can operate in either free chlorine or monochloramine modes. Only free chlorine mode has been used to date.
	The respective sample points are labelled in the diagram below.

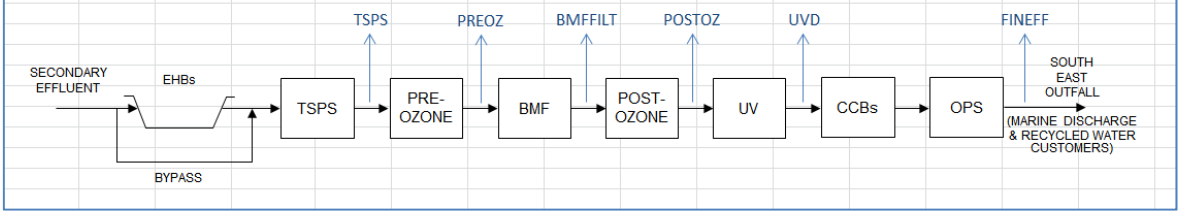


Figure 20. ETP treatment and sample collection configuration.

These data were provided in two sets, the first being 921 records for October 2013 and September 2014 with the remainder covering the subsequent period to September 2015. This split serendipitously provided us with two relatively large, unbiased, related but independent measurement sets ideal for illustrating real world (Bayesian) validation, exploring changes in treatment performance over time and demonstrating other tasks BN technology can be used for.

This case study outlines the analysis of these data and BN use for:

- Initially constructing causal BNs for the three central analytes which characterise treatment effectiveness as LRVs;
- Comparing the result of calculating LRVs using a BN v. Learning from primary data tables;

- Assessing model accuracy and hence prediction reliability;
- The use of semi-naïve BNs compared to causal BNs;
- Gathering summary statistics pertinent to LRV calculation and crediting;
- Improving understanding of system structure and function using Netica's Sensitivity (to findings) analysis tool;
- Bayesian Validation;
- The use of WEKA in data mining especially large on line data sets.

6.2. Methods

BNs were constructed reflecting all or parts of the ETP design diagram (Figure 20). LRVs were then calculated by:

- 'Learning' the characteristics of concentration nodes and subtracting (log transformed) downstream from upstream microbial concentration measurements using Netica's inbuilt algorithms to generate the CPTs;
- 'Learning' from data tables with individual in/out (log transformed) LRV data value calculated prior to the learning to produce all CPT values.

Initial preparation of worksheets for BN learning involved the following:

- Sorting the raw Melbourne Water data into database table field X record format using Pivot tables etc. and manual reorganisation and final checking;
- Conversion of arithmetic values to $\log_{10}$ values;
- Conversion of censored measurements initially into ≈detection limit values (In practice this created easily recognised slightly minus log values which could be recognized and appropriately binned by the Netica^{tm} software which also has a facility for accounting for "< value " ranges, where discretization is coded as being between the detection limit value e.g. 0 and –Infinity.
- Conversion of individual tables/worksheets into .CSV files suitable for learning (for the initial data set to September 2014 n=60 for Somatic coliphages and *E. coli* and n=29 for *C. perfringens*).

The final grab sample data set details are summarized in Table 3. The on-line data set analysed is summarized in Table 4. Note that in the present instance latter case we only selected those data points for which there was corresponding physicochemical data as this was manageable but still a large number. The negative log values indicate sample sets where there were below detection limit values.

Table 3. Summary of final historic grab sample data provided on the ETP by Melbourne Water

Station	Parameter	Units	Average	Std dev	Min	Max	n	Start date	End date
ATTP-TSPS	***C. perfringens***	Log Orgs /100mL	3.53	0.28	2.70	4.26	77	30/07/13	20/08/15
	E.coli	Log Orgs /100mL	4.33	0.47	3.36	5.38	107	30/07/13	20/08/15
	Somatic coliphage	Log pfu /100mL	4.08	0.44	2.98	5.60	104	2/08/13	20/08/15
	SS_OM	mg/L	17.55	12.38	1.00	142	663	21/02/13	11/09/15
	Alkalinity _OM	mg/L	74.15	15.34	21	112	660	26/02/13	11/09/15

Station	Parameter	Units	Average	Std dev	Min	Max	n	Start date	End date
	NH3-N _OM	mg/L	0.67	0.88	0.03	7.00	663	26/02/13	11/09/15
	UVT _OM	%	43.05	5.35	12	62.1	659	4/03/13	11/09/15
	UVTf _OM	%	47.68	5.20	23	68.2	659	4/03/13	11/09/15
	True Colour _OM	Pt/Co	86.71	17.78	6.00	137	658	5/03/13	11/09/15
	Nitrite-N _OM	mg/L	0.27	0.21	0.00	2.00	348	26/02/13	3/07/14
	Nitrate -N_OM	mg/L	9.55	3.92	1.40	20.2	347	27/02/13	3/07/14
	NH3-N _Grab	mg/L	1.22	1.28	0.06	6.50	345	4/03/13	21/07/14
	NH3-N _On-line	mg/L	1.29	1.26	0.00	6.50	336	4/03/13	21/07/14
	UVT_Grab	%	46.70	25.14	17	492	343	4/03/13	3/07/14
	UVT_Online	%	49.51	24.65	27.9	493	338	4/03/13	3/07/14
ATTP-PREOZ	***C. perfringens***	Log Orgs /100mL	3.31	0.33	2.08	4.26	78	30/07/13	20/08/15
	E.coli	Log Orgs /100mL	2.57	0.67	1.30	4.38	103	30/07/13	20/08/15
	Somatic coliphage	Log pfu /100mL	2.26	0.68	0.95	4.78	100	3/09/13	20/08/15
ATTP-BMFFILT	***C. perfringens***	Log Orgs /100mL	2.54	0.45	1.60	3.90	79	30/07/13	20/08/15
	E.coli	Log Orgs /100mL	2.31	0.56	1.00	3.84	108	30/07/13	20/08/15
	Somatic coliphage	Log pfu /100mL	1.59	0.67	-0.00	3.40	105	2/08/13	20/08/15
	Alkalinity_OM	mg/L	67.38	14.27	24.00	120	664	26/02/13	11/09/15
	UVT_OM	%	63.78	4.51	37.70	75.0	658	4/03/13	11/09/15
	UVTf_OM	%	66.00	4.08	46.50	79.3	658	4/03/13	11/09/15
	True Colour_OM	Pt/Co	18.25	6.40	6.00	72.0	657	5/03/13	11/09/15
	Nitrite-N_OM	mg/L	0.07	0.10	0.00	0.72	348	26/02/13	3/07/14
	NH3-N_Grab	mg/L	0.17	0.32	0.01	3.90	187	4/03/13	19/11/13
	NH3-N_Online	mg/L	0.27	0.62	0.04	3.40	33	3/09/13	19/11/13
	UVT_Grab	%	64.41	6.14	12.20	77.1	344	4/03/13	3/07/14
	UVT_Online	%	65.51	4.16	52.30	75.3	322	4/03/13	3/07/14
	Manganese	mg/L	0.03	0.04	0.00	0.14	28	8/05/13	19/06/13
ATTP-POSTOZ	***C. perfringens***	Log Orgs /100mL	0.46	0.55	-0.00	3.00	76	4/03/14	20/08/15
	E.coli	Log Orgs /100mL	0.22	0.57	-0.00	2.66	101	3/09/13	20/08/15
	Somatic coliphage	Log pfu /100mL	0.08	0.41	-0.00	3.04	101	3/09/13	20/08/15
ATTP-UVD	***C. perfringens***	Log Orgs /100mL	0.21	0.38	-0.00	1.68	74	4/03/14	20/08/15
	E.coli	Log Orgs /100mL	0.01	0.06	-0.00	0.30	98	3/09/13	20/08/15
	Somatic coliphage	Log pfu /100mL	-0.00	0.00	-0.00	-0.00	98	3/09/13	20/08/15
ETPFINEFF	***C. perfringens***	Log Orgs /100mL	0.02	0.08	-0.00	0.48	78	5/03/13	20/08/15
	E.coli	Log Orgs /100mL	-0.00	0.00	-0.00	0.00	138	2/01/13	20/08/15

Station	Parameter	Units	Average	Std dev	Min	Max	n	Start date	End date
	Somatic coliphage	Log pfu /100mL	-0.00	0.00	-0.00	-0.00	132	7/02/13	20/08/15

Table 4. On-line parameters monitored at 10 min intervals (selected set)

Station/parameter code	Parameter description	Units	Average	St Dev	Min	Max	n
ol_TSPS_Turbidity	Inflow turbidity	NTU	5.42	5.59	0.07	99.97	654
ol_TSPS_pH	Inflow pH	pH units	6.54	0.20	5.90	7.16	654
ol_TSPS_UVT	Inflow UV transmission	%	46.99	4.97	27.68	60.00	654
ol_TSPS_NH3	Inflow ammonia	mg/L	1.00	1.08	0.01	5.77	624
ol_Preozone_mgpL	Preozone dosage	mg/L	9.88	1.89	5.00	16.22	459
ol_Preozone_kLps	Flow rate	kL/s	4.56	0.95	0.01	7.18	487
ol_Postozone_UVT	PostOzone UVT	%	74.52	3.39	52.51	83.01	654
ol_Postozone_pH	PostOzone pH	pH units	6.23	0.33	5.00	6.91	654
ol_Postozone_NH3	PostOzone ammonia	mg/L	0.12	0.31	0.01	4.79	654
ol_Postozone_oC	PostOzone temperature	Degrees C	19.51	2.39	13.40	24.76	654
ol_BMFB21_Turbidity	Post biofilter turbidity unit 21	NTU	0.68	0.52	0.08	5.18	654
ol_BMFB22_Turbidity	Post biofilter turbidity unit 22	NTU	0.71	0.60	0.19	7.21	654
ol_BMFB23_Turbidity	Post biofilter turbidity unit 23	NTU	0.74	0.53	0.14	7.24	654
ol_BMFB24_Turbidity	Post biofilter turbidity unit 24	NTU	0.77	0.79	0.00	10.00	652
Sample_Date	Data collection period				2/1/13	11/9/15	657

BN learning employed Netica's "Incorporate Case File" and Expectation Maximization (EM) tool which are accessed through the 'Cases' menu. Both causal and semi-naïve BNs (SNBs) (Carvajal et al., 2015) were constructed.

The purpose of the causal nets was to describe how we believed concentrations and LRVs and processes related to one another and the net values functioned/varied overall in a manner reflecting the known design of the ETP recycled water system.

The purpose of the SNBs was to identify, and if possible better understand, the main influences on individual selected 'target' or 'class' nodes especially those representing LRVs.

An analogy here is the use of diagnostic tests (associated measurements) by a clinician to identify and understand a disease or disease process (target node). Such BNs function in some respects similarly to a regression analysis where the disease is the Y value. However the causal directions are relaxed and BNs allow examination of how the changes in one (disease) or more associated variables are reflected in another and how the disease states themselves are effect as well.

A related activity undertaken was Bayesian Sensitivity testing (Pollino et al., 2007).

The parameter codes above should be used in part as shorthand reference for relating Netica^{tm} nodes to variables.

In the Netica[tm] BNs shown below we have colour coded the nodes so that:

- Yellow nodes for miscellaneous (parent) nodes.
- Green nodes are sampling station concentration measurements.
- Red nodes are LRV nodes for single processes.
- Blue nodes are for multi-process LRV calculation.

We also used the data mining capacity of WEKA in the development of some SNBs (Witten et al., 2011).

6.3. Results

6.3.1. Initial BN construction

Initially a draft combined BN was constructed (Figure 21) reflecting system configuration (Figure 20). In addition to the microbial parameters, we hypothesized that season and parameter type might influence concentration nodes and hence LRVs, so we included both as “root nodes”. ‘Sensitivity to findings’ analysis suggested only a small effect, if any, of season. This primary template was used to construct further individual microorganism-focused BNs.

Overall LRVs were estimated to be >4 mainly due to the ozonation consistent with the TSPS and post UV concentration estimates changing from an average of >4 to less than 0 $\log_{10}$ units respectively. This reduction was necessarily a minimum estimate since the post ozone and post UV data were very heavily censored. But it did generate a minimum estimate of what LRV credits should be.

As expected the BN showed concisely, in a HACCP-style flow chart (technically a network):

- The summary concentration and removal statistics;
- LRVs for different processes and process combinations and how they were derived;
- The spread of the concentration data and LRV estimates;
- The impact of data censorship;
- The causal relationships believed to apply at the ETP based on the system diagram.

Node names for sampling concentration data (green) correspond to those provided by Melbourne Water. LRV node names correspond to the following:

- LRVPREO - LRV for the preozonation stage
- LRVBF or LRV Biofilt - LRV for the biological filter
- LRVPOO or LRV Postozone – LRV for postozonation process
- LRVUV - LRV for the UV systems
- LRVTB or LRV T>BMF - LRV for preozone + biological filter
- LRVTPOZ or LRV Pre>PoOZ - LRV for biological filter + post ozone
- LRVPPOZ - LRV for the preozone + biological filter + post ozone
- LRVTU or LRV T>UV - LRV for all four treatment processes i.e. TSPS to UV.

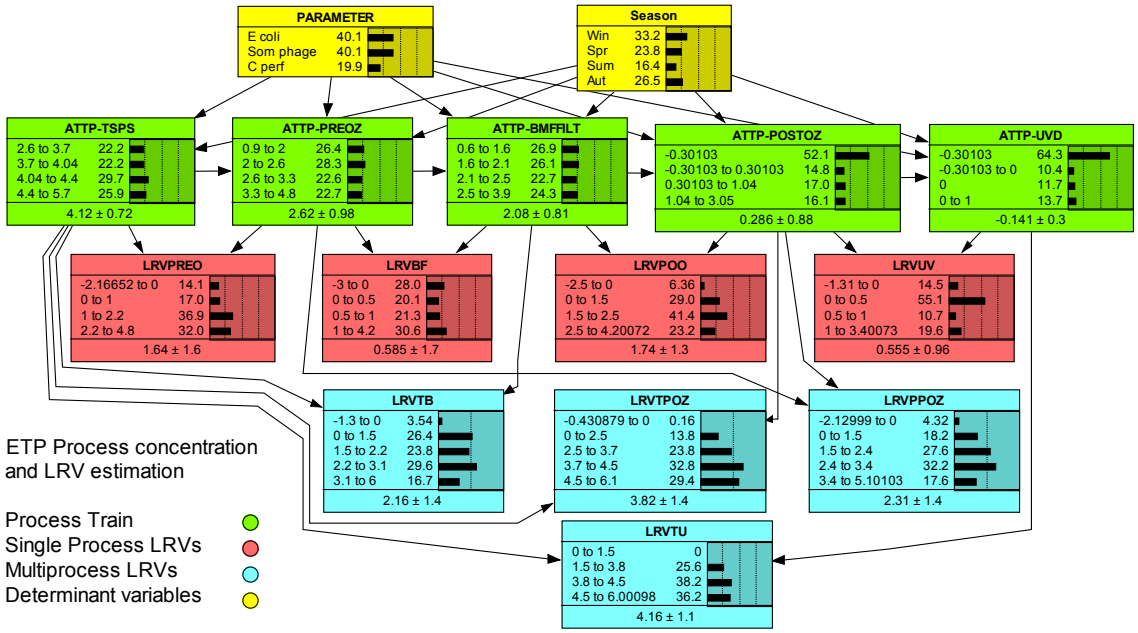


Figure 21. Combined data concentration changes and LRVs

6.3.2. *E. coli*, Somatic coliphage and *C. perfringens* BNs with calculated LRVs

As there were far fewer *C. perfringens* records in the initial pre-September 2014 data, and the effect of season appeared to be very modest, we analysed data for each microbe independently.

Figure 22, Figure 23, and Figure 24 show the individual LRV credits and causal BNs for each indicator respectively. Note that the names of the nodes have been altered to aid the reader although the arrangements are the same as for the combined BN. The first BNs compiled here were for the first half of the data to September 2014. The accompanying ones are for the whole 2 years of monitoring.

In these first causal nets (Figure 22, Figure 23, Figure 24) the LRVs were calculated by subtracting one (log transformed) probability density function from another within the BN. We routinely used 1000 samplings in a manner similar to that undertaken when using Palisade @Risk's Monte Carlo functions.

Netica[tm] includes in its suite of inbuilt functions most of these sample probability density functions for use in calculating PDF based CPTs e.g. NormalDist (average,standard deviation) for defining a Normal distribution. However the algorithms are not used directly in the manner of a spreadsheet formula but are one of several tools used to generate the CPTs which centrally underlie each node.

Subsequently, we repeated the process using the full data sets. This virtually doubled the number of concentration records used to construct the BNs. The two are shown alongside one another for comparison. It can be seen despite the year between samplings on the surface there had been little change in the BN statistics.

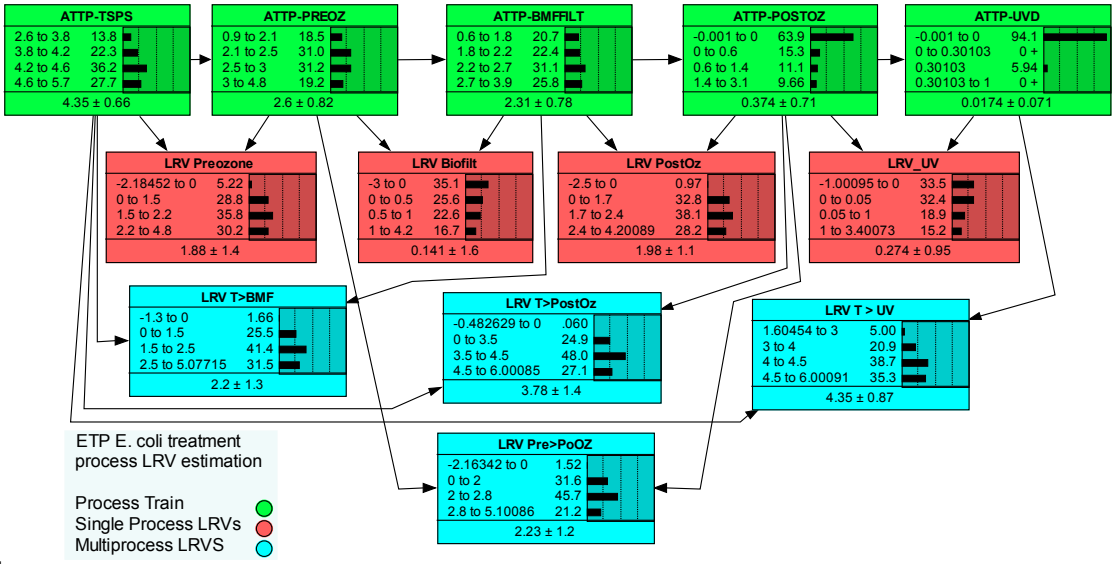
For the 3 indicators the average LRVs were > 4.35, 4.16 and 3.2 respectively. The somatic phage estimate was likely highly underestimated as most of the post ozone samples as well as the post UV samples showed no bacteriophage at all.

Prior to the UV treatment, for which post-treatment data were heavily censored, the average LRVs were already substantial at 3.87, 3.8 and 2.67 respectively.

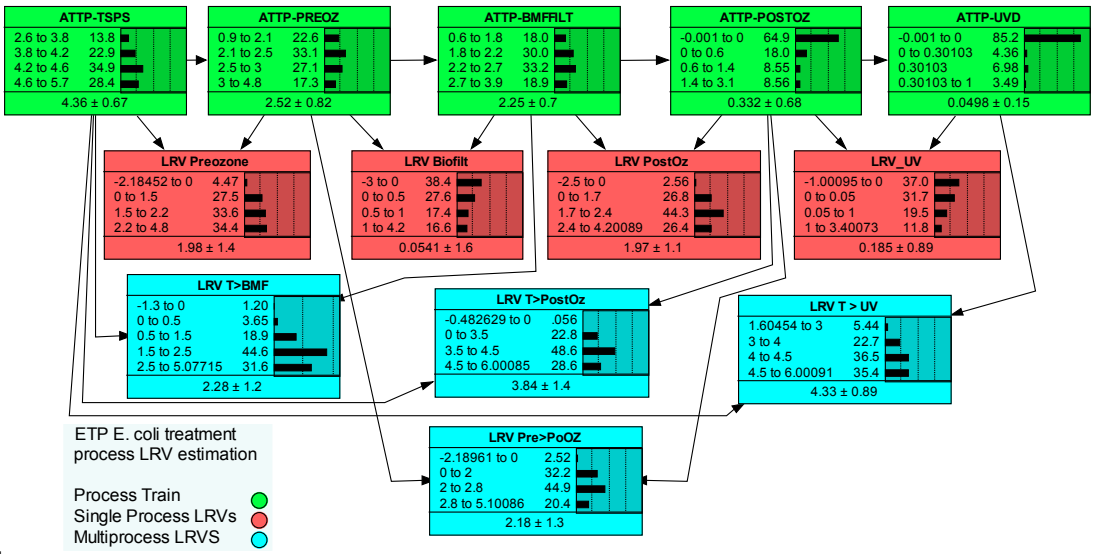
The effect of measurement censorship was highlighted by BNs in the post ozone and UV data in all cases and hence the need for developing methods for providing appropriate credits for these processes by methods other than direct measurement. These could still use BN software though as such information could be used to independently create the LRV CPTs where direct calculation or learning was less reliable. This might be done using expert opinion, estimation of LRV PDFs using methods designed for censored data sets - see (Khan, 2010), literature data and dosing experiments. In short the problem of censored data was not a limitation of BNs but rather the data. And BNs offer a defensible solution where LRV CPT probabilities are estimated by these alternative methods. Such revised LRVs can be done in a fully transparent manner which would show shows apparent LRV variation overall and for different ranges (= node states) and checking of the impact of error in LRV estimation.

In the case of *E. coli,* biofiltration appeared to have little impact though the process seemed to be more effective on phage and *C. perfringens.* A better way to assign credits may be to quantify the combination of preozonation and biofiltration together rather than each process alone as we have in fact done here i.e. the LRV T>BMF node

Irrespective of the indicator, preozonation showed high standard deviations. The reason may be that the LRV estimation via PDF algorithms may have made the value very sensitive to fluctuations in *E. coli* concentrations in the feedwater. Alternative post preozonation sampling might not have been fully representative of microbial concentrations.

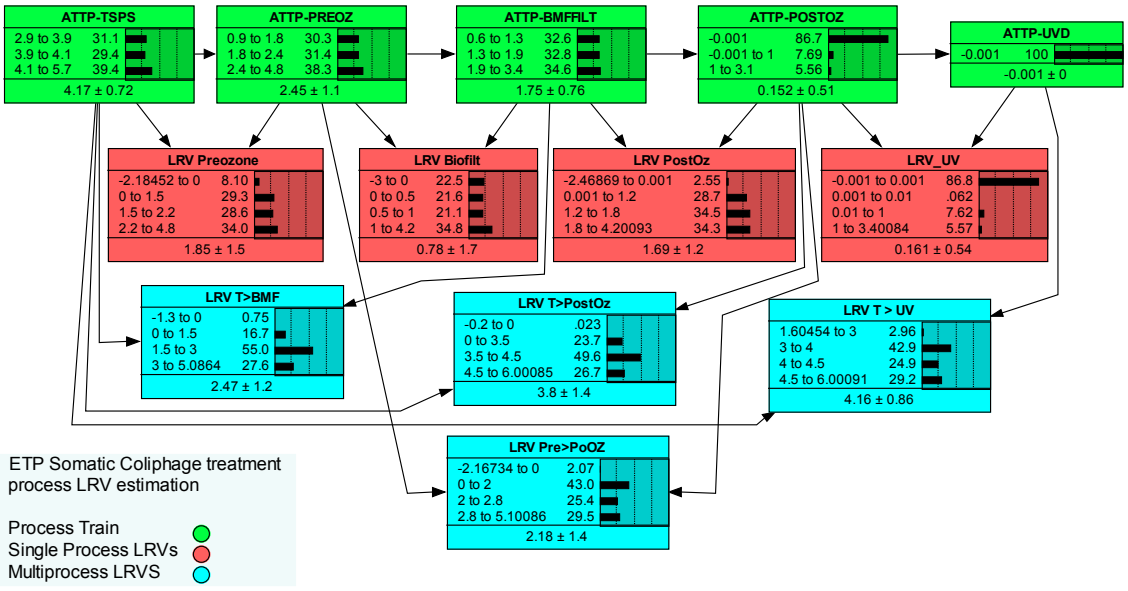


a.

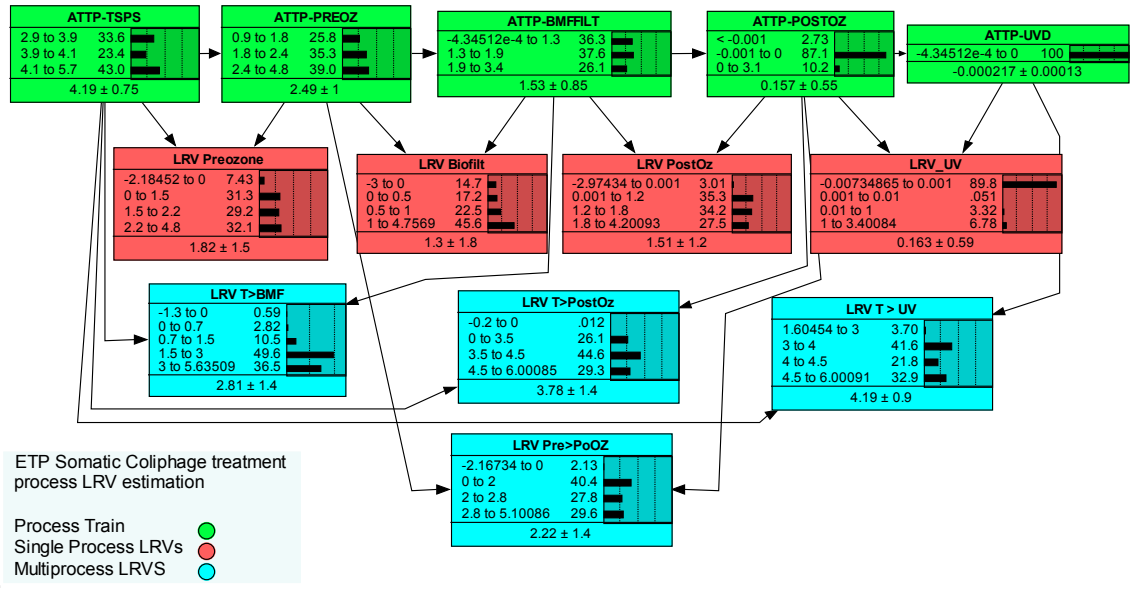


b.

Figure 22. *E. coli* ETP causal BNs a. initial data set, b. whole data set

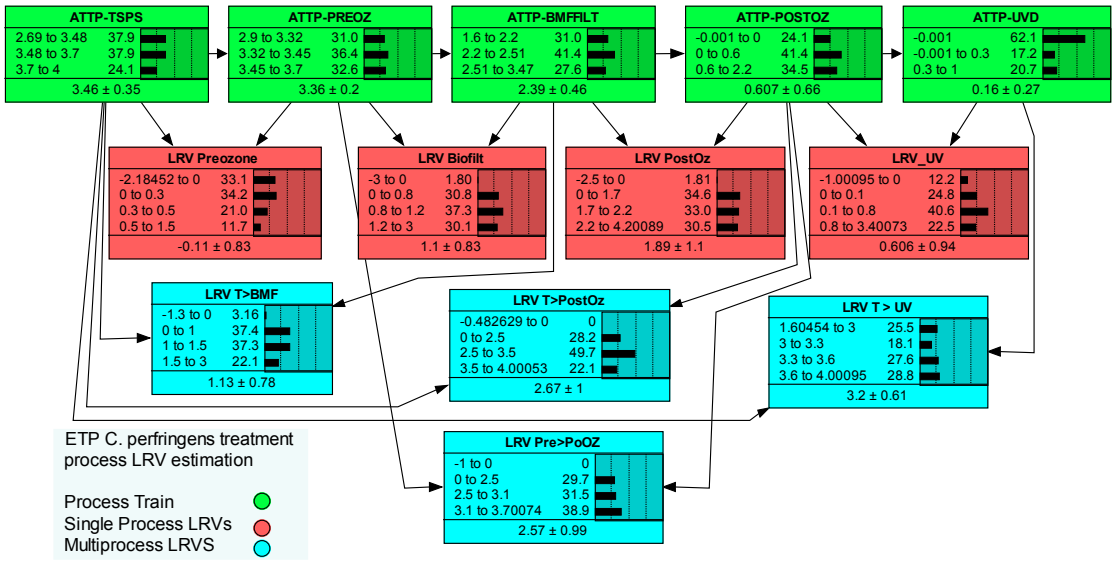


a.



b.

Figure 23. Somatic coliphage ETP causal BNs a. initial data set, b. whole data set



a.

b.

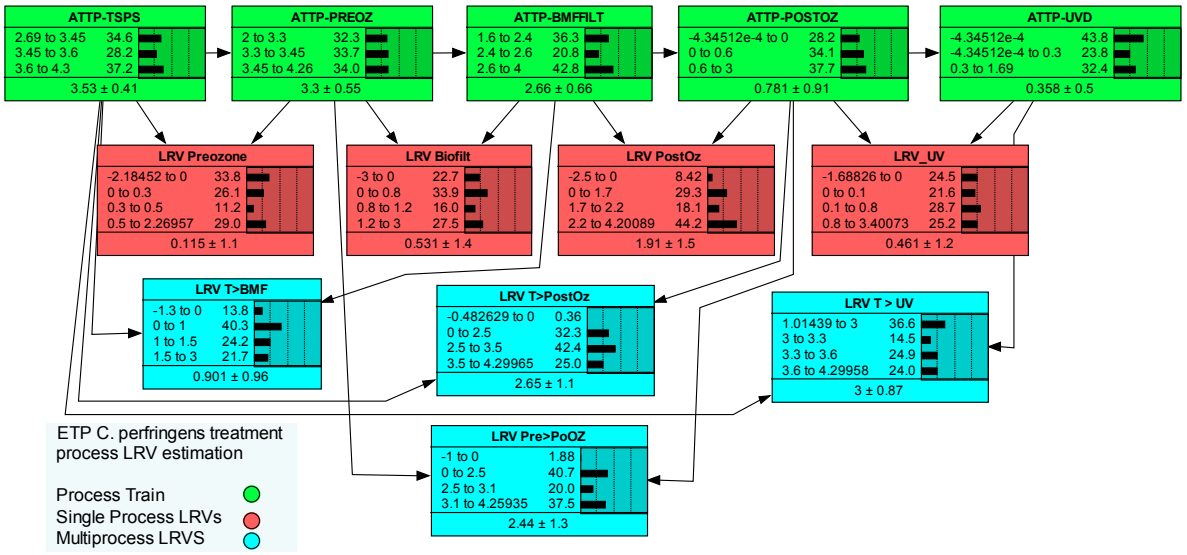


Figure 24. *C. perfringens* ETP causal BNs a. initial data set, b. whole data set

6.3.3. Learning BN LRVs

An alternative and potentially preferable method for estimating LRVs was to calculate the individual daily LRVs using a spreadsheet before the learning process was performed. This was possible because Melbourne Water had matched samples by date and time to a high degree. While it was unlikely that Melbourne Water sampled the identical slug of water as it passed through the system, all samples appeared to be taken at a similar hour and so they would have been matched in respect to the season, day of the week and general time of day, factors which are known to influence water quality.

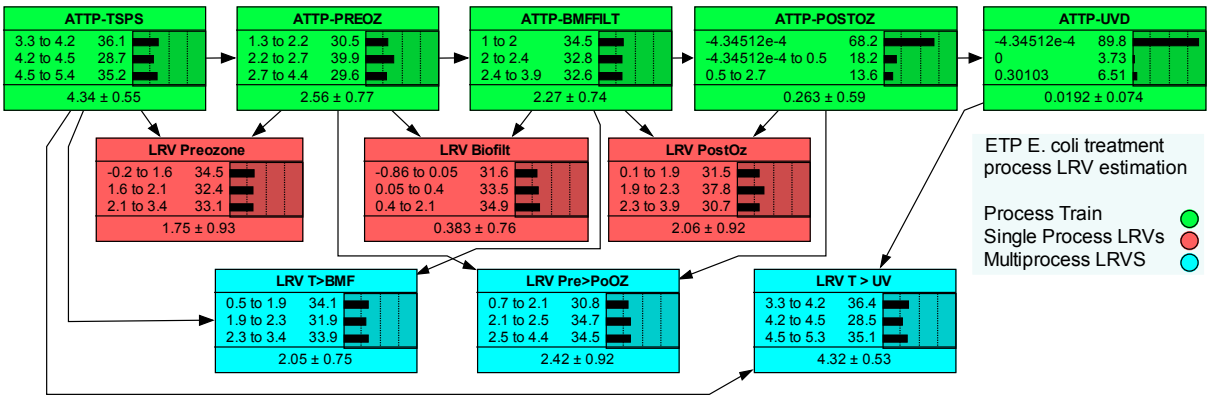
To do this all that was done was to calculate a set of LRVs using Excel for in/out data matched by date. Missing data were replaced with '*' in line with Netica^{tm} coding procedures for files to be learned.

The worksheets were then saved as updated .CSV files for subsequent learning (Though it is possible to use native .xlsx files, .CSV files were in our experience less prone to error).

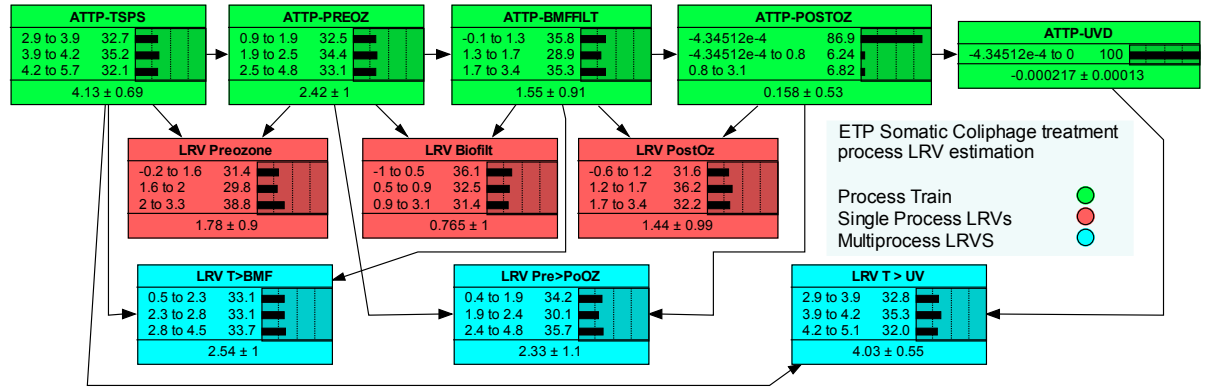
The BN structures shown in Figure 25 are essentially the same as in Figure 22, Figure 23 and Figure 24 except that we omitted the UV treatment step because the lack of uncensored data made LRV calculation problematic.

Comparison of the concentrations and LRVs shows that the summary statistics were of comparable magnitude and value to those previously collected.

a.



b.



c.

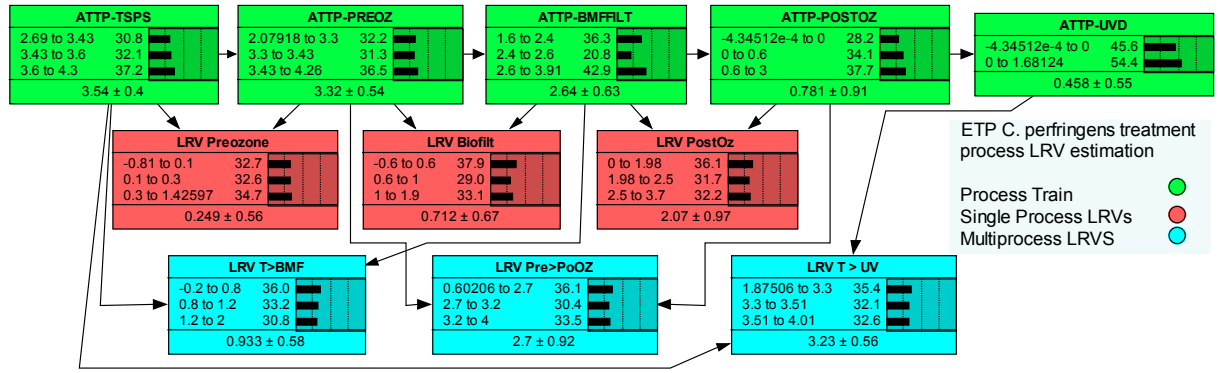


Figure 25. ETP causal BNs developed using learning for all nodes a. *E. coli,* b. Somatic Coliphage, c. *C. perfringens*

6.3.4. Calculating vs. Learning LRVs

As noted above estimating LRVs by directly subtracting one concentration PDF from another is confounded potentially by seasonal, weekly and diurnal fluctuations in the microbial concentrations which are not necessarily reflected in treatment effectiveness. To illustrate, UV

credits are judged primarily by log reduction rather than whether a given concentration is achieved.

As a result simply Monte Carlo-ing input PDFs against output PDFs to calculate LRVs could conceptually generate overly high estimates of LRV variance as standard deviations.

This effect can only be countered calculation-wise where there are process inputs and outputs which can defensibly be assumed to represent samples of the same water taken concurrently. Smeets (Smeets et al., 2008) proposed an alternative approach that in and out percentile values should be aligned but this method in our opinion unjustifiably eliminates LRV process variance which will still likely exist if only due to analysis methodology created variation as with *E. coli* mpn counts whose estimates are well known to have significant uncertainty boundaries even though these are seldom reported in the literature.

Fortunately the Melbourne Water case study data were collected as concurrently as can be undertaken practically without resorting to the use of dye or salt tracing to match sample timing precisely.

The comparison of calculated v. learned LRVs statistics is shown in Table 5. The difference on average between calculated learned LRV averages was both small and insignificant: -0.01±0.22. But consistently the Learned Standard Deviations were all lower than those calculated being only 60% on average of those obtained via the Monte Carlo approach and the outer percentiles were similarly constrained. The 10^{th}/90^{th} percentile extremes of the distribution were even more reduced by as much 2 log units an important consideration when trying to evaluate the impact of poor performance.

Table 5 also illustrates how the statistics calculated by Netica^{tm} were arguably detailed enough for claiming LRV credits for novel treatments such as preozonation and biofiltration.

Table 5. Summary LRV statistics extracted from final BNs through within BN LRV calculation and learning of pre-calculated LRV pairs

Treatment	Micro-organism	Method	Average	SD	Median	IQR	Lower 90th	Upper 90th
Preozone	*E. coli*	Calculated	1.98	1.38	1.87	1.79	.029	4.42
		Learned	1.77	0.88	1.83	1.31	0.06	3.2
	Somatic coliphage	Calculated	1.81	1.47	1.77	1.93	-0.73	4.39
		Learned	1.76	0.81	1.84	1.22	0.08	3.13
	C. perfringens	Calculated	0.12	1.08	0.18	1.31	-1.86	1.96
		Learned	0.25	0.5	0.20	0.72	-0.67	1.26
Biofiltration	*E. coli*	Calculated	0.05	1.59	0.21	1.80	-2.61	3.23
		Learned	0.36	0.68	0.24	1.02	-0.71	1.85
	Somatic coliphage	Calculated	1.30	1.78	0.90	2.39	-1.98	4.34
		Learned	0.79	0.91	0.67	1.30	-0.79	2.7
	C. perfringens	Calculated	0.53	1.38	0.64	1.30	-2.33	2.67
		Learned	0.70	0.6	0.76	1.02	-0.44	1.76
Post-ozonation	*E. coli*	Calculated	1.97	1.11	2.02	1.06	0.15	3.85
		Learned	2.02	0.84	2.09	1.06	0.38	3.6
	Somatic coliphage	Calculated	1.50	1.17	1.40	1.27	0.068	3.76
		Learned	1.41	0.9	1.45	1.25	0.31	3.13
	C. perfringens	Calculated	1.91	1.47	2.03	2.10	-1.01	3.97
		Learned	2.05	0.85	2.20	1.39	0.27	3.51
Preozone + Biofiltration	*E. coli*	Calculated	2.28	1.24	2.08	1.51	0.50	4.66
		Learned	2.03	0.69	2.09	1.06	0.70	3.2
	Somatic coliphage	Calculated	2.81	1.35	2.59	1.99	0.82	5.27
		Learned	2.53	0.91	2.55	1.38	0.77	4.24
	C. perfringens	Calculated	0.90	0.95	0.89	1.15	-0.83	2.65
		Learned	0.96	0.56	0.96	0.85	-0.06	1.87

Treatment	Micro-organism	Method	Average	SD	Median	IQR	Lower 90th	Upper 90th
Biofiltration+ Post-zonation	*E. coli*	Calculated	2.18	1.26	2.27	1.32	0.15	4.53
		Learned	2.41	0.82	2.32	1.08	0.92	4.1
	Somatic coliphage	Calculated	2.17	1.26	2.27	1.32	-1.75	5.04
		Learned	2.29	1.0	2.16	1.62	0.61	4.4
	C. perfringens	Calculated	2.43	1.26	2.72	2.06	0.19	4.10
		Learned	2.76	0.86	2.92	1.34	0.89	3.88
Overall including UV (lower limit)	*E. coli*	Calculated	4.36	0.88	4.29	1.07	2.88	5.78
		Learned	4.34	0.49	4.34	0.81	3.4	5.2
	Somatic coliphage	Calculated	4.19	0.89	4.10	1.34	1.79	5.97
		Learned	4.07	0.50	4.04	0.73	3.05	4.95
	C. perfringens	Calculated	2.99	0.87	3.27	1.21	1.28	4.15
		Learned	3.26	0.55	3.39	0.74	2.07	3.93

6.3.5. Causal model accuracy

A perennial question asked of parametric and non-parametric models is how accurate they are.

A common past practice with parametric models has been to report an R^2 value and a standard error for the Y estimate. Similar statistics are generated with more sophisticated approaches such as principle component analysis (PCA) and neural net analysis. A useful example identified repeatedly in this report which we have used as a reference is Flapper et al. (2012)where R^2 is reported for increasingly complex predictive models.

This report is worth considering in several respects:

- It illustrates the need and trend to use more sophisticated statistical analysis including the calculation of model accuracy.
- It illustrates a limitation with frequentist statistics techniques, the outputs can be very difficult to use and interpret as they lack visual cues.
- Factors influencing LRVs may be complex and their relative significance may vary depending on LRV value range.

BNs address these issues as follows:

- Unlike neural nets and PCA they generate visual representations of complex relationships.
- The need for inference to be reliable is now well recognised in the Bayesian community and diverse accuracy metrics are now routinely available in the software (Marcot, 2012) for assessing model accuracy.
- BN sensitivity analysis allows the relative significance of different ranges to be rapidly assessed (Pollino et al., 2007, Korb and Nicholson, 2011).

Note: The latter report describes alternative ways to analyse complex concentration/LRV data and is unexceptionable best practice *per se.* We have repeatedly referred to it here as it illustrates the alternative of estimating LRVs using a 'frequentist' statistics approach and by implication that should water authorities choose Bayes over the latter they will become part of the wider Bayesian v. Frequentist Statistics debate over which is the better method which has been outlined elsewhere in this report.

To undertake accuracy testing of a model one common approach is to take a whole data set and divide it into a model development set and a test set. In WEKA this process is automated such that the primary data set is split typically 80/20 with records being allocated randomly in

proportion to one or other test categories. The process is then repeated using a different random number to generate the split.

In the present instance an alternative was also available. This was to split the data into the initial and new data sets, use the old sets to parameterize the models, and then test the models against the old data sets.

Figure 26 illustrates the type of output for *E. coli* and the preozonation process. Figure 26a shows the accuracy metrics generated by Netica[tm] most notably:

- the 'Confusion Matrix', a matrix of predicted v. actual data given other new data test values;
- The overall error rate;
- A range of scoring metrics.

Although the error rate is substantial and a rate <10% would be clearly preferable, it can be seen that most errors involve a smaller deviation from expectations rather than a large one.

One way to improve model accuracy is to reduce the number of states or ranges for the target node. The improved accuracy is shown Figure 26b. Reducing the number of target node ranges to 2 also allows the calculation by Netica[tm] of further accuracy metrics (Figure 26c) – see Netica[tm] software and (Marcot, 2012) for further details.

When assessing accuracy metrics it is important to have reference values. The negative instance is known as the ZeroR case. This is essentially a BN with no links where state/range likelihood is simply the probability it will be in a predefined range. So for nodes with data in 2 or 3 equal size bins the ZeroR error rates would be *ca* 50% and 67% respectively, far greater than those observed.

Conversely the most accurate test results that could be expected would be those achieved by using the same set of data used to define a BN's bins to test its accuracy. The result of this is Figure 26d where it can be seen that given the underlying variance of all the nodes the best error rate for the LRVPREO node that could expected is about the same i.e. the accuracy estimated from the new data was as good as could be expected.

Table 11 shows the result of a number of such 'test with cases' analyses of the LRV models in Figure 25. It can be seen that generally the predictive powers of the models are respectable but not perfect suggesting there is variance in process performance we have not defined. More reassuringly the overall multi-process *E. coli* and bacteriophage LRVs appeared to be very accurate indicating the model overall still provides a tool which predicts them well. The *C. perfringens* model on the surface looked less accurate. This could be accounted for by the data set used to learn the model to be tested being based on many fewer records than in the case of *E. coli* and Somatic Coliphage (29 v 60).

a.

```
Read 47 cases, and used 47 of them to test net.

For LRVPREO:       LRV Preozone
-----------

Confusion:
  .......Predicted......
  -0.2 t  1.6 to  2.1 to    Actual
  ------  ------  ------    ------
      11       0       2    -0.2 to 1.6
       6       7       4    1.6 to 2.1
       0       0      16    2.1 to 3.4

Testing Real Value:
  Absolute error:      mean = 0.3087       max = 0.938         rms = 0.3972
  Relative error:      mean = 43.42 %      max = 1144 %
  Error / std dev:     mean = 36.69 %      max = 113.6 %
  Distribution within:  0-1 std dev      1-2 std dev     2-3 std dev     >3 std dev
                         97.8 %          2.17 %          0 %             0 %

Error rate = 26.09%

Scoring Rule Results:
  Logarithmic loss = 0.7055
  Quadratic loss   = 0.4005
  Spherical payoff = 0.7725
```

b.

```
Read 47 cases, and used 47 of them to test net.

For LRVPREO:       LRV Preozone
-----------

Confusion:
  ...Predicted..
  -0.2 t  1.9 to    Actual
  ------  ------    ------
      17       3    -0.2 to 1.9
       7      19    1.9 to 3.4

Testing Real Value:
  Absolute error:      mean = 0.3599       max = 1.089         rms = 0.4547
  Relative error:      mean = 48.69 %      max = 1181 %
  Error / std dev:     mean = 40.62 %      max = 118.5 %
  Distribution within:  0-1 std dev      1-2 std dev     2-3 std dev     >3 std dev
                         93.5 %          6.52 %          0 %             0 %

Error rate = 21.74%

Scoring Rule Results:
  Logarithmic loss = 0.4436
  Quadratic loss   = 0.2835
  Spherical payoff = 0.8405
```

c.

```
Gini coeff = 0.7558
Area under ROC = 0.8779
```

d.

```
For LRVPREO:       LRV Preozone
-----------

Confusion:
  .......Predicted......
  -0.2 t  1.6 to  2.1 to    Actual
  ------  ------  ------    ------
      20       0       2    -0.2 to 1.6
       6       5       3    1.6 to 2.1
       2       1      14    2.1 to 3.4

Testing Real Value:
  Absolute error:      mean = 0.4084       max = 1.183         rms = 0.5084
  Error / std dev:     mean = 51.17 %      max = 140.7 %
  Distribution within:  0-1 std dev      1-2 std dev     2-3 std dev     >3 std dev
                         86.8 %          13.2 %          0 %             0 %

Error rate = 26.42%

Scoring Rule Results:
  Logarithmic loss = 0.6363
  Quadratic loss   = 0.3672
  Spherical payoff = 0.7865
```

Figure 26. Illustration of Netica[tm] BN accuracy metrics printout a. for nodes with > 3 states, and b./c. for nodes with 2 states and d. for the same set as used to populate the model.

Table 6. Causal model LRV target node prediction accuracy based on Netica[tm] accuracy metrics following application of learning and the 'Test with Cases' wizard

LRV Target (3 ranges each)	Microbial parameter					
	E. coli		Somatic coliphage		*C. perfringens*	
	Error rate %	Spherical payoff	Error rate %	Spherical payoff	Error rate %	Spherical payoff
LRV Preozone	26	0.77	36	0.72	37	0.70
LRV Biofilt	44	0.69	57	0.65	38	0.63
LRV PostOz	34	0.75	6	0.93	23	0.76
LRV T>BMF	28	0.71	30	0.70	35	0.64
LRV Pre>PoOZ	27	0.77	6	0.93	29	0.72
LRV T > UV	0	0.98	0	0.99	50	0.63
[Target/(range)]	0 (0-67)	1 (0-1)	0 (0-67)	1 (0-1)	0 (0-67)	1 (0-1)

6.3.6. 'Sensitivity to findings' (STF) analysis on microbial concentration data

Varying the probability values of the BN nodes in the causal nets above individually allowed the magnitude of relationships (or lack thereof) between concentration and LRVs to be rapidly assessed. Most notably, variation in the input concentration of all three indicators of 1 to 2 $\log_{10}$ only yielded a 0.2 to 0.3 $\log_{10}$ change in the post biofilter concentration.

Variations in biofilter concentrations were analysed further using the sensitivity analysis method illustrated by Figure 27 and Table 7. Netica[tm] includes a feature, "Sensitivity to findings" using which "you can efficiently determine how much a finding at one node will likely change the beliefs at another." and which ideally should also be used with a 'Sensitivity to Parameters' assessment (Pollino et al., 2007).

Figure 27 shows a semi-naïve BN constructed using Netica's 'Learn TAN Structure' Feature using the filtration concentration as the target variable. Such semi-naïve BNs can be used for explanatory and diagnostic purposes)

In the present instance the %Variance in reduction in Table 7 was probably the most useful statistic. The STF analysis of the *E. coli* biofilter data showed that although microbial concentrations after filtration were correlated with those after preozonation and postozonation they were much less related to the initial concentration. This semi-naïve BN (discussed further below) was better for identifying the most significant relationships as it involved no initial assumptions about how the different parameters were related.

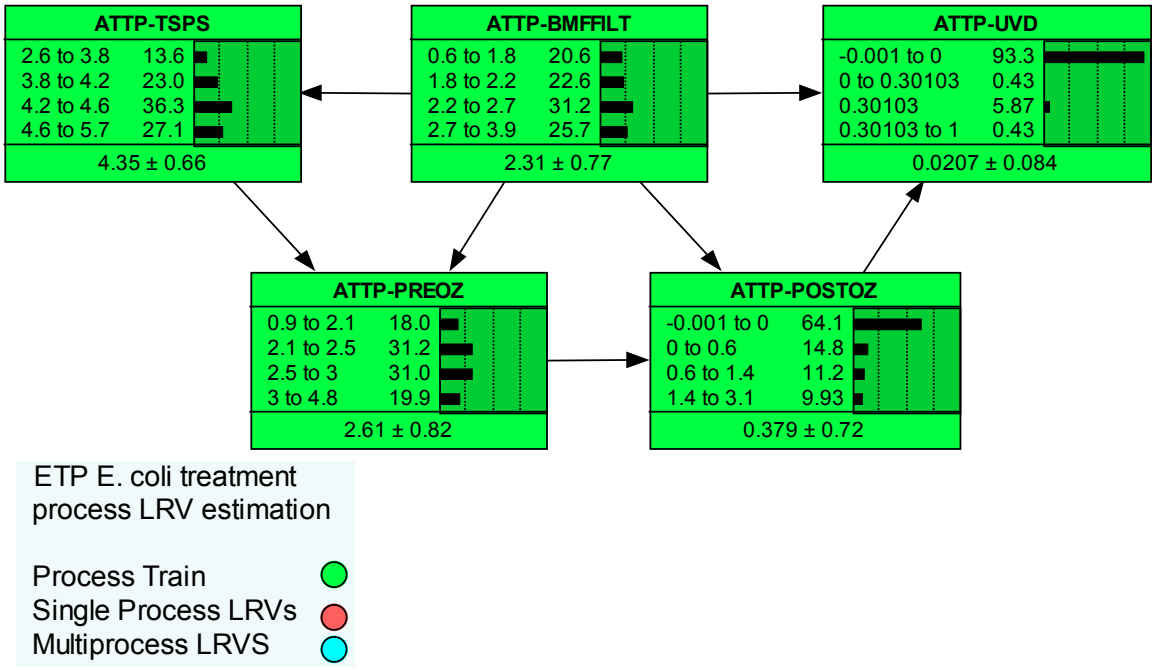


Figure 27. Semi-naive Bayesian Net derived using TAN learning to explain variation in post biological filtration concentrations of *E. coli*

Table 7. Illustration of sensitivity to findings output for 'ATTP-BMFFILT' to a finding at another explanatory node

Node	Variance Reduction	Variance Reduction as Percent	Mutual Information	Mutual Information as Percent	Variance of Beliefs
ATTP-BMFFILT	0.5981	100	1.98194	100	0.5546424
ATTP-PREOZ	0.2404	40.2	0.49122	24.8	0.0676369
ATTP-POSTOZ	0.1377	23	0.27551	13.9	0.0368298
ATTP-TSPS	0.04383	7.33	0.20842	10.5	0.0165302
ATTP-UVD	0.01053	1.76	0.05216	2.63	0.0081074

TAN learning based BN and 'Sensitivity to findings' analysis was used to measure the relationships between the three main post treatment microbial concentrations and other parameters (compare Figure 25, Table 8). The analysis suggested, bearing in mind the treatment order, that:

- Post preozonation microbial concentrations were most closely associated with the post biofilter concentrations and the preozonation LRV;
- Post biofiltration microbial concentrations were most strongly reflected not in the Preozonation and biofiltration processes individually but rather the two combined (LRV_T>BMF), and were reflected in the post postozone concentrations;
- However the Post postozone concentration variations were not measurably related to either upstream process effectiveness or upstream concentrations.

Overall these analyses indicated that although the treatment processes were effective in reducing microbial numbers across the board the variation in individual concentrations seen was

only marginally reflected in variations in concurrently measured concentrations at other stations or by the estimated LRVs.

Table 8. Sensitivity of key post treatment microbial levels to variances in other BN nodes

Possible influence	Microbe	Target Node		
		ATTP-PREOZ	ATTP-BMFFILT	ATTP-POSTOZ
ATTP-TSPS	*E. coli*	1	0	0
	Somatic coliphage	12	3	0
	C. perfringens	13	0	0
ATTP-PREOZ	*E. coli*	100	26	4
	Somatic coliphage	100	20	0
	C. perfringens	100	4	0
ATTP-BMFFILT	*E. coli*	25	100	13
	Somatic coliphage	21	100	3
	C. perfringens	4	100	1
ATTP-POSTOZ	*E. coli*	4	14	100
	Somatic coliphage	1	4	100
	C. perfringens	0	0	100
ATTP-UVD	*E. coli*	0	0	1
	Somatic coliphage	0	0	0
	C. perfringens	0	0	16
LRV_Preozone	*E. coli*	13	4	1
	Somatic coliphage	18	4	0
	C. perfringens	8	0	0
LRV_Biofilt	*E. coli*	11	0	0
	Somatic coliphage	9	2	0
	C. perfringens	6	16	0
LRV_PostOz	*E. coli*	4	13	6
	Somatic coliphage	8	35	1
	C. perfringens	0	13	19
LRV_Pre>PoOZ	*E. coli*	13	3	3
	Somatic coliphage	31	7	0
	C. perfringens	7	0	25
LRV_T>BMF	*E. coli*	3	12	2
	Somatic coliphage	2	18	1
	C. perfringens	0	24	0
LRV_T>UV	*E. coli*	0	0	0
	Somatic coliphage	9	2	0
	C. perfringens	3	0	3

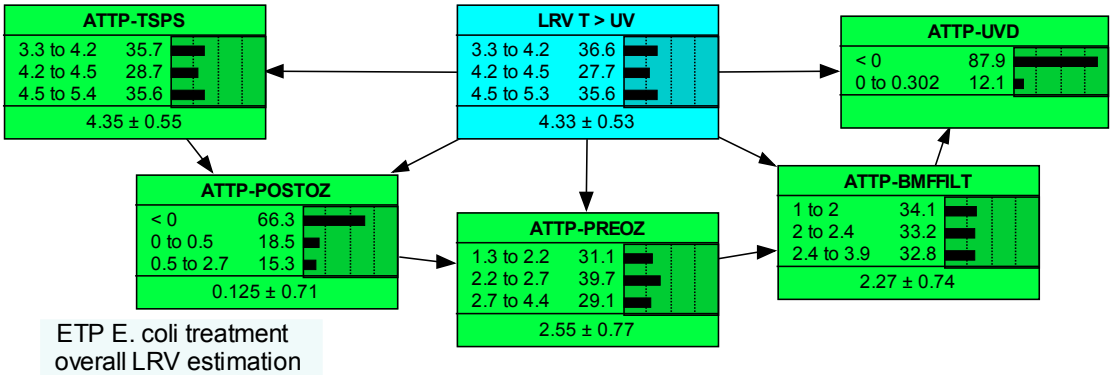
6.3.7. LRV drivers based on Semi-naïve BNs

A second set of sensitivity analyses was performed to better understand how LRVs related to or were influenced by other variables. TAN models were constructed to assess how the variances in the overall LRVs (target variable) were reflected in the different microbial concentrations (Figure 28).

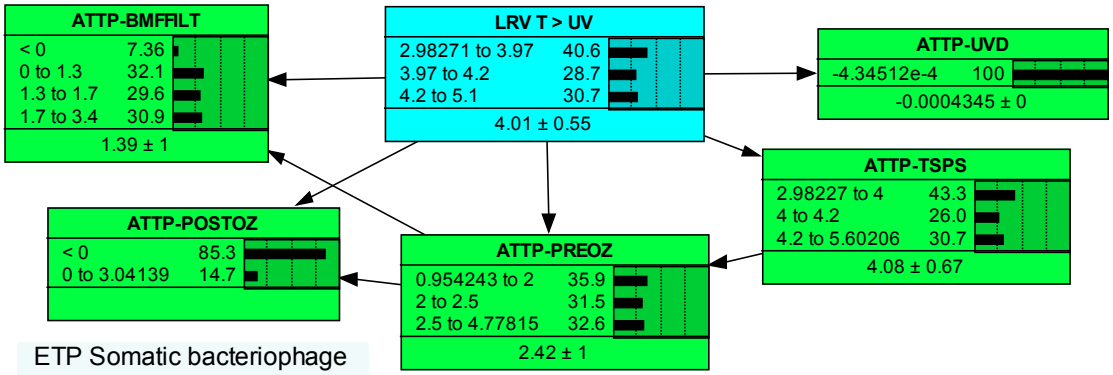
To assess the reliability of the models we first applied again the “test (the old data based BN) with (new data) cases” wizard. The accuracy metric comparison is shown in Table 9 and suggested the *E. coli* and somatic coliphage models were particularly accurate.

This sensitivity analysis undertaken on each microbial data set (Table 10) showed that the LRV estimates were particularly sensitive to the incoming and preozonation concentrations reported. This was manually confirmed by varying the node range probability and the variation was consistent with that previously found with the *E. coli* illustrative model above (Figure 27). The effect was also seen with *C. perfringens* but was less pronounced.

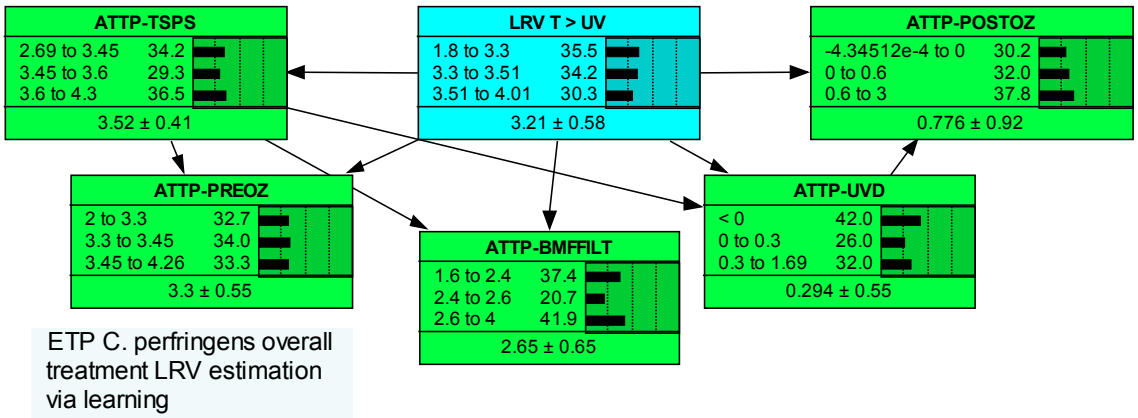
However, examination of the BNs also suggested the perceived importance of the TSPS samples may have been due to the increasing data censorship of the post-ozonation and post UV process data. As the post UV log concentrations were generally <0 it appeared that the net had created a model where the TSPS concentrations and LRV T(SPS)>UV binning and ranges were very similar. So in this instance we conclude that despite the rich data sets our efforts to identify the source and magnitude of LRV variance though theoretically sound were compromised by data censorship.



a.



b.



c.

Figure 28. Semi-naïve (TAN) models constructed to identify the more important parameters influencing overall microbial reduction

Table 9. Overall LRV TAN Model Accuracy metrics

Test with cases measures	Target/(Range of possible values)	Microbial parameter		
		E. coli	Somatic coliphage	*C. perfringens*
Error rate %	0 (0-67)	0	4	41
Spherical payoff	1 (0-1)	0.99	0.96	0.63
No. of test measurements		45	45	44

Table 10. Sensitivity as % variance reduction for overall process derived from TAN semiBayes Net for overall treatment (LRV T > UV)

Target Node	Range of possible values	Microbial parameter		
		E. coli	Somatic coliphage	*C. perfringens*
ATTP_TSPS	(0-100)	70	66	20
ATTP_PREOZ	(0-100)	6	14	6
ATTP_BMFFILT	(0-100)	4	5	6
ATTP_POSTOZ	(0-100)	2	0.01	0.6
ATTP_UVD	(0-100)	0.07	0	0.03

6.3.8. Model for LRV Bayesian validation

In another case study of the SA Water Glenelg ultrafiltration system, we used a combination of manufacturer and initial validation data to compare and contrast a new set of 'revalidation' data and determine whether the older LRVs had been revalidated. To do this we undertook a process we termed Bayesian Validation where the node probabilities based on the two different prior data sources were combined to generate posterior best estimates of treatment LRVs.

In the case of the ETP no manufacturer data were available. However, the old and new data mimicked how performance might be revalidated across the years i.e. a BN could be constructed to estimate treatment LRVs whose credit might then be claimed as part of licensing conditional on performance being demonstrated to be maintained with the latter being achieved by the follow-up monitoring data collection.

Figure 29 shows how such 'revalidation' might be undertaken in practice using somatic coliphage to illustrate the method:

- The initial data sets are used to generate by learning the main net and set estimate the concentration and LRV statistics (green and red nodes respectively).
- The model is then cloned/copied and the new 'validation' testing based model (light brown) is produced.
- The LRV estimates are then combined in relatively proportions based on regulatory requirements, expert opinion or some other rational basis. In the Figure 29 the relative importance assigned to the initial (design) model is 25% and the revised validation model is assigned 75%.
- Finally the two sets of analogous LRVs are combined into the updated violet LRV credit nodes. It can be seen that very little change has occurred and the treatment system appears to be very stable.
- The conclusion here would be that the system has been revalidated.

We found this result particularly noteworthy as the two microbial data sets were independently collected *ca* 1 year apart and the models are very similar despite the notorious variability seen in microbial concentrations in water samples.

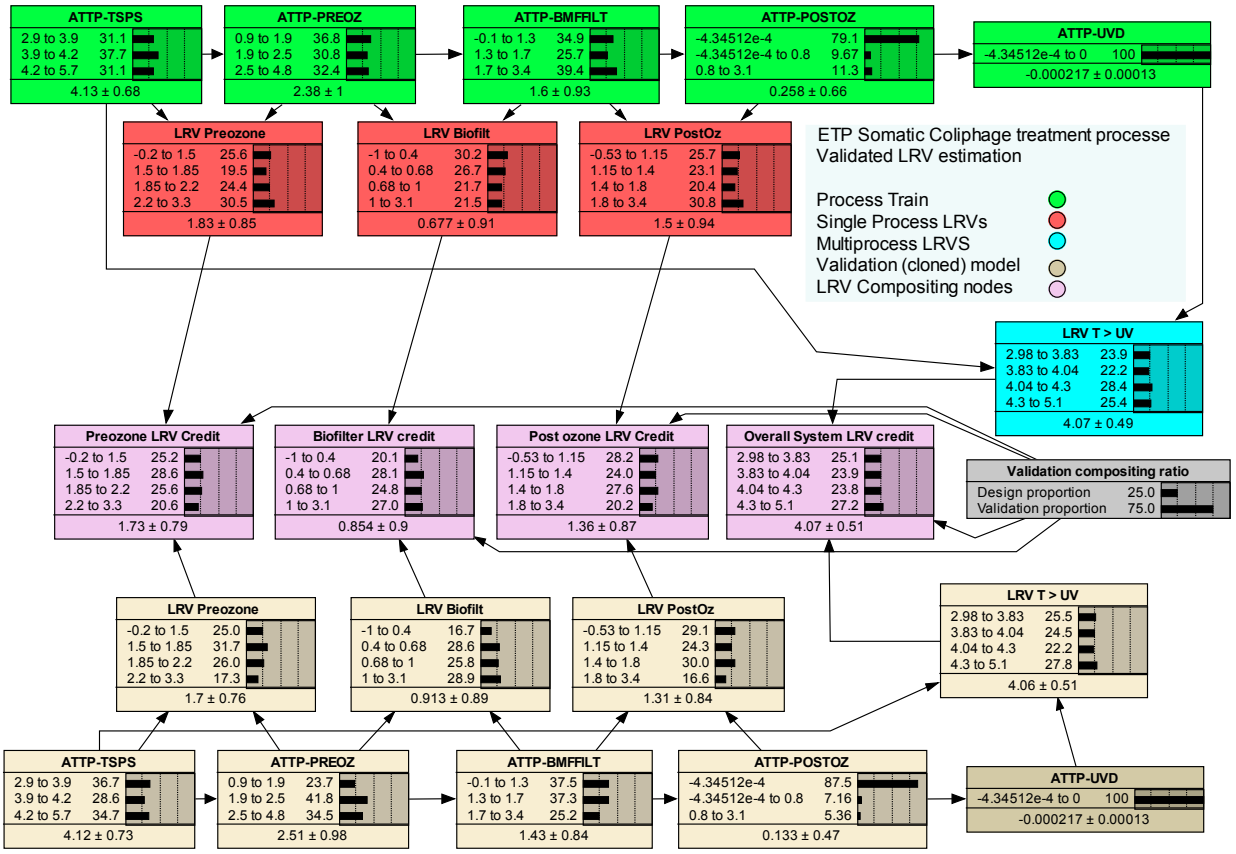


Figure 29. Illustrative LRV model Bayesian Validation for somatic coliphage using old and new data sets

6.3.9. On-line data

Like most authorities Melbourne Water collects large quantities of on-line monitoring data. The general question inevitably arose as to whether the surrogates monitored could be related to the microbial and other physicochemical parameters and possibly used to improve or fine tune treatment performance which could in turn be recognised in LRV credits.

As a test we assessed:

- The relationships of the on-line parameters to one another; and
- The relationships between the process LRVs and these parameters.

MW data included high frequency surrogate measurements of particulates (turbidity), organic matter (UV transmission), ozone and ammonia as well as standard physicochemical parameters (temperature, pH) and flow.

Melbourne Water were particularly interested in LRV credit assignment for preozone, biofiltration and postozone (potentially leading to variance in disinfection) and organic matter reduction indicating active degradation of organic matter including microorganisms, and turbidity changes indicating particle removal.

The data analyses proceeded as anticipated and some notable relationships were identified.

6.3.10. On-line parameter inter-relationships

The first assessment looked at on-line parameters alone. Seven candidate target node parameters were selected and used along with the remainder of the data to construct TAN BNs. The BNs then learned from *ca* 680 approximately daily records between 2013 and 2015.

These were mostly downstream measurements as we were looking for parameters which had undergone changes which might be related to microbial numbers.

The number of on-line records available was much higher than 680 they were collected at 10 minutes. However the latter were selected because these could be related to grab microbial and physicochemical measurements. The total of 680 was also judged to be sufficient for the detection of useful relationships based on previous experience with BNs.

Table 11 shows the higher Sensitivity of findings variance reductions identified. As expected, Turbidity in different streams was matched. Otherwise turbidity was relatively unaffected by these other on-line factors.

Temperature unsurprisingly varied with season and moderately with flow (reason unclear) and turbidity.

Postozone UVT appeared unrelated to input UVT suggesting some transformation. Plausibly this could have been ozone oxidation as indicated by a possible relationship to preozonation dosage.

Table 11. Sensitivities of main TAN Node as variance reduction by different diagnostic parameter candidates

Station_Parameter Code	Target nodes						
	Season	TSPS Turbidity	Preozone mg/L	PostOzone temperature	Postozone UVT	Postozone NH3	BMFB21 Turbidity
	% variance reduction in target node by secondary parameter						
ol_TSPS_Turbidity	**5.3**	**100**	<5	**9.7**	<5	<5	**18**
ol_TSPS_pH	<5	<5	<5	<5	<5	<5	**5**
ol_TSPS_UVT	<5	<5	**30**	<5	**12**	<5	<5
ol_TSPS_NH3	**6.5**	**9.3**	<5	**7.6**	<5	**9.7**	**10**
ol_Preozone_mgpL	**6.4**	<5	**100**	<5	<5	<5	<5
ol_Preozone_kLps	<5	**6.1**	<5	**20**	<5	<5	5.2
ol_Postozone_UVT	<5	<5	**6**	<5	**100**	<5	<5
ol_Postozone_pH	<5	<5	<5	<5	<5	<5	<5
ol_Postozone_NH3	<5	<5	<5	<5	<5	**100**	<5
ol_Postozone_oC	**23.6**	**6.9**	<5	**100**	<5	<5	<5
ol_BMFB21_Turbidity	<5	**14**	<5	<5	<5	<5	**100**
ol_BMFB22_Turbidity	<5	**13**	<5	**6.2**	<5	<5	**44**
ol_BMFB23_Turbidity	<5	**12**	<5	**6.4**	<5	<5	**40**
ol_BMFB24_Turbidity	5.5	**12**	<5	**6.9**	<5	<5	**36**
Season	**100**	**6.3**	**11**	**56**	<5	<5	<5

6.3.11. WEKA screening– Microbial LRVS v. On-line measurements

In the second assessment we used WEKA to search for relationships between each microbial LRV and all the physicochemical grab and on-line data sets.

To use WEKA it was necessary to place the LRVs into categories. For this we split the data according to whether or not each value was above or below the average LRV for that set. So the accuracy rate of the ZeroR case should have been *ca* 50%.

We then compared the actual ZeroR with the models arising from using WEKA's TAN and a simple BAN classified which used the BayesNet Simple estimator, the Hillclimber search algorithm and a maximum of 2 parent nodes per node. The procedure was repeated 5 times using different seed random numbers. This generated a range of accuracy metric data. The % accuracy metrics are tabulated in Table 12.

The Hillclimber search did not yield any better accuracy metrics than the WEKA TAN classifier. Disappointingly only four LRV parameters seemed to differ from the ZeroR estimates. Three of these were for post ozone treatment and a different 3 were for *E. coli.*

The most notable possible relationship was for the *E. coli* LRV calculated for the preozonation treatment. This net was refined by removing all variables which showed no variance in response to LRV variance. The resulting BN is shown in Figure 30.

This assessment indicated the Preozonation process might be influencing microbial concentrations and hence LRVs. But the on-line measurements added little to the earlier assessment which was based simply on our knowledge of the structure of the treatment system.

Table 12. Results of WEKA based data mining applied to microbial (target) variables and concurrently collected on-line (explanatory) variables

Target Variable	**WEKA classification model**					**Comments**
	Primary explanatory variable set			**Refined explanatory variable set**		
	Zero R	**TAN**	**'Hillclimber'-BAN with 2 parents**	**ZeroR**	**TAN**	
LRV_Ec_Preozone	53	68-75	60-70	53	68-70	Only plausible relationship
LRV_Phg_Preozone	57	49-53	nd	nd	nd	No difference to ZeroR
LRV_Cp_PreOzone	57	49-53	nd	nd	nd	No difference to ZeroR
LRV_Ec_Biofilt	51	59-65	59-65	nd	nd	Only two parameters UVT on-line –post ozone and TSPS
LRV_Phg_Biofilt	52	49-55	nd	nd	nd	No difference to ZeroR
LRV_Cp_Biofilt	50	47-53	nd	nd	nd	
LRV_Cp_PostOz	52	47-57	nd	nd	nd	
LRV_Ec_PostOz	54	53-54	53-54	nd	nd	
LRV_Phg_PostOz	56	64-68	64-68	nd	nd	Nitrite NOM and TSPS turbidity only
LRV_Cp_PreOz_BLT	50	57-63	nd	nd	nd	No difference to ZeroR
LRV_Ec_PreOz_BLT	54	50-54	nd	nd	nd	
LRV_Phg_PreOz_BLT	57	50-54	nd	nd	nd	

nd = not determined

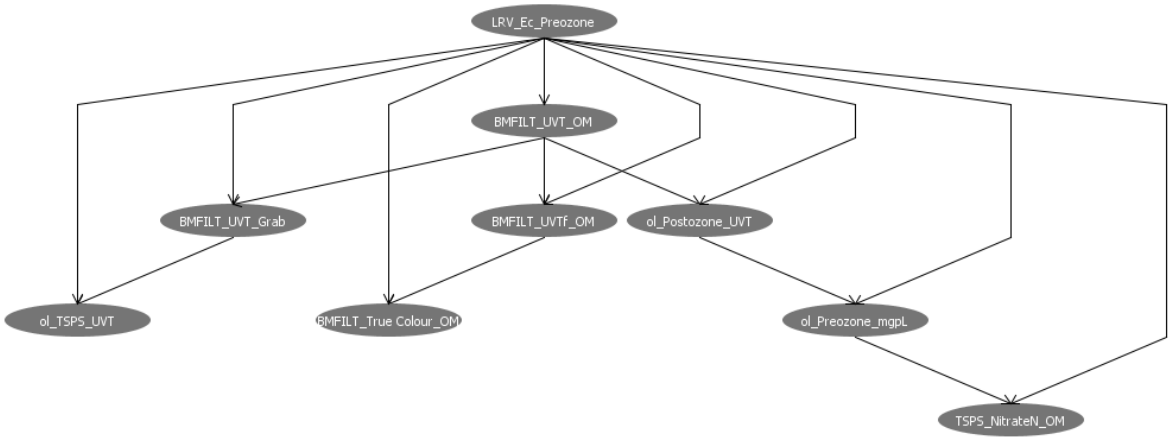


Figure 30. WEKA constructed TAN network relating *E. coli* preozone treatment LRV to on line parameters (see Table 12)

6.3.12. Supplementing and testing the initial 2013/2014 BNs

In our initial analysis of the ETP data we suggested additional data could be collected for the following purposes:

- Comparison with the primary BNs using the "test with cases" function to assess whether new data were consistent with older data or better and hence whether the system was still functioning.
- Update and improve the precision of the initial models.
- Assess the impact of proposed and implemented treatment design changes.
- Assessing model predictive capabilities.

These actions were implemented. The test with cases function appeared to provide useful accuracy metric data. We were able to update the models with larger data sets (Figure 22, Figure 23 and Figure 24) which indicated the models based on *ca* 29 and 60 records were relatively stable.

6.4. Discussion

6.4.1. Achievement of work aims

Translating the ETP data into a form suitable for 'learning' by the BN software was rapid, simple and straightforward. More time consuming was identifying appropriate LRV discretization thresholds (i.e. the boundaries between each LRV or concentration class within a node e.g. "4.6 to 5.6"). Where data is available or the BN is learnt, Netica's, 'auto-discretization' tool can be used. However, where the LRVs were calculated using equations/algorithms this process was manual. In the examples shown the aim was to quantify the likelihood of negative removal estimates compared to positive removal estimates and provide and near even spread of probabilities across the node ranges. Such data was provided by the node summary statistics.

BNs were developed for the data overall and for each microorganism. The development process proved simple, employing normal BN construction techniques based on Help file advice provided by the Netica[tm] software. The existing data and the data sets with *ca* 110 records per analyte/microbe combination appeared of sufficient size based on preliminary examination of the contingency table probabilities (good spread without missing values). Half this size was also probably sufficient and the small 29 record initial C. perfringens data set also proved large enough to generate a model that was essentially repeated when more data became available. More systematic techniques for defining LRVs are described elsewhere which are part of data mining best practice (Carvajal et al., 2015, Witten et al., 2011). But empirically the *ca* 30 measurements per station appeared sufficient to construct BN models close to the final versions. The main limitation on BN use was that the high rate of data censorship prevented the full LRV post UV treatment from being validated/verified.

The BNs facilitated the exploration of how LRVs downstream and upstream in the process train varied with one another. It was evident that input concentrations were mostly not very highly correlated or matched to downstream ones. The possible exception was the input v. post biofiltration concentration. The BN suggested that the concentrations after post-ozonation or biofilter treatment were largely uncorrelated with the TSPS levels.

6.4.2. Relevance to validation

The section 'Model for LRV Bayesian validation' above illustrates how new/validation data can be used to infer a final composite log reduction value complete with descriptive statistics. Further this section illustrates again how the LRVs can be based on one or more prior LRV estimate in addition to any data sets collected for validation or revalidation. And the relative emphasis placed on each data set could be varied interactively in the models, potentially in a workshop or proponent/regulator discussion situation.

In addition to these Bayesian validation activities, system operators may have other treatment beliefs they wish 'validated' with a view to obtain or supporting log credits for example:

- They may wish to justify a novel/unconventional/modified treatment process be used to obtain log credits and obtain data justifying these credits.
- They may wish to treat a multi-process system as a single unit for LRVs purposes.
- They may want to claim extra credits on the basis that they can maintain a particular set of operating conditions and these conditions imply a certain microbial log reduction, and need to prove these conditions are achievable and are reflected in LRVs.
- They may want to claim or include LRV credits for processes where no pathogens can be measured because of the effectiveness of previous processes.
- They may wish to claim a system continues to maintain conditions conducive to high log credit maintenance on the basis of on-line measurements being stable/relatively unchanged over several years.

The question arises how to undertake such 'validation' which may occur outside of the normal initial treatment process validation/commissioning stage?

Data mining and BN model construction, especially the learning of models from available data appears to offer solutions to all of the above.

In respect to novel treatments we have illustrated how the LRV which might be plausibly estimated and claimed for the preozonation + biological filtration and how its stability/variance could be assessed i.e. 'test with cases', along with what factors might induce variance e.g.

upstream process variance. This approach can also be applied to large on-line data sets to see if initial performance data of any kind are stable in the long term. As interesting would be to identify circumstances and meter readings where process malfunctions were taking place. The latter hazardous events could be added to BN models as distinct event nodes and used to develop rules for identifying when an even way likely or not.

We have illustrated two methods by which multi-process LRVs can be estimated with the record/case 'learning' providing data with less variance than the Monte Carlo approach achieved.

We have shown how operating parameters might be related to microbial concentrations or LRVs. Where there are no pathogens or indicators, BNs provide and auditable means of including LRV estimates based on expert opinion or tracer studies.

A final task required but not addressed in this study is the question of what assumptions can be made about actual pathogen behaviour from indicators. What is needed is the development of assumptions, expert opinion or options of how far to credit indicator reductions as equivalent or comparable to actual pathogen reductions. BNs can provide a means of incorporating belief and expert knowledge on a case by case basis. A possible model for this is studies like our work on the relationships of indicators to protozoans (Carvajal et al., 2015) using the data of (Flapper et al., 2012). It can be seen from the former study that BNs provide an option for estimating inferable LRV credits for activated sludge based on the relationships and defining the conditions under which they are believed to be applicable.

6.4.3. Censored data and improving LRVs

The precision of relationships identified can be refined/made more precise by application of other readily available software such as Palisade @Risk. For example where LRVs are derived from censored concentration data better estimates may be obtained following PDF fitting of concentration data where microorganisms have been detected.

These improved LRVs could then be incorporated into hybrid BNs (combination of expert knowledge based LRVs in effect and measured LRVs) to provide a best estimate of treatment extent. For 'improved' BNs it would then be possible to replace representative censored values with plausible estimates from the PDF models.

7. Melbourne Water ETP 2: Experimental field study

7.1. Introduction

The soundness of Bayes methods for providing useful LRV estimates and concisely integrating treatment data was demonstrated by the analysis of the historical ETP data presented in Chapter 6. However, the data and its analysis had several limitations:

- Microbial data was only collected at weekly intervals around 7 am. So it was unclear how much variance there might be between different days and over the course of individual days.
- It was unclear how well samples were matched between different treatment stages.
- Ozone generates bromate from bromide but the extent was unclear and this would impact on whether preozonation should be viewed as a valid/acceptable disinfection treatment in some circumstances.
- There were a range of other chemical contaminants especially trace organics which might be reduced or enhanced by the treatment processes especially ozonation but the extent of this was unclear.
- Several physicochemical parameters were unavailable, notably measures of organic carbon.
- Given its potential it was seen as important to confirm which/whether or not online measurements provided useful surrogate monitoring data on process effectiveness.

It was seen as desirable to obtain *de novo* a genuine model validation data set and integrate this with the historical data via 'Bayesian Validation'.

Finally we were interested in exploring the use and benefits of a more sophisticated BN Software that offered more data mining and model validation options than Netica using a realistic data set developed for validation.

Thus an experimental program was designed to address these validation issues and uncertainties. The work and data analyses were mainly undertaken by Guido Carvajal, James McDonald Santhosh Ramesh.

7.1.1. Aims

We developed and implemented a model validation monitoring plan. Specific aims were as follows:

- Collect new microbial concentration and removal data concurrently with:
    - A full range of physicochemical parameters;
    - Trace organics data;
    - Bromate and bromide data;
    - Miscellaneous in line data.
- Estimate microbial LRVs and compare with previous estimates.

- Gather information on whether preozonation and biofiltration should be considered as two individual processes or as a single unit process.
- Address the uncertainties listed above.
- Assess BayesiaLab AgenaRisk software.

7.2. Methods

7.2.1. System

Three major sampling points are referred as SP1, SP2, and SP3 (Figure 31)

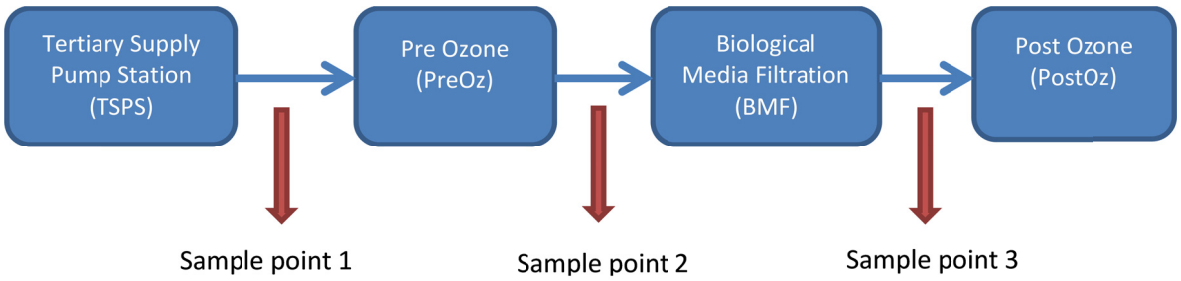


Figure 31. Front end of Eastern Treatment Plant water recycling system showing the locations of sampling points SP1, SP2 and SP3

7.2.2. Sampling program details and analytes

Collection of samples was performed over a period of 3 days from Tuesday 15th September to Thursday 17th September. The targeted processes were preozonation and biological media filtration. Figure 31 shows the selected sample points diagrammatically.

Samples were first analysed for microbial indicators (*E. coli* and *C. perfringens*) and then for organic matter (COD and TOC), bromide/bromate, nitrate, iron, manganese and trace organics detectable by LC-MS/MS. Other general parameters recorded included pH, conductivity, turbidity and temperature. The trace chemicals of particular concern were NDMA and bromate since these are known to be formed during ozonation processes.

Approximately 24 samples were collected each day over 3 days along with blanks and replicates (n=3). It was initially proposed to undertake sampling across the full 24 hours. However Occupational Health and Safety considerations prevented sampling outside of normal work hours so sample collection was limited to 5am to 5pm on each of the 3 days. As this included low plant inflow (5am) and morning peak flow (11 am) this was viewed as sufficient in the first instance.

Some samples needed to be processed onsite. These included extraction and preservation for trace chemicals and nitrite. General parameters and nitrite were measured using portable equipment. Trace organics analysis included personal care products, pharmaceuticals and estrogenic and androgenic hormones. Sample analysis was undertaken at the following locations:

- SA Water: TOC, *E.coli*, Total coliforms, SRCs *C. perfringens*
- UNSW: DOC, COD, dCOD, nitrate, total iron, total manganese, trace organics
- ALS (Melbourne): TSS, UVA
- NMI: bromide, bromate

- On-site: pH, temperature, DO (after ozonation only), conductivity, nitrite.

Samples analysed by SA Water were delivered overnight to Adelaide. Samples to ALS were delivered on the day of collection. UNSW and NMI samples were stored in eskies with ice and transported to Sydney at the conclusion of the sampling campaign.

7.2.3. Online data

ETP collects online data including applied ozone dose, transferred ozone, residual ozone concentration (only post ozone), running times (Biological media filtration), pH, turbidity, UVA, ammonia and flowrates during the sampling period.

7.3. Results and Discussion

7.3.1. Experimental outcomes

The sampling campaign proceeded nominally with no significant disruption. The microbial reductions achieved by the combination of preozonation and biofiltration were essentially the same as seen with the long term data. Feedwater and post-biofiltration microbial and chemical concentrations were remarkably consistent or constant over the three days of sampling (Table 13). The data were consistent with that obtained in routine monitoring program especially in respect to microbial reductions. In line with previous observations of the historical data set preozonation was somewhat variable and biofiltration samples showed much less variance than their preozonation sample equivalents.

E. coli were reduced by on average 2 log units across the two treatment systems. *C. perfringens* were reduced by 1.5 log units. Total coliforms were reduced by only 1 log unit on average suggesting possible growth in the filters.

Cumulative probability density function plots showed that:

- Two step reductions of *E. coli* were remarkably consistent but alone preozonation effectiveness ranged from 0.5 to 4 logs despite the small size of the data set.
- Spore reductions by preozonation were consistent but very small, *ca* 0.2-0.4 log units, with most reduction being achieved by

Bromide was present in a very narrow range. Conversion to bromate was in the order of 3 to 13%. Various trace organics were detected. Their levels and significance will be the subject of further supplementary report. Organic carbon was reduced marginally (*ca* 20%).

Turbidity was reduced markedly (80%) though suspended solids were reduced to a lesser degree suggesting the biological filter was removing finer particles of microbial size consistent with reductions in *C. perfringens* and SRC which were not inactivated by preozonation to a significant extent. Colour was reduced by preozonation and also by the biofilter. UV absorbance (organic matter) was reduced by preozonation.

Ozone dosing was relatively constant at an average of 9.1 mg/L.

Figure 32 illustrates microbial reductions. The combination of preozonation plus biofiltration appeared to be more consistent that either process alone as identified by the BN analysis of the historic data. And the extent as noted above of 2 and 1 logs for *E. coli* and clostridial spores was also the same.

Figure 33 shows changes in bromine levels. As expected bromate was formed by the preozonation but it was not at a level of concern for discharge. Conversion of bromide to bromate was largely incomplete and did vary somewhat on the different days for reasons which are unclear.

Table 13. Summary statistics for primary analytes

Parameter	Treatment Stage (sampling point)	Units	Average	Median	5th percentile	95th percentile
Total coliforms	TSPS (SP1)	mpn/ 100mL	4.8	4.8	4.4	5.0
	Post Preozonation (SP2)		3.1	3.0	1.9	5.3
	Post Biofiltration (SP3)		3.5	3.4	3.2	3.7
E. coli	TSPS (SP1)	mpn/ 100mL	3.7	3.8	3.3	3.9
	Post Preozonation (SP2)		1.6	1.5	0.3	3.0
	Post Biofiltration (SP3)		1.7	1.6	1.5	1.9
Sulphite reducing clostridia	TSPS (SP1)	cfu/ 100 mL	3.6	3.6	3.3	3.8
	Post Preozonation (SP2)		3.2	3.3	2.6	3.7
	Post Biofiltration (SP3)		1.7	1.7	1.2	2.3
C. perfringens	TSPS (SP1)	cfu/ 100 mL	3.2	3.4	2.0	3.7
	Post Preozonation (SP2)		3.0	3.1	2.3	3.5
	Post Biofiltration (SP3)		1.5	1.4	0.9	2.1
Bromide	TSPS (SP1)	mg/L	0.005	0.002	0.001	0.019
	Post Preozonation (SP2)		0.318	0.320	0.300	0.330
	Post Biofiltration (SP3)		0.318	0.320	0.300	0.340
Bromate	TSPS (SP1)	mg/L	0.319	0.320	0.308	0.330
	Post Preozonation (SP2)		0.020	0.017	0.010	0.038
	Post Biofiltration (SP3)		0.014	0.011	0.008	0.028
TOC	TSPS (SP1)	mg/L	14.7	14.9	12.2	16.8
	Post Preozonation (SP2)		13.6	13.6	12.2	14.8
	Post Biofiltration (SP3)		10.7	11.1	8.8	12.0
COD	TSPS (SP1)	mg/L	65.6	52.5	40.3	120.7
	Post Preozonation (SP2)		58.1	45.5	32.5	116.1
	Post Biofiltration (SP3)		48.6	36.5	13.3	97.7
Turbidity	TSPS (SP1)	NTU	6.7	4.3	2.2	12.6
	Post Preozonation (SP2)		6.3	3.6	1.8	13.9
	Post Biofiltration (SP3)		0.9	0.8	0.6	1.4
TSS	TSPS (SP1)	mg/L	8.1	5.0	5.0	19.8
	Post Preozonation (SP2)		10.5	5.0	5.0	21.0
	Post Biofiltration (SP3)		5.0	5.0	5.0	5.0
Apparent Colour	TSPS (SP1)	PtCo units	139.3	134.0	110.5	176.1
	Post Preozonation (SP2)		66.5	50.5	37.0	116.4
	Post Biofiltration (SP3)		22.0	21.5	14.2	32.4
UV254nm	TSPS (SP1)	OD Units	0.4	0.4	0.3	0.4
	Post Preozonation (SP2)		0.2	0.2	0.2	0.2
	Post Biofiltration (SP3)		0.2	0.2	0.1	0.2
NH_4^+N		mg/L	0.3	0.3	0.3	0.5
NO_2^-N		mg/L	0.090	0.092	0.051	0.127
NO_3^-N		mg/L	5.3	5.0	3.8	7.5
Temperature		^oC	16.8	17.4	13.6	19.0
pH		pH units	6.8	6.7	6.6	7.2
O_3 Dose		mg/L	9.1	9.2	6.9	10.6
Flow		Cumecs?	4.6	5.0	3.2	5.9

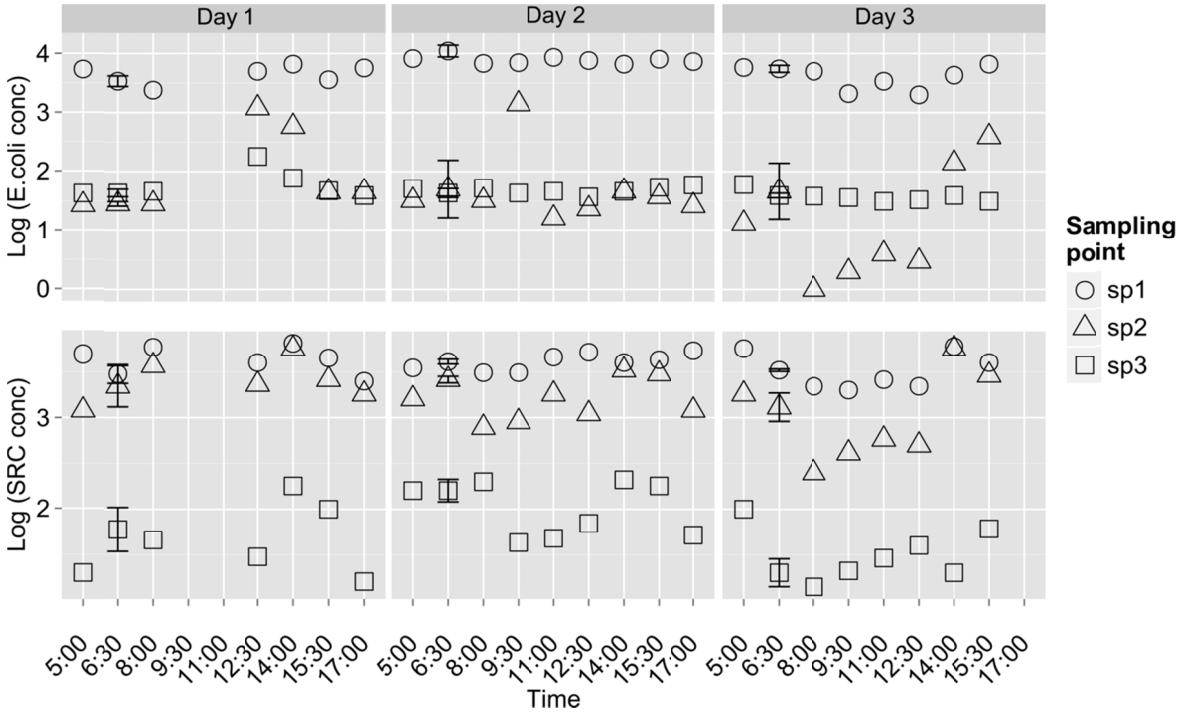


Figure 32. *E. coli* and SRC counts across three days

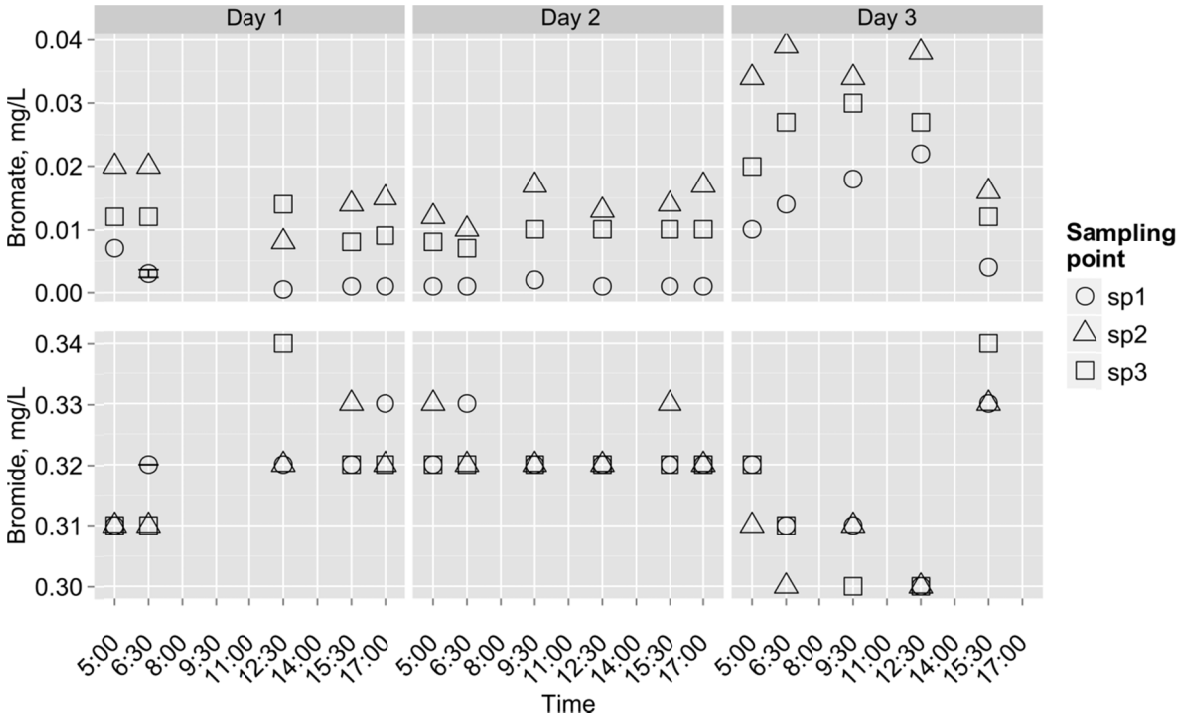


Figure 33. Bromide and bromate concentrations over the sampling period

7.3.2. PDFs

Probability density functions for microbial concentrations and LRVs are shown in Figure 34 and Figure 35, respectively. Figure 34 shows how reduction in *E. coli* concentrations across the two

processes was marked and consistent. However, preozonation reductions range over potentially 3 or more orders of magnitude. We speculated this may reflect incomplete mixing and samples being taken from streams in the process train which were less well mixed. Preozonation reduction of Clostridia was marginal at best, but reductions by the biofilter were relatively high.

Figure 35 shows the LRVs for the same data based on matched before-and-after process measurements. In the case of *E. coli* the two processes in combination produced strikingly consistent removal whereas preozonation alone was confirmed as achieving reductions as high at 4 LRV and as little as nothing. In the case of spores the best reduction achieved by preozonation was 1 LRV. However the two processes combined could achieve up to 2.5 LRV at times. Together these plots suggest:

- The two processes together can achieve as much as 4 and 2.5 LRV together.
- They are not, as yet, optimised.

All these findings were fully consistent with the provisional conclusions drawn from BN analysis of the historical data. They further supported the idea that log credits could be claimed for preozonation plus biofiltration and that the level claimable might be increased with process optimisation.

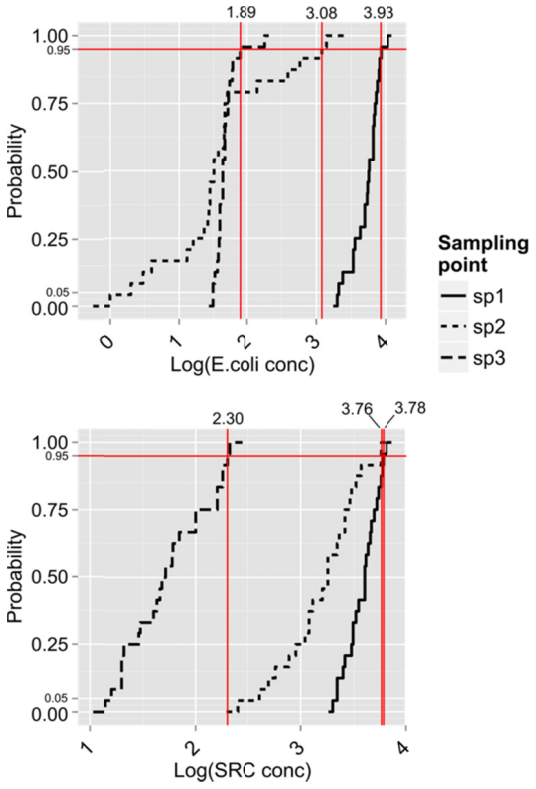


Figure 34. Cumulative probability density functions showing the 95th percentile *E. coli* and SRC concentrations at the three sampling points

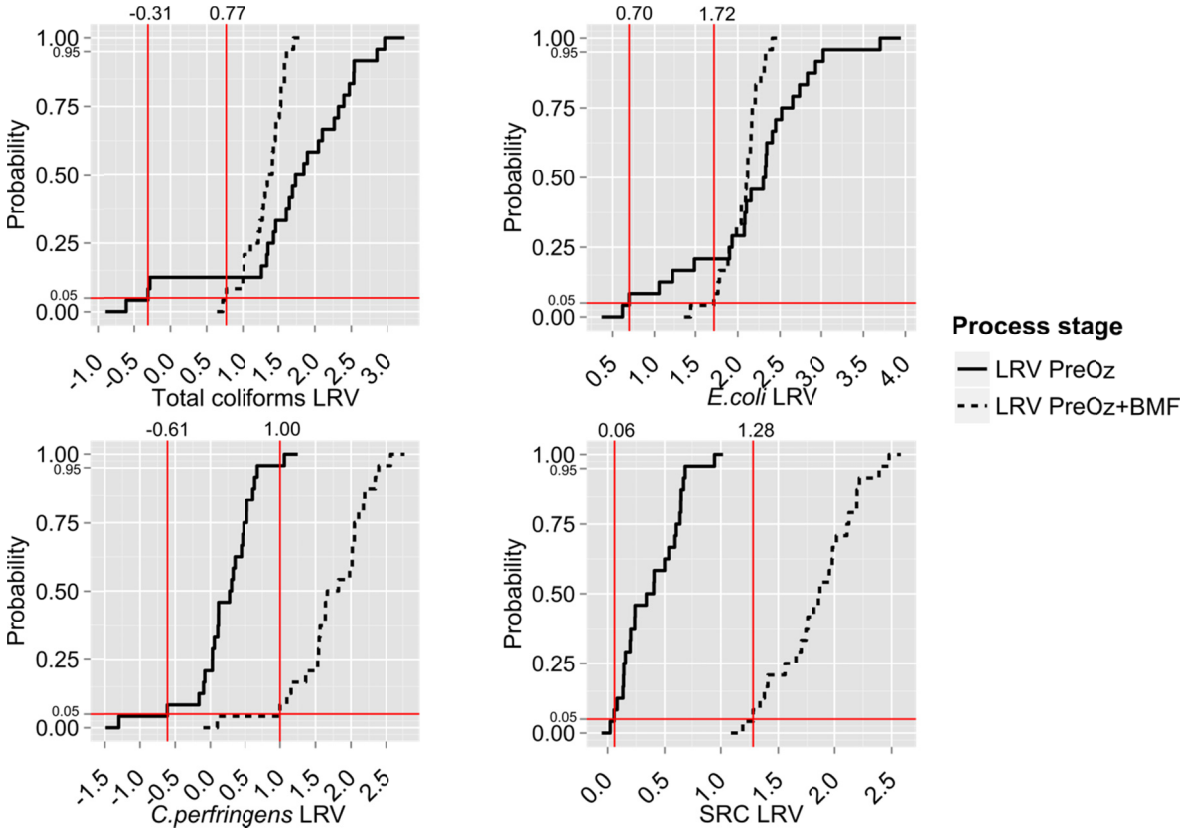


Figure 35. Cumulative probability density functions showing the 5th percentile Total coliform, *E. coli,* SRC and *C. perfringens* LRVs for preozonation and preozonation + biological filtration

7.3.3. Further BN-based analysis

The data sets obtained from Melbourne Water were used to also explore the use of more advanced software AgenaRisk.

Figure 36 illustrates how this software can also be used to generate correlations reflecting naïve BN relationships. This figure shows how *E. coli* concentrations may have been related to organic carbon and temperature, two factors not previously evaluated, but inversely related to bromate (in effect the more bromate the lower the microbial concentration which would be expected as both are expected to respond to ozone dosing).

Figure 37 illustrates that the simple naïve BN shown has a high prediction accuracy and therefore the relationships being detected are real. Figure 38 and Figure 39 show a second BN analysis focused on bromate concentrations following preozonation. Bromate formation appears linked strongly and positively to pH and turbidity and negatively to temperature, organic matter and ozone dose. Figure 38 compared to Figure 39 illustrates the similarity of the results obtain via BN sensitivity analysis (mutual information metrics) v. conventional correlation analysis respectively. Finally, Figure 40 illustrates the availability of sophisticated LRV estimation tools within AgenaRisk.

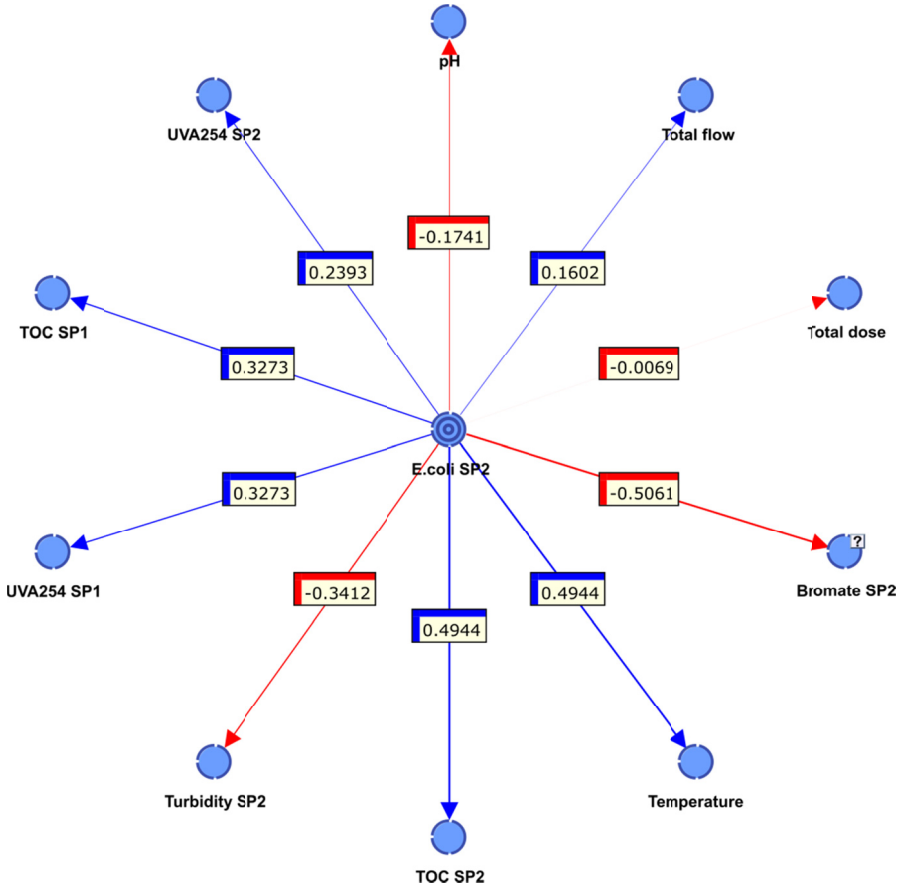


Figure 36. Pearson's R correlations showing forward (blue) and backward (red) (others are possible)

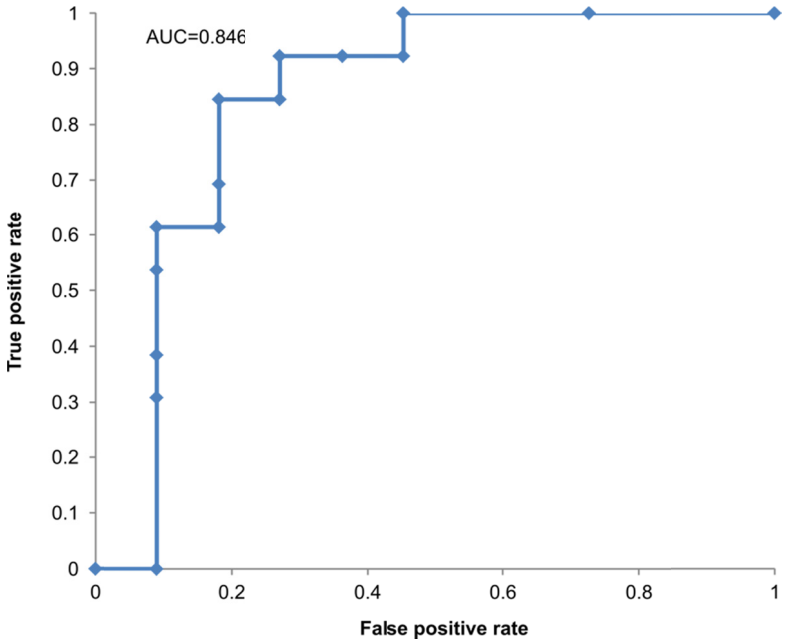


Figure 37. ROC curve for naïve Bayes model prediction

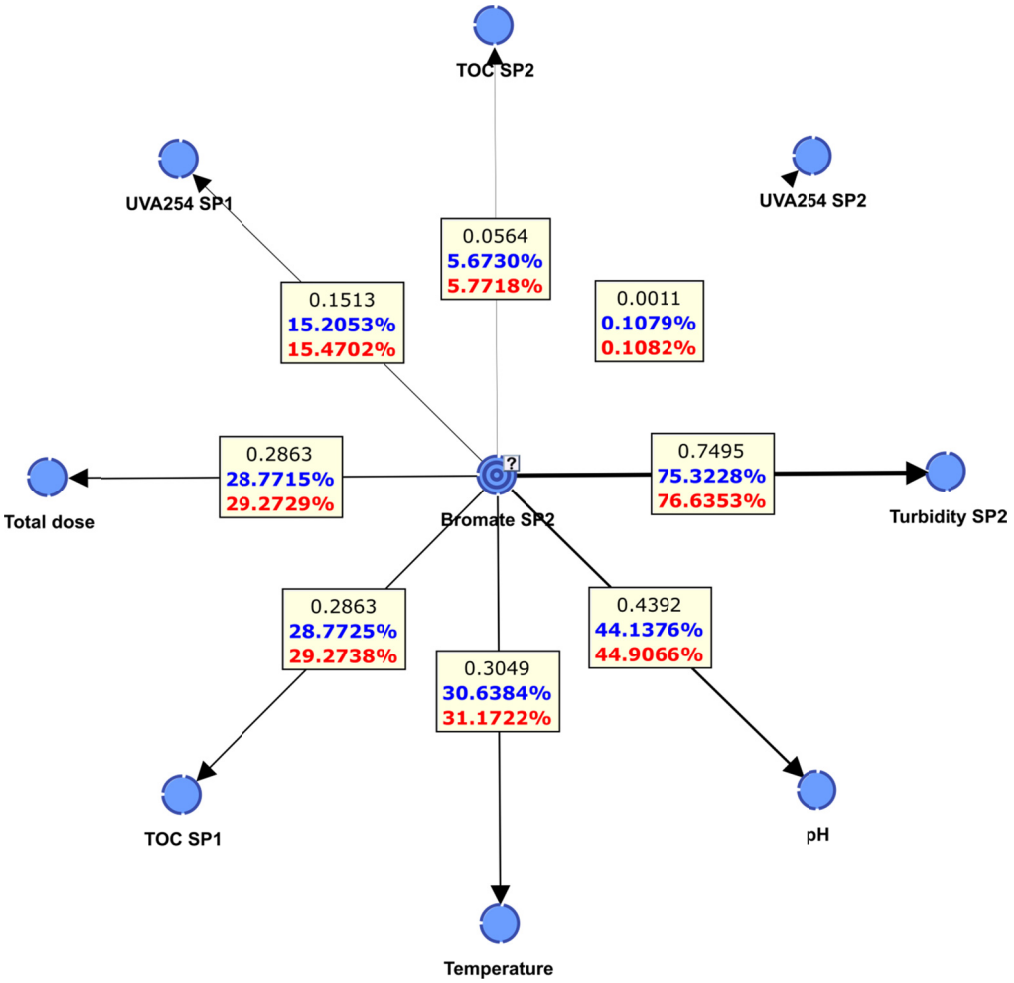


Figure 38. Mutual information for bromate at SP2 (blue = same direction calculation, red = opposite direction calculation)

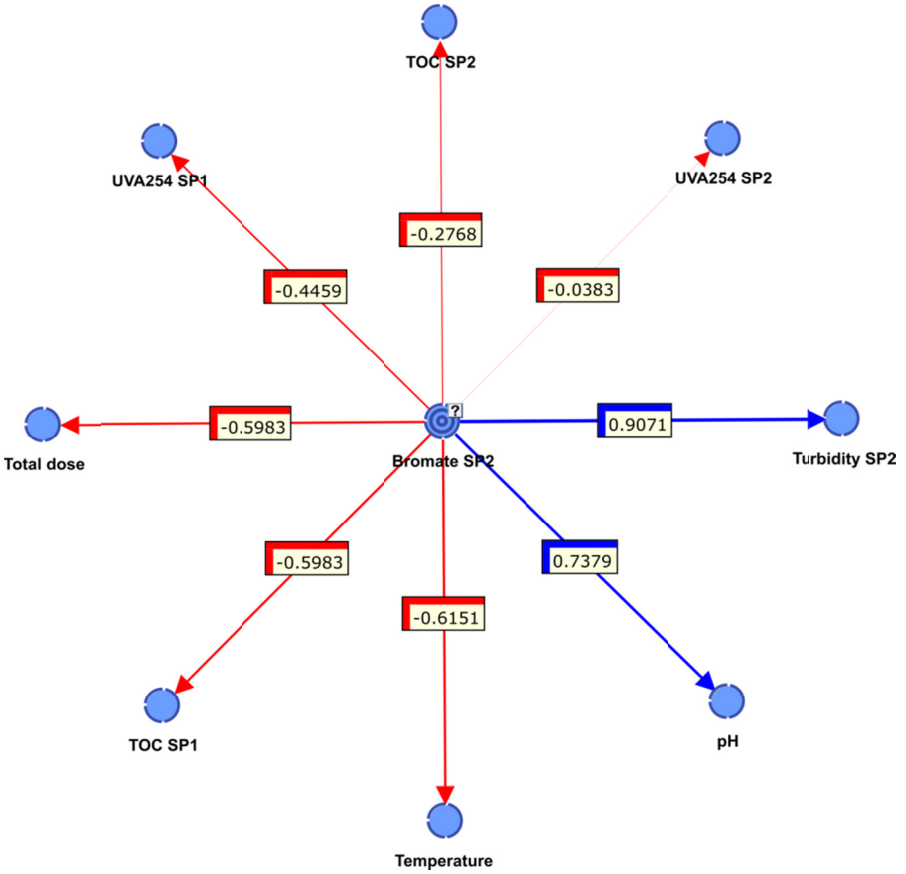


Figure 39. Pearson's R correlations for bromate at SP2 showing forward (blue) and backward (red)

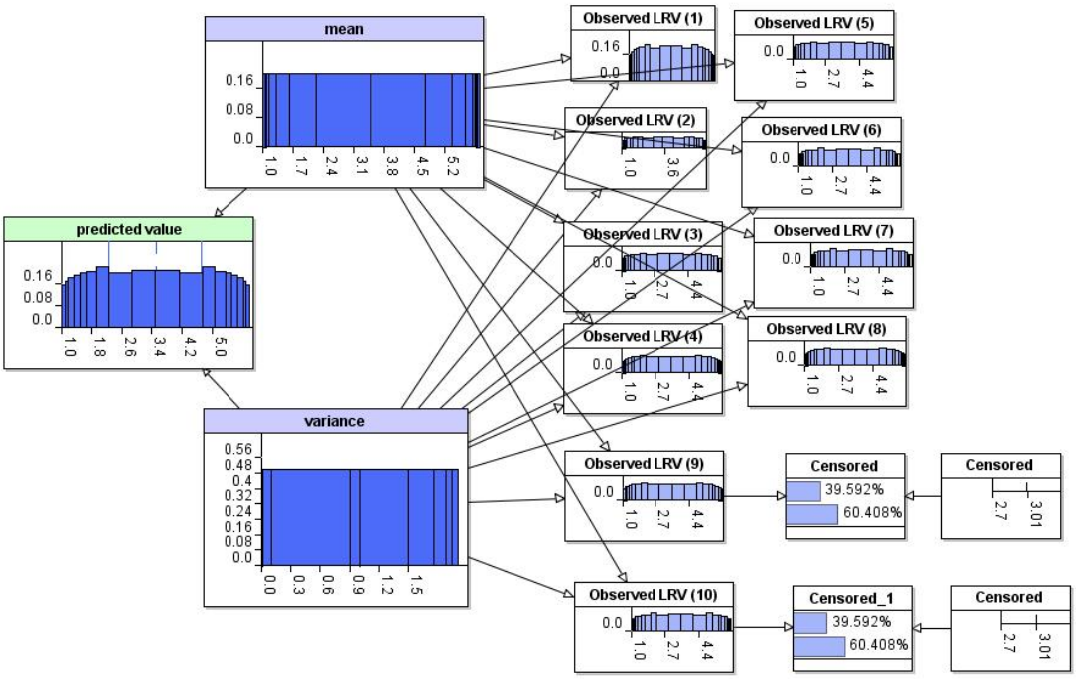


Figure 40. LRV determination with BNs where data is censored.

7.4. Conclusions

The experimental campaign confirmed the insights and estimates obtained using ETP historical data on the effectiveness and consistency of preozonation and biofiltration treatment together. Development of log credits for these systems looks feasible.

More advanced software such as AgenaRisk was able to facilitate further understanding of treatment processes by making available Bayesian data mining tools.

In respect to the uncertainties arising from the historical BN analyses, which the experimental campaign was designed to address, the following were concluded:

- *Microbial data was only collected at weekly intervals around 7 am. So it was unclear how much variance there might be between different days and over the course of individual days.*
    - Variance in reductions over short term was comparable to long term.
- *It was unclear how well samples were matched between different treatment stages.*
    - LRVs for well-matched samples were comparable to those obtained for historical data.
- *Ozone generates bromate from bromide but the extent was unclear and this would impact on whether preozonation should be viewed as a valid/acceptable disinfection treatment.*
    - Between 3 and 13% bromide converted to bromate
    - Preozonation is an effective disinfection treatment but its efficiency is very variable and needs optimization.
- *There were a range of other chemical contaminants especially trace organics which might be reduced or enhanced by the treatment processes especially ozonation but the extent of this was unclear.*
    - Trace organics data have been collected, but will be the subject of a further report.
- *Several physicochemical parameters were unavailable notably measures of organic carbon.*
    - Organic carbon, pH, turbidity data and temperature were concurrently recorded and appeared to strongly reflect *E. coli* levels and bromate
- *Given its potential it was seen as important to confirm which on line measurements provided useful surrogate monitoring data on process effectiveness.*
    - Turbidity and UV_{254} absorbance may be more useful than originally assessed.
    - These data also require further analysis.

8. Activated Sludge LRV data analysis and estimation – Based on CSIRO data

8.1. Introduction

8.1.1. Systems

CSIRO Land and Water as part of its contribution to the NatVal project has been collecting data on microbial removal by full scale Activated Sludge plants. An objective of this work has been to obtain LRV estimates that can be used to estimate water treatment credits potentially agreeable to regulators. It has focused on removal of microorganisms, which can be easily measured in both the influent and effluent, and has involved especially, concurrent input and output measurements of:

- *E. coli;*
- Adenovirus;
- Polyomavirus;
- Microviridae.

CSIRO has also obtained treatment manager collected measurements of many common water quality parameters including DO, turbidity, nitrogen, phosphorus, water temperature and pH. The treatment plants selected for this monitoring campaign were Oxley, Boneo, Beenyup and Rosney Sewage/Water Recycling Treatment Plants.

It is understood that CSIRO has attempted to analyse the complex data set using factor or principal component analysis as well as conventional parametric statistics to better quantify LRV variation and factors controlling this. However this approach has proved challenging to determine how LRVs vary in response to other parameters and one another such that indicators and surrogates can be used to predict likely LRVs in any given circumstance and how LRV credits may be assigned to one or other activated sludge plant for Validation purposes.

It was proposed to reanalyses these same data using the BN techniques developed and described in this report. 108 data records were collected and provided for this analysis. The samples had all been collected between May 2014 and June 2015. A summary of the primary microbial measurements is provided in Table 14.

These data were compiled in a single Excel data table. In the initial data obtained the In/Out data for each parameter/site/date combination were not always paired by exact date. So each record was also assigned a timestep to allow matching and estimation of individual timestep based LRVs and reductions in physicochemical parameters.

Table 14. Activated Sludge microbial measurement numbers (input, output, LRVs)

Values	Beenyup	Boneo	Oxley	Rosney	Grand Total
Count of In_E_coli	22	20	40	24	106
Count of Out_E_coli	20	20	40	24	104
Count of LRV_Ecoli	20	20	40	24	104
Count of In_Adenovirus	23	19	40	25	107
Count of Out_Adenovirus	22	19	39	25	105
Count of LRV_Adenovirus	22	19	39	25	105
Count of In_Polyomavirus	23	19	40	25	107
Count of Out_Polyomavirus	22	17	38	25	102

Values	Beenyup	Boneo	Oxley	Rosney	Grand Total
Count of LRV_Polyomavirus	22	17	38	25	102
Count of In_MicroViridae	23	19	40	25	107
Count of Out_MicroViridae	22	18	39	25	104
Count of LRV_MicroViridae	22	18	39	25	104

8.1.2. Why the AS data provided a good BN model and test for LRV estimation methods

The data set was seen as a good test of BN application for the following reasons:

- There were sufficient microbial measurements to develop models with many parameters while still maintaining a credible ratio of data points:parameters i.e. 10:1 or better - compare Table 8.3 in (Salas et al., 1980). This was comparable to that analysed previously (Carvajal et al., 2015);
- The diversity of candidate indicator and surrogate parameters was great;
- The number of microbial measurements was large enough for good prediction accuracy testing based on repeated timestep record splitting using the WEKA data mining software;
- In contrast to many parametric methods, BN models can be constructed when there are data gaps.

8.1.3. Aims of the BN analysis

The aim was to provide a provisional analysis of the AS data. In particular it aimed to illustrate how to rapidly develop and utilize naïve Bayes and Semi-naïve Bayes models (Korb and Nicholson, 2011) for the following purposes:

- Identification of the main parameters apparently influencing microbial reduction and conversely those which are less likely to;
- Identification of general model features and estimation of their accuracy compared to the ZeroR case (for discussion of ZeroR use and WEKA see (Carvajal et al., 2015, Witten et al., 2011 , Markov and Russell, 2015(accessed));
- Illustration of collection of model accuracy statistics;
- Generation of estimates of key microbial log reduction value probability density functions (LRVs) and identification of which parameters likely influence them;
- Exploration of the best models learnable by Netica^{tm};
- Assessment of whether Netica^{tm} generated TAN models are likely to be comparable to optimum models which could be generated, using other data mining settings (but less easily exported and explored than with Netica^{tm} reflecting varying file formats) .

8.1.4. Why the current analysis is necessarily provisional

BN construction is an iterative process involving data elicitation and group network design among other things (Kragt, 2009, Chen and Pollino, 2012). Time did not allow us to implement this ideal in this instance but obviously as a prelude, first cut assessments are needed / essential so our first approximation is seen as justifiable.

In the case of wastewater treatment, data elicitation is arguably not as onerous as with natural water/ecosystem management where the systems are highly variable and the data is much less complete (Pollino et al., 2007). This is because treatment systems tend to be better

understood/defined than natural aquatic systems and the data sets tend to be richer and consistent in their data collection methodology as illustrated by the CSIRO data set being used here for illustration purposes. However model design challenges still remain. For some better understood wastewater treatments such as UV irradiation the process can be straightforward and factors driving LRVs understood at the theoretical levels. But in the case of Activated Sludge the relationships between parameters are less clear. This is evidenced in (Flapper et al., 2012) where their PCA generated complex, hard to interpret/exploit combination parameters.

There is the question of causality which lies at the heart of Bayesian modelling concepts. BN algebra was developed as a result of attempts to quantify causality by Pearl and others (Pearl, 1996, Pearl, 1999 , Pearl, 2000). However in practice distinguishing cause from effect and hence the direction of links/arc in BNs with AS is more challenging. For example organic matter as BOD might be viewed as controlled by dissolved oxygen concentration or vice versa. A partial way around the causality issue discussed in our paper (Carvajal et al., 2015) and ultimately employed here is to develop naïve and semi-naïve BNs where the assumption of causality can be relaxed. Construction of this type of BN also employs machine routines to find optimal models rather than relying completely on human 'insight' and 'intuition'. This type of Bayes model has been found from experience to generate useful models consistent with qualitative patterns of relationships between variables seen in summary statistics.

The downside of this approach is that there are a wide range of machine learning methods and each in our experience can generate a number of different similar models where data sets are medium sized. This leads to a surfeit of candidate models from which preferred ones need to be chosen. Another complication is the selection of significant v. insignificant parameters which is required to avoid overfitting of models. How these issues were provisionally addressed is explained so we would not claim the method is ideal as yet but illustrative of what is possible.

Finally BN software is currently actively evolving and new and more powerful tools are being developed which have not yet found their way into the basic software package we have use, Netica^{tm}. Further versions of the latter and other packages will likely offer both greater convenience and reflect some resolution of the above issues.

8.2. Methods

The BN design method employed was based on that described previously (Carvajal et al., 2015). It was modified somewhat because:

- The larger and more complex CSIRO data set, requiring analysis, demanded a quicker screening approach.
- We wished to assess LRV variance across more than two ranges/states unlike in the original paper which considered only two alternate LRV states and assess how LRV tended to vary incrementally with incremental variations in other parameters.

The steps undertaken were as follows:

1. Organize the CSIRO data into a .CSV data table.
2. From this construct a second table where the 4 LRVs data sets of interest were one by one each assigned to just two categories (i.e. high and low representing the upper and lower 50th percentile demarked ranges - Categorization was necessary as WEKA semi-naïve Bayes analysis requires the class nodes (the LRVs) to be in category form, though Netica^{tm} does not and can tolerate continuous parametric data in a class node. Separation into only two categories was expected to maximise prediction accuracy).

3. Use WEKA to construct ZeroR, naïve and semi-naive Bayes models to determine if any/which secondary parameters could be used to predict LRVs as follows:
    a. Filter unnecessary, less useful or problematic parameters from the dataset e.g. sampling date;
    b. Iteratively construct models using WEKA Explorer's Bayesian tool suite and evaluate model credibility using cross validation to generate accuracy metrics as follows:
        i. Use the ZeroR model to measure model accuracy in the absence of any BN arcs as a reference to determine if the BN models are providing improved information;
        ii. Use the naïve BN model to establish a basic reference model and identify parameters which did not contribute anything to quantifying LRVs and discard them.
        iii. Generate and compare the accuracy statistics of a range of tree augmented network and BAN type semi-Naïve BNs.
        iv. Confirm whether a TAN semi-naïve BN would likely be close to optimal by comparing WEKA's TAN model with other configurations.
4. Based on this analysis using Netica^{tm} :
    a. Identify the most informative parameters which could be used to model LRVs and add them to an empty Netica^{tm} model template.
    b. (Using the Netica^{tm} ZeroR model) discretise the parametric nodes through the Cases>Learn>Incorporate Case file option (n=4 was used as standard in line with the recommendation by Marcot et al. (Marcot et al., 2006)) to have 5 or fewer states.
    c. Learn the Netica^{tm} TAN structure.
    d. Learn the final BN probabilities using the Expectation maximisation (EM) learning option.
    e. Identify those parameters most associated with LRVs using the Sensitive to Findings analysis and remove any remaining non or low-contributing parameters (e.g. < 1% influence on LRV variance).
    f. Assess how LRVs vary with different parameter states and range values.

Experience showed that while WEKA could generate a diversity of models, for any particular node set all models approached comparable maximum accuracy judged by the classification metrics. Thus use of the Netica TAN construction tool was seen as acceptable.

An early naïve Bayes model for Microviridae is reproduced for illustration in Figure 41. Clicking on each node within the WEKA software reveals parameter boxes like those in Figure 42. Figure 42b illustrates, surprisingly perhaps, that whether the Microviridae LRV was less than of greater than 2.46 logs was unaffected by the level of mix liquor suspended solids. Based on this the MLSS parameter was ultimately omitted from all models. By contrast the system from which the data were gathered (Figure 42a) and the inflow turbidity (Figure 42c) were potentially important variables.

More complex semi-naïve BNs were constructed after some likely redundant variables were also removed. For example in the case of Microviridae LRV there was a relationship to the input and output Microviridae concentrations. But as some relationship was to be expected and in

routine situations monitoring of Microviridae was unlikely, we removed these parameters as well, for this first modelling iteration. In selecting nodes we aimed ideally for n<10 though this proved difficult on occasion and a maximum of n=12 was settled on.

Because of the still relatively modest number of LRV records available for learning and the data gaps with some parameters the final Netica^{tm} BNs node PDFs were learnt using the EM tool rather than the 'Incorporate Case File' instance counting option to reduce the occurrence of zero probability nodes (for discussion see (Korb and Nicholson, 2011).

A final caveat on the method is that approaches for optimally selecting BNs are still under development even where they are much more widely used than is the case with wastewater treatment e.g. aquatic ecosystem management (Pollino et al., 2007).

Figure 41. Illustrative naïve Bayes model for Microviridae

a.

Probability Distribution Table For System

LRV_MicroViridae	Beenyup	Boneo	Oxley	Rosney
LT246	0.25	0.231	0.083	0.435
GT246	0.176	0.12	0.657	0.046

b.

Probability Distribution Table For MLSS

LRV_MicroViridae	'All'
LT246	1
GT246	1

c.

Probability Distribution Table For In_Turbidity

LRV_MicroViridae	'(-inf-251.5]'	'(251.5-inf)'
LT246	0.217	0.783
GT246	0.123	0.877

Figure 42. Illustration of variance in parameters in response to LRV state

Note:

1. GT246 means and LRV greater than 2.46, etc.

8.3. Results

8.3.1. NB/SNB accuracy assessment for 4 microbial LRV parameters

As expected, all the two state ZeroR models achieved accuracies of *ca* 50% reflecting the fact there were only two choices and we split the data sets *ca* 50/50.

E. coli models including the TAN models were by contrast all reasonably accurate achieving *ca* 82-83%. This included the simplest model of 12 easily monitored parameters. The true positive rate was somewhat lower when it came to predicting whether the LRV was likely to be low.

Polyomavirus and Microviridae LRV TAN models also performed comparably (80-82 % accuracy). Adenovirus performed less well with model accuracies up to *ca* 70%.

Based on these data and earlier experience (Carvajal et al., 2015) we concluded the Netica[tm] *E. coli,* Polyomavirus and Microviridae TAN models should generate satisfactory indications of how LRVs respond to changes in the major indicator and surrogate parameters identified by WEKA modelling. However, the Adenovirus model based scenarios should be interpreted with caution.

8.3.2. Differentiating key v. non-driver parameters

A number of parameters were surprising in apparently not reflecting LRVs as judged by the WEKA analysis. In particular were MLSS and total suspended solids. Other parameters generally discarded included Total Phosphorus and oxidised nitrogen parameters.

No clear influence of sampling timestep, and hence sampling date, was seen. However temperature appeared frequently to influence LRVs with low temperatures being associated with low LRVs and vice versa with all 4 models. So a season linked affect (not assessed) cannot be discounted. This caused us to closely examine when the samples were taken as discussed further below.

8.3.3. Using Netica[tm] TAN models to infer LRVs and their characteristics

The information/inferences obtainable from the Netica[tm] generated TAN models can only be appreciated fully or surveyed by examining the operating BNs using Netica[tm] itself or undertaking detailed Sensitivity to Parameter (Pollino et al., 2007) analyses. However, useful data can still be presented using static captures of each model in a given state (e.g. Figure 43, Figure 48, Figure 50, Figure 52) of examining 'Sensitivity to Findings' summary tables (e.g. Figure 44, Figure 49, Figure 51, Figure 53).

So for example Figure 43 for *E. coli* captures the overall LRV as well as the states and likelihoods of other 'diagnostic' parameters which appear to most strongly reflect the LRVs observed.

Despite the arrow directions this plot does not of course imply that the LRVs cause temperature to be in a particular range. Rather temperature records reflect the LRVs achieved. If this seems confusing a useful analogue to consider to understand what inference is happening is the classic example of the Asia/smoking cancer diagnosis net (Lauritzen and Spiegelhalter, 1988). In the latter the final nodes are an X-ray test and assessment of the presence of dypsnea (shortness of breath), i.e. given cancer they are used as diagnostic tests. The arrow goes from cancer causing the test result.

In the present instance LRV of *E. coli* is what we are interested in the same manner as a clinician is interested in cancer. The LRV can be viewed as representing the complex biomass and process environment reducing microbial numbers whose state is reflected in the other (diagnostic) BN parameters. Though backcasting from the latter the net can be used to refine our assessment of the likelihood that *E. coli* LRV is a particular value or in a particular range. In the base case for example the LRV is 2.64±0.98 given the reactor (out) water temperature is 23.5±4.8 C etc.

The fact that all nodes interact with one another raises the question of which are most indicative of LRV variance, overall. This information is provided by the order of the parameters in the accompanying figure shown, Sensitivity of Findings, and the variance or entropy (mutual information) reduction.

It can be seen that the LRV PDF is crucially influenced by which system is being considered, followed by nitrogen (mainly reduced nitrogen). Usefully it is also possible to selectively exclude specific node states and reassess the Sensitivity to Findings. In Figure 45 and Figure 46 we have excluded Rosney and Boneo which were recognised to exhibit poor LRVs. The new

Sensitivity to Findings analysis indicates there is variation between the two remaining systems but it is not as important as nitrogen. Also the LRV variance is greatly reduced and the LRV is now a much more respectable 3.28±0.58. When this sort of process is undertaken extensively and systematically it is termed 'Sensitivity to Parameters' analysis (Pollino et al., 2007).

Due to time constraints and this being a provisional evaluation we have not done exhaustive analyses but a qualitative assessment of LRV variance guided by the Sensitivity to Findings tables for each microbial analyte to illustrate what BNs promise to summarise and reveal.

A final important operational note here is in respect to summary statistics. In the BNs shown only the arithmetic mean and standard deviation are routinely displayed. However, by passing the mouse pointer over the value a range of other summary statistics for each node will be revealed on interest in risk assessment. Further the statistics displayed will change automatically as node settings are also changed (Figure 47).

8.3.4. *E. coli* (Gram negative pathogens)

The picture which emerged was as follows:

- The systems varied markedly in performance with Rosney being particularly poor, Oxley and Beenyup being much better and Boneo being an intermediate case.
- AS consistently reduced bacterial numbers and on average in all systems.
- The overall 10th percentile removal was 0.67 and for the best 2 systems 1.80.

LRVs for Beenyup + Oxley were far better at 3.37±0.44 compared to the poorest LRV bin, 1.38±0.58, which was 86% associated with Rosney data and to a minor extent Boneo.

The LRV sensitivity analysis pointed to the different performances being accounted for by factors related to nitrogen, oxygen and temperature. Comparison of individual systems suggested the conditions prevailing during each sampling campaign may have been critical.

Generally high reactor nitrogen and low DO were associated with poor performance – suggesting either overloading or poor performance for other reasons such as the low temperature supporting only low sludge activity. The absence of MLSS as a factor indicated it was not biomass presence but its activity which was important.

Poor LRV appeared closely associated with high reduced nitrogen – 20/33 mg/L whereas satisfactory LRV (>2.3) was associated with much lower average NH_4^+-N and TN of 2.2 and 7.4 mg/L.

Perhaps critically the Rosney system was operating at the lowest average temperature of all (18.6 °C). Conversely the Oxley system was operating at an average reactor (out) temperature of 26.9 °C. Given rates of nutrient processing, as well as microbial inactivation, are temperature-dependent this may have been critical. Re-examination of the raw data confirmed these summary statistics and indicated that Oxley sampling was heavily weighted towards the warmer months whereas there were only 4 summer samplings at Rosney. Provisionally the BN indicated LRVs were 1.65±0.75 when temperature was in the range 14 to 20 °C and 3.05±0.62 when the temperature was in the range 26.6 to 29.5 °C.

Overall the following was provisionally concluded in regard to LRVs:

- AS can consistently achieve significant average bacterial removal of 1.3 to 3.5 logs comparable to that reported by Flapper et al. (Flapper et al., 2012, Flapper et al., 2010) of 2.8±0.52 logs.

- Individual systems show similar standard deviations of *ca* 0.5. However across all four systems the performance was more highly variable at *ca* 1.0 standard deviation.
- The high overall LRV standard deviation reflected the marked differences in performance between different AS systems.
- The cause of this variance is not at this stage clear. However, there appeared to be an association with Nitrogen and reactor temperature.
- A high input BOD concentration (>430 mg/L) appeared to be associated with reducing the LRV by 1.3 logs.
- Nitrogen, temperature and oxygen levels in AS may be indicators of how well bacterial LRVs

Conversely the BN based analysis eliminated a range of parameters as contributing little to AS LRV values and variance notably Turbidity, TSS, Oxidised Nitrogen and pH. This contrasts with the findings of our analysis of Flapper et al. (2012).

Overall these results suggest that LRV allowable might vary according to the climatic zone and A was located in.

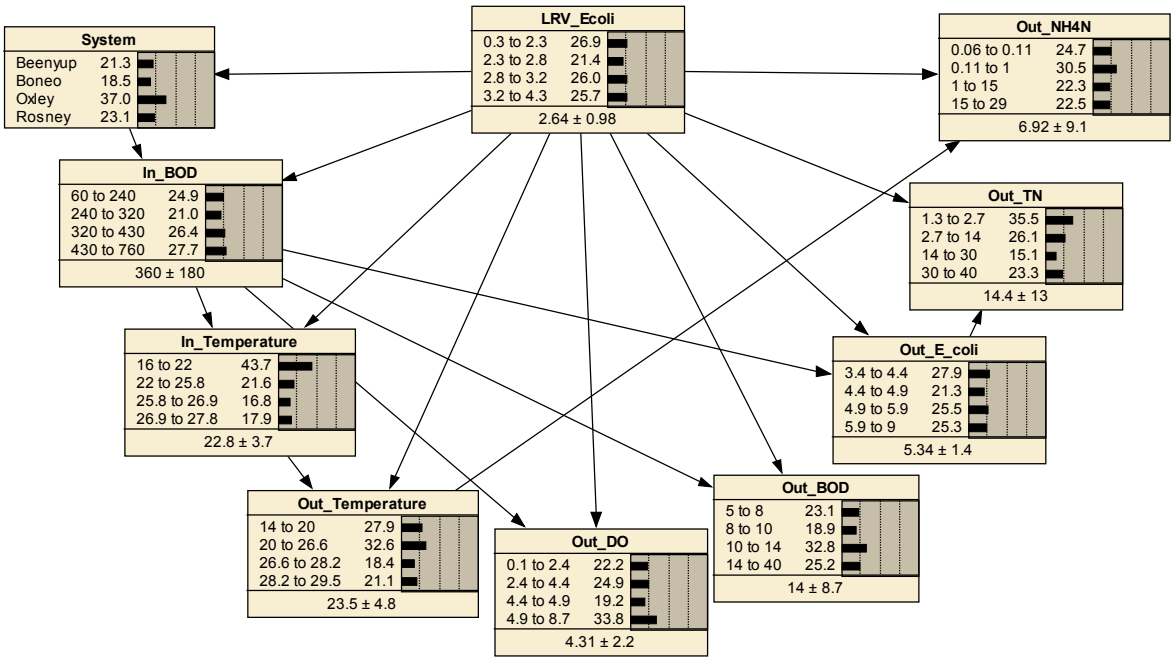


Figure 43. TAN LRV BN for *E. coli*

Sensitivity of 'LRV_Ecoli' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV_Ecoli	0.9582	100	1.99461	100	0.5601995
System	0.6557	68.4	0.93296	46.8	0.1912092
Out_TN	0.5606	58.5	0.78985	39.6	0.1418690
Out_NH4N	0.5441	56.8	0.69203	34.7	0.1273128
Out_DO	0.5335	55.7	0.70456	35.3	0.1483591
Out_Temperature	0.3821	39.9	0.50962	25.5	0.0578548
In_BOD	0.2526	26.4	0.35847	18	0.0475705
In_Temperature	0.2479	25.9	0.37343	18.7	0.0360942
Out_E_coli	0.1478	15.4	0.49341	24.7	0.0485198
Out_BOD	0.06336	6.61	0.14989	7.51	0.0098536

Figure 44. TAN *E. coli* LRV BN Node sensitivity analysis results

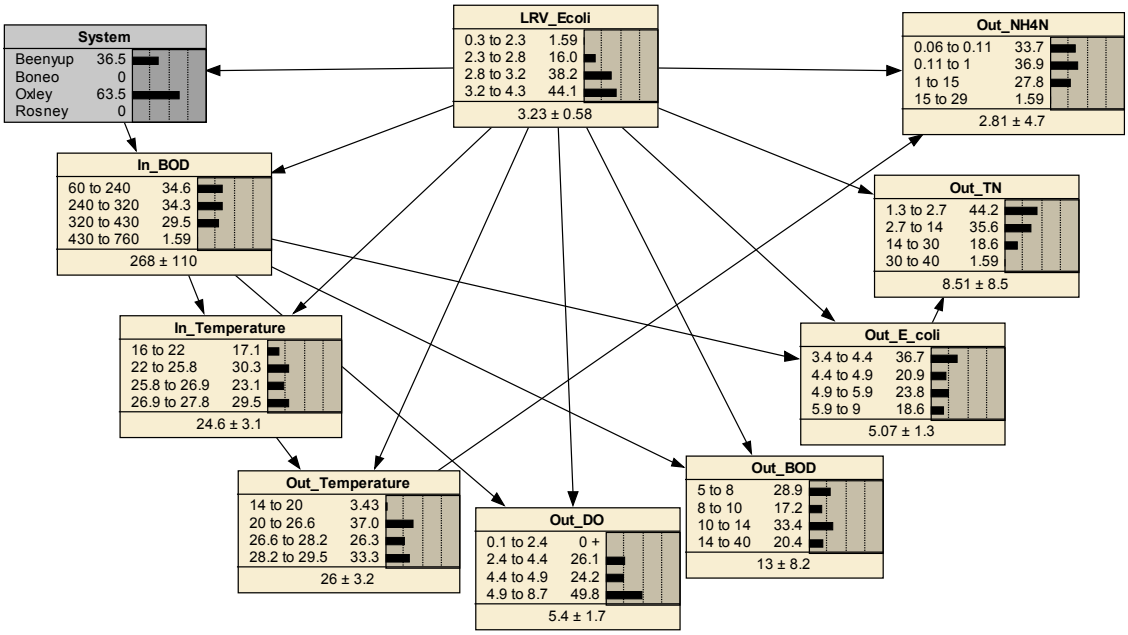


Figure 45. TAN LRV BN for *E. coli* without Rosney and Boneo data

```
Sensitivity of 'LRV_Ecoli' to a finding at another node:
```

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV_Ecoli	0.331	100	1.56930	100	0.4119607
Out_TN	0.08273	25	0.21108	13.5	0.0333429
Out_NH4N	0.06852	20.7	0.15703	10	0.0225278
Out_E_coli	0.05239	15.8	0.33492	21.3	0.0263956
System	0.04994	15.1	0.15534	9.9	0.0326784
Out_BOD	0.0426	12.9	0.22801	14.5	0.0435774
In_BOD	0.04023	12.2	0.17847	11.4	0.0283008
Out_Temperature	0.03841	11.6	0.23141	14.7	0.0502280
In_Temperature	0.02013	6.08	0.13265	8.45	0.0236070
Out_DO	0.01987	6	0.12867	8.2	0.0175940

Figure 46. TAN *E. coli* LRV BN Node sensitivity analysis results without Rosney and Boneo data

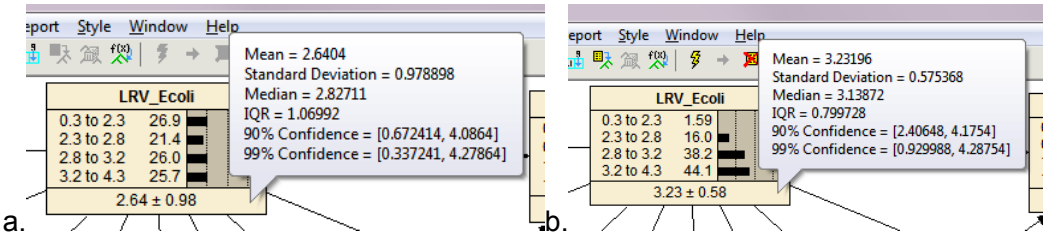


Figure 47. TAN LRV BN Node for *E. coli* with (a.) and without (b.) Rosney and Boneo data illustrating the extended statistics available and how they change with new scenarios.

8.3.5. Polyomavirus

In finalizing the provisional net the input temperature and DO and *E. coli* concentrations were removed as these parameters were judged to be covered by the reactor measurements which

would also be more representative. The BN inferences and summary statistics were essentially the same with or without this data.

Polyomavirus LRVs were overall comparable on average to *E. coli* (2.75 logs v. 2.69 logs) but more variable (SD = 1.5!). Indeed all three viruses showed these high SDs. High SD was also evident with 3 of the 4 individual systems, the exception being Oxley, indicating its source lay with the analytical methodology or was inherent in removal rather than reflecting differences between systems alone. Independent assays of replicate samples or repeat assays would be a useful way to assess whether methodology played a role.

In contrast to *E. coli* Beenyup was not one of the two superior systems though it averaged 0.5 logs more removal than Rosney which was again the worst performer.

Lower temperatures below 20 °C were again associated with a 50% reduction in LRV. Again decreasing nitrogen was associated with increasing LRV. DO was associated with LRV but no clear trend was apparent.

E. coli concentrations and removal were somewhat associated with removal though not as markedly as the 3 physicochemical parameters.

The best LRV removal was obtained with the Boneo + Oxley system combination (3.45±1.0 logs). Discounting high nitrogen reactor content (>14 mg/L) samples yielded a somewhat better 3.59±0.82 logs.

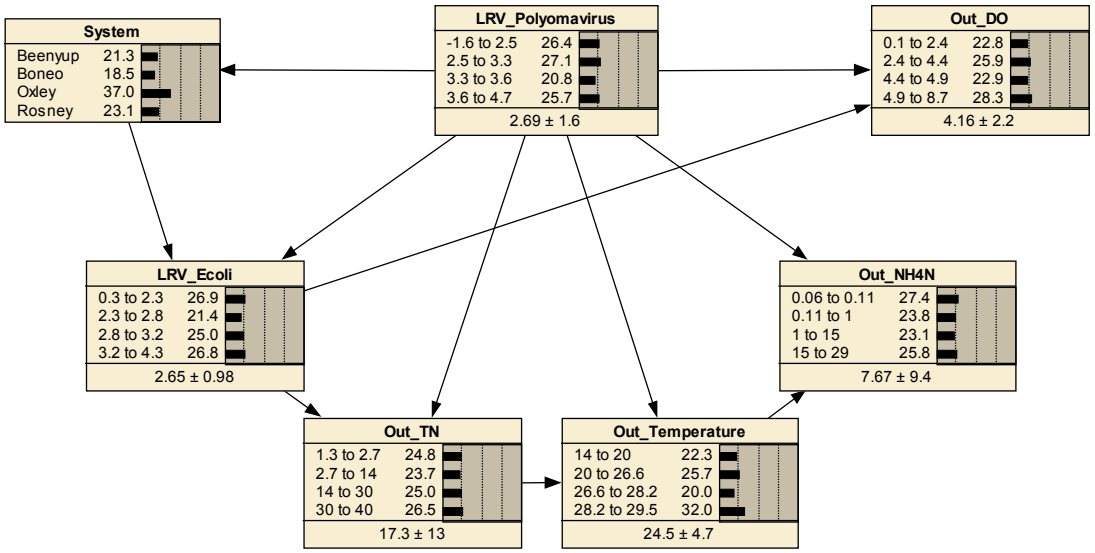


Figure 48. TAN LRV BN for Polyomavirus

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV_Polyomavirus	2.413	100	1.99283	100	0.5594438
Out_NH4N	0.8896	36.9	0.50420	25.3	0.0614440
System	0.8154	33.8	0.50699	25.4	0.0334885
Out_DO	0.6081	25.2	0.39559	19.9	0.0235657
Out_Temperature	0.444	18.4	0.29985	15	0.0224634
LRV_Ecoli	0.443	18.4	0.24610	12.3	0.0227629
Out_TN	0.2502	10.4	0.14595	7.32	0.0109042

Figure 49. TAN Polyomavirus LRV BN Node sensitivity analysis results

8.3.6. Microviridae

Microviridae LRVs were again of a similar order to *E. coli* but highly variable. Oxley performed well with an average reduction of 2.91 but the other three only yielded averages between 1.39 and 1.57.

Temperature again influenced LRV which increased by a factor of 1.5 between the coldest and hottest ranges. Low ammonium/TN levels again appeared associated with higher LRVs.

LRV appeared much less influenced by the other reactor (Out) parameters than *E. coli* and Polyomavirus and the influence of nitrogen was marginal.

Unexpectedly input measurements were more closely associated with LRVs than Out measurements. A check of the primary database suggested BN reliability might be constrained by the significant number of missing values for physicochemical parameters measured at systems other than at Oxley. This may have reduced the accuracy of the final provisional net which utilized the outflow associated measurements.

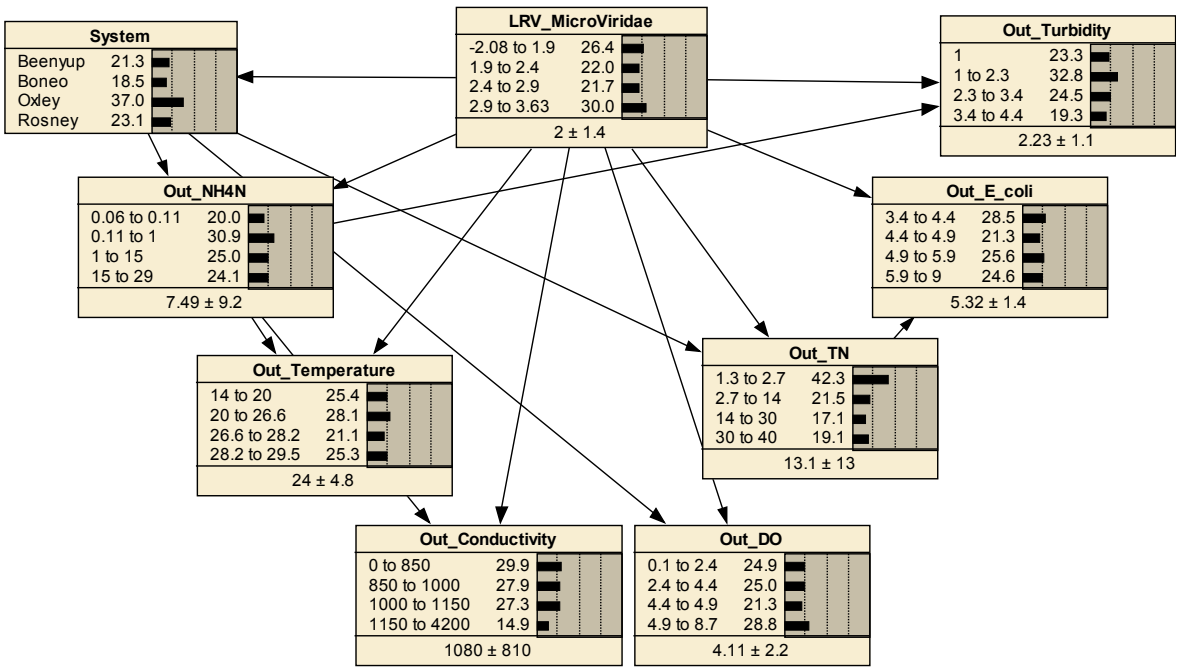


Figure 50. TAN LRV BN for MicroViridae

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV_MicroViridae	2.1	100	1.98666	100	0.5567082
System	0.4933	23.5	0.38679	19.5	0.0410019
Out_Temperature	0.4402	21	0.41123	20.7	0.0373944
Out_Turbidity	0.2861	13.6	0.25296	12.7	0.0393122
Out_Conductivity	0.2197	10.5	0.13158	6.62	0.0115469
Out_E_coli	0.201	9.57	0.17553	8.84	0.0174452
Out_NH4N	0.1816	8.64	0.16467	8.29	0.0165732
Out_TN	0.1195	5.69	0.22494	11.3	0.0315581
Out_DO	0.1127	5.37	0.20391	10.3	0.0188362

Figure 51. TAN MicroViridae LRV BN Node sensitivity analysis results

8.3.7. Adenovirus

Finally Adenovirus displayed an average reduction of 2.07 logs and high SD like Microviridae. Beenyup and Rosney were again the poorest performers with average LRVs of only 1.2 and very high SDs (1.6 To 1.8). Conversely Oxley achieved an LRV of 3.06±1.1.

Unlike the previous microbial analytes Adenovirus LRV did not appear to be influenced by nitrogen and the increase with increasing temperature was only *ca* 1 log unit.

Low *E. coli* removal was associated with low Adenovirus removal.

The most influential parameters appeared to be the conductivity and turbidity

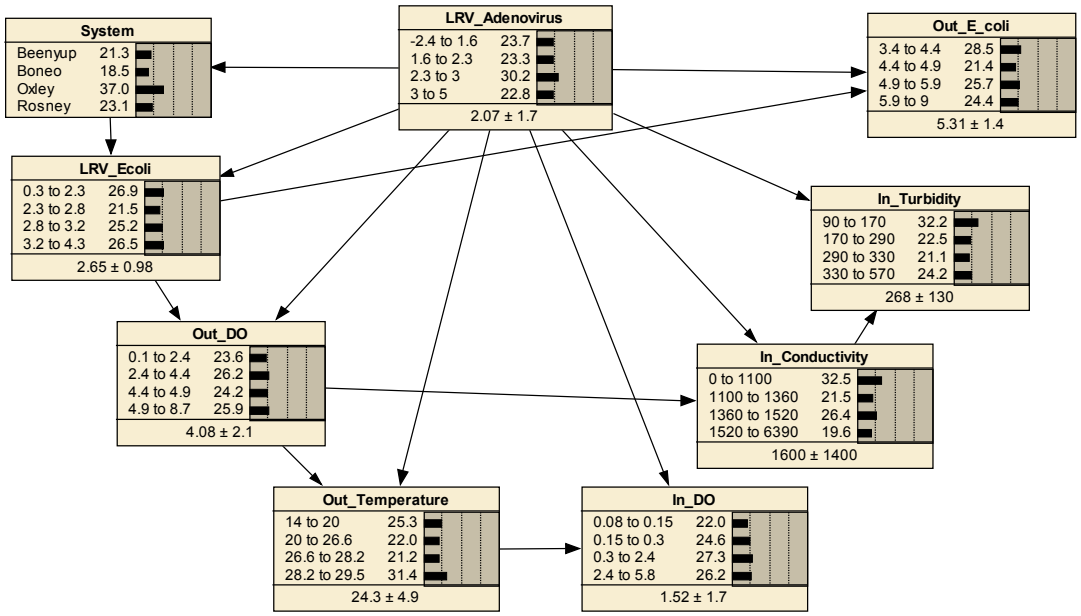


Figure 52. TAN LRV BN for Adenovirus

Sensitivity of 'LRV_Adenovirus' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV_Adenovirus	2.811	100	1.98971	100	0.5579801
In_Turbidity	0.9413	33.5	0.33057	16.6	0.0399951
In_Conductivity	0.8624	30.7	0.36146	18.2	0.0326096
In_DO	0.734	26.1	0.30700	15.4	0.0324236
Out_E_coli	0.6857	24.4	0.25579	12.9	0.0194679
System	0.6647	23.6	0.31289	15.7	0.0200062
Out_DO	0.4333	15.4	0.21377	10.7	0.0172214
Out_Temperature	0.2372	8.44	0.14038	7.06	0.0144548
LRV_Ecoli	0.1734	6.17	0.08775	4.41	0.0082122

Figure 53. TAN Adenovirus LRV BN Node sensitivity analysis results

8.3.8. Further scenario exploration and simple validation

A great benefit of the BNs is it allows interactive exploration of how the microbial LRVs varied with reactor operating conditions while efficiently providing LRV estimates (credits) reflecting these conditions. Further BNs provide a system for relating non-microbial operating parameters to one another and potentially obtaining insights on how to optimise a treatment plant or how far this is possible given their competing functions.

The main caveat on such exercises is uncertainty about how much credit we can plausibly ascribe to the models and how far we can extrapolate the future or more generally to other STPs based on the data at hand. Model accuracy testing provides a means for evaluating this. But in the end a degree of human belief in these causal and semi-naive networks and the sufficiency of the available data will always be required.

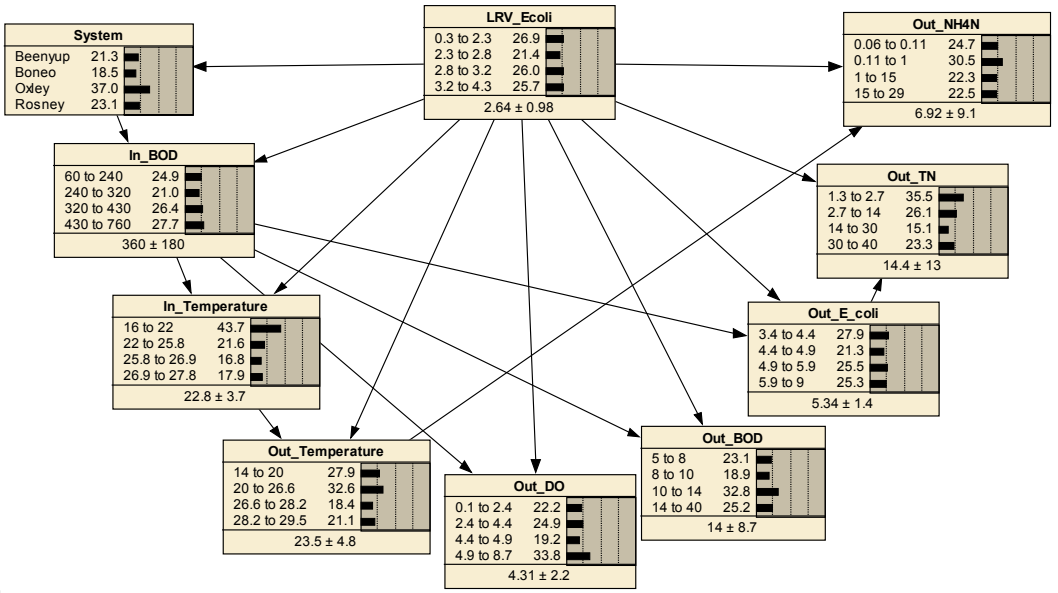
We have presented a range of basic inferences in the results section above. As a final exercise below using the provisional *E. coli* and Polyomavirus LRV models we illustrate how LRV credits might be estimated based on:

- 'validation data' alone;
- a combination of historic experience and new validation data.

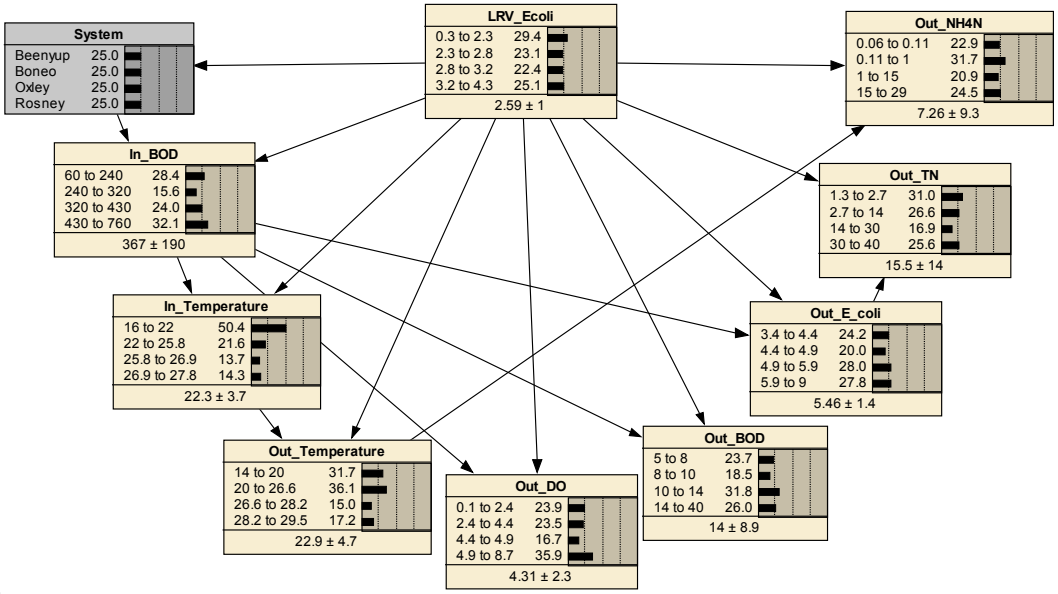
Figure 54a reproduces the *E. coli* LRV model. While this provides an estimate of the expected LRV this value is also heavily weighted towards the Oxley data set. The BN provides a method for assigning equal weighting (Figure 54b) to the different systems – the calibration option where alternative probabilities are entered. The net then adjusts in response. The result is a slight drop in the *E. coli* LRV estimate.

Other combinations are of course possible e.g. assigning different more appropriate weightings to the reactor (out) temperature distributions. Figure 54c and Figure 54d show how this can be undertaken further using the Polyomavirus LRV model. In this case we have adjusted the temperature ranges by changing the discretization thresholds to equal intervals and relearned the BN. It can be seen that the data is not fully representative of the full range of temperatures at which the Ass operated but heavily weighted toward the high temperature range. When the temperature probabilities are then rebalanced to a more even spread the average LRV is reduced from 2.71 to 2.16! This demonstrates clearly that before LRVs are applied to treatment systems temperature/seasonal variance needs to be better understood unless a minimum credit is to be applied.

An interesting possibility would be to substitute the temperature CPT in this reference BN and replace it with one for a new recycled water plant location. This could provide a 'literature' LRV which could in turn be combined with LRVs developed specifically for validation testing.

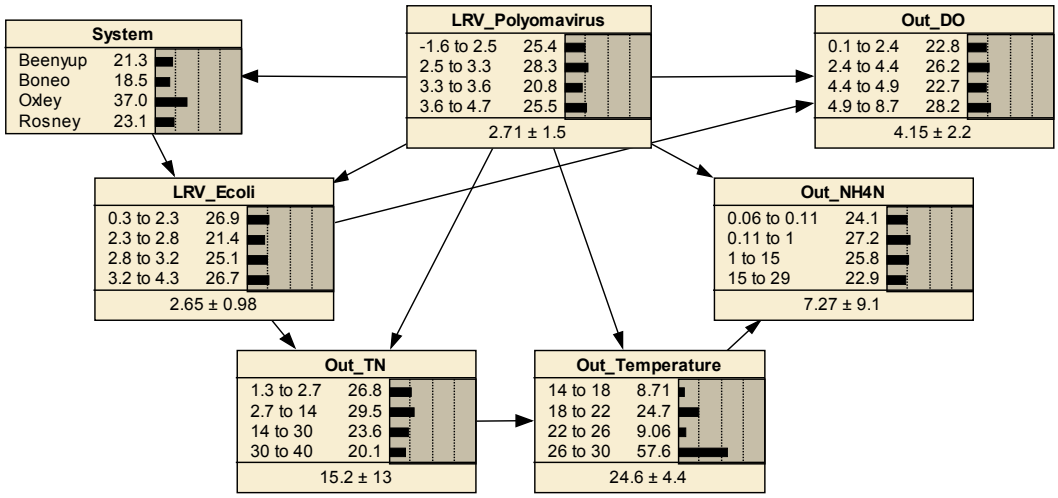


a.



b.

c.



d.

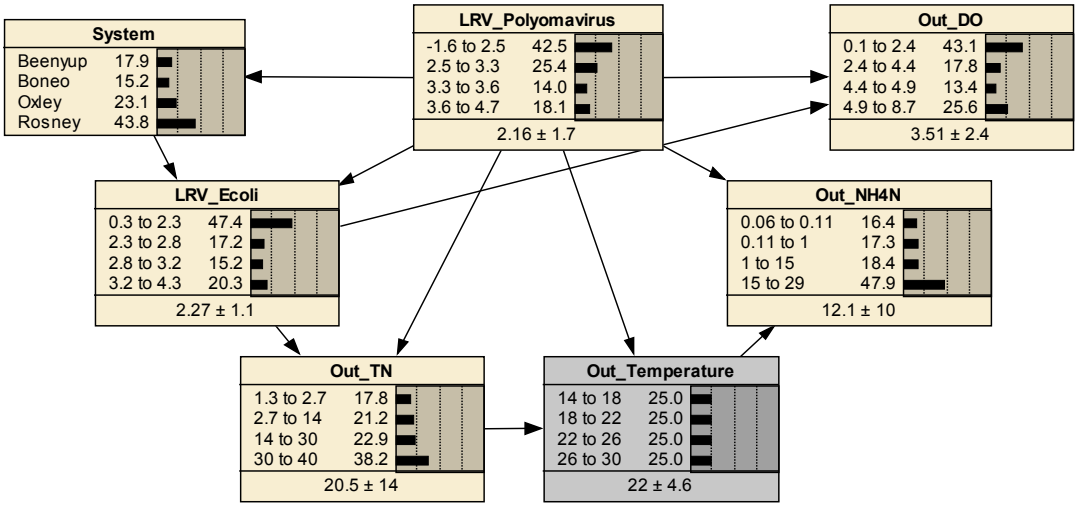


Figure 54. Effect of rebalancing contributions from each system (a.,b.) and optimising TN and normalising Temperature ranges (c.,d.)

8.3.9. Bayesian Validation

Figure 55 illustrates the 'Bayesian Validation' technique described in detail for the Glenelg case study. In this case we have used the Flapper et al. (2010) table 2 data to generate a log normal PDF LRV prior for *E. coli* and combined this with the CSIRO model based estimate weighting their credit 1:2 to yield a composite *E. coli* LRV of 2.68±1.2.

Both this example and the previous illustrate the potential of BN model based data analysis especially validation. Some similar actions are possible using parametric models but we have not encountered the same degree of potential previously which captures and presents uncertainty in an efficient manner e.g. the ability to rapidly allow for data sets of different sizes and weight inferences accordingly and vary scenarios according to variations in expert opinion.

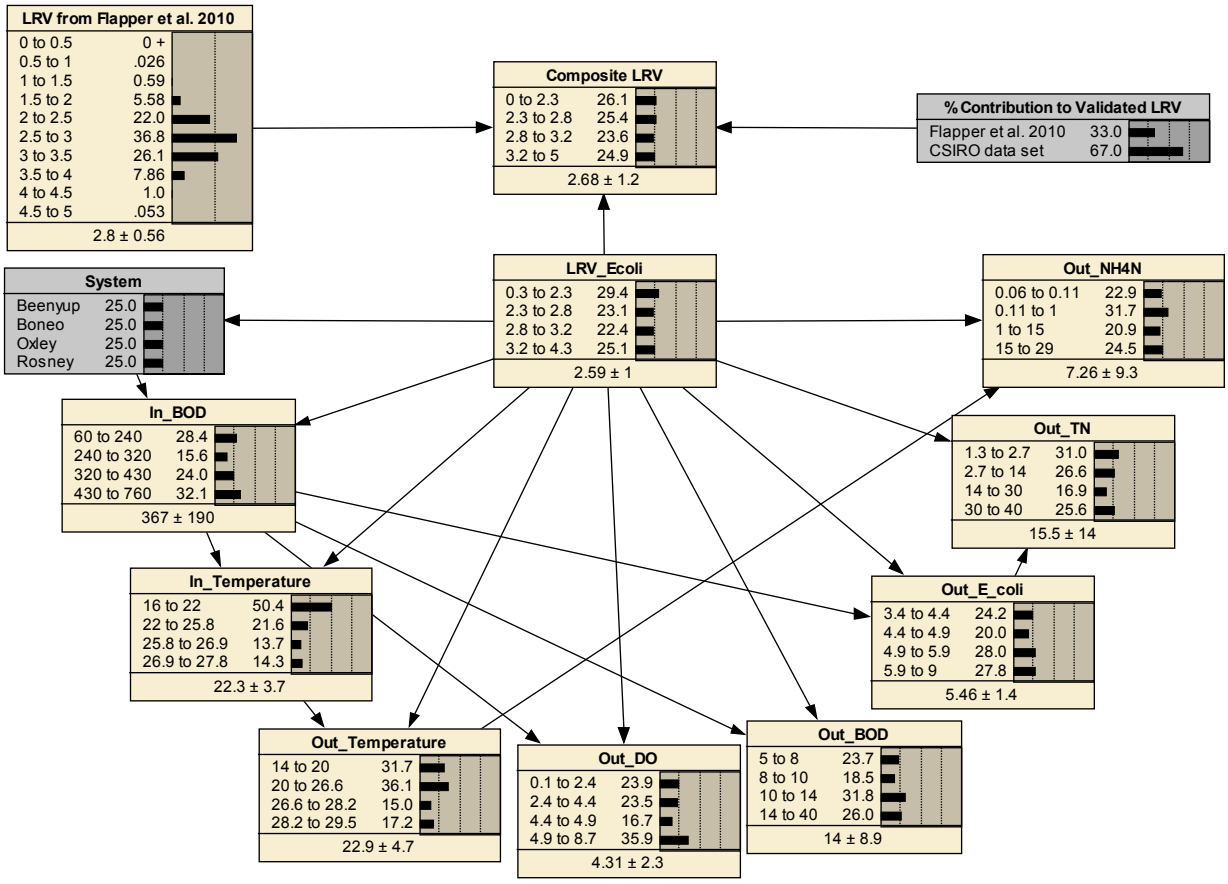


Figure 55. Bayesian Validation of *E. coli* LRV based on Flapper et al. (2010) prior (33% contribution) and CSIRO data (66.7% contribution with all systems given equal weighting)

8.4. Discussion

8.4.1. AS LRV credits

Flapper et al.'s (2010) review indicated that Activated Sludge treatment units were capable of reducing bacteria and viruses by *ca* 2-3 and 1.5-4 log units on average with standard deviation estimates ranging from 0.2 to 1.9.

LRV estimates obtained here using BNs were consistent with these reported values. Unfortunately the high variation in removal rates was also evident again.

But on the plus side analysis provided much more insight or at the least defensible testable hypotheses. So the CSIRO data, especially for Oxley STP, suggested that an average of 3 log reductions might be achievable under optimal conditions. It also indicated under what conditions this might occur and hence when log credits >2 might be assigned e.g.:

- AS systems designed and operating similar to Oxley;
- High operating temperatures and low nitrogen content.

Conversely systems comparable to the Rosney system appeared less promising especially when treating cooler water during winter months and when overloaded such that high nitrogen levels were present.

8.4.2. Benefits of using BNs

The benefits of using a BN based analysis illustrated by this case study included:

- A relatively objective way to identify candidate monitoring parameters, estimate they operating ranges and infer what they should test us;
- Rapid exploration of 'what-if' scenarios addressing more unusual issues e.g. 'what-if' data sets from similar systems are given equal weighting, what if models relationships are adjusted/weighted towards more realistic environmental variance?
- Definition of likely AS process conditions/parameters ranges associated with good and conversely poorer bacterial and viral LRVs;
- Efficient integration and communication of data and findings from complex data analysis in a simple standard format;
- Easy and auditable incorporation of Validation data and priors into models and production of auditable (posterior) LRVs which could be used to obtain LRV credits.

These benefits are similar to those we identified earlier when considering the removal of *Cryptosporidium* and *Giardia* and bacterial and phage indicators (Carvajal et al., 2015). And the two study data sets are complementary to one another.

This new analysis though illustrates that BN application to full scale water treatment systems is no more onerous.

8.4.3. Monitoring and control of AS processes

Historically AS has been managed using a somewhat limited range of monitoring tools and techniques which are far fewer than the tools now available. But the explosion in analytical techniques, databases and on line monitoring begs the question of how to integrate old and new approaches and routinely make sense of the extensive data generated?

The conceptual answer is knowledge management and data mining but like the term 'validation' this begs the question of operational method. There is also the question arising in this project of how to relate validation to other testing activities. The methods applied here illustrate what is possible using a combination of WEKA type data mining and BN based inference/analysis.

8.4.4. Further work

As noted initially the process of developing BNs needs to be iterative and so the nets constructed here much be seen as provisional or initial versions. This said the data appeared to be readily analysed using data mining and BNs and the output estimates were consistent with existing beliefs regarding the magnitude of AS microbial LRVs. So a second round of refinement appears fully justifiable.

While the accuracies achieved were not as great as previously where the *Cryptosporidium* model accuracy exceeded 90% (Carvajal et al., 2015) the latter study analysed a single much more tightly controlled system where the input concentrations of *Cryptosporidium* were also much more tightly defined by virtue of seeding. Additionally the LRV estimates were based on single value in/out calculations which were likely not so we matched in time as with the earlier study. How representative each single inflow measurement was is unclear. These might be resolved by more intensive spot sampling.

Beyond this there is probably a great deal of other AS data available which might be mined bearing in mind other considerations such as systems design and hydraulic parameters which were unavailable in the present instance.

8.4.5. Uncertainties

The data had a number of significant gaps. Some might be easily filled by for example obtaining more of the relevant data from the AS managers for these systems notably temperature oxygen and pH which should be available from in line monitoring. The gap locations were evident in the .CSV file.

'Activated Sludge' describes many different design approaches. But it was unclear how this might have affected the results obtained e.g. how old were plants and were they over or under loaded compared with their design specifications? Were they designed for different retention and hence times? Such information was not available in this first instance.

A number of parameters appeared to reduce or enhance LRVs notably temperature and wastewater nitrogen content – in effect extent of digestion. Intuitively the former seems critical. Unfortunately the data available suggested different plants were sampled at different times of the year confounding the separation especially Oxley (26.9±3.0 C) v. Rosney (18.5±3.5 C).

Further analysis of the available data could benefit from discussions with BN theoreticians as to how far the data could be used for inferential purposes. A clear trap with BN construction is overfitting of data. This can take the form of using too many parameters and too great discretization.

Nevertheless in principle it is clear that BN based analysis can generate LRV probability density functions and these can be used to support decision making in their basic form, after scenario exploration and via using Bayesian Validation techniques.

8.5. Conclusions

- Analysis of the CSIRO AS data proved feasible and the results were informative in the fashion anticipated i.e. the BNs generated well defined LRVs, they facilitated Bayesian Validation and exploration generated useful and testable hypotheses of importance to Validation e.g. the LRV credit assigned will likely depend on specific system design and may vary with temperature (i.e. season).
- The data analyses generated first cut models suitable for further consideration and refinement.
- The BNs were based on variables (nodes) which when combined using semi-naïve Bayesian approaches generated models having respectable accuracy of *ca* 80% in 3 of the 4 LRV sets.
- Manipulation of the BNs by way of exploring AS LRV behaviour was straightforward.
- The BNs generated LRVs consistent with the literature.

9. Glenelg water recycling plant ultrafiltration validation

9.1. Introduction

In June 2015 SA Water undertook to 'revalidate' its water recycling plant at Glenelg. This system is designed to take chlorinated activated sludge supernatant water, store it and finally pass the water through ultrafiltration (UF) membrane assemblies known as 'skids' prior to reuse. There are 9 skids in total, each of which comprises 100 UF modules.

The revalidation involved directly measuring the reduction in feedwater of seeded MS-2 bacteriophage numbers, under typical membrane pressure and other operating conditions, over the course of one *ca* 30 minute filtration cycle per skid tested. Testing was based on a representative number of skids.

For the purposes of this BN analysis, SA Water also measured the reduction in total coliform numbers and concurrently a range of other physico-chemical parameters as grab samples and on-line parameters.

The revalidation looked at 5 skids and measured in/out parameters at 2 (coliform) to 4 (bacteriophage) time-steps in the case of grab samples, and at 30 second intervals in the case of on-line parameters over the course of one filtration cycle per skid.

The Glenelg system was earlier identified as having a rich and diverse initial validation data set and so was likely optimal for an industry based full scale demonstration case study application of Bayesian inference and BN application to recycled water validation.

9.1.1. Why Glenelg revalidation data provided a model case study for validation

The following were the reasons why Glenelg revalidation was viewed as a suitable model for trialling the concept of Bayesian Validation, the application of Bayesian inference and BNs to recycled water validation data, and using the model outputs to estimate final composite LRVs:

- The recycled water treatment system design was relatively simple and well suited to summary description, and hence communication of validation details to third parties.
- It was a full scale system comprising multiple filtration units and subunits.
- It was comparable in design to those understood to be operated by other project partners and water authorities notably in Western Australia.
- The system had been comparably validated in 2010 on three different dates. Further, a small set of manufacturer data on the performance of similar UF membranes had been provided by other project partners. Thus we had 3 independent UF data sets which could be used for modelling trials.
- Nearly the same skids had been used by SA Water in 2010 and 2015 trials and so direct before and after comparisons, all else being equal, was defensible.
- The data sets available were substantial. For example the total number of distinct bacteriophage records (experiments X Skids X timesteps X replicates) was *ca* 200 for the SA validation and revalidation work combined and most of the outlet bacteriophage measurement data were uncensored by virtue of the trials using a phage seeding methodology.
- Each bacteriophage measurement involved concurrent in/out samplings making it possible to obtain a series of individual and replicate LRV estimates directly as well as

calculating them from the in and out concentration measurements (see Melbourne Water ETP case study for a subsequent detailed comparison of the different methods of calculating LRVs).

The overall data was seen as providing a model for:

- Validation data collation and management;
- Developing, comparing and otherwise exploring causal and semi-naïve BN development methods.
- Obtaining primary data on the system's LRV data which SA Water could use to assess potential benefits for them from using BNs.
- Estimating LRVs, and hence appropriate credits, which could be compared with those obtained via standard statistical approaches.
- Identifying critical sources of LRV variance in a quantitative manner.
- Assessing the strengths and limitation of BNs in this role.

9.1.2. Aims of the current document

The aims of this analysis were as follows:

- Document data collation and analysis methodology particularly in respect to BN development;
- Construct and compare causal and semi-naïve BNs;
- Compare qualitative impressions from BN exploration (i.e. varying node range probabilities in a what-if Scenario X applies) with what was indicated by BN and data mining metrics;
- Trial the key software selected for the report, WEKA and Netica, on real data and assess how far water managers trained in using these readily available programs might be able to go with data analysis;
- Provide recommendations in respect to Best Practice validation;
- Demonstrate the strengths and limitations of Bayesian Validation.

9.2. Methods

9.2.1. Framing the validation data analysis task

Our earlier evaluation of how to estimate LRVs (Carvajal et al., 2015) indicated BNs could be used in two ways to characterize water treatment:

- Causal BNs to define and clarify beliefs about how the systems behaved overall and the factors believed to be determining LRVs.
- Naïve and semi-naïve BNs to characterize LRVs and influence one by one.

Semi-naïve BNs were seen as providing the final models for LRV treatment process validation with causal nets being developed for comparison purposes.

Once data had been tabulated, our first task was to identify and test an optimal or near optimal model for quantifying LRVs. Then the model would be cloned and used to individually estimate validation, revalidation and specification LRVs. Finally the LRVs were then related to one another and their outputs composited interactively – which we term Bayesian Validation.

Validation involves “the confirmation that the treatment technology meets the specified performance targets.” We interpreted this to mean operationally: using BNs to learn, define and estimate:

- Prior probabilities – especially the probabilities of different LRV bins in the same and different membrane units based on the data from the initial 2010 validation LRVs and manufacturer claims.
- New (prior probability) evidence in the same fashion - the 2015 revalidation LRVs.
- Revised posterior probabilities for LRV performance e.g.
    - Using the validation data + manufacturer specifications as priors determine whether the revalidation study data achieved the expected treatment degree as judged by comparison and combination (calculation of combined posterior probabilities) of the revalidation data with these other sets.
    - Using the specification data as references, whether the validation study data achieved expectations. Also undertake this for comparison using the 2010 initial validation data.
    - The best system LRVs, based on different apportioning of prior probabilities from the specification, validation and revalidation data sets calculate final LRVs which might be used for claiming LRV credits.
    - (using hypothesis testing) Assess whether the composite LRVs were comparable to or better than specification LRVs and whether the revalidation LRV was comparable or different to the primary validation LRV.

9.2.2. Initial data collation

Data were obtained as spreadsheets and reformatted as a single table with the following fields:

- Experiment (e.g. Validation run 2 = V2);
- Station (sampling station such as Skid outlet);
- Date+time (of measurement);
- Timestep (minutes since beginning of filtration cycle);
- Parameter (e.g. bacteriophage, transmembrane pressure, turbidity);
- Units (pfu/mL);
- Sign (e.g. LT = Less than);
- Measurement (numerical value).

The data collated is summarized in Table 15.

The data fields described in this report were coded by:

- Experiment (Revalidation = RV1; Validation runs = V1, V2, V3; Specification data = M1);
- (Treatment) Unit Skid number 2,4,5,6,8 and Module number 1,2,3,4);
- Timestep (minutes in the range 0 to 30);
- Replicate (1 to 3);
- Inlet concentration - log pfu/mL;
- Outlet concentration - log pfu/mL;
- Point estimate bacteriophage LRV (logs).

The summary statistics for each parameter are tabulated and enumerated in Table 16. An illustrative snapshot of the data as it appears in Excel is shown in Figure 56. As samples were collected at similar timesteps across each filtration cycle but not at identical times they were sorted and matched by timestep. Timestep was coded as a continuous variable to allow for any time of collection related changes.

Table 15. Glenelg UF 2010 Validation (V) and 2015 Revalidation (RV) Parameters and Total Measurements

Experiment	Parameter Code	Number of measurements per station											
		Feedwater B1	Skid2In	Skid2Out	Skid4In	Skid4Out	Skid5In	Skid5Out	Skid6In	Skid6Out	Skid8In	Skid8Out	FiltWater
RV1	Conc_MS2		4	4	4	4	4	4	4	4	4	4	
	LRV_MS2			4		4		4		4		4	
	Conc_coliforms		2	2	2	2	2	2	2	2	2	2	
	LRV_colforms			2		2		2		2		2	
	DOC		2	2	2	2	2	2	2	2	2	2	
	TOC		2	2	2	2	2	2	2	2	2	2	
	UVt		2	2	2	2	2	2	2	2	2	2	961
	pH	961											
	Cl2	961											
	Turbidity	961											961
	Level	961											
	Temperature	961											
	Volume	961											
	FFI			961		961		961		961		961	
	Flow			961		961		961		961		961	
	R			961		961		961		961		961	
	TMP			961		961		961		961		961	
	Fwsetpoint			961		961		961		961		961	
	Cvalve			961		961		961		961		961	
	Tsbackwash			961		961		961		961		961	
V1	Conc_MS2		4	4	4	4	3	3	3	3			
	LRV_MS2			4		4		3		3			
	FCV			31		30		31		21			
	TOC		2	2	2	2	2	2	1	1			
	UVt												1650
	pH	1650											
	Cl2	1650											
	Turbidity	1650	4	4	4	4	4	4	3	3			1650
	Level	1650											
	Temperature	1650											
	Volume	1650											
	FFI			1681		1680		1681		1671		1650	
	Flow			1681		1680		1681		1671		1650	
	R			1681		1681		1681		1671		1650	
	TMP			1681		1681		1681		1671		1650	
	TSS		4	4	4	4	4	4	3	3			
	Fwsetpoint			1650		1650		1650		1650		1650	
	Cvalve			1650		1650		1650		1650		1650	
	Tsbackwash			1650		1650		1650		1650		1650	
V2	Conc_MS2		4	4	4	4	4	4	4	4			
	LRV_MS2			4		4		4		4			
	FCV			30		31		31		31			
	TOC		2	2	2	2	2	2	2	2			
	UVt												1650
	pH	1650											
	Cl2	1650											
	Turbidity	1650											1650
	Level	1650											
	Temperature	1650											
	Volume	1650											
	FFI			1680		1681		1681		1681		1650	
	Flow			1680		1681		1681		1681		1650	

Experiment	Parameter Code	Number of measurements per station											
		Feedwater B1	Skid2In	Skid2Out	Skid4In	Skid4Out	Skid5In	Skid5Out	Skid6In	Skid6Out	Skid8In	Skid8Out	FiltWater
	R			1681		1681		1681		1681		1650	
	TMP			1681		1681		1681		1681		1650	
	TSS		4	4	4	4	4	4	4	4			
	Fwsetpoint			1650		1650		1650		1650		1650	
	Cvalve			1650		1650		1650		1650		1650	
	Tsbackwash			1650		1650		1650		1650		1650	
V3	Conc_MS2		4	4	4	4	4	4	4	4			
	LRV_MS2			4		4		4		4			
	FCV			31		31		31		31			
	UVt												1650
	pH	1650											
	Cl2	1650											
	Turbidity	1650	4	4	4	4	4	4	4	4			1650
	Level	1650											
	Temperature	1650											
	Volume	1650											
	FFI			1681		1681		1681		1681		1650	
	Flow			1681		1681		1681		1681		1650	
	R			1681		1681		1681		1681		1650	
	TMP			1681		1681		1681		1681		1650	
	TSS		4	4	4	4	4	4	4	4			
	Fwsetpoint			1650		1650		1650		1650		1650	
	Cvalve			1650		1650		1650		1650		1650	
	Tsbackwash			1650		1650		1650		1650		1650	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Experiment	Unit	Timestep	Replicate	Inlet	Outlet	LRV	Experiment1	Unit1	Timestep1	Replicate1	Inlet1	Outlet1	LRV1	Experiment2	
2	V1	Skid_2	1	1	5.19	2.32	2.87	RV1	Skid_2	1	1	3.9	1.6	2.3	M1	Module_1
3	V1	Skid_2	1	2	5.23	2.08	3.15	RV1	Skid_2	1	2	3.7	1.48	2.22	M1	Module_1
4	V1	Skid_2	1	3	5.07	2.11	2.96	RV1	Skid_2	7	1	5.05	2.71	2.35	M1	Module_1
5	V1	Skid_2	12	1	5.14	1.78	3.36	RV1	Skid_2	7	2	4.98	2.88	2.11	M1	Module_1
6	V1	Skid_2	12	2	5.18	2.04	3.14	RV1	Skid_2	7	3	5.1	2.79	2.32	M1	Module_2
7	V1	Skid_2	12	3	5.19	1.9	3.29	RV1	Skid_2	14	1	5	2.98	2.02	M1	Module_2
8	V1	Skid_2	20	1	5.3	1	4.3	RV1	Skid_2	14	2	4.96	2.95	2.01	M1	Module_2
9	V1	Skid_2	20	2	5.27	1.78	3.49	RV1	Skid_2	14	3	5.03	2.96	2.07	M1	Module_2
10	V1	Skid_2	20	3	5.31	1.6	3.71	RV1	Skid_2	20	1	5.06	2.94	2.12	M1	Module_3
11	V1	Skid_2	28	1	5.43	1.6	3.83	RV1	Skid_2	20	2	5	2.99	2.01	M1	Module_3
12	V1	Skid_2	28	2	5.59	1	4.59	RV1	Skid_2	20	3	5.01	2.97	2.04	M1	Module_3
13	V1	Skid_2	28	3	5.59	1.48	4.11	RV1	Skid_4	1	1	5.02	3.14	1.88	M1	Module_3
14	V1	Skid_4	1	1	5.15	2.38	2.77	RV1	Skid_4	1	2	4.72	3.17	1.54	M1	Module_4
15	V1	Skid_4	1	2	5.2	2.36	2.84	RV1	Skid_4	1	3	4.91	3.29	1.63	M1	Module_4
16	V1	Skid_4	1	3	5.24	2.2	3.04	RV1	Skid_4	7	1	4.71	3.09	1.62	M1	Module_4
17	V1	Skid_4	12	1	5.48	1.48	4	RV1	Skid_4	7	2	4.8	3.17	1.63	M1	Module_4

Figure 56. Subsection of .CSV file used in learning BNs illustrating datatable format

Table 16. UF Bacteriophage input/output concentrations and LRV estimates (Pivotable)

Data_set	Experiment	Unit	Average of Inlet	Average of Outlet	Average of LRV	StdDev of LRV
Manufacturer specifications	M1	Module_1	6.52	4.01	2.51	0.26
		Module_2	6.48	3.80	2.68	0.29
		Module_3	6.80	4.55	2.25	0.34
		Module_4	7.27	5.28	2.03	0.17
	M1 Average		6.77	4.41	2.37	2.37
	M1 Count		16	16	16	16
	M1 StdDev		0.36	0.62	0.36	0.36
Manufacturer specifications Count			16	16	16	16
Validation	V1	Skid_2	5.29	1.72	3.57	0.55
		Skid_4	5.31	1.68	3.61	0.50
		Skid_5	6.11	4.43	1.80	0.69
		Skid_6	5.70	2.77	3.24	1.05
	V1 Average		5.52	2.64	3.22	3.22
	V1 Count		38	44	38	38
	V1 StdDev		0.36	1.27	0.93	0.93
	V2	Skid_2	5.28	2.50	2.74	0.28
		Skid_4	5.37	2.52	2.86	0.28
		Skid_5	5.32	2.63	2.68	0.40
		Skid_6	5.36	3.03	2.34	0.18
	V2 Average		5.33	2.68	2.65	2.65
	V2 Count		47	46	45	45
	V2 StdDev		0.18	0.36	0.35	0.35
	V3	Skid_2	4.75	1.39	3.36	0.45
		Skid_4	4.73	1.42	3.31	0.24
		Skid_5	4.63	1.06	3.57	0.21
		Skid_6	4.68	1.64	3.04	0.37
	V3 Average		4.70	1.38	3.32	3.32
	V3 Count		48	48	48	48
	V3 StdDev		0.11	0.37	0.37	0.37
Validation Count			133	138	131	131
Revalidation	RV1	Skid_2	4.80	2.66	2.14	0.13
		Skid_4	4.90	3.13	1.77	0.16
		Skid_5	5.16	3.19	1.97	0.12
		Skid_6	5.04	3.00	2.04	0.08
		Skid_8	5.04	3.13	1.91	0.11
	RV1 Average		4.99	3.03	1.96	1.96
	RV1 Count		59	59	59	59
	RV1 StdDev		0.26	0.31	0.17	0.17
Revalidation Count			59	59	59	59

9.2.3. Data table format and BN learning wizard use

Specification and local validation data were able to be tabulated together in a common format. These different data were assembled using MS Excel and saved as .CSV files. Although Netica^{tm} can learn variable (node) names, ranges/states and data bin probabilities using .xlsx files .CSV is less error prone. In line with standard Netica^{tm} practice missing values were coded as '*'.

For details of Netica^{tm} (v 5.18) 'Learning' readers should refer to the Help files with the software or on the Norsys web site (Norsys Software Corporation, 2013). The following outlines some important features of learning which we employed in developing BNs in the present case study.

Among other things Netica[tm] learning facilitates the following, using properly set up .CSV files, mostly via features listed under the 'Cases' menu:

- Node names, data types and (interactively) discretisation can be learnt semiautomatically speeding up coding.
- The nodes can be discretised to any extent manually or automatically (typically the auto-discretisation command divides continuous data into equal percentile interval bins). This is in effect the ZeroR reference model which is used for comparison with naïve Bayes and semi-Naïve Bayes models.
- The construction of one type of semi-naïve Bayes (TAN or Tree Augmented Network).
- The node state/range value bin probabilities can be learnt via simply counting the number of instances of each CPT cell combination and directly calculating the likelihood of a particular combination (i.e. a CPT probability table entry) or indirectly using the 'expectation maximisation' or 'gradient' tools which interpolate CPT probabilities for the general 'Markov Blanket' where limited data is available.
- Where a BN has already been constructed additional, data with the same field names can be learned in order to:
    - Modify the model state/data range probabilities in a statistically reproducible way.
    - Used via the 'Test with cases' to assess whether additional data is consistent with the initial data by seeing if the new data inputs reliably predict a selected target node i.e. undertake model accuracy testing.

Using these tools, changes can be made to the BN *en masse* or for groups of nodes in a given BN. And revision can be done in minutes.

Each change to nodes in effect creates a new/modified model. Such flexibility has the downside that the diverse models can be hard to keep track of and the potential for constructing one or more similar models leads to the question of which is best? A similar problem is encountered when doing multiple regression analysis where different combinations and interactions of variables are trialled to improve model fits.

Some useful principles to minimise the explosion in candidate models and which we employed (see for example discussion of good practice in (Marcot et al., 2006)) are as follows:

- Avoidance of more than 3 parent nodes per child node.
- Avoidance of excessive discretization (Where net probabilities are learned a maximum of 5 and preferably less is advised. The exception is where the node represents a continuous probability distribution, emulating a given PDF like a normal distribution, is being recreated or the node reflects a very large data set).
- Removal of uninformative links (identifiable by for example 'Sensitivity to Findings' analysis).
- Model achieves marked improvement compared with the ZeroR and naiveBayes cousins.

Compared to BNs constructed on the basis of expert opinion of causal links and directions, semi-naïve BNs look for optimal explanatory nets. Like other semi-naïve BNs, the TAN networks introduce new additional links between nodes other than the target node. One great benefit of using the TAN construction tool is that it takes the decision of what is optimal out of the hands of the user and in our experience often yields useful improvements over basic naïve Bayes models. For comparison the naïve Bayes model can then be constructed by simply disconnecting all links other than those from the target node to the other candidate subsidiary nodes and recompiling the now naïve BN.

In respect to learning a BN via instance counting, EM or gradient learning, the greater the number of measurements, the more these three possible approaches generate similar nets and CPTs. EM and gradient learning will tend to produce slightly different nets each run as they use and iterative testing procedure to refine candidate models which can have different starting point. But the differences proved relatively minor i.e. similar accuracy, and predictions

9.2.4. Model LRVs

The LRV statistics illustrated in the model nodes were learned using EM learning as there was insufficient data to generate different concentration PDFs for each experiment. For this reason the alternative of calculating the LRV PDFs directly using the equation and PDF function facilities in Netica[tm] was not undertaken.

The LRVs were treated as being conditionally independent on the inputs and output concentrations though they were estimated individually by subtraction.

For comparison we developed both causal and semi-naïve models which yielded similar LRVs. The semi-naïve TAN models were preferred for final Bayesian validation because they allow the assumption of node independence to be relaxed (Korb and Nicholson, 2011).

9.2.5. WEKA use

As with other case studies our use of Netica[tm] was supplemented with the use of the general data mining software WEKA (Witten et al., 2011).

WEKA includes a wide range of methods for data mining including finding optimal semi naïve Bayesian models which are more powerful than those built in to Netica. However, WEKA also has constraints. The main strengths and limitations of both packages based on our experience reflecting the need for practical uncomplicated Validation are outlined in Table 17. The differences in part likely reflect our inevitable selection biases derived from use and familiarity with these packages and separately the fact they are still being under development. For example Norsys assert they are developing further automatic net construction tools comparable to the TAN tool.

Our conclusion based on the experiences captured in Table 17 was that WEKA was a far more powerful system for identifying the optimum machine learning based model. However, the nets of interest could not be rapidly converted into a working (Netica) BN and WEKA had some further limitations. So we used WEKA in this case study to supplement Netica[tm] modelling and check if the TAN models were close to optimum as judged by accuracy metrics. In another case study for Melbourne Water ETP where there were many more candidate variables WEKA proved useful for identifying those which did not influence target variables of interest.

In applying the methods presented here, the following are emphasised:

- There are other BN packages which are variously more powerful and flexible. The reason for selecting Netica[tm] was its relative simplicity in our hands compared to these more powerful programs and low price which made it a practical entry level program for Bayesian Validation and analysis generally.
- BN methodology and theory are both rapidly evolving so the methods here are seen as effective but not the last word.
- There are reportedly many BN concepts and abilities which are incompletely understood theoretically so data interpretation is necessarily incomplete.
- What metric values reflect ‘good’ ‘fair or problematic models is not immediately clear so these and other quantitative model outputs should be seen as providing decision support

and a consistent basis for inferences e.g. what LRV credits should be allowed, rather than the last word in validation.

- The case study models were first cut/draft models. A key recommendation in Bayesian modelling is that eliciting model data and structure should be an iterative and group effort so our models here are an attempt to do a first cut.

This said as a tool for quantitatively systematizing water treatment process validation Netica^{tm} appeared to be very fit for purpose and clarified process performance very effectively and transparently.

Table 17. Comparison of key Netica^{tm} and WEKA features based on case study modelling experiences

Attribute	Netica	WEKA
Construction of working BNs	+	-
Target nodes can be numerical	+	-
Construction of any BN for testing	+	-
Easy direct conversion to Netica	n/a	-
Netica^{tm} model can be rapidly revised based on evaluation	++	-
Diverse learning algorithms	+	+++
Arc reversal (arc from subsidiary node to target node)	-	+
Diverse semi-naïve machine learnt BNs possible	TAN only (v 5.18)	+++
Accuracy metrics	+	+++
Optimal accuracy metrics	+	++
Accuracy testing – model v. test data set	+	+++
Sensitivity to parameters	±(manual)	++
Sensitivity to findings	+	?
Automated splitting into model building and test data sets	-	+++
Other data mining tools and programs	-	+++

9.3. Results and Discussion

9.3.1. Causal Model

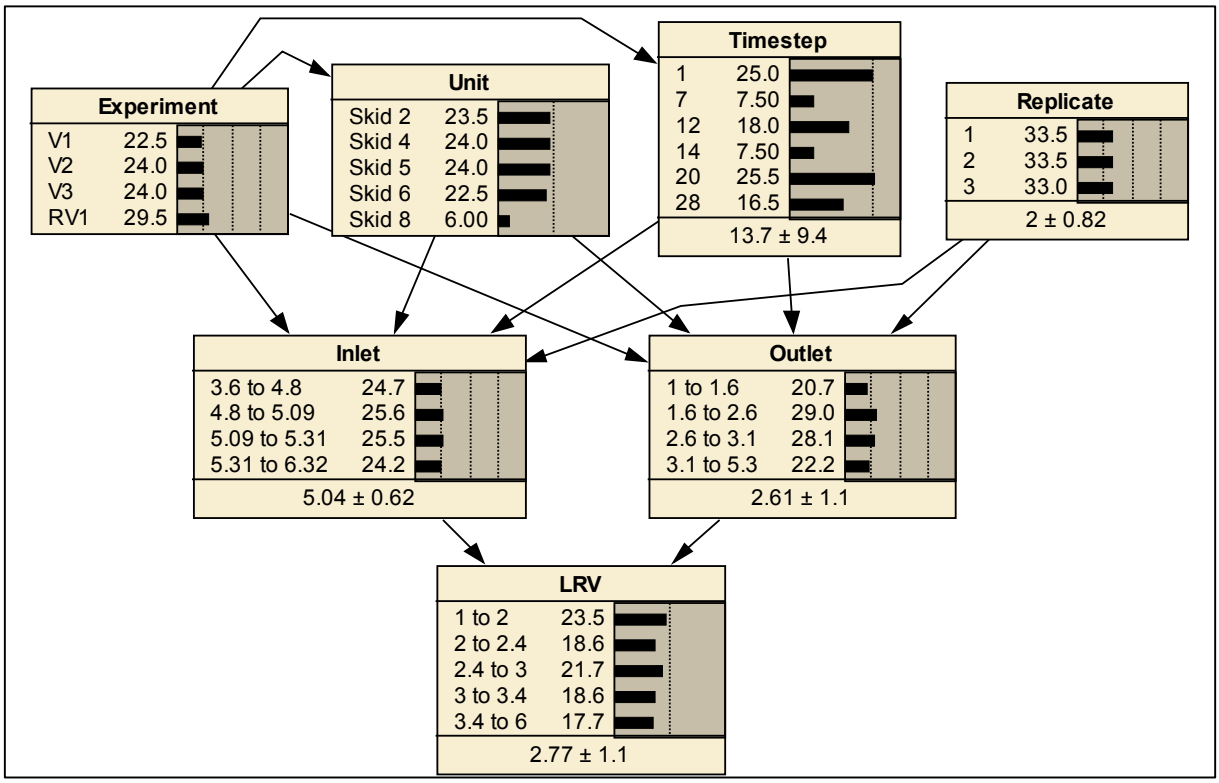
Validation v. Revalidation

Figure 57a shows the basic causal model constructed based on the assumption that the different experiments, units, timestep and replicate measurements could independently influence inlet and outlet bacteriophage concentrations, and from there LRVs, in different ways and extents. The LRV average can be seen to be 2.77.

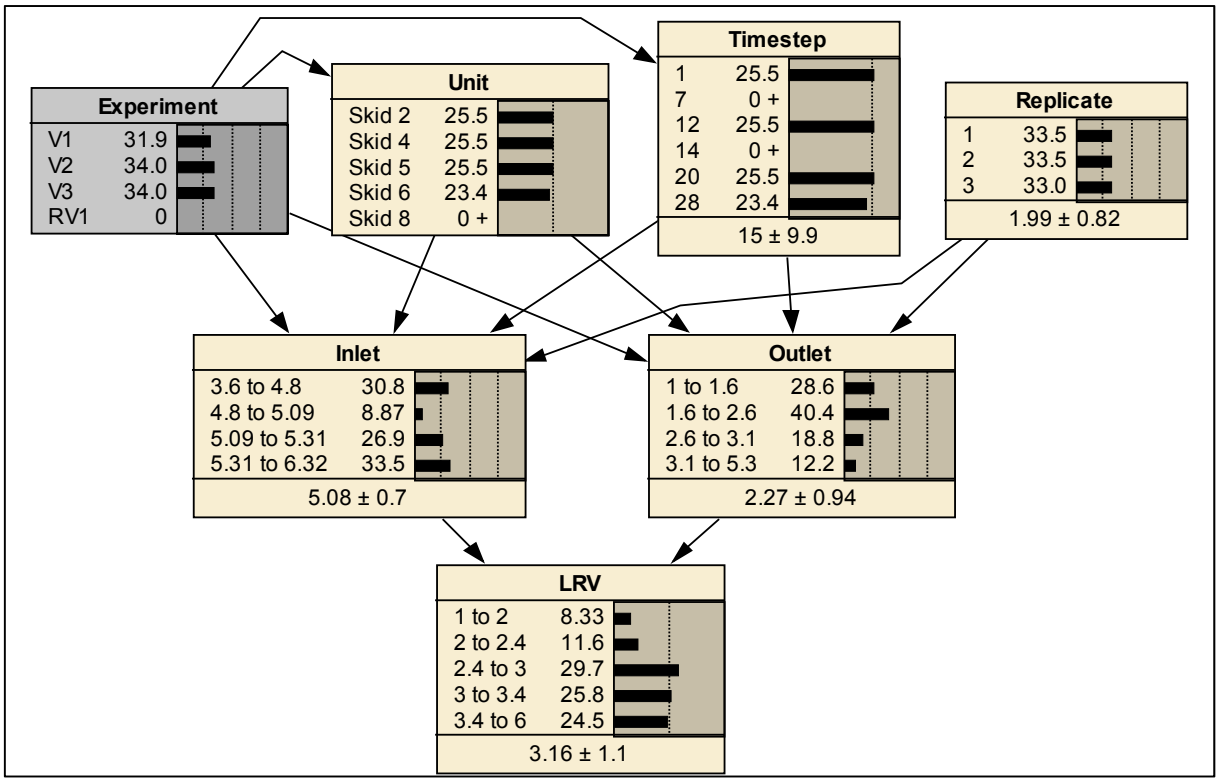
In Figure 57b and Figure 57c we have respectively excluded and isolated the revalidation data set. The nets suggest that very different removal was occurring in 2010 v. 2015 though the input concentrations were similar.

These differences were similar to those seen in summary statistics (Table 16). The main difference was that the information could be accessed interactively and the fine details in terms of bin probabilities were straightforward to see and compare.

a.


Timestep
1 25.0
7 7.50
12 18.0
14 7.50
20 25.5
28 16.5
13.7 ± 9.4
Experiment
V1 22.5
V2 24.0
V3 24.0
RV1 29.5
Unit
Skid 2 23.5
Skid 4 24.0
Skid 5 24.0
Skid 6 22.5
Skid 8 6.00
Replicate
1 33.5
2 33.5
3 33.0
2 ± 0.82
Inlet
3.6 to 4.8 24.7
4.8 to 5.09 25.6
5.09 to 5.31 25.5
5.31 to 6.32 24.2
5.04 ± 0.62
Outlet
1 to 1.6 20.7
1.6 to 2.6 29.0
2.6 to 3.1 28.1
3.1 to 5.3 22.2
2.61 ± 1.1
LRV
1 to 2 23.5
2 to 2.4 18.6
2.4 to 3 21.7
3 to 3.4 18.6
3.4 to 6 17.7
2.77 ± 1.1

b.


Timestep
1 25.5
7 0 +
12 25.5
14 0 +
20 25.5
28 23.4
15 ± 9.9
Experiment
V1 31.9
V2 34.0
V3 34.0
RV1 0
Unit
Skid 2 25.5
Skid 4 25.5
Skid 5 25.5
Skid 6 23.4
Skid 8 0 +
Replicate
1 33.5
2 33.5
3 33.0
1.99 ± 0.82
Inlet
3.6 to 4.8 30.8
4.8 to 5.09 8.87
5.09 to 5.31 26.9
5.31 to 6.32 33.5
5.08 ± 0.7
Outlet
1 to 1.6 28.6
1.6 to 2.6 40.4
2.6 to 3.1 18.8
3.1 to 5.3 12.2
2.27 ± 0.94
LRV
1 to 2 8.33
2 to 2.4 11.6
2.4 to 3 29.7
3 to 3.4 25.8
3.4 to 6 24.5
3.16 ± 1.1

c.

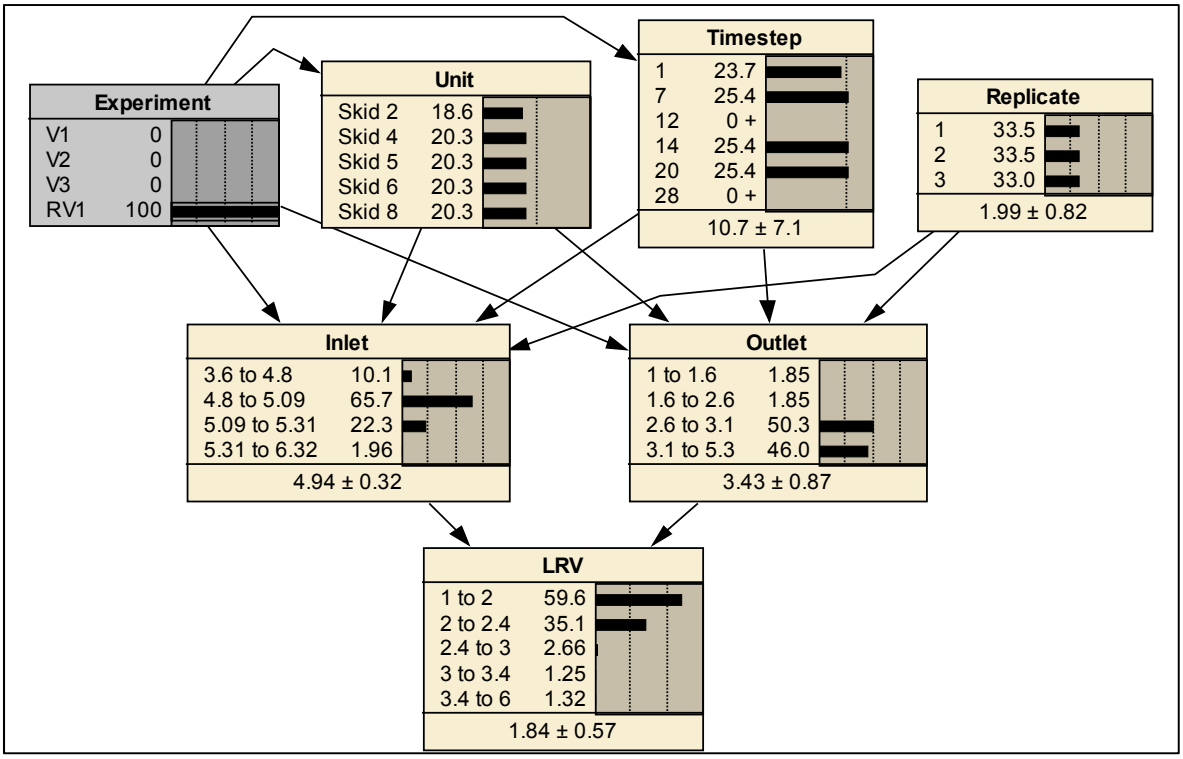


Figure 57. Primary causal BN for 2010 and 2015 data sets combined (a.) and individually (b.,c.)

Qualitative assessment

From varying the root nodes and individual probability values the following were provisionally inferred:

- LRVs varied markedly between different experiments and particularly between 2010 and 2015 with V3 being superior to RV1 by *ca* 1.8 logs.
- Skids 2 and 4 appeared to perform markedly better than 5 and 6 and especially unit 8. Most skids displayed high variance as judged by the LRV standard deviation being typically >0.5 log units.
- Timestep LRV variance change was modest and did not show any clear decrease or increase. In the 2010 validation there appeared to be improvement with time but this was not seen with the revalidation data.
- Replicates were no different to one another.

Sensitivity Analyses

Sensitivity to Findings analysis allowed the relative magnitude of factors associated with LRV variance to be compared overall (Figure 58a) and for subsets most notably 2010 (Figure 58b) v. 2015 (Figure 58c).

The qualitative patterns were seen with the Sensitivity to Findings overall (Figure 58a). LRV variance was particularly strongly associated with different Experiments. That said most of the LRV variance was accounted for by outlet concentration variance.

The order of parent variable importance/influence was Experiment > Timestep > Unit.

Replicates again had no influence suggesting future validation could reduce replicate numbers and reassign monitoring resources to quantifying other variance.

The ‘Experiment’ variance reduction appeared most associated with the difference between the 2010 and 2015 data.

Once 2010 v. 2015 data had been separate as a consideration, timesteps became essentially irrelevant.

In the 2010 data the influence of between unit variance also reduced greatly, however unit variance remained relatively high in the case of the revalidation data.

Our plain English interpretation was as follows:

- There was marked variance in LRVs reflected in the standard deviations of 1.1 log units.
- This variance reflected variance in the outlet concentrations much more than inlet concentrations and so presumably reflected variable membrane performance.
- After this, the LRV varied most according to Experiment with the revalidation data being most concerning with an average reduction in performance between 2010 and 2015 or 1.3 log units.
- The poorer revalidation performance was seen in all 5 skids tested.
- There were marked differences in performance on different occasions for reasons which were unclear. This was not solely due to different systems as the LRV for V2 and V3 differed by 0.9 log units.
- After accounting for other factors sampling replication – timestep and replicates were assessed as minor influences on LRV.

a.

Sensitivity of 'LRV' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV	1.263	100	2.31321	100	0.6365906
Outlet	0.8693	68.8	1.04947	45.4	0.1328973
Experiment	0.4638	36.7	0.54058	23.4	0.0413679
Timestep	0.2612	20.7	0.27826	12	0.0239228
Inlet	0.2195	17.4	0.25003	10.8	0.0178364
Unit	0.1376	10.9	0.19139	8.27	0.0228436
Replicate	0.000932	0.0738	0.00252	0.109	0.0001509

b.

Sensitivity of 'LRV' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV	1.14	100	2.18173	100	0.5894540
Outlet	0.6898	60.5	0.94755	43.4	0.1106685
Experiment	0.1441	12.6	0.20098	9.21	0.0111264
Unit	0.08358	7.33	0.14273	6.54	0.0070945
Inlet	0.07998	7.01	0.13250	6.07	0.0121056
Timestep	0.0401	3.52	0.04015	1.84	0.0027381
Replicate	0.001188	0.104	0.00261	0.12	0.0001988

c.

Sensitivity of 'LRV' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV	0.3277	100	1.27529	100	0.2951683
Outlet	0.1583	48.3	0.38840	30.5	0.0667233
Unit	0.05855	17.9	0.20043	15.7	0.0325560
Inlet	0.01058	3.23	0.14410	11.3	0.0026729
Timestep	0.01025	3.13	0.10554	8.28	0.0066290
Replicate	0.001449	0.442	0.03535	2.77	0.0016566
Experiment	0	0	0.00000	0	0.0000000

Figure 58. Sensitivity of LRV (causal net) to other parameters for a. All data, b. 2010 Validation data only c. 2015 Revalidation data only

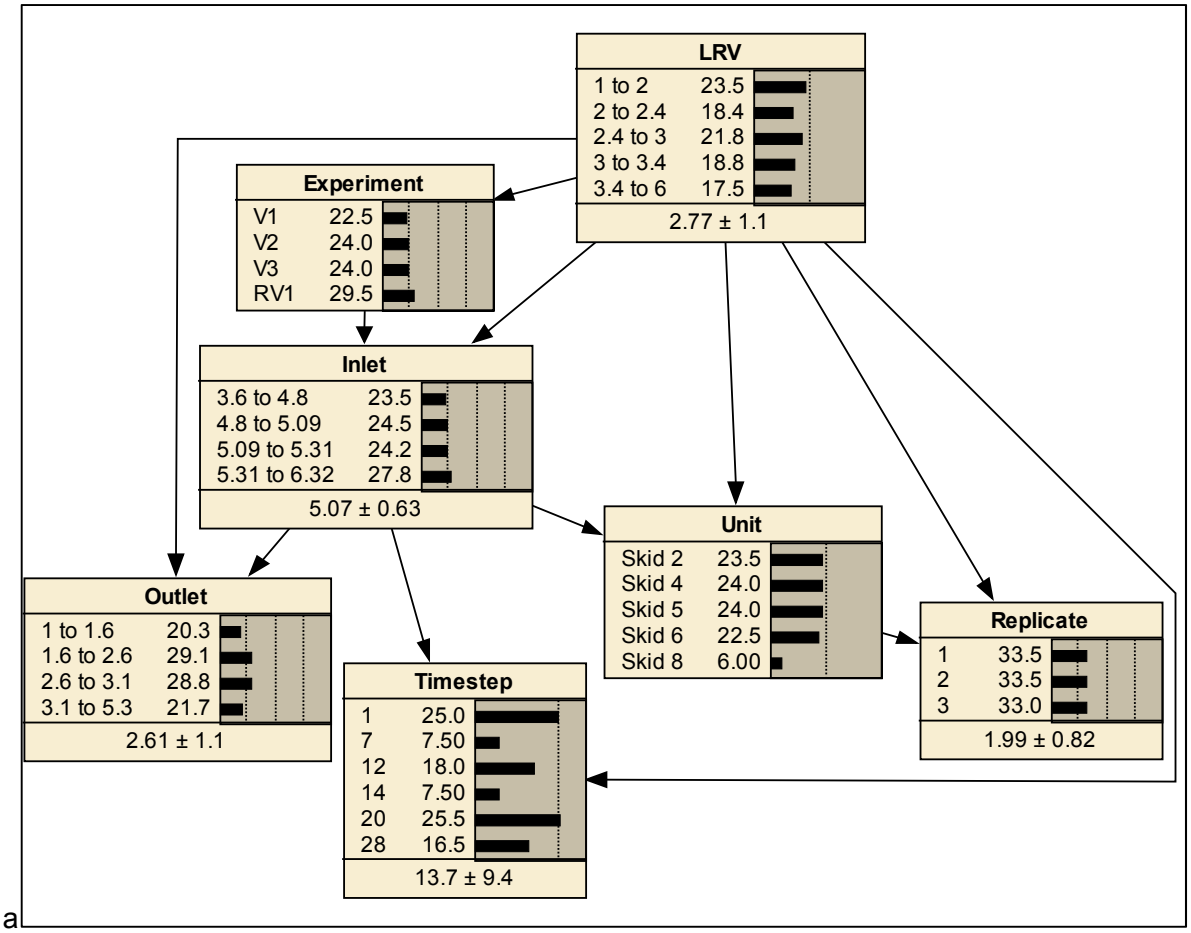
9.3.2. Semi-Naïve Modelling

Netica^{tm} TAN model

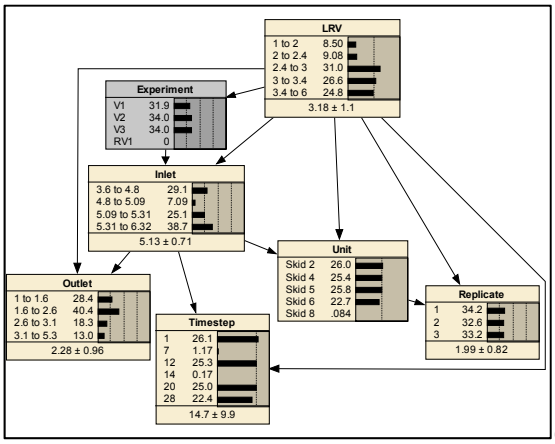
Characterising/explaining a variable is often better achieved by naïve and semi-naïve BNs. So we constructed a TAN model using LRV as the target node. The models and sensitivity analyses are shown in Figure 59 and Figure 60 respectively. From Figure 60, approximately the same pattern as with the causal nets can be seen. Some other features noted were:

- In a given experiment, individual unit LRVs were prone to vary markedly.
- This was most noticeable with V1 where the LRVs ranged from 3.96 to 1.73.
- This between unit variance was in fact least between revalidation nodes. So although the revalidation skids performed less satisfactorily overall their performance appeared to be much more homogenous/consistent.

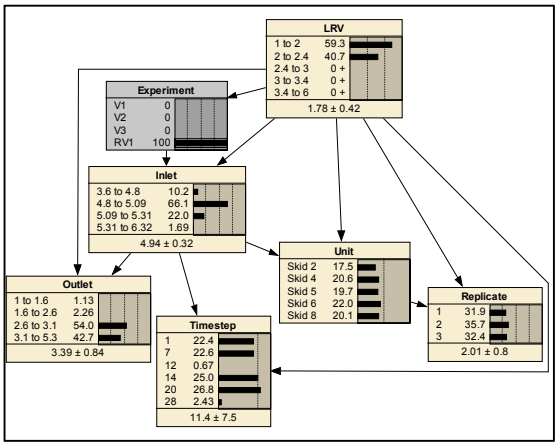
Many of the statistics and observations above could be generated using conventional means. However the BNs allowed the whole LRV picture to be captured in one platform and in a clear graphic format. In our opinion this is invaluable for communication, discussion and decision support in respect to achieving concurrence on LRV credits provided the statistics are representative and not to a major degree artefacts arising from mistakes such as model overfitting.



a



b.



c.

Figure 59. TAN BN learnt from a. combined validation and revalidation data sets and b. validation and c. revalidation sets alone

a.

Sensitivity of 'LRV' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV	1.256	100	2.31276	100	0.6364154
Outlet	0.863	68.7	1.05046	45.4	0.1359544
Experiment	0.5408	43.1	0.81714	35.3	0.0517544
Timestep	0.2874	22.9	0.35419	15.3	0.0345201
Inlet	0.2289	18.2	0.25455	11	0.0176715
Unit	0.1406	11.2	0.34497	14.9	0.0293741
Replicate	3.27e-05	0.0026	0.01475	0.638	0.0008939

b.

Sensitivity of 'LRV' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV	1.133	100	2.14763	100	0.5774482
Outlet	0.707	62.4	1.01003	47	0.1227096
Experiment	0.1909	16.9	0.43408	20.2	0.0204279
Inlet	0.1444	12.7	0.24337	11.3	0.0185029
Unit	0.08301	7.33	0.26001	12.1	0.0183052
Timestep	0.07338	6.48	0.16062	7.48	0.0190687
Replicate	0.00076	0.0671	0.01951	0.909	0.0021148

c.

Sensitivity of 'LRV' to a finding at another node:

Node	Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
LRV	0.1731	100	0.97479	100	0.2413106
Unit	0.06405	37	0.48043	49.3	0.1307114
Outlet	0.0397	22.9	0.27847	28.6	0.0810274
Timestep	0.01604	9.27	0.10685	11	0.0327423
Inlet	0.003249	1.88	0.02402	2.46	0.0066291
Replicate	0.0002872	0.166	0.00175	0.179	0.0005860
Experiment	0	0	0.00000	0	0.0000000

Figure 60. Sensitivity of TAN model LRV to other controlling parameters a. Overall, and b. validation and c. revalidation sets alone

Semi/naïve Bayes model assessment – WEKA

To assess semi/naïve model accuracy we used WEKA to create ZeroR (no links), naïve Bayes, TAN and a BAN model and then assessed accuracy using the standard cross fold approach available in WEKA explorer.

Prior to doing this we divided the LRVs into 4 categories representing the range bands corresponding to 0-25th, 25-50th, 50-75th and 75-100th percentiles. This is because WEKA BN classification cannot handle numerical formats with the target node. These classes were comparable to the auto-discretised ranges generated previously.

The detailed WEKA results are shown in the supplementary information section. The ZeroR model only yielded an accuracy of 28% close to the expected 25%. The naïve Bayes however achieved 75% with the best semi-naïve BNs models achieving 79%. The TAN model achieved 76% accuracy. The main statistics are shown in Figure 61.

```
=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.943     0.059      0.862     0.943     0.901      0.977    LT210
                 0.605     0.055      0.765     0.605     0.675      0.924    EQ210_274
                 0.596     0.141      0.583     0.596     0.589      0.865    EQ274_328
                 0.848     0.07       0.796     0.848     0.821      0.968    GT328
Weighted Avg.    0.757     0.081      0.755     0.757     0.753      0.935

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 50  2  1  0 |  a = LT210
  5 26 12  0 |  b = EQ210_274
  3  6 28 10 |  c = EQ274_328
  0  0  7 39 |  d = GT328
```

Figure 61. WEKA cross validation statistics for TAN model

Note: LT210 means LRVs less than 2.10, EQ210_274 means LRVs in the range 2.10 to 2.74 etc.

It can be seen in addition to achieving respectable accuracy that the average ROC is high and in the confusion matrix (predicted v. actual LRVs) most predictions are incorrect by only 1 step.

We also looked at model accuracy if LRVs were assigned to only two categories. The TAN and BAN models performed essentially identical and accuracy was 90% in both cases.

From this we concluded the TAN models were a good representation of LRV levels and variance in response to different experiments and skids etc.

9.3.3. Bayesian Validation model net

Figure 62 summarises Bayesian Validation using BNs. The TAN model is reproduced three times, one to represent the 2010 LRV and other initial validation data (orange), on to represent the 2015 revalidation data (violet) and a third to represent manufacturer LRV data (yellow).

The 2010 and 2015 models were constructed simply by selecting the appropriate ‘Experiment’ categories from the TAN model above. For the manufacturer data we have assumed the LRVs reflect the same factors and analogous measurements (in fact the manufacturer LRV could have been entered here in isolation but we were interested in exploring how varying the settings in this node might change the corresponding LRV bin probabilities).

The manufacturer subnet probabilities were learned using the primary data available and the “Cases > Learn > Learn using EM” wizard after modifying the Experiment, Treatment Unit and Replicate node states and modifying the discretization thresholds.

In short we copied the TAN model previously developed and used it to provide the common design for different validation LRV definition.

The different LRVs were then combined into a (posterior) composite LRV (blue) in proportion to the weighting node (Data Relative Contributions %). In the example shown “manufacturer specifications”, “Validation results”, and “Revalidation results” were assigned relative weightings of 7.4%, 65% and 23% respectively. These proportions reflect the number of measurements in each of the 3 different data sets.

The algorithm and coding used in the Composite LRV is shown in **Figure 63**. Essentially what the node does is add the relative proportions in each bin coming from the submodel LRVs and adjust for their relative contribution.

Compositing could also have been done via options inbuilt in Netica^{tm} e.g. 'Fading', variations on 'Learning' not discussed here; but the method shown proved more flexible so we have only shown it only.

Once a composite 'posterior' LRV node had been created it could also be compared with the individual submodel LRVs. In this case we included the manufacturer LRV alone and the revalidation LRV alone along with a test of how much they differed probability wise.

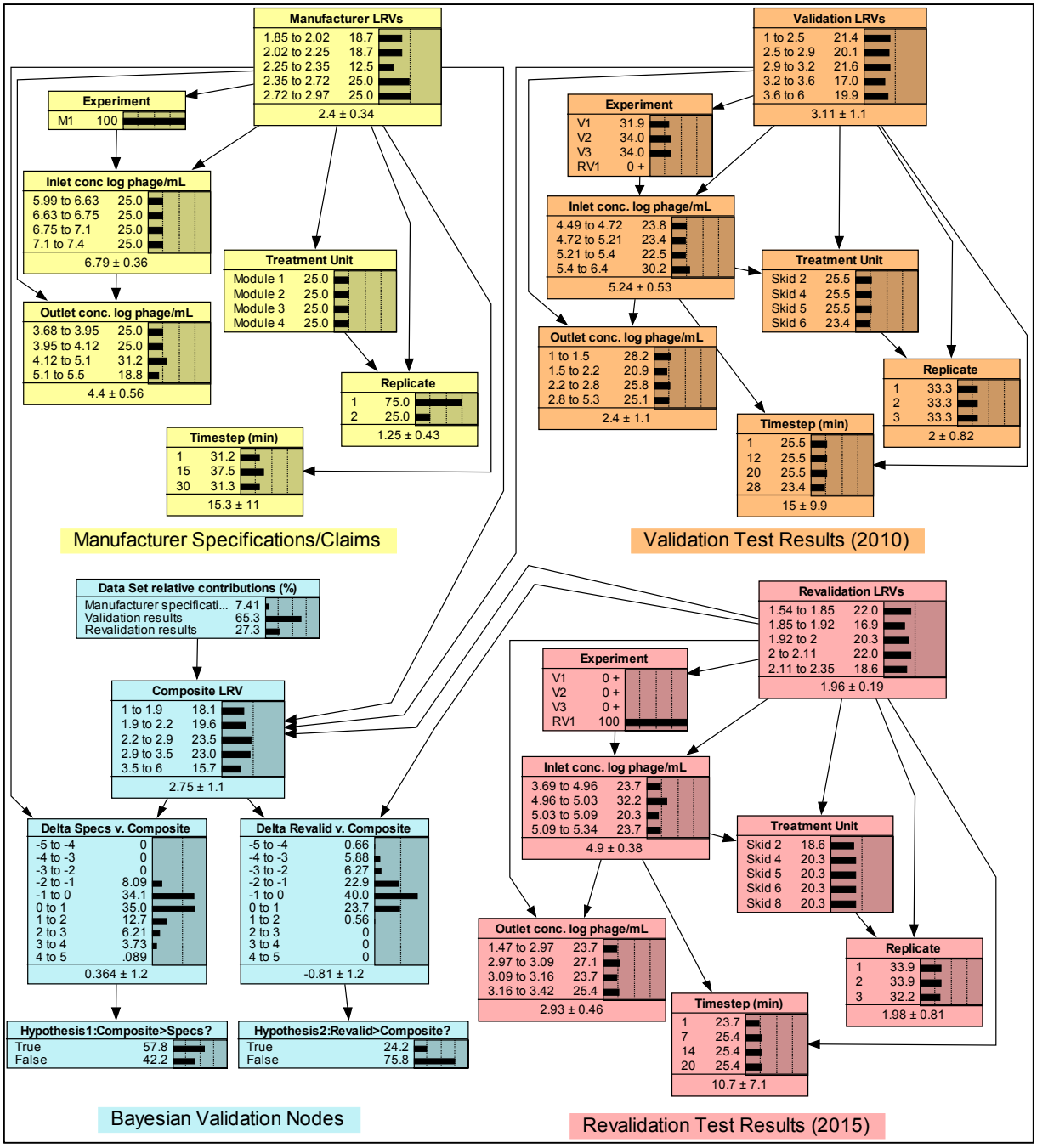


Figure 62. Primary UF 'Bayesian Validation' BN

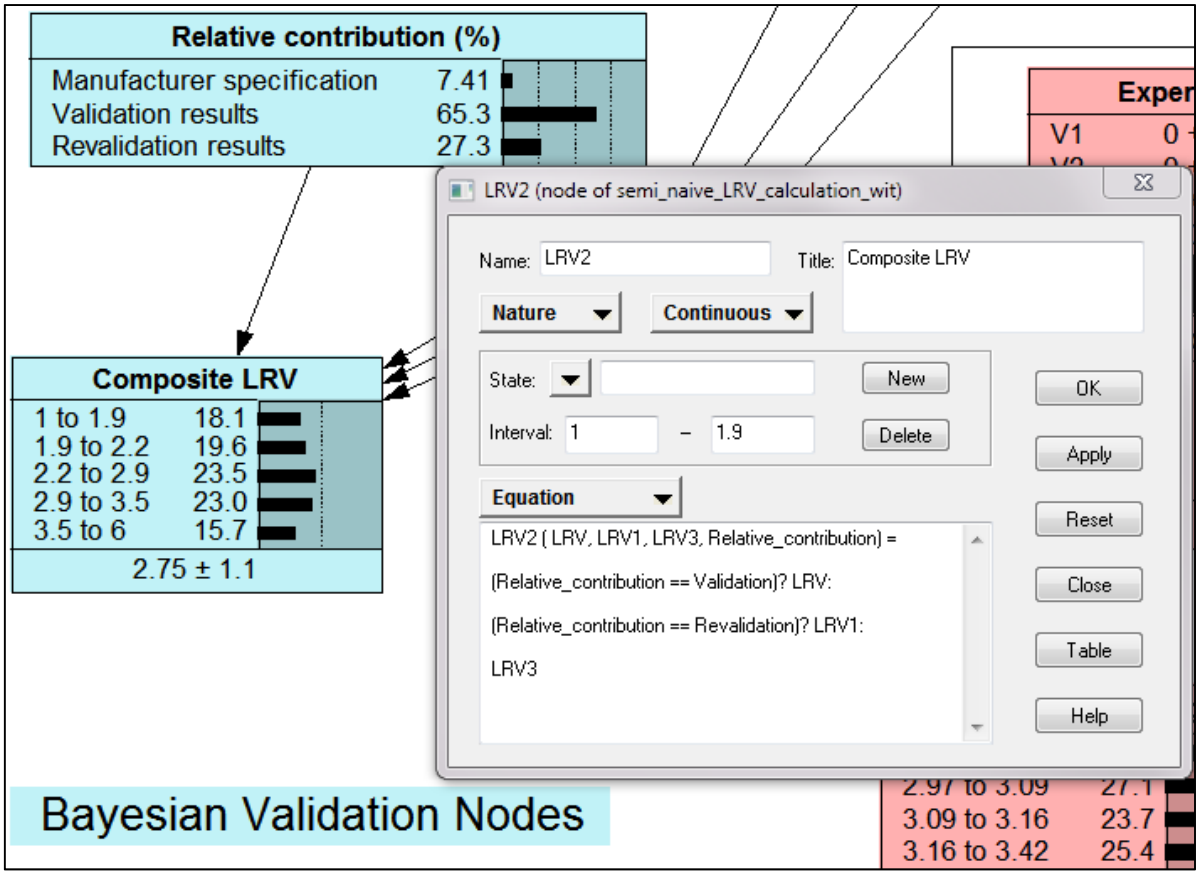


Figure 63. LRV compositing algorithm

Influence of relative importance values

Central to calculating the final posterior validation LRVs was the relative weighting of the submodel LRVs (Figure 64a). While assigning emphasis on the basis of numbers of records is clearly defensible process, skid behaviour may change over time or opinions on how individual modules reflect full skids may vary.

Two other proportioning options are shown in Figure 64b (equal weighting) and Figure 64c (increasing emphasis on more recent revalidation work and larger scale). Figure 64a also shows the accessibility of detailed summary statistics which vary with each scenario being examined.

Moving toward the latter allocation schemes can be seen to reduce the prospective LRV credit from 2.75 logs to 2.32 logs.

Figure 65 illustrates some features of the apportioning process:

- Figure 65a shows where the original apportioning may be set. This is in a table, different to the CPT but accessed also via the node dialogue box and 'Table' option. It can be seen here how the number of records underlying each submodel was 16, 141 and 59. These were converted automatically into the %s seen.
- Figure 65c shows the (right mouse click accessed) 'Enter finding > calibration' option. This yields dialogue boxes requesting a new probability for each of these (3) relative contribution states.

- Figure 65b shows the last one of these for revalidation results – 33.4% as the change is being made.

Thus it is possible to rapidly explore the effect of varying beliefs of how much emphasis should be placed on different information sources.

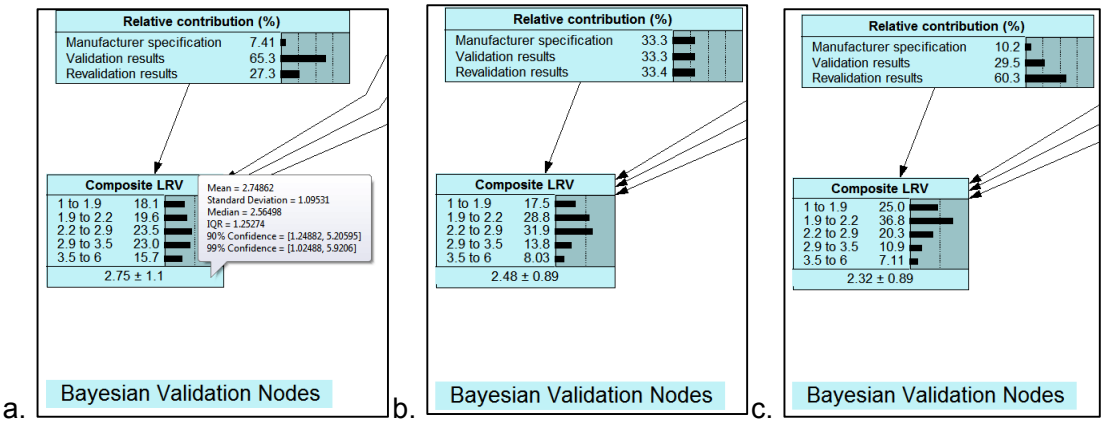
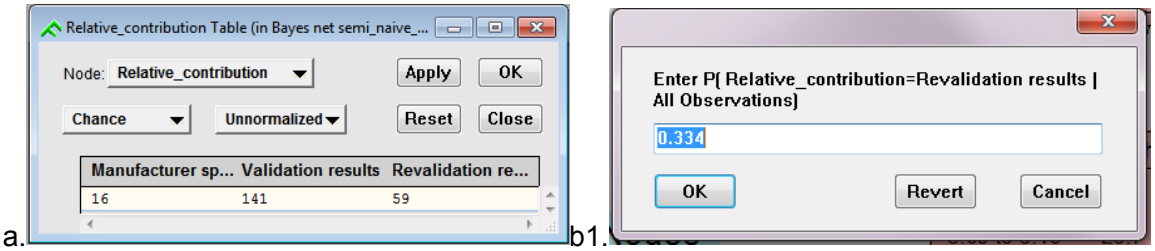


Figure 64. Three illustrative weightings from different data sources



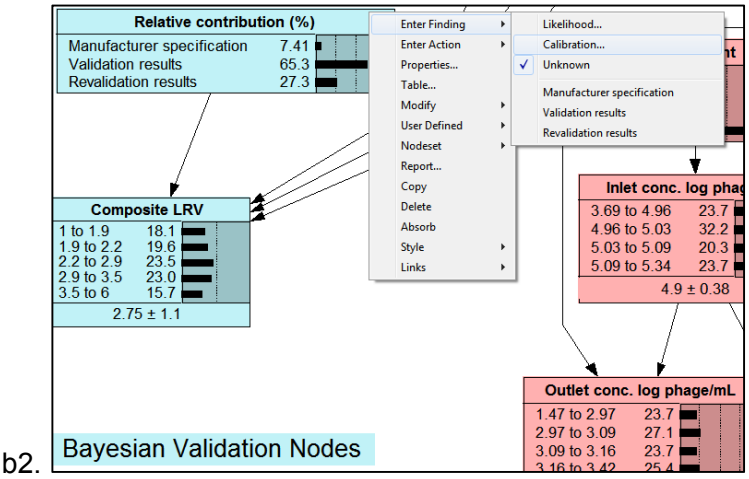


Figure 65. Illustration of weighting method

Hypothesis testing

A final question likely to arise as part of validation is given old priors and new evidence how different is or how likely there is a difference between old or expected LRVs and new LRVs collected for example via revalidation? BNs make answering these questions simple by virtue of their ability to combine different node probabilities.

In Figure 66a the question asked is whether the composite LRV was less than, greater than or comparable to the LRV calculated from manufacturer specifications and the new revalidation data (2 lower branches respectively). The differences (delta) in the LRVs are calculated and the likelihood they were < 0 was calculated. It can be seen that the composite and manufacturer specification LRVs were virtually identical but the revalidation data LRVs were less with a likelihood of 75%.

Similarly in Figure 66b the composite LRV was set to be identical to the 2010 validation data to ask the same questions but in relation to the 2010 data alone. It can be seen that the 2010 performance likely exceeded the performance indicated by the manufacturer but revalidation performance was likely to be less than during the initial validation trials.

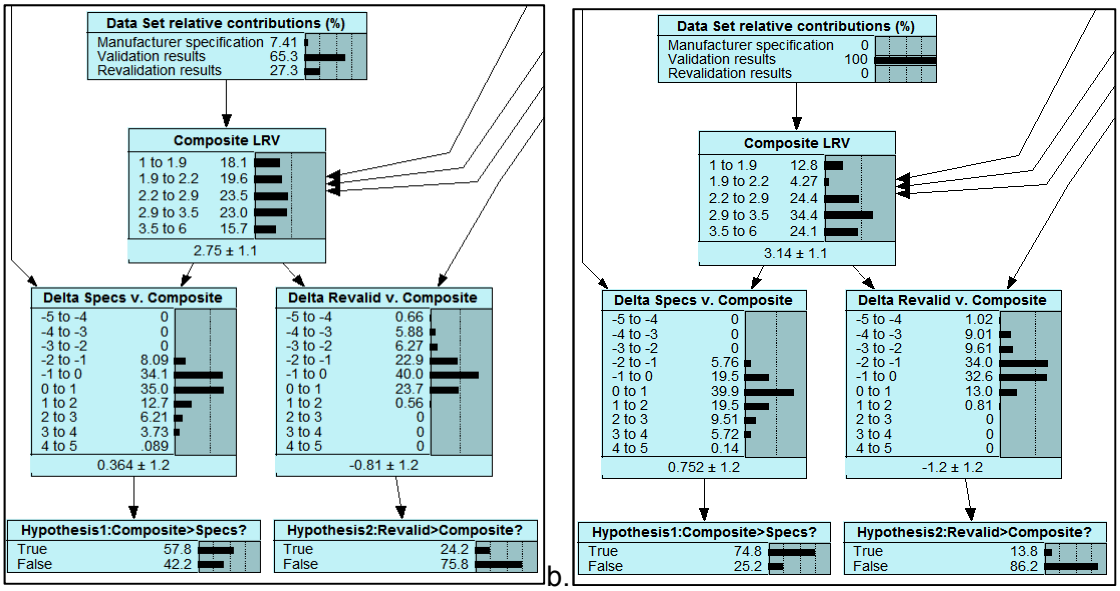


Figure 66. Hypothesis testing – comparison of a. overall composite and specification LRV and b. initial validation and revalidation LRVs

10. Validation, hazardous events scoping and Fault Tree Analysis

10.1. Introduction

NatVal 2.2 Deliverable 4 is the *"Development of a rigorous basis for the incorporation of potential hazardous events (i.e., non-ideal operational conditions) and performance failures in the validation process."* From our general analysis of the uses of Bayes Nets we proposed:

- Hazardous events, breakdowns etc. can be viewed as simply extremes in the probability ranges of Bayes Net nodes (variables); or
- Hazardous events can be introduced into BNs as separate nodes (variables) so as to affect risk probabilities to the appropriate extent.
- As with all scenario analysis with BN Hazardous event incorporation simply involves the introduction of '**new evidence'** into a starting Bayes Net constructed using **prior** information to calculate the **posterior** BN states such as elevated risks and the impact of management implementation.
- Management options can also be added to a primary net or explored to assess how best to address such events e.g. what is most cost effective. Indeed BNs include the option of including special optimization nodes (Decision and Utility nodes) specifically developed with this in mind.

10.1.1. Hazardous scenarios

These potential uses beg the question of how to characterize hazardous events, place them in context and relate them to one another and normal water recycling system operation in the first place.

BNs can aid this aspect of recycling process validation. Prior to incorporating consideration of hazardous events in validation the following activities must be undertaken:

- The vulnerabilities of a recycled water system to hazardous events must be identified and assessed especially though the identification of critical control points;
- Plausible events need to be identified, characterized, quantitatively if possible, and prioritized with a view to formal management and risk assessment.

One screening approach is application of Consequence X Likelihood analysis (Nadebaum et al., 2004 Tool B29, IEC/ISO, 2009) which is well suited for scoping hazardous events. As illustrated elsewhere this matrix analysis can be undertaken using Bayes Nets.

Beyond this 'Tier 1' approach are various (ISO approved) risk scenario analysis tools, including Failure mode effect analysis (B13), Fault tree analysis (B14) (FTA) and Event (*sic*) tree analysis (B15) (ETA) (IEC/ISO, 2009). FTA and ETA are particularly notable as they involve the construction of tree style flow diagrams of the kind which BNs are ideal for and they involve probability calculations.

10.1.2. FTA Analysis

"FTA is a technique for identifying and analysing factors that can contribute to a specified undesired event (called the top event). Causal factors are deductively identified, organized in a logical manner and represented pictorially in a tree diagram which depicts causal factors and their logical relationship to the top event."(IEC/ISO, 2009). ISO's illustrated example is reproduced in Figure 67. The key features this diagram illustrates include:

- How the FTA integrates and relates diverse faults (hazardous events) influencing the key management or system activity, in this case emergency generator failure.
- The use of 'AND gates' where multiple faults must occur concurrently.
- The use 'OR gates' where the top, or an intermediary event, may have more than one cause and the probabilities must be combined in a probabilistically coherent manner.
- The cascading that occurs from some events e.g. circuit A fault > control module fault > fault in reception of signal > no start-up signal.
- How data may be introduced from other sources (e.g. other events).

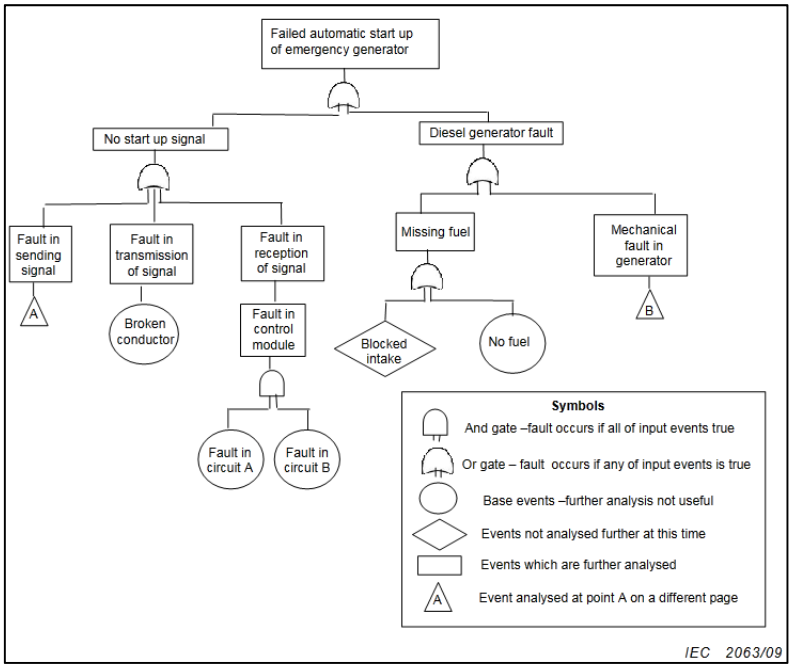


Figure 67. FTA Example - Figure B.2 Example of an FTA from IEC 60300-3-9 (IEC/ISO, 2009)

This current chapter/section illustrates how BNs can be used to relate, quantify and prioritize recycled water treatment system fault (=potentially hazardous event) occurrence using an indirect potable recycled water FTA case study. Specifically it reproduces a basic, classical FTA developed by Lindhe et al.(2012b) using a Bayes Net and shows how the latter can be used to define, scope and prioritize system failure events.

10.2. Methods

10.2.1. A case study in FTA analysis - the Gothenburg, Sweden water supply system

Gothenburg's water supply has been the subject of much study of water quality risks. We were familiar with it from a risk assessment project undertaken in 2006 (Nilsson, 2006) associated with the EU MicroRisk project (Medema et al., 2006).

Among other things this system has been the subject of detailed conventional FTA and the reports and papers arising provide extensive details on hazardous event types, their likelihoods and interrelationships (Lindhe et al., 2009, Lindhe et al., 2010, Swartz et al., 2010, Lindhe et al., 2012b).

The Gothenburg water supply system (Nilsson, 2006) is, as in many places in Europe, involves a degree of unintentional indirect potable water recycling. Thus it is appropriate as a case study here.

The primary water supply is a river bank (Göta älv.) offtake located at Lat. 57.764093° N, Long.12.004452° E, 6.5 km north of the city centre. The river receives discharges from communities, industries and agriculture, and functions to supply raw water supply for about 700 000 people. The system at Gothenburg itself is focused around two water treatment plants Alelyckan and Lackereback. Depending on demand and quality, water can be transferred between reservoirs, treatment plants and the river.

Though the catchment above Gothenberg is mainly wooded there are anthropogenic contaminant sources upstream. A general view of river water contamination can be found here (Swedish EPA, 2009).

Conventional and enhanced FTA has been undertaken by Lindhe et al.(2012b). Figure 68 shows their first of three basic FTAs involving AND and OR gates where:

- *An AND-gate is used to model events that must occur simultaneously in order for the output event to occur. The AND-gate corresponds to a parallel system where the probability of failure is calculated as the product of the n independent events' probabilities* (Lindhe et al., 2009).
- *An OR-gate occurs if at least one of the input events occurs. The OR-gate corresponds to a series system with n independent events* (Lindhe et al., 2009)

The algorithms for calculating basic AND and OR gate probabilities assuming events are ergodic (https://en.wikipedia.org/wiki/Ergodicity) are described by Lindhe et al.(Lindhe et al., 2009, Lindhe et al., 2012b):

- $P(F_i) = \frac{\lambda_i}{\lambda_i+\mu_i}$ (Equation 1)
- $P(F)$ (at basic AND gate) = $\prod_i P(F_i)$ (Equation 2)
- $P(F)$ (at basic OR gate) = 1 - $\prod_i(1-P(F_i))$ (Equation 3)

Where:

- $P(F_i)$ = Probability of (a specific) failure (event);
- $P(F)$ = Probability of multiple (i) antecedent independent events, any of which (OR gate) may induce the higher tier failure event, or all (AND gate) of which are required to induce the higher tier failure event.

- λ = a constant = mean failure rate assuming failure times follow exponential PDFs
- $1/\lambda$ = mean time to failure
- μ = repair rate assuming repair times follow exponential PDFs
- $1/\mu$ = mean downtime

For further information see Lindhe et al. (2012b, 2009).

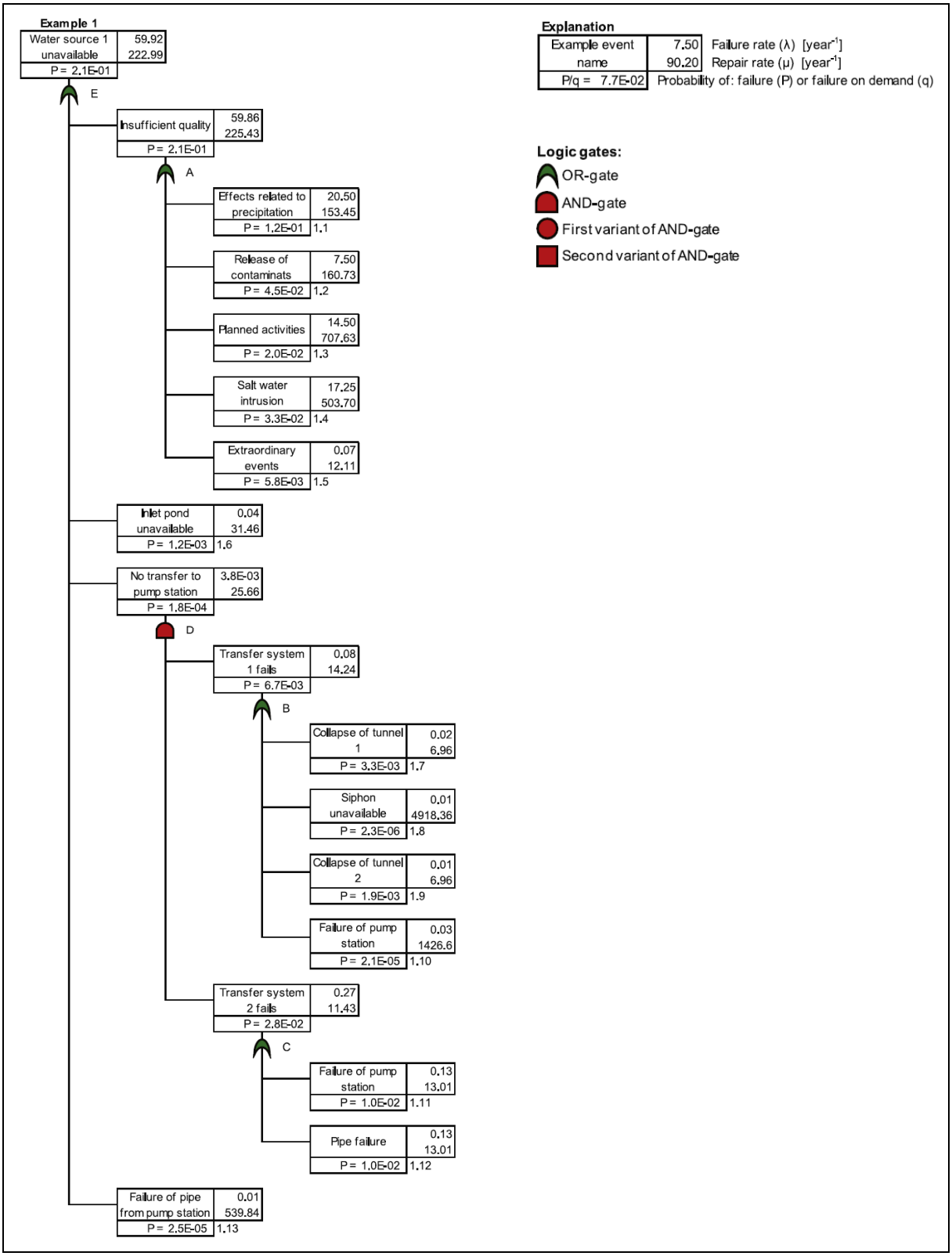


Figure 68. FTA Analysis of Goteborg water supply (Lindhe et al., 2012b Example 1)

10.2.2. Software, BN construction and scenario analysis

The software used in the present instance was Norsys Netica. The equation syntax for basic AND and OR gates for FTA were easily mimicked in Netica. Equations 1, 2 and 3 codings are illustrated respectively in Figure 69a, b. and c. respectively.

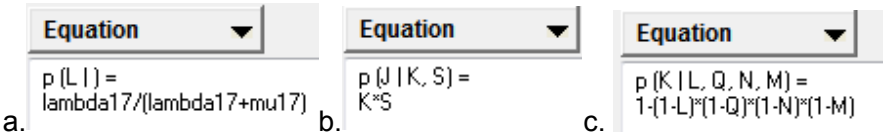


Figure 69. Illustrations of Netica equations for 'Equation to Table' calculation

A BN was constructed to mimic Example 1 of Lindhe et al.(2012b) reproduced above in Figure 68. Compared to normal BN construction practice the final node was located at the vertex in line with FTA design. The BN looks similar to that constructed for naïve and semi-naïve Bayes Nets. However the arcs converge on the higher tier variable/node. The primary failure rate and repair rate values are contained in nodes labelled 'lambda(n)' and 'mu(n)' e.g. lambda17 and mu17.

10.2.3. Input data

The trial input data were those developed by Lindhe et al. (2012b) for their first example shown in Figure 70. These numbers can be seen to correspond to those in Figure 68.

n.b. These numbers use European convention for 'thousands' i.e. they use '.' Instead of ',' as with item 1.8

Ref.	Basic event	λ (year^{-1})			μ (year^{-1})		
		Mean	P05	P95	Mean	P05	P95
1.1	Effects related to precipitation	20.5	16.9	24.4	153	127	182
1.2	Release of contaminats	7.5	5.4	9.9	161	116	212
1.3	Planned activities	14.5	11.5	17.8	708	562	867
1.4	Saltwater intrusion	17.3	14.0	20.8	504	408	607
1.5	Extraordinary events	0.070	0.005	0.200	12.1	3.0	26.1
1.6	Inlet pond unavailable	0.039	0.005	0.100	31.5	6.1	73.0
1.7	Collapse of tunnel 1	0.023	0.010	0.040	7.0	3.0	12.2
1.8	Siphon unavailable	0.011	0.005	0.020	4.918	52	17.520
1.9	Collapse of tunnel 2	0.013	0.005	0.025	7.0	3.0	12.2
1.10	Failure of pump station	0.029	0.020	0.040	1.427	183	3.650
1.11	Failure of pump station	0.134	0.050	0.250	13.0	4.1	26.1
1.12	Pipe failure	0.134	0.050	0.250	13.0	4.1	26.1
1.13	Failure of pipe from pump station	0.013	0.005	0.025	540	52	1.460

Figure 70. Data table of event likelihoods for FTA shown in Figure 68 (Lindhe et al., 2012b)

10.2.4. Sensitivity analysis

Netica was used to undertake 'Sensitivity to Findings' and 'Sensitivity to Parameters' analyses.

"Sensitivity analysis is used to measure the sensitivity of changes in probabilities of query nodes when parameters and inputs are changed. The query nodes in this study (are) model endpoints. Two types of sensitivity analyses (are possible) in evaluating the BNs. The first, ''Sensitivity to Findings'', considers how the BN's posterior distributions change under different conditions, while the second, ''Sensitivity to parameters'', considers how the BN's posterior distributions change when parameters are altered." (slightly modified from wording in Pollino et al., 2007)

Netica includes an inbuilt “Sensitivity to Findings’ calculation tool (Norsys Software Corporation, 2013). This produces 3 measures where the sensitivity of a selected/target node is compared numerically to all other nodes in the BN. As explained in the Netica help files:

> *“you can identify which are the most important questions to ask at each point (to provide information on the variables of interest), based on the answers to questions already received, so as to avoid asking unnecessary or irrelevant questions…In real-world modelling, such as environmental modelling, you can determine which parts of the model most affect the variables of interest; thereby identifying which parts should be made the most carefully and accurately.*
>
> *(You) select a node (called the "query node") and choose Network → Sensitivity to Findings from the menu. A report displays how much the beliefs, expected value, etc. of the query node would be influenced by a single finding at each of the other nodes (each is called a "varying node").*
>
> *The first part of the report has a section for each varying node, showing how much it can affect the query node using several different sensitivity measures. The second part is a summary table which compares the sensitivities for each of the varying nodes. To limit the report to a few varying nodes, you select the query node, and then use ctrl-select to add the desired varying nodes to the selection. Then choose Network → Sensitivity to Findings..”*

The “Sensitivity to Findings’ tabulated statistics are (Norsys Software Corporation, 2013):

- *Variance Reduction - Definition: The expected reduction in variance of the expected real value of Q due to a finding at F. This turns out to be the square of RMS Change of Real.*
- *Entropy Reduction (Mutual Information) - Definition: The mutual information between Q and F (measured in bits). The expected reduction in entropy of Q (measured in bits) due to a finding at F.*
- *"Variance" of Node Belief - Definition: The expected change squared of the beliefs of Q, taken over all of its states, due to a finding at F.*

Where Q is the query variable e.g. water supply availability and F is the varying variable e.g. water quality or water quality impacting event such as rainfall.

For reference the metric for Q is also calculated against itself to act as a reference and this is compared to that of F variables.

In plain English these metrics show how much the variance in the values the target or query node of primary interest take can be accounted for by other node, or in the case of a the Gothenberg water treatment, how much the likelihood of a (top) failure event in the FTA can be accounted for by other failure events.

‘Sensitivity to parameters’ is assessed by altering parameter values for different nodes. Sensitivity to parameters in illustrated in Korb and Nicholson (2011 p 391 and Table 11.6). The process can be automated or undertaken manually.

10.2.5. Most probable explanation

This is another feature in Netica which provides useful information on how BN nodes interrelate.

“Given findings for some nodes, you may want to find the most probable configuration of values for the rest of the nodes. This can be thought of as providing a plausible explanation for the

observed findings, and is called the most probable explanation or MPE (it is a special case of the maximum a-posteriori probability, or MAP)".

After updating, each node will have a belief-bar at the 100% level, and usually some bars at lower levels. You can read off the most probable configuration by taking for each node the state with the bar at the 100% level. The shorter bars indicate the relative probabilities of the other states given that the other nodes are in the most probable configuration (scaled by the same factor used to bring the longest bar to 100%).(Norsys Software Corporation, 2013 Help files).

The MPE can be displayed by activating the menu tool for that purpose.

10.3. Results and Discussion

10.3.1. Building the BN FTA

Figure 71 illustrates the BN FTA emulating that in Figure 68 and using the data in Figure 70. Comparison shows that the top event and post AND and OR gate probabilities are identical e.g. the probability of water source 1 being unavailable is about 21%. The most influential events appear to be those relating to water quality.

10.3.2. Scenario analysis

While Figure 71a illustrates that water quality is most likely a reflection of precipitation, setting unavailability to 100% (Figure 71b) shows clearly that 57% of unavailability is due to precipitation and this statistic can be compared with all other event.

Figure 71 v. Figure 72 shows the impact of removing all water quality event types other than routine maintenance = planned activities. The latter now dominates unavailability compared with other sporadic events.

10.3.3. Sensitivity analysis

The results of illustrative 'Sensitivity to Findings' analyses performed on the Top event (availability) are shown in the first set of statistics in Table 18. This again shows clearly the most important event group in water quality in particular the impact of precipitation.

Once most of the water events are discounted it can be seen that water transport pumping is the most important. This can be seen in the second set of statistics where most water quality events have been discounted by setting their no event probabilities to 100% likelihood as seen in Figure 72.

10.3.4. Most Probable Explanation

Figure 73 illustrates the most probable explanation overall and for situations when events occur.

Figure 73a shows simply that most of the time water is available and critical points are at their nominal settings.

Figure 73b shows in a different fashion that when water is unavailable it is most likely due to water quality problems and the most likely cause is precipitation.

10.3.5. Beyond Basic FTA gates

A further feature of the studies of Lindhe et al.(2012b, 2009) is the use of more complex dynamic FTA based around the use of first and second 'variant' AND gates. These cannot be constructed using single conventional BN nodes.

However, Dynamic Bayes Nets are also possible to construct with some software notably BayesiaLab's AgenaRisk. Details are provided in Fenton and Neil (2012).

Further Fenton and Neil (2012 p. 352) identify several advantages to doing basic FTA using BNs:

- BNs can be diagnostic as well as predictive because they allow full backcasting and forward casting and inputting of scenarios which diagnosis is desired for e.g. if there is failure in the water supply above what are the potential causes?
- Classic FTA assumes events are independent but BNs allow other interrelationships to be included;
- Specification of component states is more flexible with BNs;
- Calculations in discrete BNs are exact whereas classical fault tree calculations are approximate.

10.3.6. Dynamic Bayesian Networks

Dynamic gates in a fault tree analysis employ Markov models as a method to deal with the dynamic behaviour of fault tolerant systems. In recent years, dynamic Bayesian networks have been successfully used to encode Markov models in reliability analysis (Weber and Jouffe, 2003, Portinale et al., 2010). This method offers advantages over Markov models such as the avoidance of large number of states and model complexity. This section presents two examples of dynamic gates based on the examples in Linde et al. (Lindhe et al., 2012a) which were transformed to dynamic Bayesian networks.

Example 1- Second variant AND gate (See Example 2 in Lindhe et al., 2012b)

The fault tree in Figure 74 shows the model we encoded through Bayesian network. This model utilises the second variant of the AND-gate as shown in Lindhe et al.(2012b). It defines a quality failure as the outcome of either two potential cases; detection of a quality deviation but no action was possible, or no detection of a failure. Each of these two dynamic gates is mapped into a conditional probability table as presented in Figure 75 using the approach from (Weber and Jouffe, 2003). The original units of failure rates were transformed from $year^{-1}$ to h^{-1} as shown in Table 19.

The evolution of the values over time when the starting state is S0 (failure of detection, and no action possible event when detected, respectively) for both sub events (or quality failure=TRUE), are presented in Figure 76. As can be seen, the system is very resilient to this initial state, rapidly returning to a non-failure mode. The long term probability of the system for the working state (Quality failure=FALSE) is close to 100% (99.95%).

Example 2 – First variant AND gate (Higher tier portion of Example 3 in Lindhe et al., 2012b)

The second example of mapping Markov models to a dynamic Bayesian network is depicted in Figure 78 (right). This corresponds to the first variant of AND-gate as analysed in Lindhe et al. (2012b)'s Example 3. Reliability parameters used on this model are presented in Table 20. The Markov model was encoded through the conditional probability tables shown on Figure 77. The Bayesian network model is presented on Figure 78 (left).

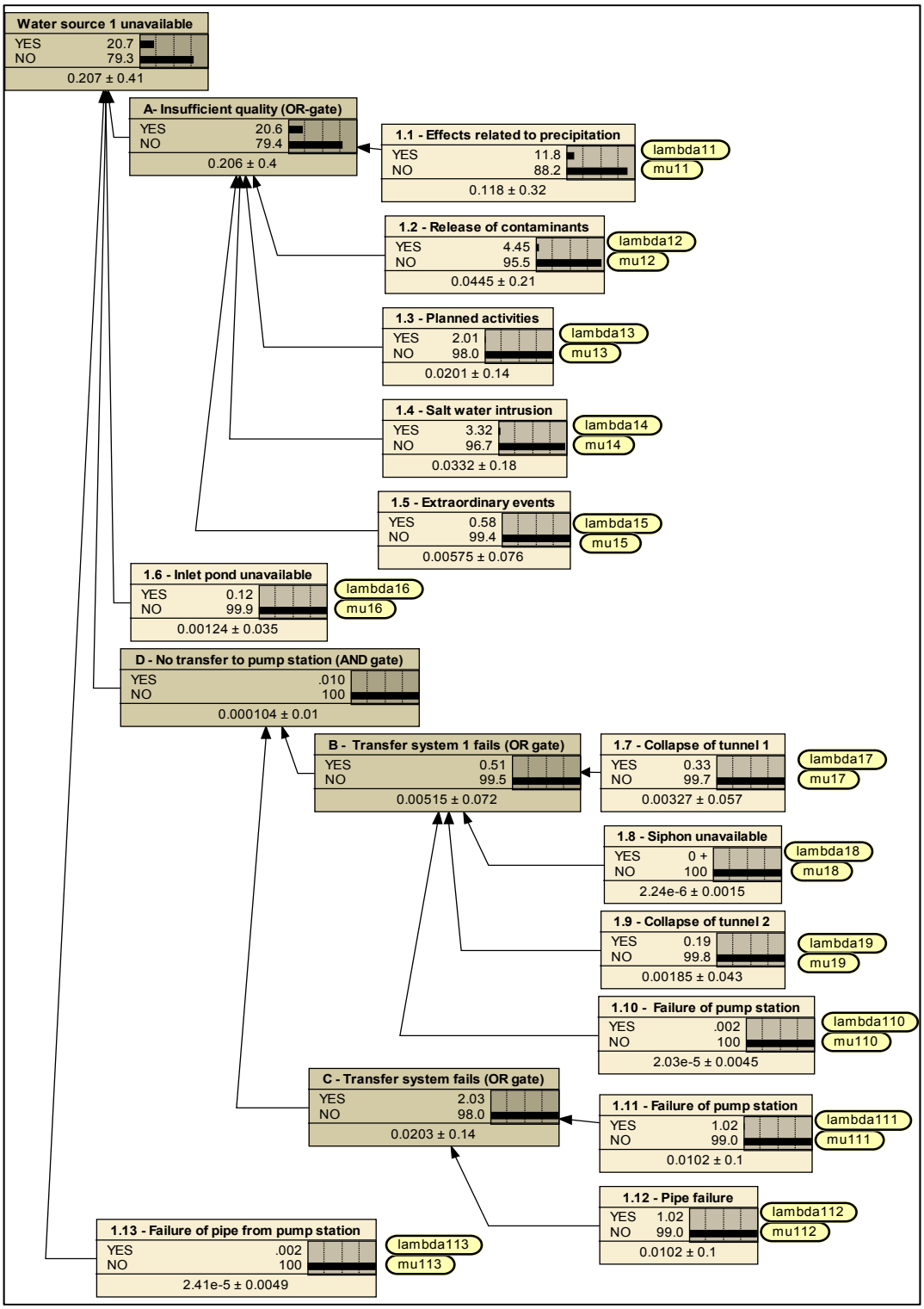
This model shows similar behaviour to our Example 1 above. As it can be seen in Figure 79, the system is resilient to an initial failure mode (S0) state, returning to fully working condition in a short period of time. The long term probability of the system in functional mode (S1) is almost 100% (99.57%).

10.4. Conclusions

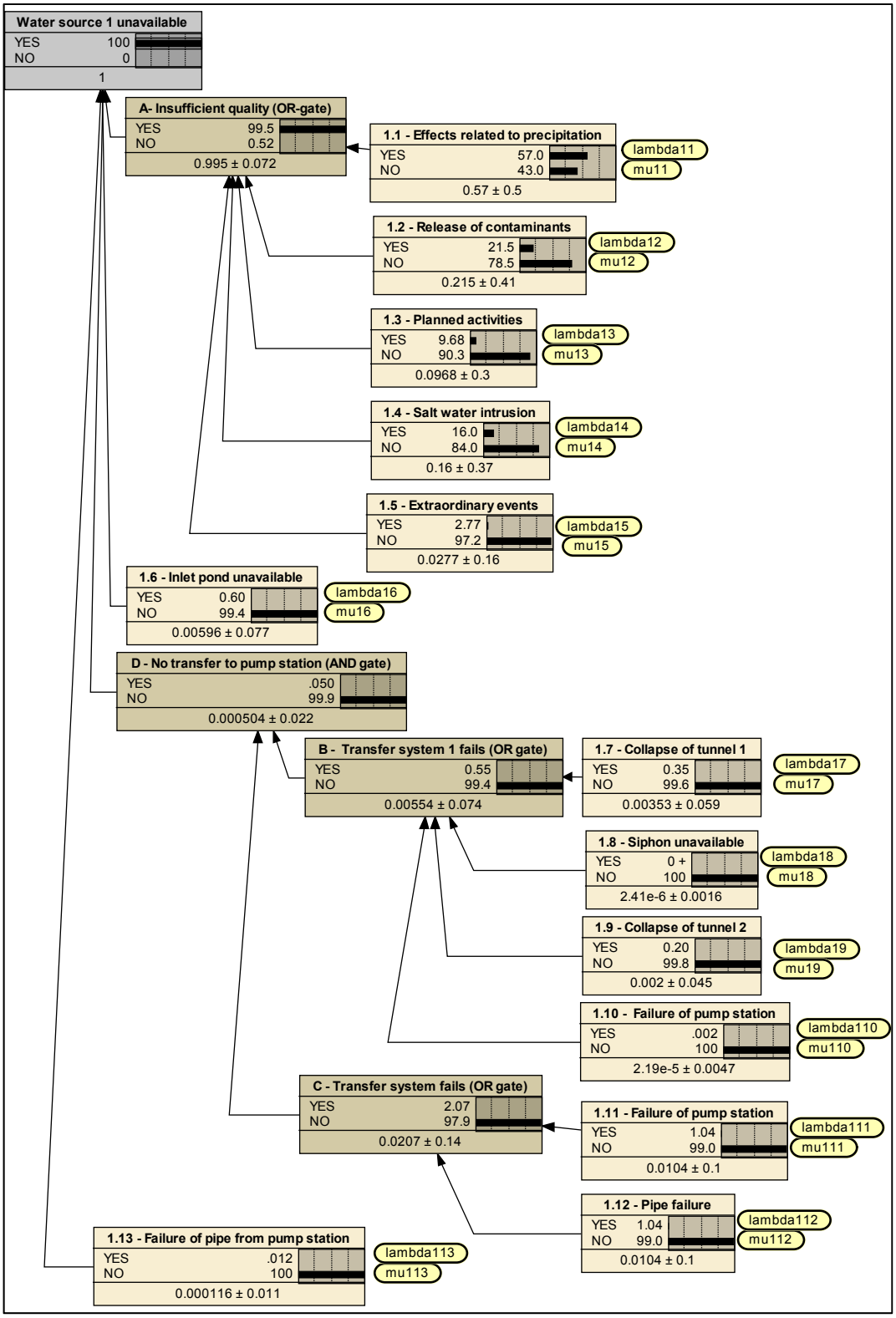
FTA and other event analysis (Fenton and Neil, 2012) can easily be undertaken using the same BN software and logic as can be applied to QMRA, basic likelihood/consequence analysis etc.

As a result Hazardous Event analysis requires no new methodology but merely an extension of the causal logic and inference of Bayes.

10.5. Example tables and figures



a.



b.

Figure 71. Scenario 1 - Reference Scenario FTA recast in Bayes Net format a. Primary BN, b. Water Source unavailability set to 100%

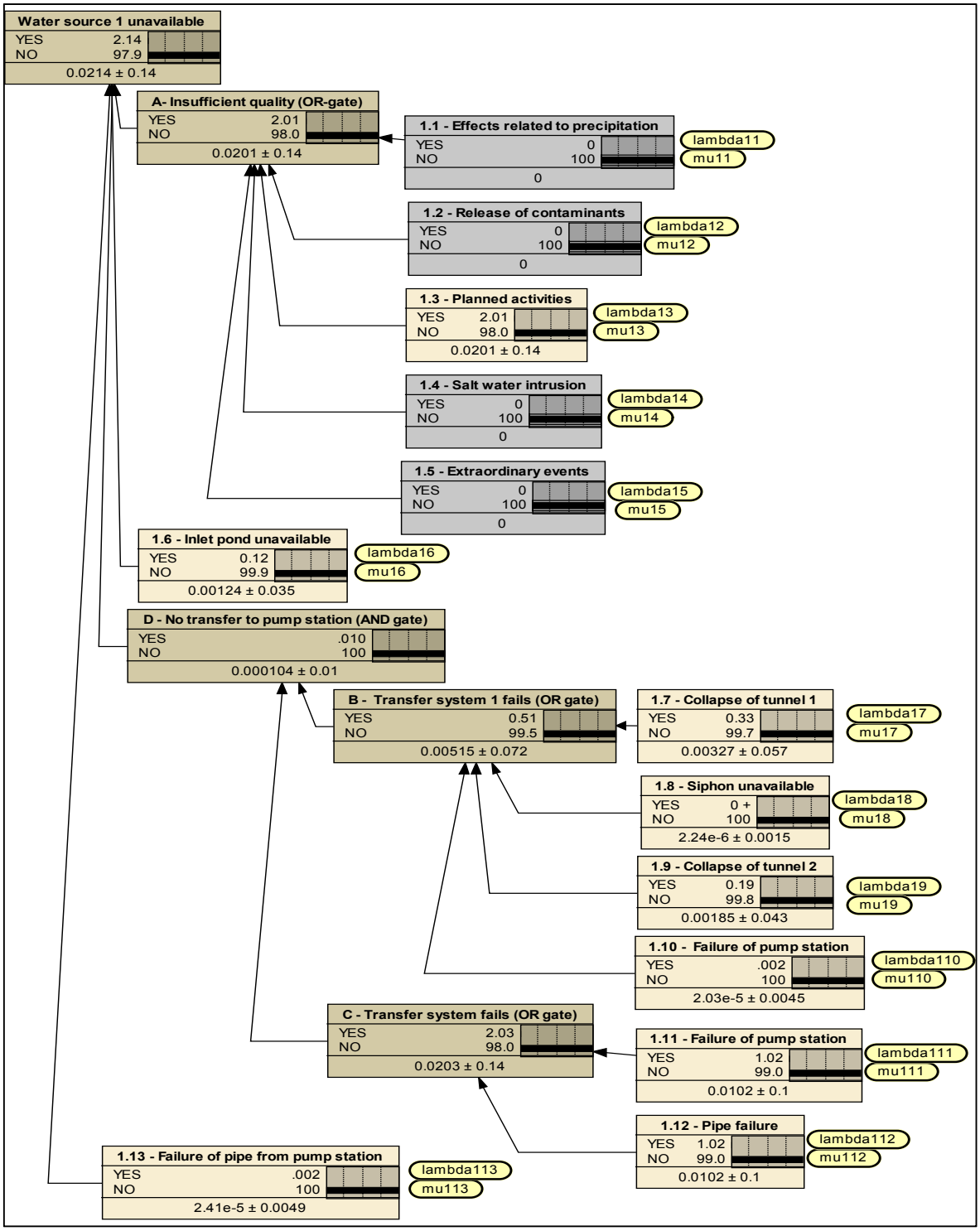
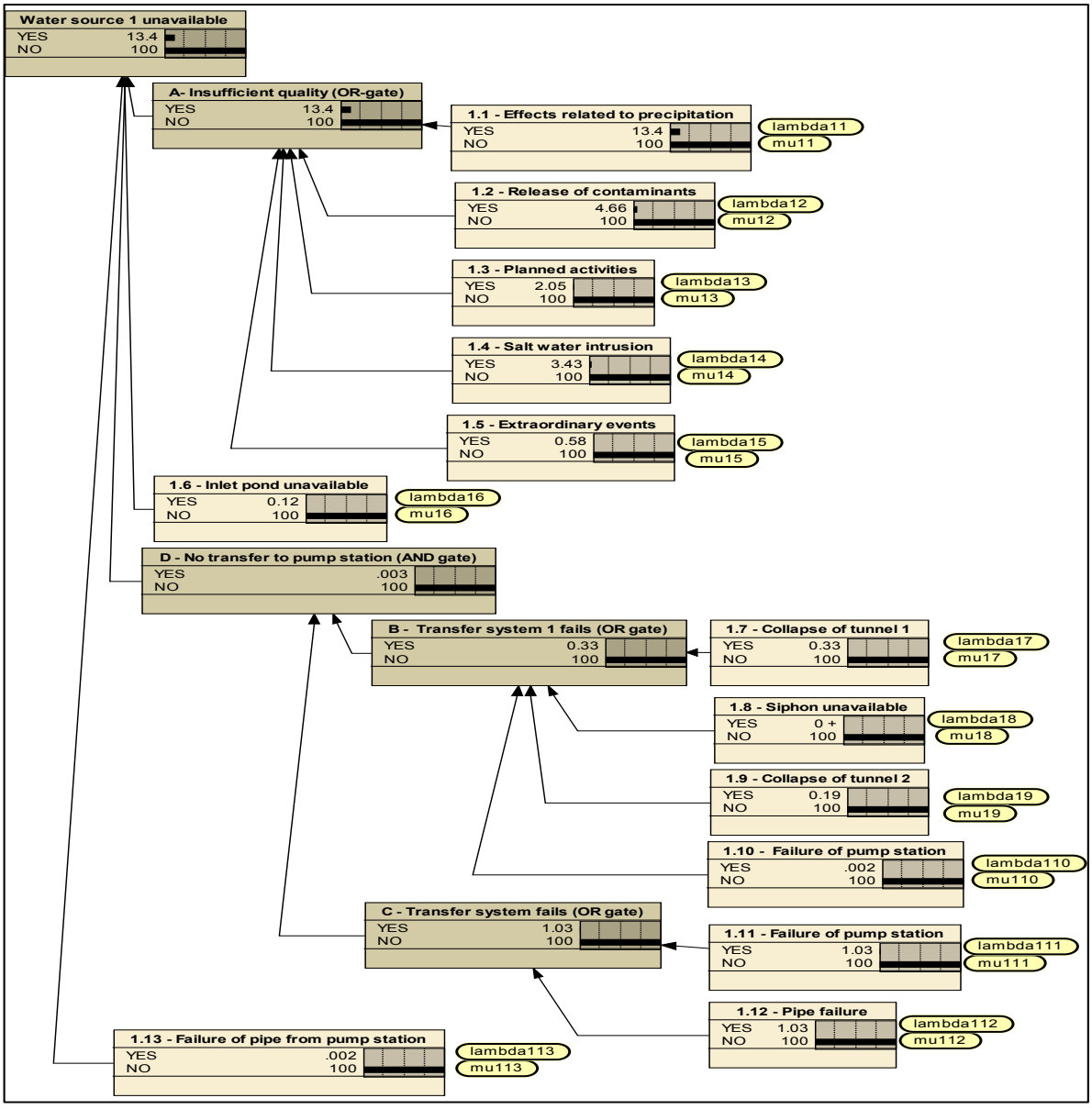


Figure 72. Scenario 2 – FTA excluding major unpredictable water quality impacts

Table 18. Results of Sensitivity to Findings Analyses performed on 'Water Source 1 unavailable' Node

Scenario	Node	Variance Reduction (Percent)	Mutual information (Percent)	Variance of Beliefs
Reference	***(Water source 1 unavailable***	100	100	0.1643813
	A- Insufficient quality	99.3	98.4	0.1632994
	1.1 - Effects related to precipitation	51.2	43.4	0.084174
	1.2 - Release of contaminants	17.8	14.5	0.0292652
	1.4 - Salt water intrusion	13.1	10.7	0.0215641
	1.3 - Planned activities	7.83	6.34	0.0128662
	1.5 - Extraordinary events	2.21	1.79	0.0036344
	1.6 - Inlet pond unavailable	0.473	0.382	0.0007778
	D - No transfer to pump (AND Gate)	0.0399	0.0322	0.0000656
	1.13 - Failure of pipe from station	0.0092	0.00742	0.0000151
	B - Transfer system 1 fails (OR gate)	0.000805	0.000769	0.0000013
	C - Transfer system fails (OR gate)	0.000201	0.000197	0.0000003
Quality failure modes omitted except for 1.3 Planned Activities	***Water source 1 unavailable***	100	100	0.0209485
	A- Insufficient quality	93.6	90.2	0.0196128
	1.3 - Planned activities	93.6	90.2	0.0196128
	1.6 - Inlet pond unavailable	5.66	4.63	0.0011857
	D - No transfer to pump (AND Gate)	0.478	0.388	0.0001
	1.13 - Failure of pipe from station	0.11	0.0895	0.0000231
	B - Transfer system 1 fails	0.00965	0.0368	0.000002
	C - Transfer system 2 fails	0.00241	0.0109	0.0000005


Water source 1 unavailable
YES 13.4
NO 100
A- Insufficient quality (OR-gate)
YES 13.4
NO 100
1.1 - Effects related to precipitation
YES 13.4
NO 100
lambda11
mu11
1.2 - Release of contaminants
YES 4.66
NO 100
lambda12
mu12
1.3 - Planned activities
YES 2.05
NO 100
lambda13
mu13
1.4 - Salt water intrusion
YES 3.43
NO 100
lambda14
mu14
1.5 - Extraordinary events
YES 0.58
NO 100
lambda15
mu15
1.6 - Inlet pond unavailable
YES 0.12
NO 100
lambda16
mu16
D - No transfer to pump station (AND gate)
YES .003
NO 100
B - Transfer system 1 fails (OR gate)
YES 0.33
NO 100
1.7 - Collapse of tunnel 1
YES 0.33
NO 100
lambda17
mu17
1.8 - Siphon unavailable
YES 0 +
NO 100
lambda18
mu18
1.9 - Collapse of tunnel 2
YES 0.19
NO 100
lambda19
mu19
1.10 - Failure of pump station
YES .002
NO 100
lambda110
mu110
C - Transfer system fails (OR gate)
YES 1.03
NO 100
1.11 - Failure of pump station
YES 1.03
NO 100
lambda111
mu111
1.12 - Pipe failure
YES 1.03
NO 100
lambda112
mu112
1.13 - Failure of pipe from pump station
YES .002
NO 100
lambda113
mu113

a.

b

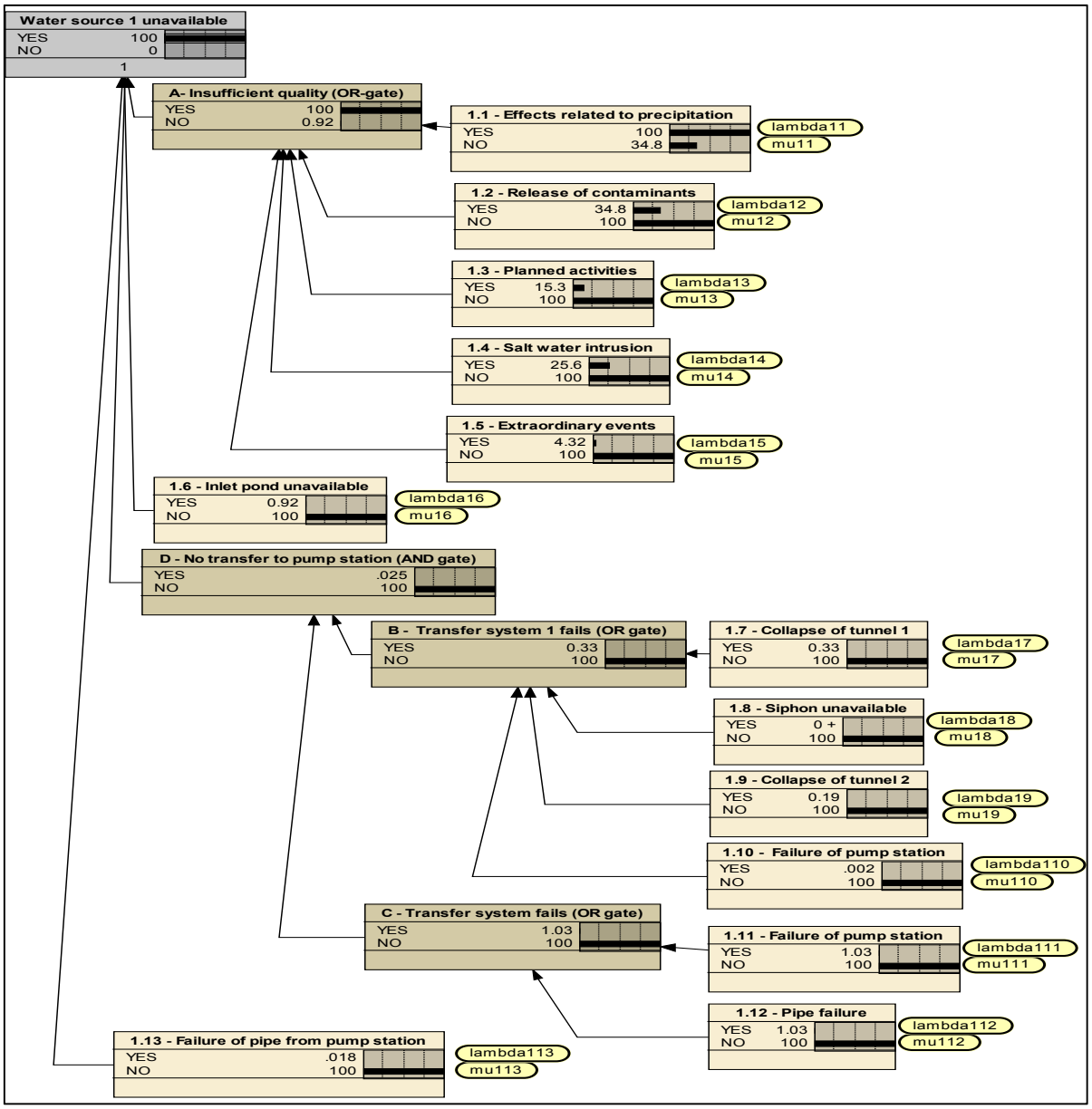


Figure 73. Most probable explanation format a. Overall, b. Water unavailable

Table 19. Reliability parameters used on the Markov model for Example 1 – 2nd variant AND gate

Parameter	Gate F	Event	Gate G	Event	Units
lambda1	3.08E-05	2.1	2.74E-05	2.3	1/h
mu 1	0.0062	2.1	0.0203	2.3	1/h
lambda2	0.00071	2.2	0.00071	2.4	1/h
mu2	0.308	2.2	0.308	2.4	1/h
q2	0.095	2.2	0.095	2.4	

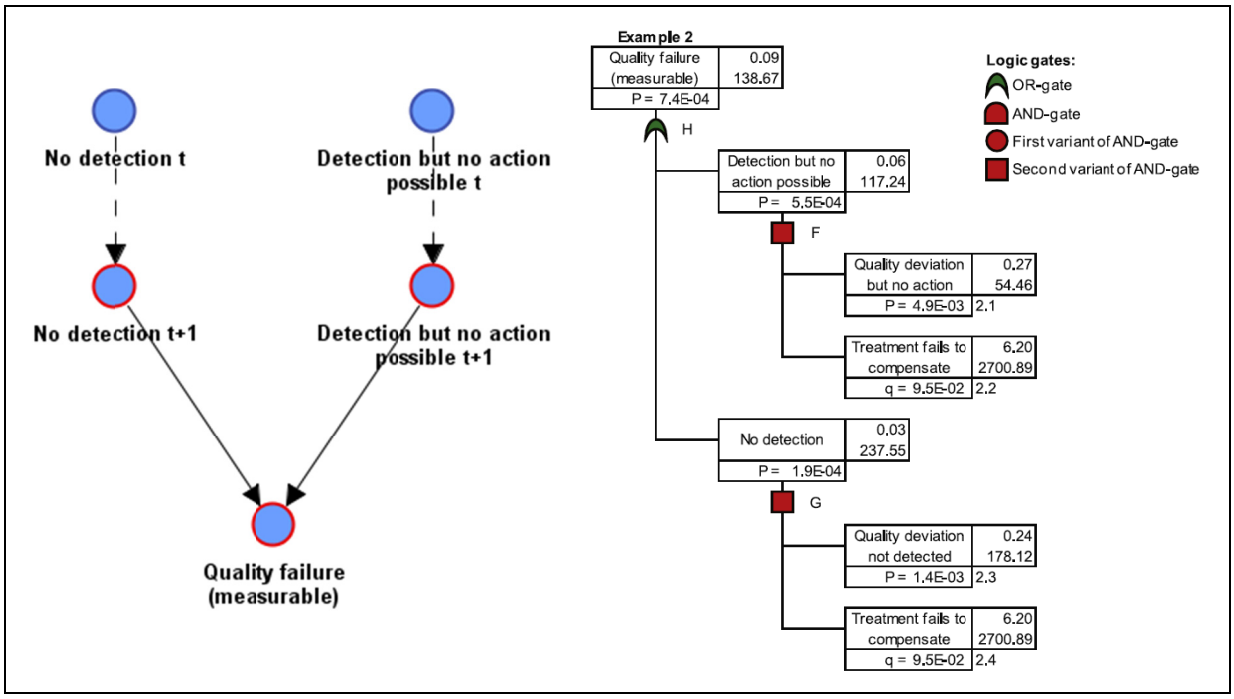
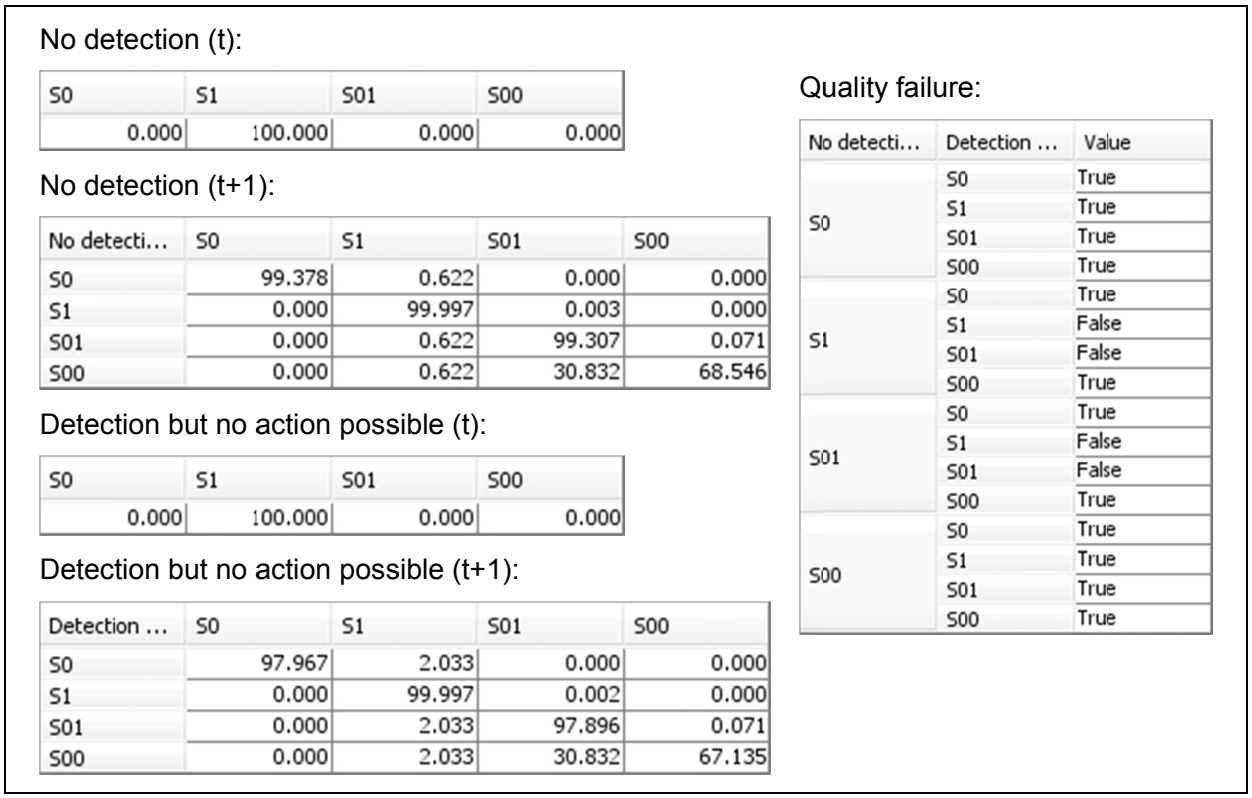


Figure 74: Dynamic Bayesian network model (left) and fault tree analysis (right) for example 1

Table 20: Reliability parameters used on the Markov model (Example 2 – 1st variant AND gate)

Event	Parameter	Value	Units
Failure at WTP1	mu1	0.0937	1/h
Failure at WTP1	lambda1	0.00041	1/h
Event 3.8	lambda2	0.1829	1/h
Event 3.9	lambda3	0.0880	1/h
Event 3.8	q2	0.011	
Event 3.9	q3	0.027	



No detection (t):

S0	S1	S01	S00
0.000	100.000	0.000	0.000

No detection (t+1):

No detecti...	S0	S1	S01	S00
S0	99.378	0.622	0.000	0.000
S1	0.000	99.997	0.003	0.000
S01	0.000	0.622	99.307	0.071
S00	0.000	0.622	30.832	68.546

Detection but no action possible (t):

S0	S1	S01	S00
0.000	100.000	0.000	0.000

Detection but no action possible (t+1):

Detection ...	S0	S1	S01	S00
S0	97.967	2.033	0.000	0.000
S1	0.000	99.997	0.002	0.000
S01	0.000	2.033	97.896	0.071
S00	0.000	2.033	30.832	67.135

Quality failure:

No detecti...	Detection ...	Value
S0	S0	True
	S1	True
	S01	True
	S00	True
S1	S0	True
	S1	False
	S01	False
	S00	True
S01	S0	True
	S1	False
	S01	False
	S00	True
S00	S0	True
	S1	True
	S01	True
	S00	True

Figure 75: conditional probability tables for the dynamic Bayesian network, example 1

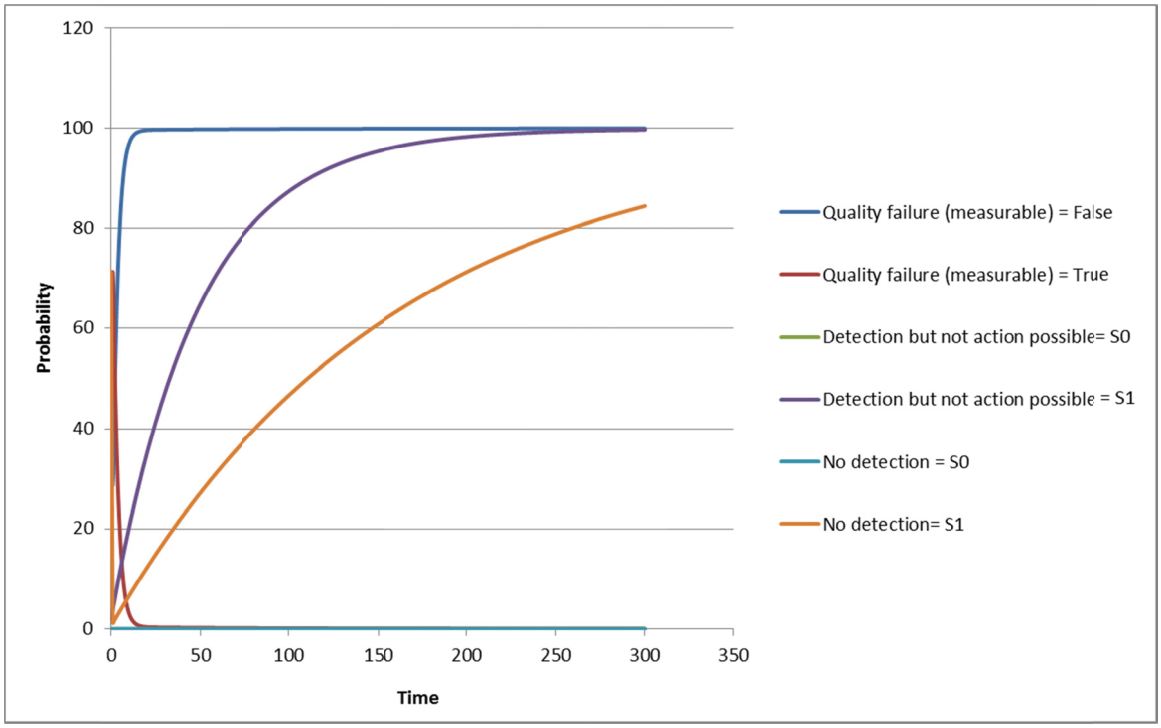


Figure 76: temporal graph of quality failure event and sub events. S0 indicates the state in which the system is in failure mode and S1 in working condition without the use of back-up.

Quantity failure due to WTP 1 (t):

S0	S1	S01	S001
0.000	100.000	0.000	0.000

Quantity failure due to WTP 1 (t+1):

Quantity f...	S0	S1	S01	S001
S0	90.621	9.379	0.000	0.000
S1	0.000	99.959	0.040	0.000
S01	0.494	9.379	72.324	17.804
S001	8.802	9.379	0.000	81.819

Figure 77: Conditional probability tables for the dynamic Bayesian network, example 2

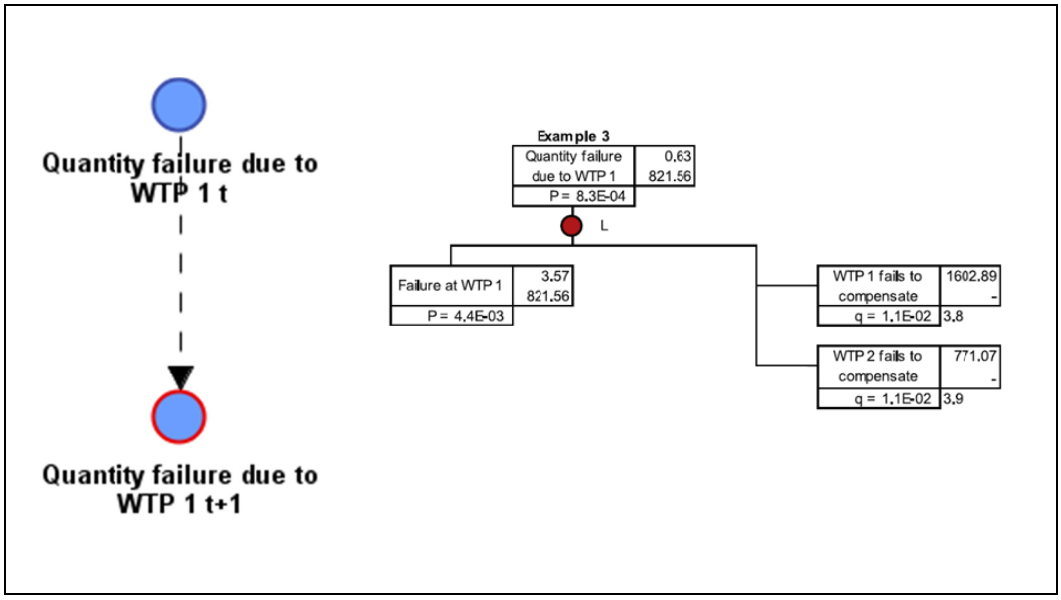


Figure 78: Dynamic Bayesian network model (left) and fault tree analysis (right) for example 1

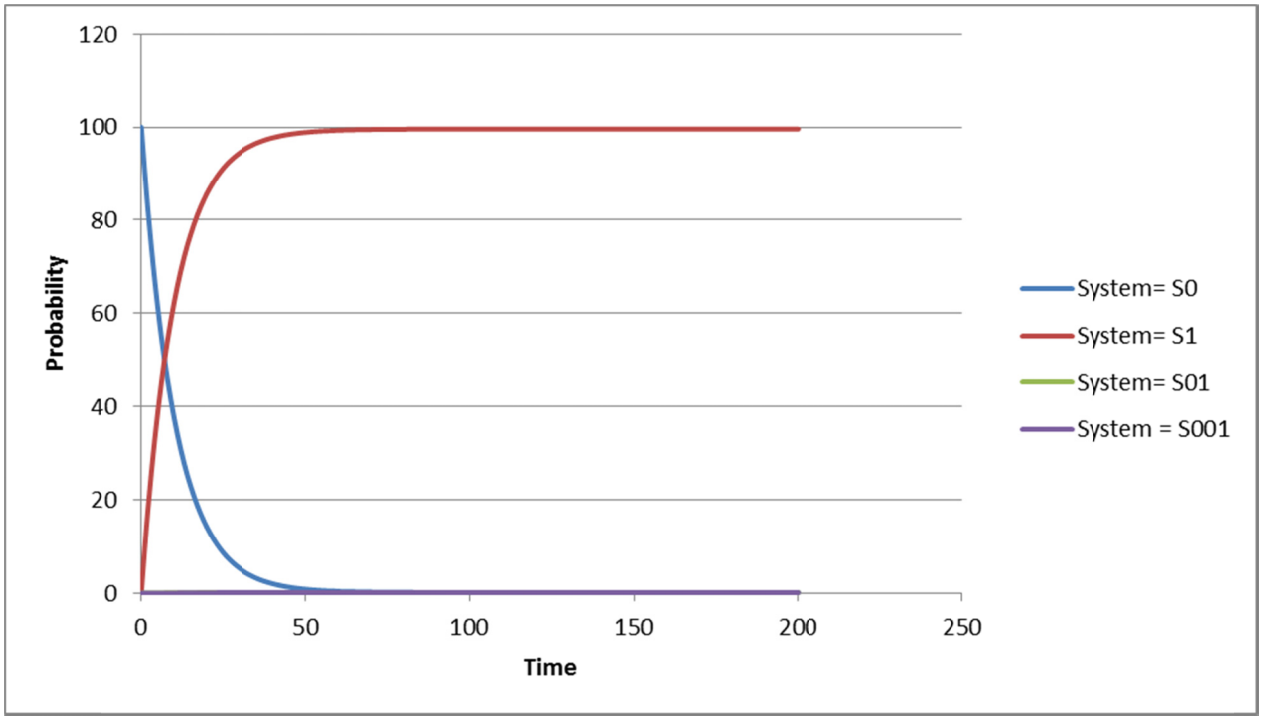


Figure 79: conditional probability tables for the dynamic Bayesian network

11. Assessment of Compounding Conservativeness of LRV Attribution for Increasing Numbers of Treatment Barriers

11.1. Introduction

Log reduction values (LRVs) for pathogens are attributed to numerous individual barriers in a multiple-barrier treatment train. An overall LRV can then be attributed to the combined barriers in a multiple-barrier treatment train. The current Victorian State guidelines for validating treatment processes for pathogen reduction state that "*In general, a conservative approach is taken to analysing validation data to establish the challenge test LRV. Unless otherwise specified in this guidance, the lower 5th percentile LRV established during challenge testing must be used*" (Department of Health Victoria, 2013).

A limitation of this approach is that it becomes increasingly conservative with the number of independent barriers in the multiple barrier treatment train. A conceptual assessment of multiple barrier combinations was undertaken to investigate the effect of this increasing conservativeness.

An alternative to this approach is to employ full Monte Carlo style modelling along the lines used for Quantitative Microbial Risk Assessment (QMRA) for the past 15-20 years (Haas and Trussell, 1998). This is now more feasible than in the past due to the increasing availability of suitable generic (@Risk add on to Excel) and purpose designed (e.g. QSPOT) software and the greater familiarity generally of graduate engineers with both the concepts and sophisticated software tools (e.g. Matlab, Mathematica) which include Monte Carlo tools in their suites. Furthermore, as demonstrated here, BN techniques can easily provide the same analysis.

11.2. Method

Monte Carlo modelling was used to simulate and compare combined multiple barrier system LRVs with summed individual LRVs. Conceptual treatment trains for recycled water were composed of varying numbers of individual barriers from 1 to 8. Each individual barrier was assumed to achieve LRVs described by one of three lognormal probability density functions (PDFs). In all cases, the mean barrier LRV was assumed to be 3. However, in order to assess the effect of performance variability (or uncertainty), the three lognormal PDFs differed in terms of their attributed standard deviations. The three attributed standard deviations were selected to represent a very tight LRV distribution (s.d.=0.1, Figure 80), a medium distribution (s.d.=0.5, Figure 81) and a very broad distribution (s.d.=1, Figure 82). The 5th values for these three PDFs are presented in Table 21.

Two programming approaches were explored, conventional full @Risk Monte Carlo simulation and BN based simulation using Agena Risk software.

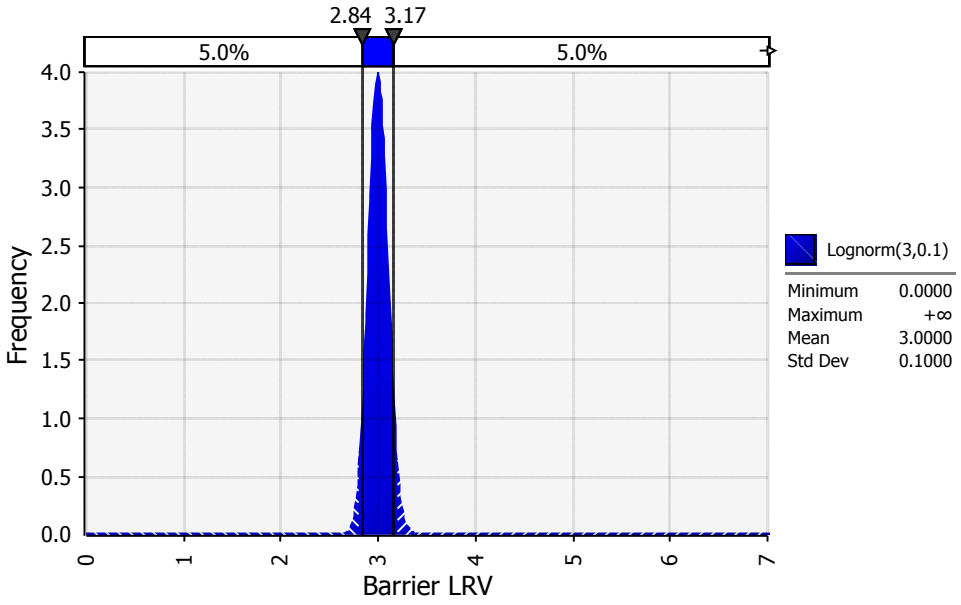


Figure 80. LRV for a single barrier with μ=3, s.d=0.1

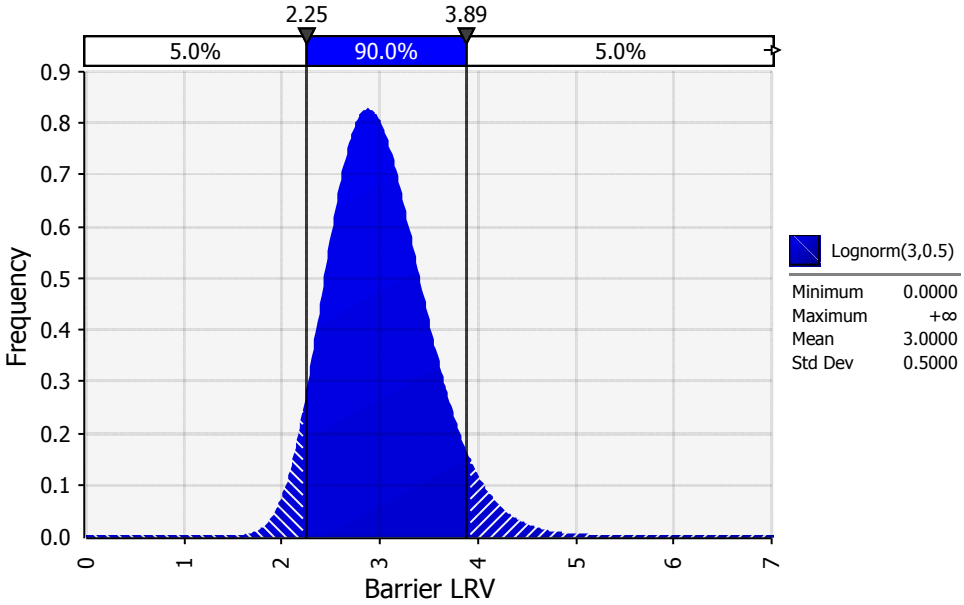


Figure 81. LRV for a single barrier with μ=3, s.d=0.5

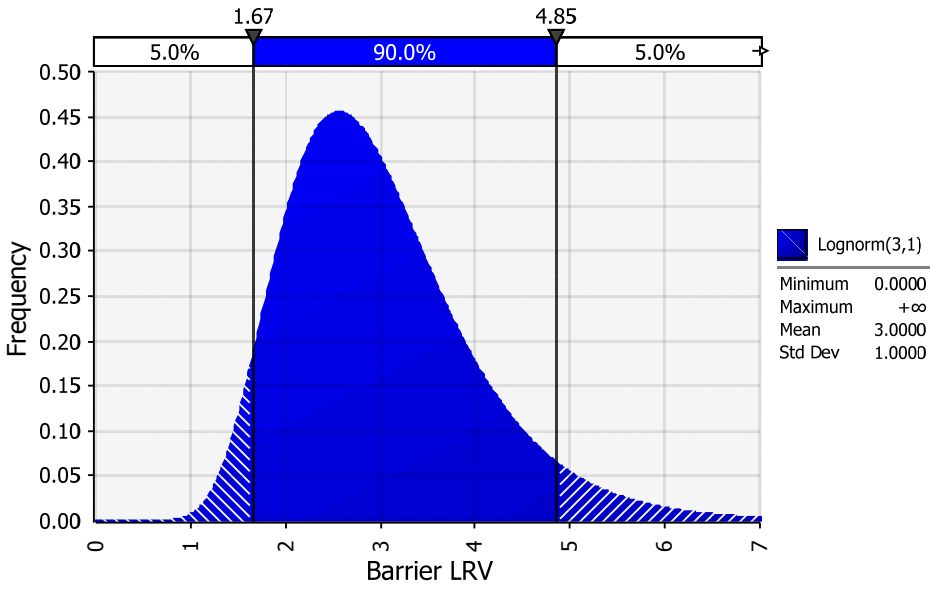


Figure 82. LRV for a single barrier with μ=3, s.d=1

Table 21. 5th percentile values for LRV Lognormal distributions with μ=3 and s.d=0.1, 0.5, 1.

Mean LRV	Std Dev	5th Percentile
3	0.1	2.8
3	0.5	2.3
3	1	1.7

11.3. Results

11.3.1. Simulations using conventional Monte Carlo Software

PDFs for LRVs achieved by individual treatment barriers may be mathematically combined to an overall LRV for a treatment train by means of a Monte Carlo simulation. For example, an overall LRV PDF for a treatment train consisting of four independent barriers, each with LRV μ=3, sd=0.5 is presented in Figure 83. The 5th percentile of the combined LRV PDF is 10.4, which is somewhat larger than the number that would be obtained from the sum of four 5th percentile barriers (2.3+2.3+2.3+2.3=9.2). In fact, the sum of the four 5th percentiles is roughly equal to the 0.05th percentile. Therefore this value represents a highly unlikely number, applied with significantly greater conservatism that would be used with a smaller number of barriers.

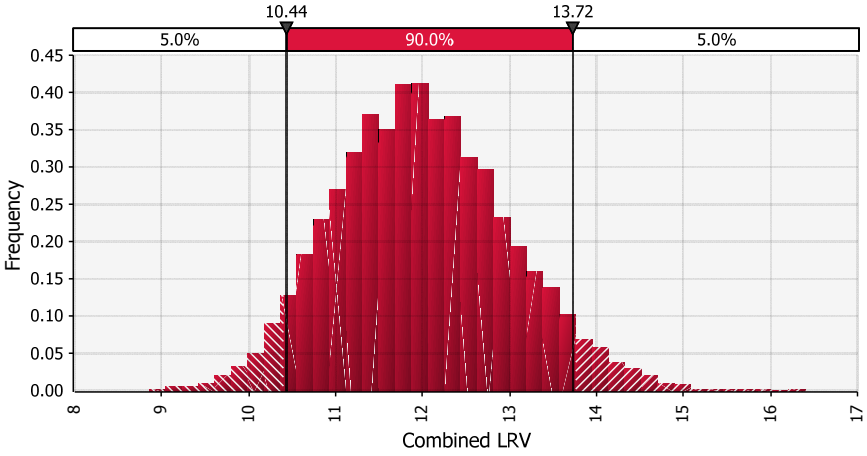


Figure 83. Combined 4 barriers with LRV m=3, sd=0.5. 5th percentile = 10.4

The degree of conservatism achieved by summing individual 5th percentiles increases with the number of independent barriers. This is demonstrated in Figure 84, which shows the true multiple barrier percentile value of the summed 5th percentiles of 1-8 individual barriers. For individual barrier with mean = 3 and s.d= 0.1, the summed 5th percentile value is approximately equivalent to the 1st percentile for two barriers, the 0.2th percentile for three barriers, the 0.05th percentile for four barriers and the 0.02th percentile for 5 barriers.

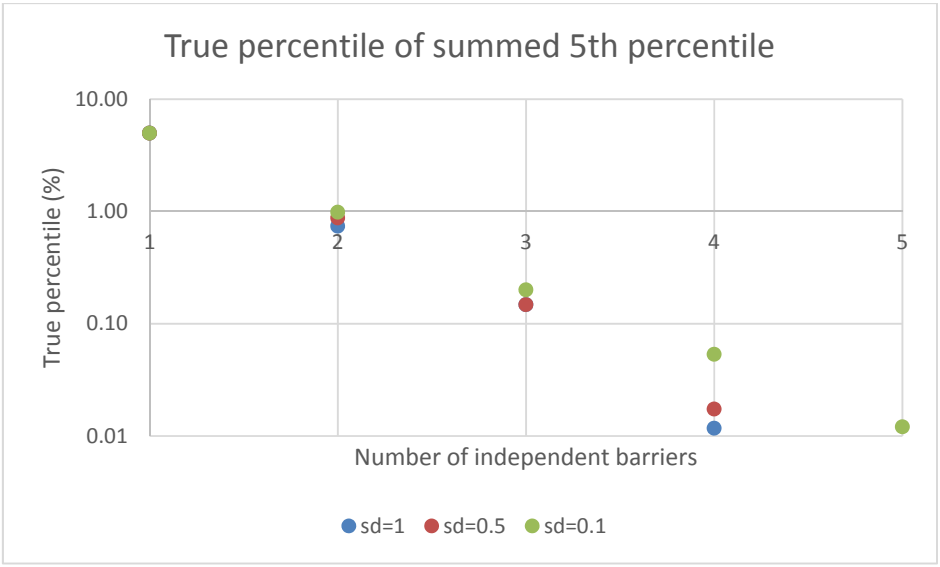


Figure 84. True multiple barrier percentile value of the summed 5th percentiles of 1-5 individual barriers

A consequence of this increasing conservatism for increasing numbers of barriers is that, in some cases, additional LRVs could be attributed to multiple barrier systems while maintaining the same level of conservatism that would be required for systems with fewer independent barriers. The degree to which this could be achieved is dependent upon the number of sequential independent barriers, as well as the relative variability (e.g., LRV standard deviation) of the individual barriers.

For barriers with very tight LRV distributions (in this case s.d.=0.1), the advantage achieved by using a Monte Carlo simulation to combine the barriers, compared to summing the individual 5th percentiles is minimal, as indicated in Figure 85. For a distributions with medium spread (in this case, s.d.=0.5), the advantage can be moderate for a large number of barriers, as indicated in Figure 86. For broad distributions (in this case, s.d.=1), the advantage can be more significant, as indicated in Figure 87.

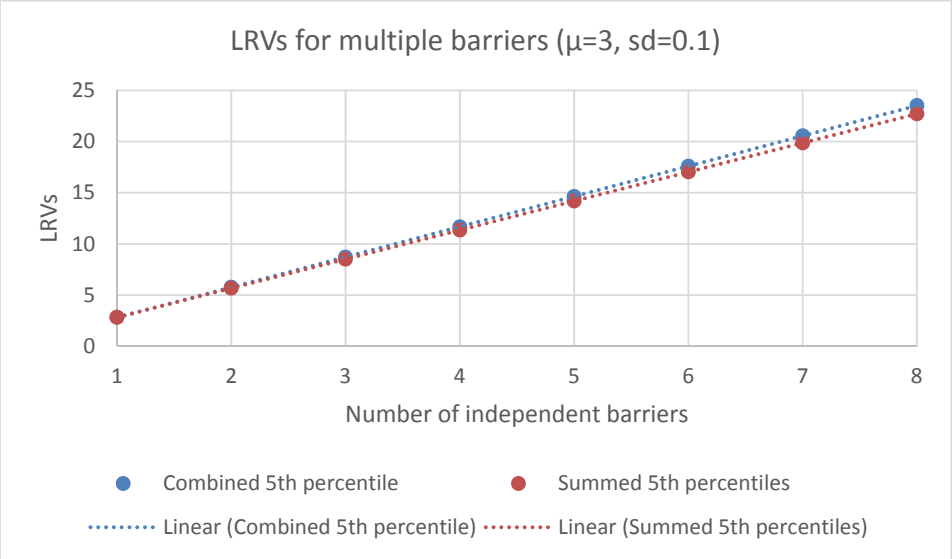


Figure 85. Comparison of LRV achieved by a combined 5th percentile compared to summed 5th percentiles for 1-8 barriers, s.d=0.1.

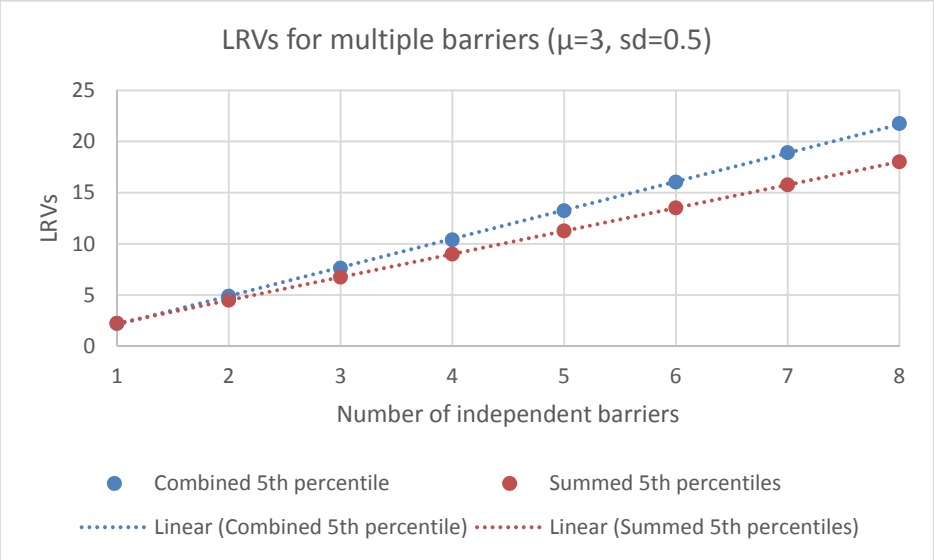


Figure 86. Comparison of LRV achieved by a combined 5th percentile compared to summed 5th percentiles for 1-8 barriers, s.d=0.5.

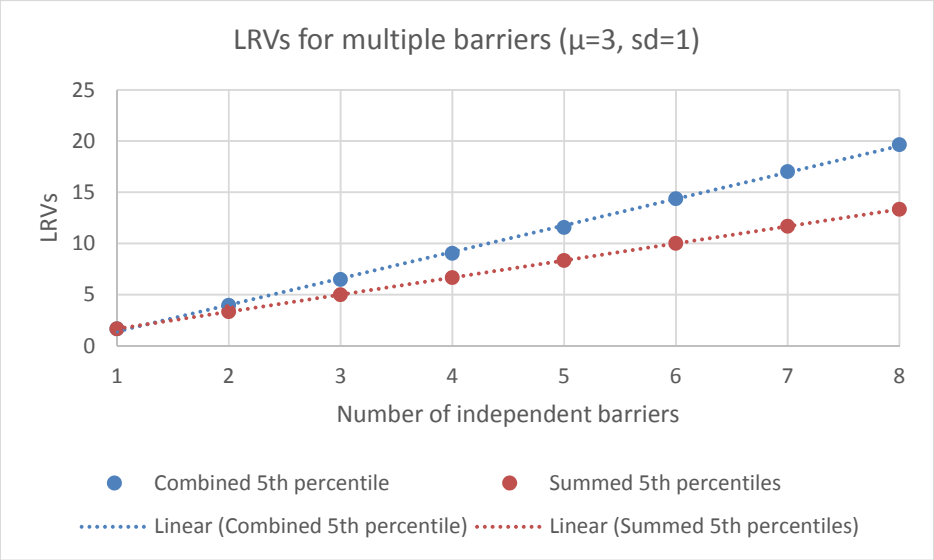


Figure 87. Comparison of LRV achieved by a combined 5th percentile compared to summed 5th percentiles for 1-8 barriers, s.d=1.

The relative increase in attributable LRV for a combined multiple barrier 5th percentile, compared to summed individual barrier 5th percentiles is presented for the three individual barrier types (s.d.=0.1, 0.5 and 1) assessed in the above examples (Figure 88).

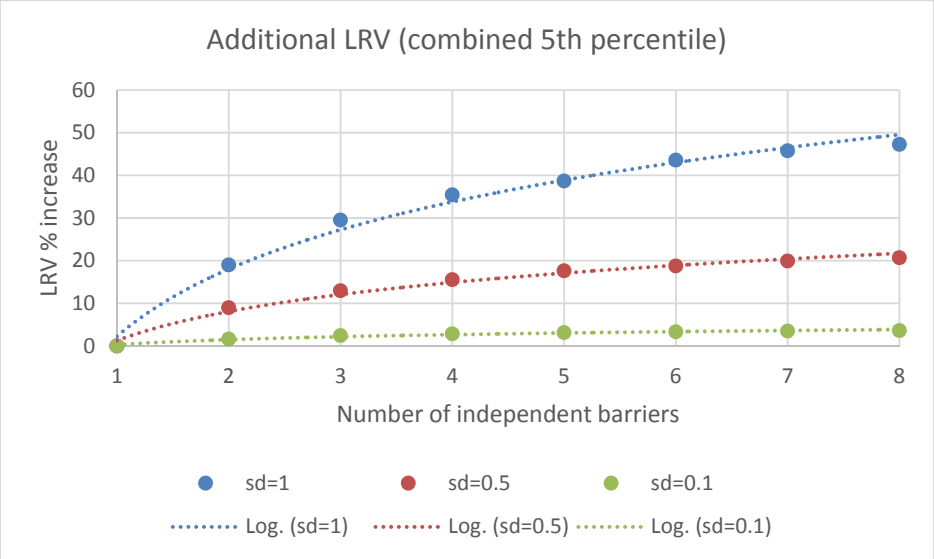


Figure 88. Relative increase in attributable LRV for a combined multiple barrier 5th percentile, compared to summed individual barrier 5th percentiles.

In order to assess this effect using realistic water treatment plant LRV data, two contrasting multiple barriers systems were conceptually composed as described in Table 22 and Table 23. These two tables represent two different treatment trains.

Table 22 provides PDFs for Giardia LRVs for what is commonly reffered to as a 'full advanced treatment train' featuring microfiltration, reverse osmosis, UV-advanced oxidation and

chlorination. The PDFs were adopted from a recent WateReuse Foundation study that collected long-term operating data from full-scale plants to develop them (Walker et al., 2016).

Table 23 provides PDFs for Giardia LRVs for a non-membrane based treatment train. This train includes sand filtration, ozonation, UV-disinfection and chlorination. As above, The PDFs were adopted from a recent WateReuse Foundation study that collected long-term operating data from full-scale plants to develop them (Walker et al., 2016).

Table 22 PDF for Giardia LRV used for monte carlo simulation 1 – Full advanced treatment train

Barrier	**Giardia LRV**	**5th percentile**
Microfiltration	Normal(4.637796,0.022429)	4.61
Reverse Osmosis	Weibull(9.3,5.65)	4.11
UV-AOP	BetaGeneral(34.696,18.92,7.03077,7.99704)	7.55
Chlorination	Lognorm(3.244,3.3382,Shift(0.66014))	1.22
Summed 5th percentile		**17.49**

Table 23 PDF for Giardia LRV used for monte carlo simulation 2 – Non-membranes based treatment train

Barrier	**Giardia LRV**	**5th percentile**
Sand filtration	Lognorm(2.2307,0.18667,Shift(0.017559)	1.96
Ozonation	Lognorm(16.771,8.1651,Shift(0.2789))	7.34
UV-Disinfection	BetaGeneral(2.8076,2.0897,4.93709,5.69425)	5.11
Chlorination	Lognorm(3.244,3.3382,Shift(0.66014))	1.22
Summed 5th percentile		**15.63**

The 5th percentile values for each of the individual treatment barrier processes are shown in the last column of the tables. These are summed to give 17.49 LRV for the full advanced treatment train and 15.63 LRV for the non-membranes based train.

Monte Carlo simulation was used to generate PDFs for overall Giardia LRV across each of these two treatment trains. The results reveal combined 5th percentile Giardia LRV of 18.46 for full advanced treatment (Figure 89) and 17.8 for the non-membranes based treatment (Figure 90). This represents an increase of approximately 1 LRV for the full advanced treatment and 2 LRV for the non-membranes based treatment train.

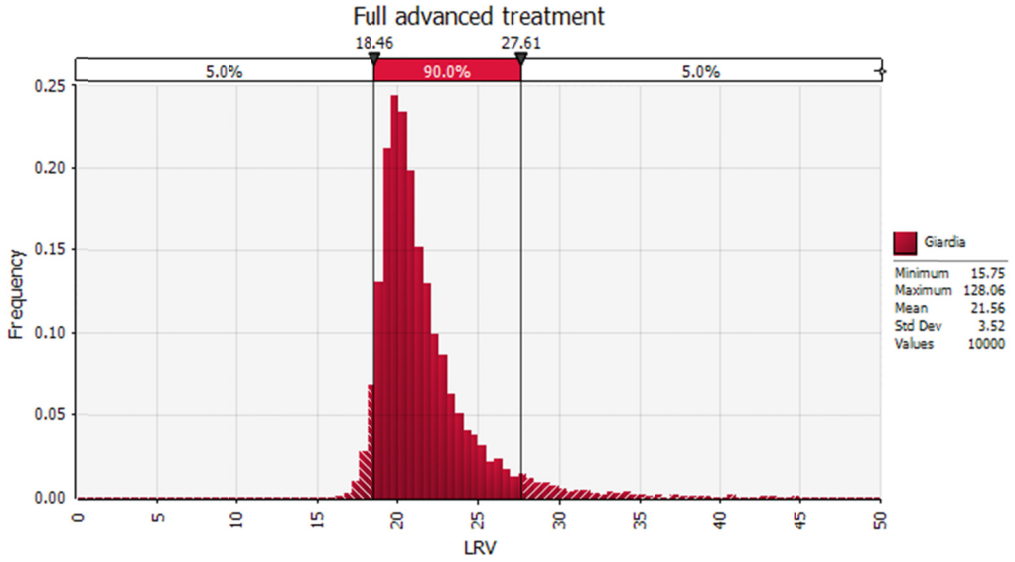


Figure 89 Combined multiple barrier Giardia LRV for a MF-RO-UV/AOP-Cl treatment train.

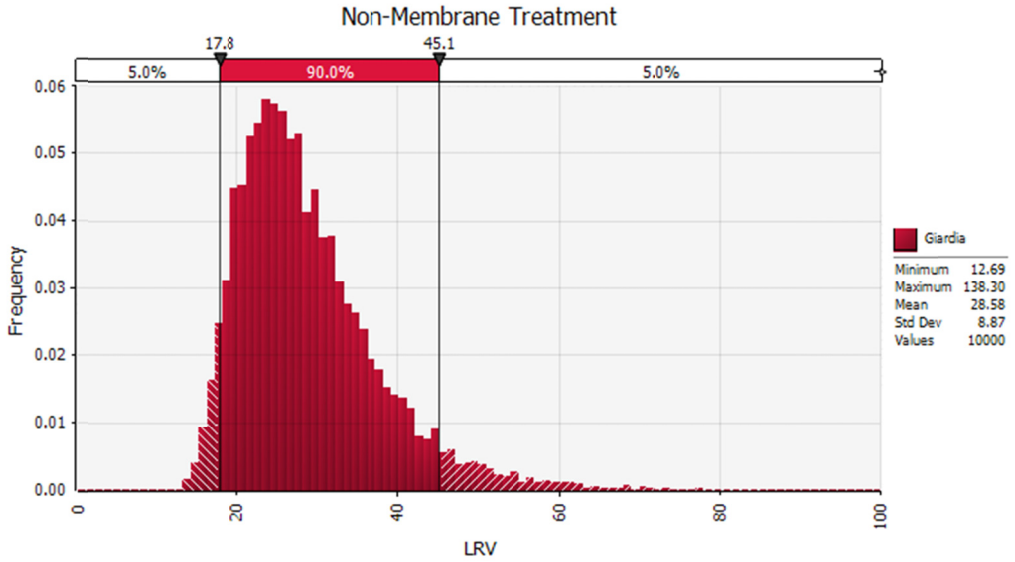


Figure 90 Combined multiple barrier Giardia LRV for a sand filtration-ozone-UV-chlorination treatment train.

11.3.2. Bayesian Belief Net simulations

Combined multiple barrier system LRVs derived by Bayes Networks

The same two systems can also be modelled in a Bayesian network. These are shown for the full advanced treatment train (Figure 91) and the non-membranes based treatment train (Figure 92). The 5th percentile can be read directly from the summary tables showing for the “combined LRV” nodes as the parameter “lower percentile”. Due to the probabilistic nature of these calculations, the figures are not identical to those derived by Monte Carlo simulation, but they are very close.

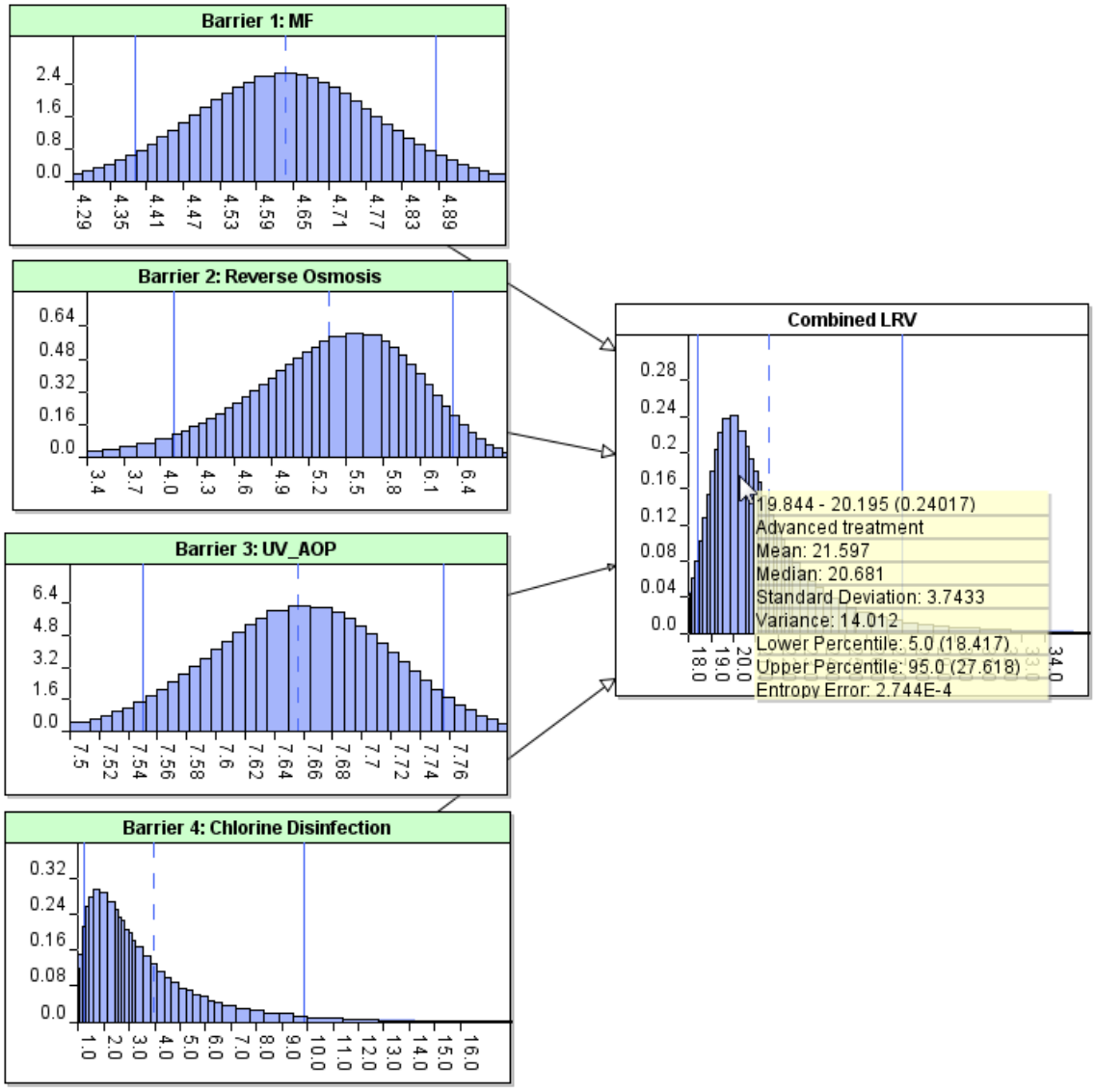


Figure 91 Combined multiple barrier Giardia LRV for a MF-RO-UV/AOP-Cl treatment train modelled as a BN

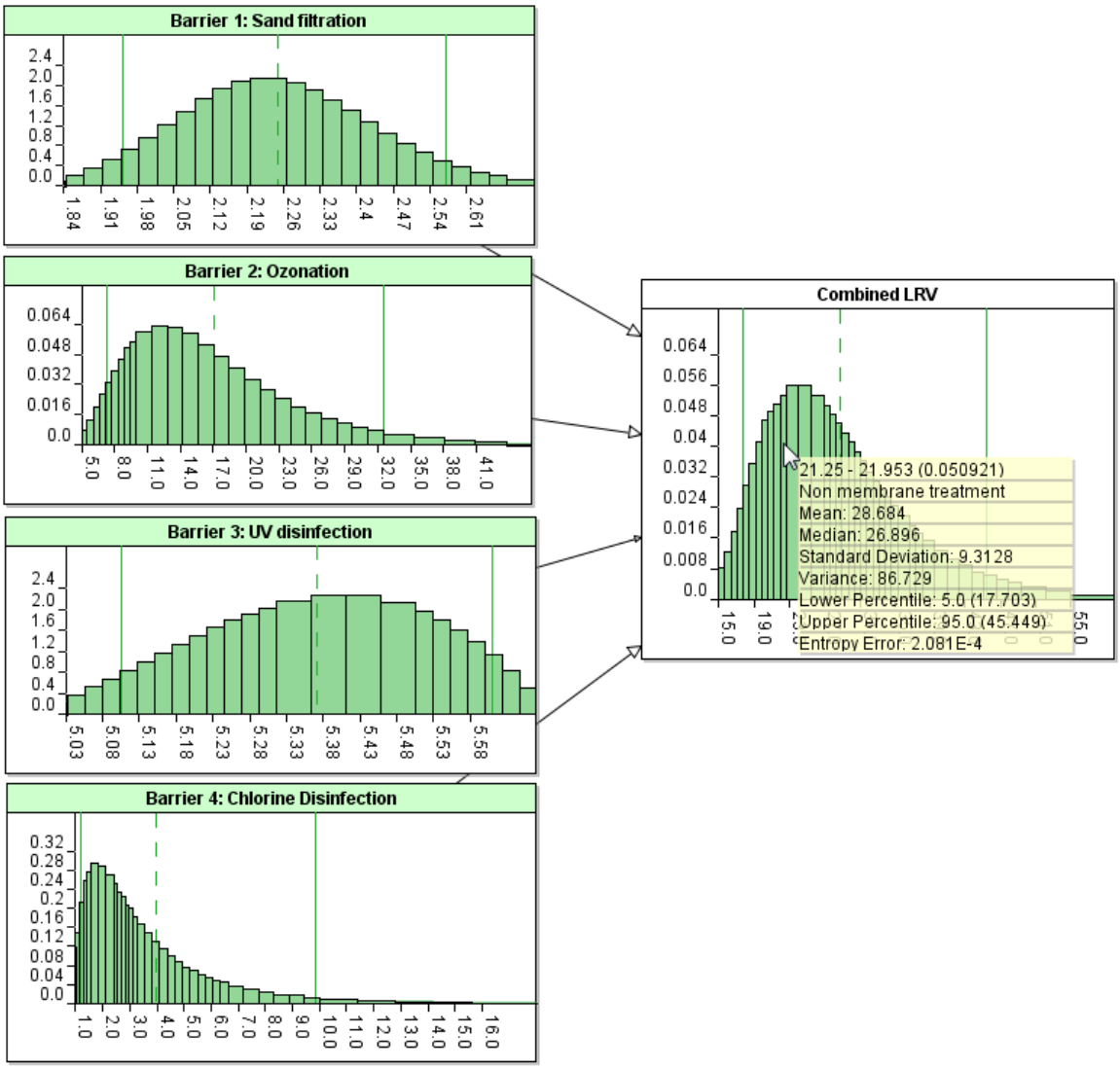


Figure 92 Combined multiple barrier Giardia LRV for a sand filtration-ozone-UV-chlorination treatment train modelled as a BN.

11.4. Conclusions

For multiple-barrier systems, there is an inherent increasing conservativeness associated with the approach to summing 5th percentile LRV values. The level of conservativeness increases with increasing numbers of independent multiple barriers. In some cases, the actual impacts to attributable LRVs may be very minor, even negligible. However, in others, it may be significant and worth exploring for addition LRVs that might be attributed at an accepted level of conservatism.

Overall system performance variation can be modelled in a number of ways. Here we have described the use of Monte Carlo simulations (using @Risk software) and Bayesian networks (using Agena Risk software). Both approaches are satisfactory and gave near identical results. However, a number of additional valuable functions were also demonstrated for the Bayesian networks approach. These included:

- Ease of construction
- Ease of system visualization
- Ease of model manipulation (e.g., to investigate scenarios)

Other advantages of the BN approach (not shown here, but described elsewhere in the report) include:

- Ease of comparison with validation targets
- Ease of model updating with new (even unrelated) data

For all of the above advantages, we consider that the use of BN as a general platform for water recycling process validation is a logical and worthwhile recommendation.

12. Additional aspects of Bayesian Validation

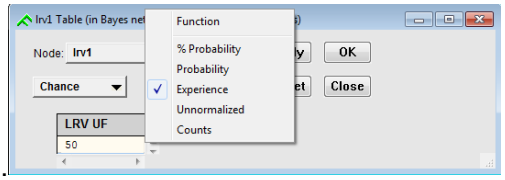
12.1. Sequential learning

The 'Learning' features of BN software allow a model to sequentially learn new data without removing existing parameters in the nodes. When Netica[tm] learns parameters from a case file, probabilities are calculated as well the extent of the experience (partly the number of records, partly a weighting factor). Experience corresponds to the number of cases that have been seen when using EM algorithm. In the case of "Counting" algorithm, the experience is equal to the number of cases plus the number of states. This is the result of adding a single count to every state and it is done to avoid zero or impossible probabilities, technique known as "Laplace smoothing". This effect vanishes as more data is inputted. However, with a limited dataset it will affect our probability distribution mean and standard deviation. It should be noted that the "real" number of values in the experience value is obtained through the EM algorithm. The gradient learning system though does not provide this estimate.

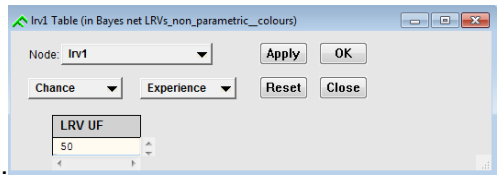
Figure 93 illustrates some of the dialogue boxes used. Figure 93a. and b. show where the extent of experience (e.g. number of records) going into a learned table can be entered.

Figure 93c shows how the underlying LRV CPT can be greatly altered.

a.



b.



c.

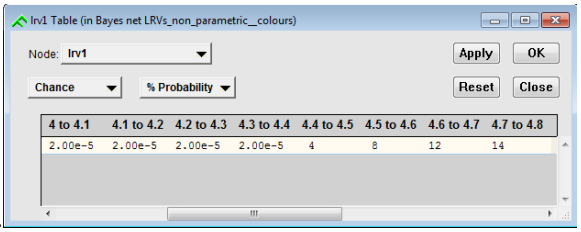
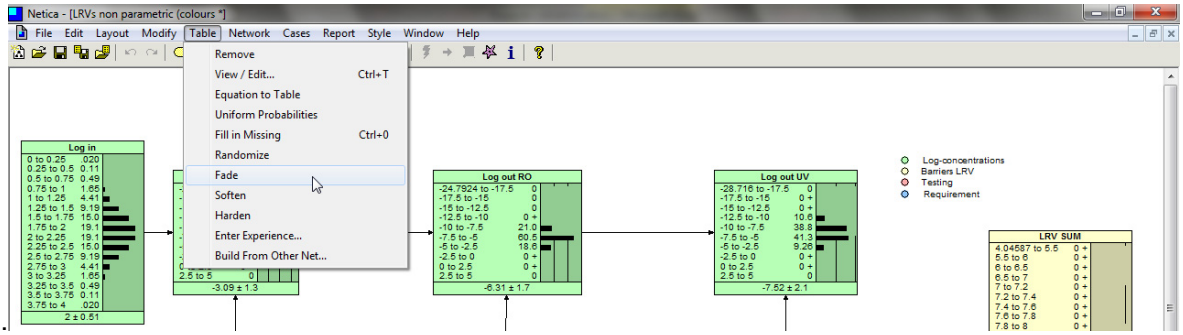


Figure 93. Dialogue boxes used during 'Learning'

12.2. Adaptation

Adaptation of the network to more recent conditions is performed through the "fade" option in Netica[tm] (Section of **Error! Reference source not found.** shown in Figure 94). Clicking this option prompts a request for a "degree of fading". This degree ranges between 0 to 1, with 0 having no effect, and 1 creating uniform distributions with no experience (thereby undoing all previous learning).

a.



b.

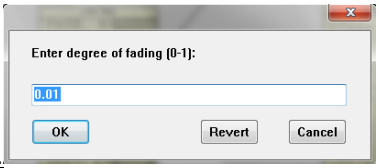


Figure 94. Use of Netica 'Fade' option

The effect is that the targeted node probabilities are flattened simulating increasing uncertainty about the central tendency. Hardening and softening produce similar changes though more in the vary of decreasing of increasing the typical spread as might be expressed by the node standard deviation.

The Fade calculation of the conditional probabilities with a specified degree of fading uses the following equation.

$$\boldsymbol{prob'} = \boldsymbol{normalize}(\boldsymbol{prob} \cdot \boldsymbol{exper} \cdot (\mathbf{1} - \boldsymbol{degree}) + \boldsymbol{degree})$$

The degree of fading is chosen by the user, but it can be specified according to the amount of time since the last fading was done (Δt), and how quickly the process is changing (r). The parameter r is a positive number less than 1 but close to 1. Different nodes may require different values of r.

$$\boldsymbol{degree} = \mathbf{1} - \boldsymbol{r}^{\Delta t}$$

Adaptation has the benefit of not dismissing all past information which can improve the reliability of the distributions and their percentiles.

12.3. Consequence X Likelihood Matrices

Current water quality management often involves the use of consequence X likelihood matrices. Figure 95 shows how their equivalent can be constructed using BNs. Indeed the matrix is comparable a Bayes contingency table. This allows child node management response to risks such as alerting different personnel within and external to an organisation to be linked logically to parent node risk estimates. Conversely the (parent) criteria for different (child) consequence and likelihood can be documented as can the (parent) real world events that correspond to these triggers. Such protocols systems can of course be constructed long hand. The benefits of Bayes though are:

The QRA process of Hazard ID + Dose Response + Exposure Assessment => Risk Characterisation => Risk Management can be concisely defined and constructed.

Development of prioritization rules is straightforward as is rule revision.

For a valid Bayesian system the network must have valid links and logically consistent input assumptions.

Extraction and documentation of these assumptions is straightforward.

Korb and Nicholson (2011 Mistake 8 Ch 10) questioned this use of a BN. They considered this represented confusion about what a node represents. However we suggest the BN provides a convenient way of capturing the logic of decision making.

We have also found this to be an effective tool for illustrating how BNs can support standard operating procedures.

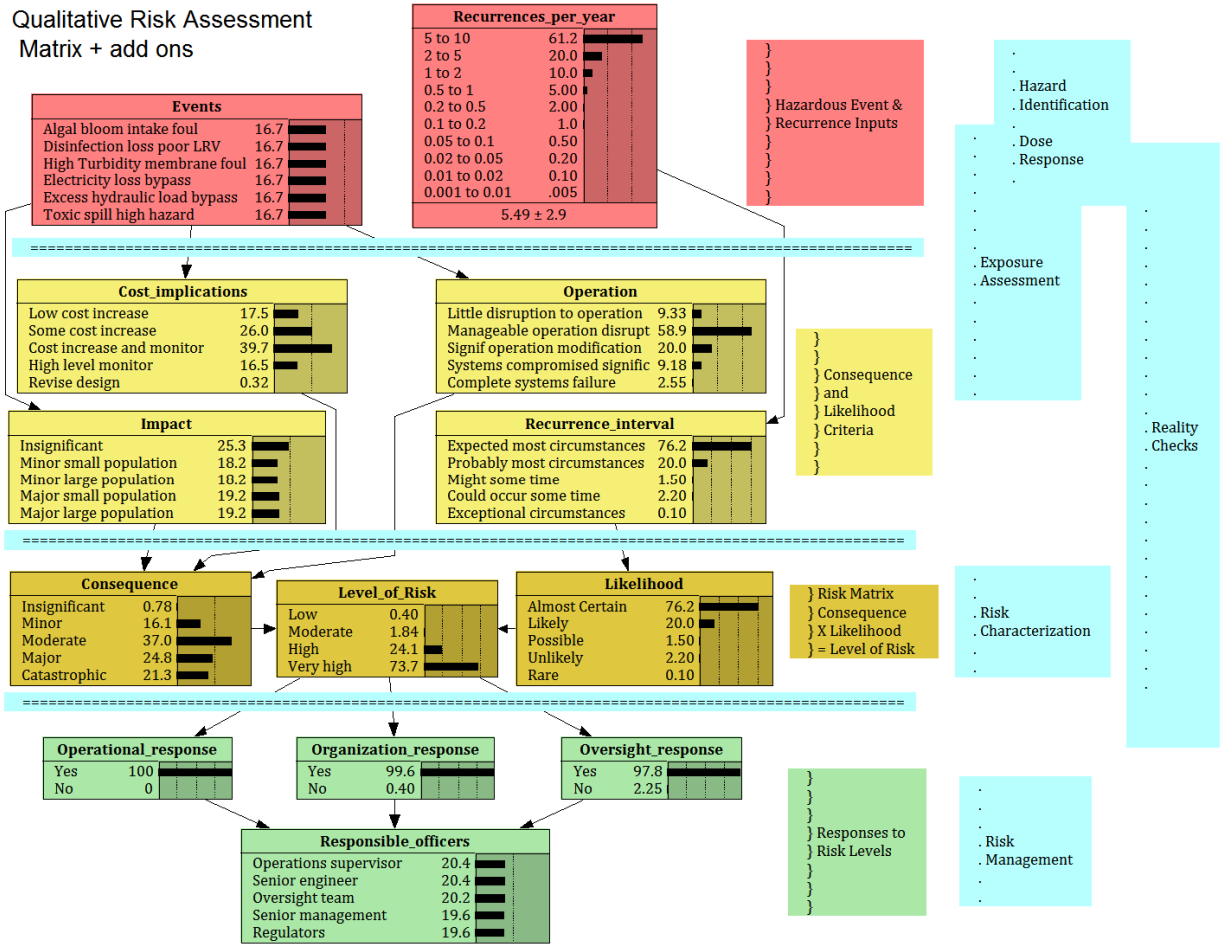


Figure 95. Extension of Consequence/Likelihood Risk assessment to Management and Predisposing Events

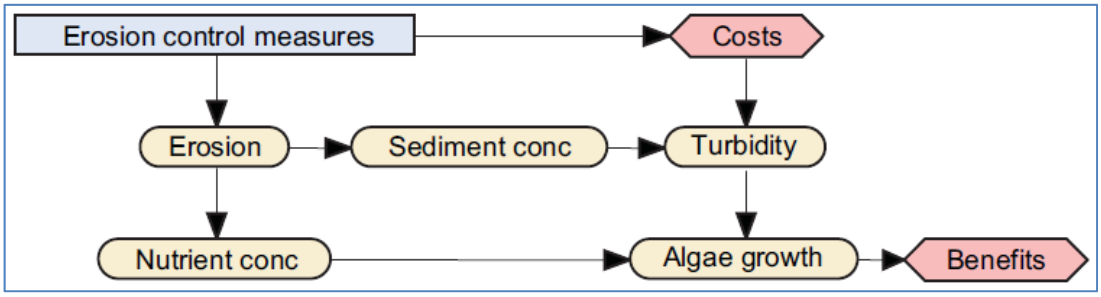
12.4. Overall system definition prior to Validation assessment – prior to high resolution Monte Carlo analysis

A feature of note in the paper by Kragt (2009) and many of the case studies on the Netica[tm] web site is that BNs nodes are largely non-parametric.

Rather they are used to work through the logic of decision sequences and what they imply for management. An example of a large and accessible water management case study is presented by Cain (2001).

In Netica[tm] there are many examples in the Netica[tm] library including one from Cain. Illustrative examples relating to water management from Kragt (2009) and Cain (2001) are shown in Figure 96. These show how there are many uses complementing validation for BN. The place for these in validation is making the process transparent and explaining the reasoning.

a.



b.

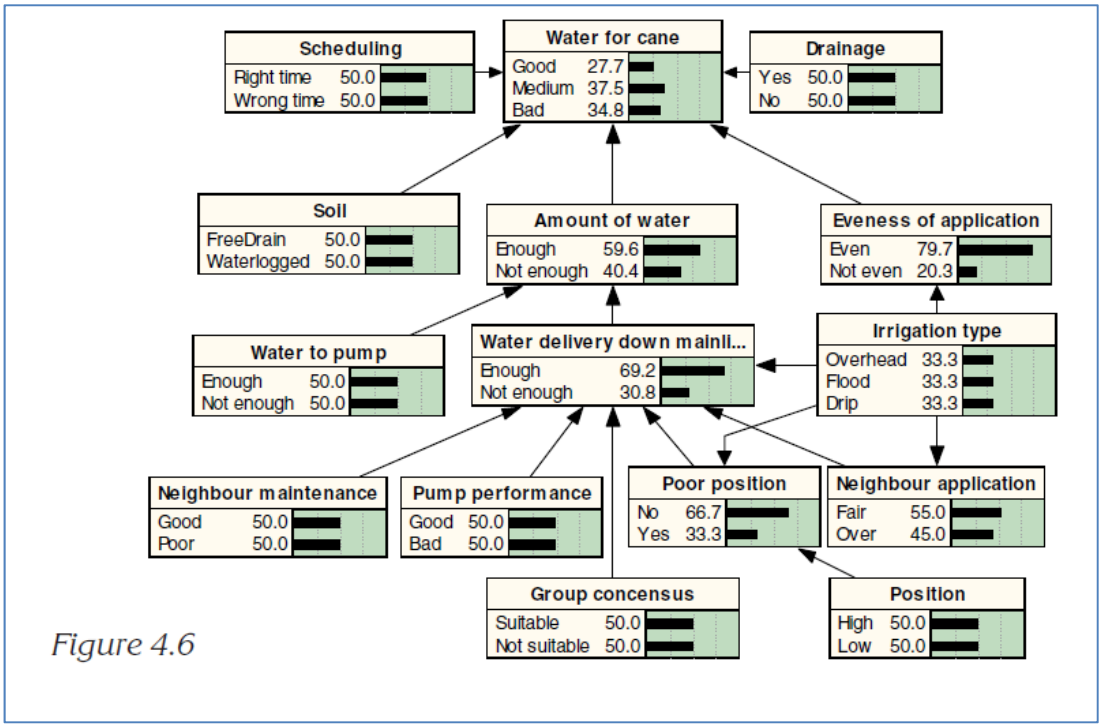


Figure 96. Illustrative decision supporting water management BNs – see Kragt (2009) and Cain (2001)

12.5. Validation sample sizes

BNs and Bayesian inference provide the proposed framework but they do not directly address the question of how many samples are needed for a given degree of measurement precision.

12.5.1. Sample size determination

Sampling design is a critical step in the validation of a treatment barrier. Determining sample size requires defining whether the objective is to infer the population statistic by the degree of power for a hypothesis test or a confidence interval with a specified width. In this case, the task consists of the estimation of log concentrations considering either the mean or 95th percentile without the inclusion of an alternative hypothesis. As the concept of an alternative hypothesis is not used in this situation, controlling the power by the sample size would not be appropriate. This section presents various approaches for computing the sample size for the estimation of the mean and 95th percentile. The methods herein assume that the data are normally distributed.

12.5.2. Sample size based on confidence interval/ Sample size based on mean (Variance known)

The width of the confidence interval is given by the difference between the upper and lower limits (UL and LL). As the statistic is positioned in the middle of this interval, the maximum error of estimation is E=W/2, with probability 1-α. Thus, sample size (n) could be computed for a specified maximum error of estimation.

$$\begin{aligned} W &= UL - LL \\ &= \bar{X} + \frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}} - \left(\bar{X} - \frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}} \right) \\ &= 2 \cdot \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}} \right) = 2 \cdot E \end{aligned}$$

Equation 1

If the variance for the population is known from previous studies (n≥30) the following equation can be used:

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{E} \right)^2$$

Equation 2

When the number of samples from previous studies is limited (n<30) then the t-distribution can be used in this case:

$$n = \left(\frac{t_{\frac{\alpha}{2}, n-1} \cdot S}{E} \right)^2$$

Equation 3

To obtain the t statistic in this equation we require the degrees of freedom (n-1), therefore this equation must be solved by trial and error.

12.5.3. Sample size based on percentiles

Percentiles are common as a measure of reliability as they denote the value of a variable that has a specific percentage of the distribution at or below it. In this sense, percentiles can be more meaningful and useful than the mean for compliance and performance assessment.

We present two different methods (Asymptotic Normality interval and Exact equal-tailed T-interval) for computing the sample size of a percentile which were adapted from the confidence intervals of a normal percentile developed by (Chakraborti and Li, 2007). It should be noted the t-statistic is used, so that an iterative procedure must be used to find the sample size.

A third approach will be considered which is based on Equation 1 and an uncertainty factor (Ellis, 1989). This methodology is recommended by the (NHMRC, 2011). We compare the sample sizes obtained by the three methods for different hypothetical cases. This will serve as a comparison point to assess the suitability of the 20 samples (over 12 months) proposed by the (Department of Health Victoria, 2013).

Equation 4 was adapted from the “Asymptotic Normality interval”, whereas Equation 5 was obtained from the “Exact equal-tailed T-interval” which utilises the gamma function.

$$n = \left(\frac{t_{1-\frac{\alpha}{2},n-1} \cdot S \cdot e^{Z_p^2/2}}{E} \right)^2 \cdot 2 \cdot \pi \cdot p \cdot (1-p)$$

Equation 4

$$n = \left(\frac{t_{\frac{\alpha}{2},n-1} \cdot S}{E} \right)^2 \cdot \left(1 + n \cdot Z_p^2 \cdot (C^2 - 1)\right)$$

Equation 5

$$C = \frac{\sqrt{\frac{n-1}{2}} \cdot \Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

Reflecting these equations/methods sample sizes for different levels of precision and standard deviations are presented in Table 24, Table 25, and Table 26.

Table 24. Sample size based on "Asymptotic Normality interval"

	Precision (E)				
Standard deviation	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
0.1	20	7	5	4	4
0.15	42	13	7	5	5
0.2	72	20	11	7	6
0.25	110	30	15	10	7
0.3	157	42	20	13	9
0.35	213	55	26	16	11
0.4	277	72	33	20	14
0.45	350	90	42	25	17
0.5	432	110	51	30	20
0.55	522	133	61	35	24
0.6	620	157	72	42	28

Table 25. Sample size based on "Exact equal-tailed T-interval"

	Precision (E)				
Standard deviation	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
0.1	10	5	4	3	3
0.15	18	7	5	4	4
0.2	29	10	6	5	4
0.25	44	13	8	6	5
0.3	61	18	10	7	6
0.35	82	23	12	8	6
0.4	106	29	15	10	7
0.45	134	36	18	12	9
0.5	164	44	21	13	10
0.55	198	52	25	16	11
0.6	235	61	29	18	13

Table 26. Sample size based on normal distribution and uncertainty factor

	Precision (E)				
Standard deviation	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
0.1	11	3	2	1	1
0.15	24	6	3	2	1
0.2	42	11	5	3	2
0.25	65	17	8	5	3
0.3	94	24	11	6	4
0.35	128	32	15	8	6
0.4	167	42	19	11	7
0.45	211	53	24	14	9
0.5	260	65	29	17	11
0.55	315	79	35	20	13
0.6	375	94	42	24	15

12.5.4. Example

A lagoon system is being validated for Adenovirus removal. The results of a previous validation study showed that the standard deviation for the $\log_{10}$ concentration in the inlet is 0.50 and for the outlet 0.25. It is decided that the sampling will be defined according to a maximum error of 0.3 log.

According to the three methods previously presented, the required number of samples would be as it is shown in Table 27.

Table 27. Number of samples required according to methods 1 to 3

Method	Number of samples (inlet)	Number of samples (outlet)
Method 1	51	15
Method 2	21	8
Method 3	29	8

As it can be seen in Table 27, due to the higher variability in the influent results the number of samples for the inlet is larger than for the outlet. Methods 2 and 3 returned very similar results. However, method 1 was highly conservative which is in agreement with the results obtained by (Chakraborti and Li, 2007). Considering this results, 20 samples as recommended by the (Department of Health Victoria, 2013) would be an appropriate number.

Using Equations 2 or 3 based on the mean instead of percentiles would give smaller sample sizes.

13. Conclusions

As evidenced by current state recycled water validation guidelines (Queensland Water Supply Regulator et al., 2013, Department of Health Victoria, 2013) there are many considerations that go into the validation of water recycling schemes. Not least of all are the challenges of coherently and quantitatively combining:

- Different quantitative data sets;
- Other relevant information including expert opinion and regulator aims.

The combination of worst case/95th percentile performance is inefficient in that it does not properly credit periods when systems perform nominal or even better than expected.

13.1. In summary

Our proposal is that the process of inference itself needs to be made more systematic, quantitative and credible through being based on fundamental statistical theory in particular Bayesian inference. Complementing this, operational validation should exploit a recent technology reflecting the logic of Bayes' Theorem, Bayes Nets.

The diverse case studies presented in this report illustrate and support our suggestion that BNs can be used to integrate any well definable (in effect quantifiable) influence on water recycling and use this in LRV estimation in assigning defensible LRV credits for treatment systems and also the barrier effects of downstream buffer zones.

Validation is "the confirmation that the treatment technology meets the specified performance targets." We interpreted this to operationally mean using BNs to learn, define and estimate:

- Prior (LRV) probabilities which define what performance is expected of one or more processes expressed in the form of PDFs – such as the knowledge and data from:
    - earlier test results (in the case of a revalidation trial);
    - manufacturer claims;
    - expert opinion;
    - performance of similar systems operated by the same authority at similar plants under similar circumstances;
    - refereed and high quality grey literature especially meta-analyses which combine the data for several comparable systems.
- New (prior probability) evidence, in practice the validation data set – as illustrated by the 2015 revalidation LRVs from the Glenelg and ETP case studies.
- A combined final LRV credit based on all input data available – in Bayes terms – a set of revised posterior probabilities for LRV performance of the specific system of interest.

13.2. The Benefits of Bayesian methods

The benefits of doing this in a Bayesian fashion are as follows:

- LRV estimation automatically becomes probabilistic taking into account uncertainty and variance to an extent determined by the data available and opinions of the managers/decisions makers. This can be as little or as much as desired.
- Point estimates/guesses/simple opinions (100% probability of LRVs being a particular value) are possible to include.
- All such inputs can be audited and readily modified as desired including in real time in cooperative settings (workshops, proponent/regulator/auditor meetings).

- Contingency factors are simple to include.
- Virtually all ISO 31010 risk management approaches can be emulated in a single platform.
- Bayes probability provides a rational method for combining all data inputs in a probabilistic fashion and yes/no decisions (essentially 100% acceptable/unacceptable) can be coded using the same system.
- Hazardous events can be considered and incorporated easily by treating them as a new node or an extreme range value or state. Multiple hazardous events and their interaction can be characterized by Fault Tree Analysis.
- The process is underpinned by Bayesian inference, causality and belief concepts and thus dispenses with the need for what has seemed in past *ad hoc* decision making or conclusions – estimating LRV credits without specifying where a selected value comes from.

Many of the statistical manipulations reported here might also have been done using more conventional summary statistics and regression analyses. However BNs provided many additional benefits.

BNs are grounded in a theory of causality and probabilistic inference. This has not been so much the case with historical use of frequentist statistical approaches beyond basic hypothesis testing which does not capture the complexity of systems and how their different components are likely to interact. This appears to be because purely frequentist statisticians tend to reject complex inference.

The format of BNs is identical to that of HACCP diagrams which is already used for water safety plans. Conditional Probability Tables which are central to Bayesian inference are comparable to the qualitative risk matrix widely used in the water industry. And BNs can be used in virtually the same way as Monte Carlo programs to calculate risks. Thus the water industry is already moving in the direction of BN use if not explicitly. But this implies its thinking about risk is ripe for adopting Bayesian methods if it so chooses and this transition should be smooth rather than disruptive i.e. BN provide and theoretical framework and techniques for doing what the water industry is already doing only more transparently and efficiently.

The rise in a risk assessment based framework in water management has led to validation and other water risk management work being inherently about asking questions and addressing them through scenario exploration. BNs are ideal for this purpose. BNs appear to provide a means for grounding expert opinion/intuition.

The graphic format that BNs use concisely summarizes the central information on a recycled water system often in one page, coming close to the ideal of being able to develop one page summaries for senior management which do not need to omit details while the software can make more details readily accessible if so desired.

13.3. Addressing communication and other human factors

The findings and input assumptions in BNs are easily audited provided such auditors or regulators understand the Bayesian terminology concepts and approach that could be called Bayesian thinking. The latter may prove challenging for those unfamiliar with this approach. However this communications barrier seems possible to overcome. And once this is done BNs appear to offer a common interface/language at least as intelligible as spreadsheets if not more for many diverse tasks. Further, the reasoning underlying BNs is in fact what engineers scientists or medical epidemiologists should already be familiar from applying risk principles to

water management, the only things lacking being the formal processes and familiarity with the algebra.

The outputs and data which may be incorporated in BN and data learning processes seem more limited by imagination and availability that inherent impracticality. And there appears no barrier to reality checking of models. Indeed testing of accuracy has for some time been a major feature of best practice BN application.

BNs design can produce many different but related models. However the model construction process can still be managed. Two complementary approaches are available. On one hand causal models capture how we believe a system works. In the case of water recycling this is probably much better grounded than in medicine or ecosystem management where BNs are used increasingly in that engineers have generally designed the systems in the first place based on intuitive inference and long experience in wastewater treatment which shows their primary beliefs are well justified.

Separately 'semi/naïve' Bayes models can be constructed which reduce the emphasis on belief and increase the relative influence of mathematical optimisation techniques in the identification of variable/node relationships. Being machine generated, such BNs are arguably much less biased. Both kinds were demonstrated to be applicable and reassuringly generated similar LRVs and influences when applied to the same input data sets.

BN construction can be done by individuals or specialists or BN consultants in combination with water specialists. But for optimum communication and learning, some understanding of BNs operation and BN principles, on the part of the water experts is essential for all stakeholders including regulators and auditors. This needs to include understanding Bayesian inference and causality and key model features such as, as well as limitations. Ideally it would include some diverse experience in using BN software and these limitations.

In our experience it is not possible for a BN to be easily constructed using simply a consultant with little knowledge of the problem at hand and a water expert with little knowledge of how BNs and Bayesian inference works. Compartmentalisation does not work.

The use of manufacturer data raises questions of 'commercial in confidence' when it comes to using LRV data and other knowledge of operating conditions. Discussions with manufacturers of treatment systems are probably needed to find a balance between transparency and confidentiality.

A possible compromise would be to require manufacturers to supply, in addition to mid-range estimates, simple PDF functions describing LRVs and other relevant information e.g. operating conditions such as transmembrane pressures. Point value data will likely be insufficient.

13.4. Bayesian Validation

A particular recycled water system validation challenge BNs can address well is that of integrating multiple disparate information sources provided they can be framed quantitatively. The Queensland Guidelines section 3.8.7 (Queensland Water Supply Regulator et al., 2013) propose consideration of all sources of information during the validation process but the integration process advanced is a relatively crude semi-quantitative one that does not yield LRVs. Bayes provides a more quantitative approach which we suggest be call 'Bayesian Validation' and involves the incorporation and integration of multiple priors.

13.4.1. 'Enter Findings' method

The first approach, assumes a ***prior*** parametric distribution for the LRVs. ***Priors*** in the first approach are specified for the parameters of the distribution e.g. mean and standard deviation for normal distribution in the example shown.

The nodes for the ***new evidence*** validation test data to be entered are shown in green. One datum is entered at a time which accounts for the large number of validation nodes.

The overall result of this analysis is a group of distributions for each subset of parameters, which gives a higher estimated LRV variability. The revised BN in turn can be used for example to assess whether there has been an improvement on the original design expectation.

13.4.2. Distribution integration method

The second method uses a nonparametric approach. It incorporates the ***prior*** data directly into the distribution or interest. This requires assigning relative weightings to the earlier ***prior*** and the ***new evidence*** validation data set. The outcome in this case is a single distribution. This process is automated by software such as Netica^{tm}.

This process may sound complicated but in practice it takes a few seconds to implement.

13.4.3. Interactive distribution integration method

The final method builds on the previous one. It includes nodes which allow the user to interactively change the weighting given to each of the different priors and new evidence using another belief bar node.

The method is most clearly illustrated in the Glenelg case study. This method allows experts in a workshop to interactively come to a quantitative consensus on how effective a process or treatment train will be based on all the data they have available and clearly document the resulting beliefs and final LRVs.

13.5. General use recommendations

Finally it can be seen that there are in effect 2 types of BNs which can be used for different purposes.

As a guide we suggest:

Use Causal BNs for well-defined systems, that is:

- Cases where a validation approach (what to monitor and how to interpret it) is already well developed and accepted;
- Cases where the system is well defined such as a treatment train;
- Simple systems where cause-effect relationships can be confidently introduced by expert knowledge;
- Systems where there are strong causal beliefs to explore.

Naïve/Semi-Naïve BNs may be needed also for less well-defined systems:

- In particular, biological systems where many mechanistic knowledge gaps remain;
- Activated sludge and MBRs seem to be appropriate candidates;
- Simpler systems where there is a desire to identify alternative monitoring parameters.

Where appropriate use both approaches in combination especially where there are large data sets e.g.:

- Use data mining semi/naïve BNs to first evaluate if there are useful relationships that can be identified in a data set prior to trying to developing causal BNs;
- To estimate likely BN accuracy;

Employ BN best practice as recommended for example by:

- (Marcot, 2012, Marcot et al., 2006);
- (Korb and Nicholson, 2011);
- (Pollino et al., 2007, Pollino et al., 2012, Chen and Pollino, 2012).

14. Glossary

AI	Artificial Intelligence. Though relevant to, AI is not just about robots. It is firstly about the mathematization of reasoning based on data especially the modern large data set now available through computer and communications technology. cf (Korb and Nicholson, 2011) which despite its name is about BN directed data collection and manipulation, not about SF style robots.
AIC	Akaike Information Criterion score,
ASP/AS	Activated Sludge Process/Activated Sludge (treatment)
Assessment (Bayesian)	(As used here) Probability assessment is the process of humans determining the probabilistic or deterministic relationships between nodes and their parents (usually in the form of conditional probability tables) after all the nodes and the link structure have been created. Alternatively, they can be determined automatically by some learning procedure.
AUC	Area Under the Curve for the receiver operating characteristic curve
BAN	Bayesian network augmented naïve Bayes
BN	A Bayes net (also known as a belief net) is composed of a set of nodes representing variables of interest, connected by links to indicate dependencies, and containing information about the relationships between the nodes (often in the form of conditional probabilities). Usages include prediction, diagnosis, probabilistic modelling, learning from data and forming a basis for building decision nets.
BN/BBN	Bayesian Belief Network – we have standardized on 'BN' but it is essential to not forget the involvement of 'Belief' in the construction of BNs and the subtle traps it lays.
Belief	The belief of a node is the set of probabilities (one for each of its possible states), taking into account the currently entered findings by using the knowledge encoded in the Bayes net. Technically, it is the marginal posterior probability distribution of the node, given the findings and the BN model. Sometimes the plural form "beliefs" is used to mean each of the probabilities in the set.
Belief updating	Belief updating is the process of finding new beliefs for the nodes of a BN to account for the findings that are currently known. It is a form of probabilistic inference. During belief updating the BN model (in particular, the conditional probability tables between the nodes) is not modified at all; for that probability revision is used.
ca	*Circa about* = approximately
Case	A case is a set of findings that go together to provide information on one object, event, history, person, or other thing.
cf.	Compare for example
Chance node	A chance node is a nature node whose relationship to its parents is probabilistic (i.e. not deterministic). If its parents' values are all known, and there is no further information, then its value can only be inferred as a probability distribution over possible values. Compare with deterministic node.
Child node	BNs are directional. If there is a link going from node A to node B, then B is said to be a child node of A. Some people refer to it as a direct successor.
Conditional probability	The conditional probability of an event is the probability of the event occurring under certain given conditions.
COPC	Contaminant of Potential Concern. Typically pathogens and toxic or carcinogenic chemicals which may be present in recycled water.
CPT	CPT is an abbreviation for conditional probability table (also known as "link matrix"), which is the contingency table of conditional probabilities stored at each node, containing the probabilities of the node given each configuration of parent values. Sometimes CPT is used to refer to the deterministic function table of a node, since the node's conditional probabilities can easily be found from that. It is a form of node relation, so you use the table dialog box to change or view it.
.CSV	File format standing for Comma Separated Values. These are a standard data storage file format suitable for use by many software packages including Excel, Netica^{tm} and WEKA.
DAG	Directed Acyclic Graph
Decision net	If decision nodes (representing variables that can be controlled) and utility nodes (representing variables to be optimized) are added to a BN, then a decision net (also known as an "influence diagram") is formed.
Decision node	A decision node is a node in a decision net which represents a variable (or choice) under the control of the decision maker. When the net is solved, a decision rule is found for the node

	which optimizes the expected utility (EU). Decision nodes are normally drawn as rectangles (without rounded corners).
Decision theory	Decision theory is a normative theory which indicates how a single agent should best make decisions to maximize his expected utility (EU). It considers sequences of decisions, what information the agent will have when he makes the decisions, uncertainties in the beliefs of the agent, and complex probabilistic interactions in the environment in which the agent is operating.
Deterministic node	A deterministic node is a nature node whose relationship with its parents is given as a function of the parent values (i.e. deterministic rather than probabilistic). If the parent values are all known, its value can be determined with certainty. Compare with chance node.
Entering findings	When a BN is applied to a particular situation, or case, then the known information about that case is entered into the BN by assigning values (called "findings", or "evidence") to the known variables (i.e. nodes), and that process is known as entering findings into the nodes. Entering a finding into a particular node does not retract existing findings at that node or other nodes (but for convenience, at least in Netica^{tm} applications, if the new finding for a node directly contradicts a previously entered finding for that node, the previous finding will be retracted first).
Expected value	The expected value (also known as mean value) is not the value you “expect” to see, and usually it isn’t even the value most likely to occur. This term, from probability theory, means the average value that will occur, where the average is weighted by the probability of occurrence. For example if a value will be 3 with probability 0.2 and 9 with probability 0.8, then the expected value is (0.2 x 3) + (0.8 x 9) = 7.8.
Finding	A finding (also known as “evidence”) is a value for one of the nodes (i.e. variables) of a BN when it is applied to a particular situation.
FNR	False negative rate
FPR	False positive rate
Function table	When the relationship between a node and its parents is deterministic, rather than probabilistic, then instead of a CPT a node may have function table, in which each row corresponds to a configuration of parent values, and the row provides a single output value for the child node. If a function table is converted to a CPT, then each row of the resulting CPT will consist only of zeroes, with a single 1 (or 100%) positioned at the state that was the function table's value for that row.
GUI	Graphical User Interface
IDEA	Intermittently Decanted Extended Aeration
Informational link	Any link entering a decision node is known as an informational link, and indicates that the decision maker will know the value of the parent node when he must make that decision.
KS	Kappa statistic
Leaf node	A leaf node is a node with no children.
Link	A link (also known as an “arc” or an “edge”) is a connection between two nodes indicating dependence, and is usually drawn as a line with an arrow at one end.
LL	Log-Likelihood score
LRV	Log_{10} Reduction Value. Other acronyms used are DEC and DR for Decimal reduction. This value describes extent to which a process of barrier reduces a contaminant level. It is useful because the reductions on microbial numbers typically desired are quantified in logarithms. It is useful as for most purposes an LRV of 1 implies a 1 log risk reduction as well.
Nature node	A nature node in a BN represents some variable of interest. It may also appear in a decision net in which case it is a variable that cannot be directly controlled by the decision maker (i.e. it is determined by nature). If a nature node has a functional relationship with its parents, it is called a deterministic node, whereas if the relationship is probabilistic, it is called a chance node. The characteristic shape for a nature node is an ellipse, or a rectangle with rounded corners.
NB	Naïve BN
Net	In Netica^{tm} documentation, the word net is used to mean a BN or a decision net.
Netica^{tm}	Netica^{tm} is a program created by Norsys for working with BNs and decision nets.
Node	A node is a component of a BN or decision net used to represent a variable (i.e. scalar quantity) of interest, and in Netica^{tm} is usually drawn as a rectangle, rounded rectangle, circle or flattened hexagon.
Node relationship	A node relationship, or node relation for short, is the relationship between a node and its parents. It may provide the value of the node as a function of its parents’ values, or it may provide a probability distribution for the node depending on its parents’ values. It is often

	expressed as a CPT in which case it can be viewed or edited using the table dialog box. Alternately, it may be expressed as a probabilistic or deterministic equation.
No-forgetting links	If a decision maker remembers the decisions they made at an earlier time, and also the knowledge they had available at that time, then in his decision net there will be informational links going from earlier decision nodes and their parents, to later decision nodes. These are called no-forgetting links.
Outcome	The outcome is the result of an event, or series of events, that could have turned out in one of several ways.
PA	Prediction Accuracy
Parent node	If there is a link going from node A to node B, then A is said to be a parent node of B. Some people refer to it as a "direct predecessor".
Probabilistic inference	Probabilistic inference is the process of calculating new beliefs for a set of variables, given some findings. Technically speaking, it is the process of finding a posterior distribution, given a prior distribution, a model and some observations.
Root node	A root node is a node with no parents. See also leaf node.
SNB	Semi-Naïve Bayesian Network
States	A discrete variable can take on one of several values, and these values are called states. For example the states may be "female, male", or they might be "US, Europe, Japan, China", or "True, False". With Netica^{tm} you can just let the states of a node be numbered, but usually you give them meaningful names.
TAN	Tree Augmented naïve Bayes
TNR	True negative rate
TPR	True positive rate
User reports	A Netica^{tm} mechanism which displays customized information pertaining to a node, group of nodes, or to an entire net. The user report could be as simple as a text message giving a more detailed description of what a node means. Or it could be more complex, such as the current belief probabilities of the nodes, or a sensitivity analysis of the net.
Utility node	A utility node (also known as a "value node") is a node in a decision net whose expected value is to be maximized while searching for the best decision rule for each of the decision nodes. It is usually drawn as a flattened hexagon or a diamond.
WEKA	Waikato Environment for Knowledge Analysis
ZeroR	-

15. References

BOLLEN, K. A. (2002) Latent variables in psychology and the social sciences. *Annual review of psychology,* 53**,** 605-634.

CAIN, J. (2001) Planning improvements in natural resources management: Guidelines for using Bayesian networks to support the planning and management of development programmes in the water sector and beyond Centre for Ecology and Hydrology Oxford UK ISBN 0 903741 00 9.

CARVAJAL, G., ROSER, D. J., SISSON, S. A., KEEGAN, A. & KHAN, S. J. (2015) Modelling pathogen log10 reduction values achieved by activated sludge treatment using naïve and semi naïve Bayes network models. *Water Research (in press)*.

CHAKRABORTI, S. & LI, J. (2007) Confidence interval estimation of a normal percentile. *The American Statistician,* 61**,** 331-336.

CHEN, S. H. & POLLINO, C. A. (2012) Good practice in Bayesian network modelling. *Environmental Modelling & Software,* 37 134-145.

CLARK, J. S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters,* 8**,** 2-14.

COOK, A., DEVINE, B., RODRIGUEZ, C., ROSER, D., KHAN, S., MCGUINNESS, N., ASHBOLT, N. & WEINSTEIN, P. (2013 (accessed)) ASSESSING THE PUBLIC HEALTH IMPACTS OF RECYCLED WATER USE Interim report.

DEPARTMENT OF HEALTH VICTORIA (2013) Guidelines for validating treatment processes for pathogen reduction Supporting Class A recycled water schemes in Victoria February 2013. http://docs.health.vic.gov.au/docs/doc/41FDFAB63422E5CFCA257B180077F489/$FILE/DHD%2013%202708%20%20Guidelines%20for%20validating%20treatment%20processes%20for%20pathogen%20reduction%20(Web%20version).pdf (accessed Sept 2013).

DONALD, M., COOK, A. & MENGERSEN, K. (2009) Bayesian Network for Risk of Diarrhea Associated with the Use of Recycled Water. *Risk Analysis,* 29**,** 1672-1685.

DONALD, M., MENGERSEN, K., TOZE, S., SIDHU, J. & COOK, A. (2010) Incorporating parameter uncertainty into Quantitative Microbial Risk Assessment (QMRA). *Journal of Water and Health,* 9**,** 10-26.

DOWIE, J. (2006) The Bayesian approach to decision making. IN KILLORAN, A., SWAN, C. & KELLY, M. (Eds.) *Evidence at the crossroads: new directions in changing the health of the public: a manual.* Oxford University Press, Oxford.

ELLIS, J. (1989) Handbook on the design and interpretation of monitoring programmes,. Water Research Centre.

ELLISON, A. M. (1996) An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications,* 6**,** 1036-1046.

FENTON, N. & NEIL, M. (2012) *Risk assessment and decision analysis with Bayesian networks,* Boca Raton, London, NY, CRC Press.

FLAPPER, T., CAMPBELL, B., O'CONNOR, N. & KEEGAN, A. (2012) Quantification of pathogen removal in Australian Activated sludge plants (Phase 1 and 2). Projects 512-001 and 72M-7104. WQRA.

FLAPPER, T. G., CAMPBELL, B., DEERE, D., BLACKBEARD, J. & HALLIWELL, D. (2010) QUANTIFYING PATHOGEN LOG REDUCTION IN AUSTRALIAN ACTIVATED SLUDGE PLANTS. *Water (Australian Water Association),* FEBRUARY**,** 56-62.

HAAS, C. N. & EISENBERG, J. N. S. (2001) Risk assessment. IN FEWTRELL, L., BARTRAM J., (Ed.) *Water Quality Guidelines, Standards and Health: Assessment of risk and risk management for water-related infectious disease.* London UK., IWA Publishing/WHO.

HAAS, C. N., ROSE, J. B. & GERBA, C. P. (1999) *Quantitative Microbial Risk Assessment,* New York, John Wiley and Sons, Inc.

HAAS, C. N. & TRUSSELL, R. R. (1998) Frameworks for assessing reliability of multiple, independent barriers in potable water reuse. *Water Sci. Technol.,* 38**,** 1-8.

HINCKS, T. K., KOMOROWSKI, J.-C., SPARKS, S. R. & ASPINALL, W. P. (2014) Retrospective analysis of uncertain eruption precursors at La Soufrière volcano, Guadeloupe, 1975–77: volcanic hazard assessment using a Bayesian Belief Network approach. *Journal of Applied Volcanology,* 3**,** 1-26.

IEC/ISO (2009) IEC/ISO 31010 Risk management - Risk assessment techniques Edition 1.0 2009-11.

ISO (2009) Risk management - Principles and guidelines: INTERNATIONAL STANDARD ISO 31000 First edition. ISO.

KHAN, S. & ROSER, D. J. (2007) Risk Assessment and Health Effects Studies of Indirect Potable Reuse Schemes Final Report.

KHAN, S. J. (2010) Quantitative Chemical Exposure Assessment for Water Recycling Schemes. IN GOVERNMENT, N. W. C. A. (Ed.) Canberra, Waterlines.

KHAN, S. J., ROSER, D. J., ASHBOLT, N. J., LOVELL, A. & ANGLES, M. (2007) Health risk assessment for recycling for replacement river flows. *3rd AWA Water Recycling and Reuse Conference.* Australian Water Association, Sydney, NSW.

KORB, K. B. & NICHOLSON, A. E. (2011) *Bayesian artificial intelligence*, CRC press.

KRAGT, M. E. (2009) A beginners guide to Bayesian network modelling for integrated catchment management. Technical Report No. 9. http://www.landscapelogic.org.au/publications/Technical_Reports/No_9_BNs_for_Integrated_Catchment_Management.pdf.

LAURITZEN, S. L. & SPIEGELHALTER, D. J. (1988) Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 157-224.

LINDHE, A., NORBERG, T. & ROSEN, L. (2012a) Approximate dynamic fault tree calculations for modelling water supply risks. *Reliability Engineering & System Safety,* 106**,** 61-71.

LINDHE, A., NORBERG, T. & ROSEN, L. (2012b) Approximate dynamic fault tree calculations for modelling water supply risks. *Reliability Engineering and System Safety,* 106 61-71.

LINDHE, A., ROSÉN, L. & HOKSTAD, P. (2010) Risk Evaluation and Decision Support for Drinking Water Systems Deliverables number D4.2.5 and D4.4.2. *Risk (quality).*

LINDHE, A., ROSEN, L., NORBERG, T. & BERGSTEDT, O. (2009) Fault tree analysis for integrated and probabilistic risk analysis of drinking water systems *Water Research,* 43**,** 1641-1653.

MARCOT, B. G. (2012) Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling,* 230 50 - 62.

MARCOT, B. G., STEVENTON, J. D., SUTHERLAND, G. D. & MCCANN, R. K. (2006) Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research,* 36**,** 3063-3074.

MARKOV, Z. & RUSSELL, I. (2015(accessed)) An Introduction to the WEKA Data Mining System.

MEDEMA, G., LORET, J.-C., STENSTRÖM, T. A. & ASHBOLT, N. (2006) Quantitative Microbial Risk Assessment in the Water Safety Plan. Final report on the EU MicroRisk Project.

NADEBAUM, P., CHAPMAN, M., MORDEN, R. & RIZAK, S. (2004) A Guide To Hazard Identification and Risk Assessment For Drinking Water Supplies. Adelaide, CRC for Water Quality and Treatment Research.

NHMRC, N. (2011) Australian drinking water guidelines paper 6 national water quality management strategy. National Health and Medical Research Council, National Resource Management Ministerial Council, Commonwealth of Australia: Canberra.

NILSSON, P. (2006) The use of water treatment SCADA data to quantify microbiological hazardous events and risks arising – A case study from Sweden. *Department of Design Sciences, Division of Ergonomics and Aerosol Technology, Faculty of Engineering.* Box 118 SE-221 00 Lund, Lund University.

NORSYS SOFTWARE CORPORATION (2013) Netica http://www.norsys.com/.

PAGE, D., DILLON, P., VANDERZALM, J., TOZE, S., SIDHU, J., BARRY, K., LEVETT, K., KREMER, S. & REGEL, R. (2010) Risk assessment of aquifer storage transfer and recovery with urban stormwater for producing water of a potable quality. *Journal of Environmental Quality,* 39**,** 2029-2039.

PEARL, J. (1996) The Art and Science of Cause and Effect UCLA Lecture Series http://bayes.cs.ucla.edu/jp_home.html.

PEARL, J. (1999) REASONING WITH CAUSE AND EFFECT International Joint Conference in Artifical Intelligence 1999 Research Excellence Award Lecture http://singapore.cs.ucla.edu/IJCAI99/index.html.

PEARL, J. (2000) *Causality http://bayes.cs.ucla.edu/BOOK-99/book-toc.html*.

POLLINO, C. A., THOMAS, C. R. & HART, B. T. (2012) Introduction to Models and Risk Assessment. *Human and Ecological Risk Assessment: An International Journal,* 18**,** 13-15.

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 & Software,* 22**,** 1140-1152.

PORTINALE, L., RAITERI, D. C. & MONTANI, S. (2010) Supporting reliability engineers in exploiting the power of Dynamic Bayesian Networks. *International journal of approximate reasoning,* 51**,** 179-195.

QUEENSLAND WATER SUPPLY REGULATOR, WATER SUPPLY AND SEWERAGE SERVICES & SUPPLY, D. O. E. A. W. (2013) Recycled water management plan and validation guidelines: November 2008.

ROSER, D., KHAN, S., DAVIES, C., SIGNOR, R., PETTERSON, S. & ASHBOLT, N. (2006) Screening Health Risk Assessment for the Use of Microfiltration-Reverse Osmosis Treated Tertiary Effluent for Replacement of Environmental Flows. Centre for Water and Waste Technology, University of New South Wales.

SALAS, J. D., DELLEUR, J. W., YEVJEVICH, V. & LANE, W. L. (1980) *Applied Modeling of Hydrologic Time Series*, Bookcrafters Inc. Chelsea Michigan.

SMEETS, P. W. M. H., Y. J. DULLEMONT, P. H. A. J. M. VAN GELDER, DIJK, J. C. V. & MEDEMA, G. J. (2008) Improved methods for modelling drinking water treatment in quantitative microbial risk assessment; a case study of Campylobacter reduction by filtration and ozonation. *Journal of Water and Health,* 6**,** 301-314.

STANDARDS AUSTRALIA & STANDARDS NEW ZEALAND (2009) AS/NZS ISO 31000:2009 Australian/New Zealand Standard TM. Risk management—Principles and Guidelines.Onginated as AS/NZS 43601 995. Third edition 2004. Revised asid redesignated asASNZS ISO 3100G.2009. ISBN 0733792898.

STANDARDS AUSTRALIA & STANDARDS NEW ZEALAND (2013) Risk Management-Guidelines on Risk Assessment Techniques. SA/SNZ HB:892013.

SWARTZ, C., PETTERSSON, T. & LINDHE, A. (2010) Risk assessment and risk management in water supply systems: State-of-the-art and case studies in southern Africa. *WISA (Water Institute of South Africa) 2010 Biennial Conference & Exhibition, Durban, April.*

SWEDISH EPA (2009) Wastewater treatment in Sweden. ISBN 978-91-620-8416-5.

TWARDY, C. R., NICHOLSON, A. E., KORB, K. B. & MCNEIL, J. (2006) Epidemiological data mining of cardiovascular Bayesian networks. *electronic Journal of Health Informatics* http://www.ejhi.net *ISSN: 1446-4381,* 1**,** e3.

U.S. ENVIRONMENTAL PROTECTION AGENCY & U.S. DEPARTMENT OF AGRICULTURE (2012) MICROBIAL RISK ASSESSMENT GUIDELINE PATHOGENIC MICROORGANISMS WITH FOCUS ON FOOD AND WATER EPA/100/J-12/001; USDA/FSIS/2012-001. Interagency Microbiological Risk Assessment Guideline Workgroup.

WALKER, T., STANFORD, B. D., KHAN, S. J., ROBILLOT, C., SNYDER, S., VALERDI, R., DWIVEDI, S. & J., V. (2016) Critical Control Point Assessment to Quantify Robustness and Reliability of Multiple Treatment Barriers of a DPR Scheme. WateReuse Foundation Project Report WateReuse-13-03.

WEBER, P. & JOUFFE, L. (2003) Reliability modelling with dynamic bayesian networks. *In 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS'03), Washington, DC, USA.* IFAC.

WITTEN, I. H., FRANK, E. & HALL, M. A. (2011) *Data Mining Practical Machine Learning Tools and Techniques Third Edition* Elsevier Inc.

16. Appendix A: Carvajal et al (2015).

Water Research 85 (2015) 304–315

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Modelling pathogen $\log_{10}$ reduction values achieved by activated sludge treatment using naïve and semi naïve Bayes network models

Guido Carvajal [a], David J. Roser [a], Scott A. Sisson [b], Alexandra Keegan [c], Stuart J. Khan [a, *]

[a] UNSW Water Research Centre, School of Civil & Environmental Engineering, University of New South Wales, NSW, 2052, Australia
[b] School of Mathematics & Statistics, University of New South Wales, NSW, 2052, Australia
[c] Australian Water Quality Centre, SA Water Corporation, Adelaide, SA, 5000, Australia

ARTICLE INFO

Article history:
Received 9 June 2015
Received in revised form
3 August 2015
Accepted 19 August 2015
Available online 21 August 2015

Keywords:
Naïve Bayes
Bayesian belief networks
Activated sludge
Removal efficiency
Modelling
Pathogens

ABSTRACT

Risk management for wastewater treatment and reuse have led to growing interest in understanding and optimising pathogen reduction during biological treatment processes. However, modelling pathogen reduction is often limited by poor characterization of the relationships between variables and incomplete knowledge of removal mechanisms. The aim of this paper was to assess the applicability of Bayesian belief network models to represent associations between pathogen reduction, and operating conditions and monitoring parameters and predict AS performance. Naïve Bayes and semi-naïve Bayes networks were constructed from an activated sludge dataset including operating and monitoring parameters, and removal efficiencies for two pathogens (native *Giardia lamblia* and seeded *Cryptosporidium parvum*) and five native microbial indicators (F-RNA bacteriophage, *Clostridium perfringens*, *Escherichia coli*, coliforms and enterococci). First we defined the Bayesian network structures for the two pathogen $\log_{10}$ reduction values (LRVs) class nodes discretized into two states ($<$ and $\geq$1 LRV) using two different learning algorithms. Eight metrics, such as Prediction Accuracy (PA) and Area Under the receiver operating Curve (AUC), provided a comparison of model prediction performance, certainty and goodness of fit. This comparison was used to select the optimum models. The optimum Tree Augmented naïve models predicted removal efficiency with high AUC when all system parameters were used simultaneously (AUCs for *C. parvum* and *G. lamblia* LRVs of 0.95 and 0.87 respectively). However, metrics for individual system parameters showed only the *C. parvum* model was reliable. By contrast individual parameters for *G. lamblia* LRV prediction typically obtained low AUC scores (AUC $<$ 0.81). Useful predictors for *C. parvum* LRV included solids retention time, turbidity and total coliform LRV. The methodology developed appears applicable for predicting pathogen removal efficiency in water treatment systems generally.

1. Introduction

Activated Sludge (AS) is widely employed at municipal wastewater treatment plants to achieve secondary treated effluent quality sufficient for environmental discharge or further treatment. The primary objective of AS is large reductions in biochemical oxygen demand (BOD_5). Concurrent nitrogen removal has also been targeted by AS. Key process control parameters include solids retention time (SRT), mixed liquor suspended solids (MLSS), hydraulic retention time (HRT) and temperature. Performance verification is focused on water quality parameters including BOD_5, chemical oxygen demand (COD) ammonium (NH_4^+), nitrite (NO_2^-) and nitrate (NO_3^-), total Kjeldahl nitrogen (TKN), alkalinity, pH, turbidity and total Suspended Solids (S) (Metcalf and Eddy Inc. et al., 2014).

Pathogen reduction has not generally been a key aim of AS. However, with increased interest in water reuse, there has been growing interest in understanding and optimising the performance of AS for the improvement of microbial water quality (Wen et al., 2009).

Contemporary water reuse guidelines, such as the Australian Guidelines for Water Recycling (NHMRC et al., 2006) promote the attribution of pathogen $\log_{10}$ reduction values (LRVs) to diverse treatment barriers with a view to minimising exposure risks. Microorganisms for which secondary treatment LRVs have been

* Corresponding author.
E-mail addresses: guido.carvajal@unsw.edu.au (G. Carvajal), djroser@unsw.edu.au (D.J. Roser), scott.sisson@unsw.edu.au (S.A. Sisson), Alex.Keegan@sawater.com.au (A. Keegan), s.khan@unsw.edu.au (S.J. Khan).

http://dx.doi.org/10.1016/j.watres.2015.08.035
0043-1354/

proposed include *Escherichia coli* (1.0–3.0), bacteriophage (0.5–2.5), *Clostridium perfringens* (0.5–1.0), *Giardia lamblia* (0.5–1.5) and *Cryptosporidium parvum* (0.5–1.0). However, the guidelines emphasise these ranges are 'indicative' and assignment of 'LRV credits' requires further supporting evidence during system validation.

Several studies have now reported qualitative and semi-quantitative relationships between AS operational performance monitoring data and pathogen LRVs (Robertson et al., 2000; Stadterman et al., 1995; Suwa and Suzuki, 2001). However, there is still no consistent methodology for quantitatively relating process performance parameter data to LRVs, such as when assigning and validating LRV credits. A likely reason is the limited degree to which the relationships between AS operational parameters and LRV outcomes have been defined.

Archetypal disinfection processes, including for example photochemical and chemical inactivation, are well characterised and microbial reductions are consistently correlated with disinfectant doses. By contrast, AS removal mechanisms are incompletely understood and likely involve multiple competing processes whose relative contributions are unknown or variable (Flapper et al., 2010; Wen et al., 2009). Consequently, AS LRVs are also more difficult to model and predict with conventional statistical tools. The relationships between AS operational parameters and the prediction of conventional removal efficiencies for BOD_5 and N have been modelled through approaches such as the 'Activated Sludge Model' (ASM) of Henze et al. (2006). Other modelling has employed Artificial Neural Networks to capture the relationships between operational parameter variables and water quality outcomes (Cote et al., 1995; Flapper et al., 2012; Lee et al., 2005). However, these have tended to be "black box" models which do not clarify dependencies between variables or generate probabilistic predictions (Pittman, 2008). Separately, conventional parameters used to control and monitor AS processes (e.g. MLSS, SRT, HRT, SS, COD) have not been able to successfully explain observed LRVs.

As a solution we proposed that the causality based reasoning and techniques of Bayesian Belief Networks (BBNs) might more successfully explain and predict LRVs. BBNs are probabilistic graphical models represented by 'Directed Acyclic Graphs', which can model non-recursive causal relationships in complex systems and facilitate inferential reasoning. They emerged from artificial intelligence research and have been applied to medical diagnosis, resource management, reliability and risk assessment, and robotics (IEC/ISO, 2009; Liu et al., 2012; Smid et al., 2010). BBN model design can be both causal and non-causal, as with 'naïve' and 'semi-naïve' Bayes models (NB and SNB respectively) (see Supplementary information A.1 for an explanation of the difference). Other attractive BBN characteristics include a capability for incorporating expert knowledge, and automated learning of relationship structures and conditional probabilities from databases which may include missing values.

A BBN structure is defined by directional connections, known as 'arcs', which specify the dependence and independence assumptions between random variables, termed 'nodes'. These interdependencies determine what information is required to specify the probability distribution of the random variables of a network. Two variables are identified as 'parent' and 'child' nodes if there is an arc from the former to the latter (Korb and Nicholson, 2011). When a variable has parents, a set of conditional probabilities must be defined for the child node for each combination of parent node 'states' which may be categories or value ranges. Nodes without parents (root nodes) only require marginal probabilities. BBNs reduce the quantity of information required to define a joint probability distribution through factorisation conducted using the chain rule (Eq. (1)):

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \middle| X_{pa[i]}\right) \quad (1)$$

Where $P(X_1,X_2,\ldots,X_n)$ is the joint probability distribution of variables $(X_1, X_2,\ldots,X_n)$, X_i corresponding to a random variable represented by the node i in (1,…,n), $pa[i]$ denotes the parents of node i, and $X_{pa[i]}$ indicates a set of random variables associated with $pa[i]$.

BBNs have previously been used to predict process upsets and water quality (Chong and Walley, 1996; Li et al., 2013; Sahely and Bagley, 2001). These studies successfully predicted removal performance for conventional chemical and physical parameters, illustrating the applicability of BBNs to wastewater management. However, no application to predicting pathogen LRV has been reported to our knowledge.

This paper explores the use of 'naïve' and 'semi naïve' Bayes networks as tools for explaining, quantifying and predicting AS pathogen removal efficiency where a substantial operating and water quality parameter data set is available. Naïve Bayes models (NB) are non-causal BBN models commonly used for classification problems (Kjræulff and Madsen, 2008). They often provide good accuracy, while offering simplicity and efficiency. Their construction employs a range of objective rules and tests, which address modelling traps including the use of inappropriate variables, modeller bias and over-fitting. By definition, the structure of an NB model always employs a "class node" which is the only parent of each other node (attribute nodes), all of which are conditionally independent given the class node. For naïve Bayes models Eq. (1) becomes:

$$P(\mathrm{A}_1 \ldots, A_n, C) = \mathrm{P(C)} \prod_{i=1}^{n} P(A_i|C) \quad (2)$$

Where: A_i indicates the ith of n attribute nodes and C indicates the class variable.

In the case of the related SNBs, the independence assumption is relaxed by allowing some arcs between the attribute nodes using link selection rules not necessarily involving a choice by the investigator. Examples of SNBs include Tree Augmented Naïve (TAN) Bayes models in which the nodes depend on the class node and at most one other node (Korb and Nicholson, 2011), and Bayesian network Augmented Naïve Bayes (BAN) models, where two or more arcs are allowed between nodes additional to the class node (Cheng and Greiner, 1999).

In this study we first constructed naïve and semi-naïve models using *C. parvum* and *G. lamblia* LRVs as class nodes and evaluated their predictive capacity using various performance metrics and identified the best model describing LRV variance in response to reactor operating conditions. We then identified which relevant AS operating and monitoring parameters could be used as predictors for pathogen LRV. Finally, we analysed the practical use of NBs and SNBs and explored how to interpret and apply the model outputs.

2. Methods

The stepwise modelling approach developed for this study is summarized in Fig. 1. This approach is expected to be generally applicable to data collected for other water treatment processes. A central aim was to assess the predictability of the class nodes/variables of interest (pathogen LRVs in the present case) based on water treatment system control parameters, water quality monitoring and derived parameters. This procedure provides a basis for constructing models dispassionately and selecting the best from those available after implementing step 4. Common terminology, acronyms and abbreviations are presented in Table 1.

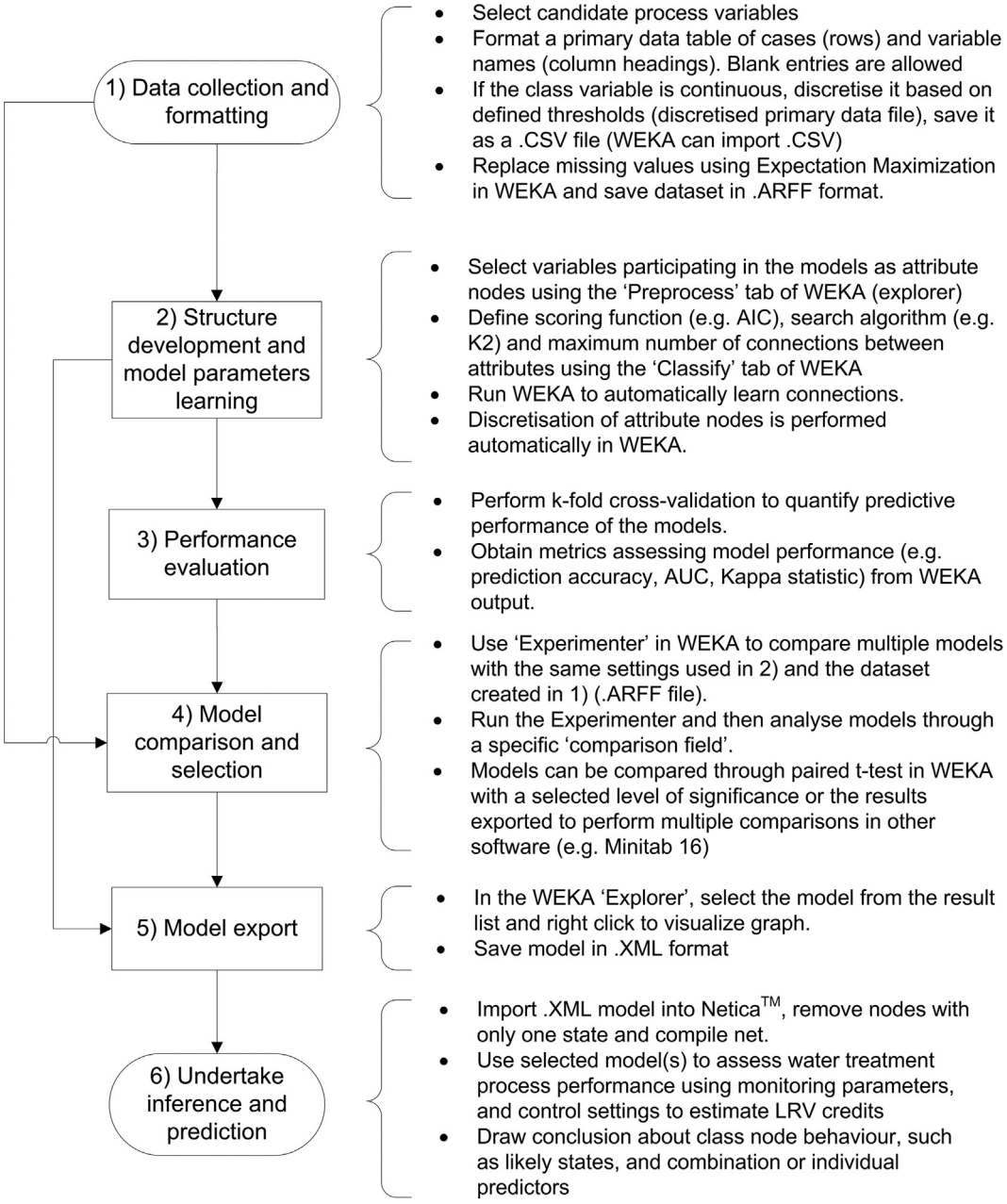


Fig. 1. Flowchart for model development, evaluation and selection.

2.1. AS system

The data we used were obtained from a study of a 150 L working volume Intermittently Decanted Extended Aeration (IDEA) AS pilot plant (Fig. 2) (Flapper et al., 2012). During that study many physicochemical, microbial and operational data were collected.

Key features of the pilot plant study were as follows: Primary effluent was collected from a full scale wastewater treatment plant in Victoria, Australia and used as influent for the pilot plant. The reactor was operated in a three stage cycle comprising: i) (top up) influent feed and aeration, ii) settling and iii) (partial) supernatant decanting. The reference operating conditions for the reactor were HRT = 24 h and SRT = 15 days (three experimental runs). Additional operating conditions investigated were: HRT = 24 h and SRT = 10 days (three experimental runs); HRT = 24 h and SRT = 20 days (one experimental run); HRT = 7.5 h and SRT = 15 days (two experimental runs) (Flapper et al., 2012). *C. parvum* oocysts were added to the influent tank (3.6 $\log_{10}$ oocysts/L) to ensure effluent concentration data were uncensored and sufficient for estimating LRVs. The mixed liquor dissolved oxygen (DO) concentration was maintained at 1.5 mg L^{-1} and the reactor was operated at 14.6–27.1 °C (Flapper et al., 2012). Key operating and water quality parameters, measured or controlled in this study, included: three reactor operating parameters (SRT, MLSS, HRT), seven microbial water quality parameters (F-RNA bacteriophage, *E. coli*, Total coliforms, enterococci, *C. perfringens*, *G. lamblia*, *C. parvum*) and eleven physicochemical parameters (COD, BOD_5, NH_4^+, NO_2^-, NO_3^-, TKN, Alkalinity, pH, Turbidity, SS, Temperature). Pathogen and indicator LRVs were computed from temporally matched concentrations in the reactor inlet and outlet. A total of 98 records were available for

Table 1
Key Bayesian belief network abbreviations and terminology relevant to model validation.

Abbreviation /acronym	Meaning	Explanation/use/comments	Reference
PA	Prediction accuracy	Quantifies the number of correctly predicted values divided by the total number of cases.	(Witten and Frank, 2005)
KS	Kappa statistic	Measures the agreement between model predictions and actual values as a metric in the range [−1,1]. KS = 1 means perfect agreement, KS = 0 means that agreement is equal to chance, and KS = −1 means "perfect" disagreement.	(Marcot, 2012)
AUC	Area under the curve for the receiver operating characteristic curve	AUC ranges between 0 and 1, where 1 represents perfect matching, 0.5 reflects totally random models, and <0.5 indicates models generating predominantly inaccurate predictions.	(Korb and Nicholson, 2011)
LL	Log-likelihood score	Measures how well the data fit each model. Used to compare models with the same variables and dataset but different node/arc structure. Higher scores reflect a better fit.	(Koller and Friedman, 2009)
TPR	True positive rate	Rate of correct positive predictions (high reductions).	
FPR	False positive rate	Failure to detect low reductions when they occurred.	
TNR	True negative rate	Rate of correct negative predictions (low reductions).	
FNR	False negative rate	Failure to detect high reductions when they occurred.	
BBN	Bayesian belief network	Probabilistic graphical models formed by nodes (variables) and arcs (connections) in a directed acyclic graph.	
NB	Naïve Bayesian network	Bayesian network with a class node as the only parent of the remaining nodes.	
SNB	Semi-naïve Bayesian network	Naïve Bayesian network in which attribute nodes are allowed to be connected one another.	
AIC	Akaike information criterion score,	Information-theoretic scoring function, which trades off the model's goodness of fit with its complexity.	(Kjræulff and Madsen, 2008)
ZeroR	–	Baseline model, it can be seen as a network without arcs.	
BAN	Bayesian network augmented naïve Bayes	Semi-naïve Bayes model. Two or more arcs between attributes are allowed.	
TAN	Tree augmented naïve Bayes	Semi-naïve Bayes model. At most one arc between attributes is allowed.	
WEKA	Waikato environment for knowledge analysis	Data mining software.	(Hall et al., 2009)
Netica™		Bayesian belief network modelling software.	(Norsys, 2015)
LRV	Log reduction value	Logarithmic transformation of the percentage reduction.	
IDEA	Intermittently decanted extended aeration	Semi-batch activated sludge reactor with cycles of influent feeding, aeration, decanting and effluent withdrawal.	
Attributes	–	Variables hypothesized as related to a class node in a NB and SNB models.	
Nodes	–	Variables formatted in Bayesian belief net format.	

the BBN analysis.

2.2. Modelling software

Four candidate models were constructed and evaluated using a variety of performance measures. The models were designed to quantify the influence of operating and water quality parameters on confirmed *C. parvum* and *G. lamblia* LRVs. Models were designed and evaluated using the Waikato Environment for Knowledge Analysis (WEKA) data mining software v. 3.6.11 (Hall et al., 2009). Final model usage was performed in Netica™ Bayesian modelling software (Norsys, 2015). WEKA includes machine learning algorithms for data mining and provides various tools for data processing and evaluation of algorithm optimality (Witten and Frank, 2005). Netica™ provides a popular and simple graphical interface for building and working with BBNs (Norsys, 2015).

The database was formatted to facilitate processing using WEKA and Netica™. The data were first compiled in a single spreadsheet table comprising records (rows) and variables (columns). For each model, initial WEKA processing then involved selection of a class node and its manual discretization into 2 states. Because WEKA ignores missing values for the class node, records lacking *C. parvum* and *G. lamblia* data were removed when learning NB and SNB structures.

The final LRV datasets consisted of 88 and 75 records for *C. parvum* and *G. lamblia* respectively. The remaining data records still included some missing values for other reactor operational and water quality parameters: MLSS and Temperature (5–16% of records) and SS, pH, NO_3^-, TKN and COD (2–3% of records). The Expectation Maximization (EM) imputation method in WEKA was used to replace the missing values. EM uses a multivariate normal model to impute missing values. The reliability of WEKA was confirmed by also running the EM multiple imputation methods in AMELIA II package in R (Honaker et al., 2011) which offers several options for data pre-processing.

The model designs generated by WEKA were exported as .XML files and imported into Netica™.

2.3. Model design

An NB model and three SNB (two TAN and one BAN) models were constructed for each of the two pathogens. WEKA's automated structure learning tool defined the arcs and nodes' states in the networks, using the dataset of Flapper et al. (2012). The two TAN model structures were developed by applying the Chow and Liu (1968) (TAN (1) model) and the K2 Hill Climbing (TAN (2) model) (Cooper and Herskovits, 1992) algorithms. K2 algorithm requires a fixed ordering of the variables in the dataset as input. The

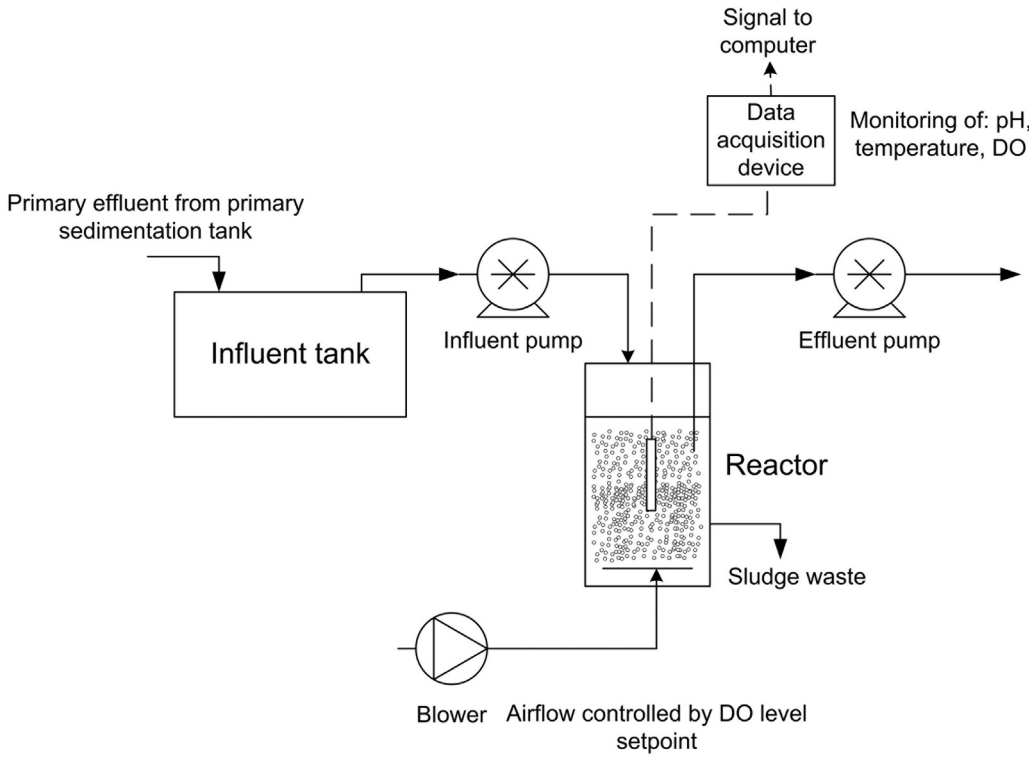


Fig. 2. Scheme of the activated sludge reactor.

variable ordering and the explanation for such selection can be found in Supplementary Table A. 1. The BAN model was constructed in the same manner as the TAN (2) model but up to two nodes were allowed as parents in addition to the class node. In this approach the learning processes are treated as optimisation problems where WEKA's search algorithms maximise a scoring function applicable to BBNs, in this case, the Akaike Information Criterion (AIC) (Kjræulff and Madsen, 2008). The AIC scoring function trades off the model's goodness of fit with its complexity. "Structured learning" as employed by WEKA is designed to find the BBN structure that best describes the statistical relationship between variables. Our rationale for using structured learning was to develop models in a systematic and objective fashion. These models were also compared with their baseline model (ZeroR) equivalents. ZeroR models predict the mode for a nominal class or the mean for a numeric class (Witten and Frank, 2005). When viewed as a Bayesian network a ZeroR model appears as a set of nodes without connections. So for example, in the case of *C. parvum*, the class node had two states (LRV < 1 and LRV ≥ 1) which were distributed 51.1% (LRV < 1) and 48.9% (LRV ≥ 1) and ZeroR would always predict LRV < 1.

2.4. Model parameters, discretization and learning

The most appropriate number of states for the remaining nodes (i.e. their discretization) of each model was also determined by WEKA. WEKA optimises thresholds of the attributes based on the class variable (i.e. *C. parvum* and *G. lamblia* LRV node) using the minimum description length principle (Fayyad and Irani, 1993). Where only one state was defined for a variable, the corresponding node was concluded not to contribute to the classification process and was discarded from the NB model. The same nodes and states were also used by WEKA to define the TAN and BAN models.

Although Netica™ has a TAN learning wizard based on the Chow and Liu algorithm, this was not used because Netica™ did not permit comparison to other models and cross-validation. The final nodes and states of the two NB models are presented in Fig. 3. The best SNB (TAN (2)) models are shown in Fig. 4. The remaining SNB (TAN (1), BAN) models are presented in Fig. A. 2 and Fig. A. 3 of the Supplementary information A.2. Models are available in their native format, from the corresponding author.

2.5. Model evaluation and validation

Due to the low ratio of data records to nodes (*ca* 5:1 and 10:1 for *C. parvum* and *G. lamblia* respectively), stratified 10-fold cross-validation was performed to confirm model stability when undertaking validation using WEKA. This approach randomizes and partitions the data into 10 equally sized sets and then 10 validations are made using 9/10th and 1/10th of the data for training and testing, respectively, every time with each portion of the data (Koller and Friedman, 2009). Cross-validation is performed to mitigate any bias produced by a particular sample chosen for training and testing. Randomised stratification means that the proportions within the classes of the class node are approximately the same in each fold.

During the cross-validation test, WEKA updated the probabilities of the network with each case, except for the unobserved class nodes (LRV nodes), and then generated state probabilities for those nodes which were then compared against their actual values. The output of this analysis was a comparison of predicted and the real data (metrics in Table 2 and Table 3). Of the two possible LRV states the higher value was taken as the 'positive' result for error calculation purposes. WEKA was also used to estimate prediction accuracy for LRV nodes using single (Table 3) and multiple node groups, such as the different coloured operational, control and monitoring groups in Fig. 3.

Eight different performance gauging measures recommended

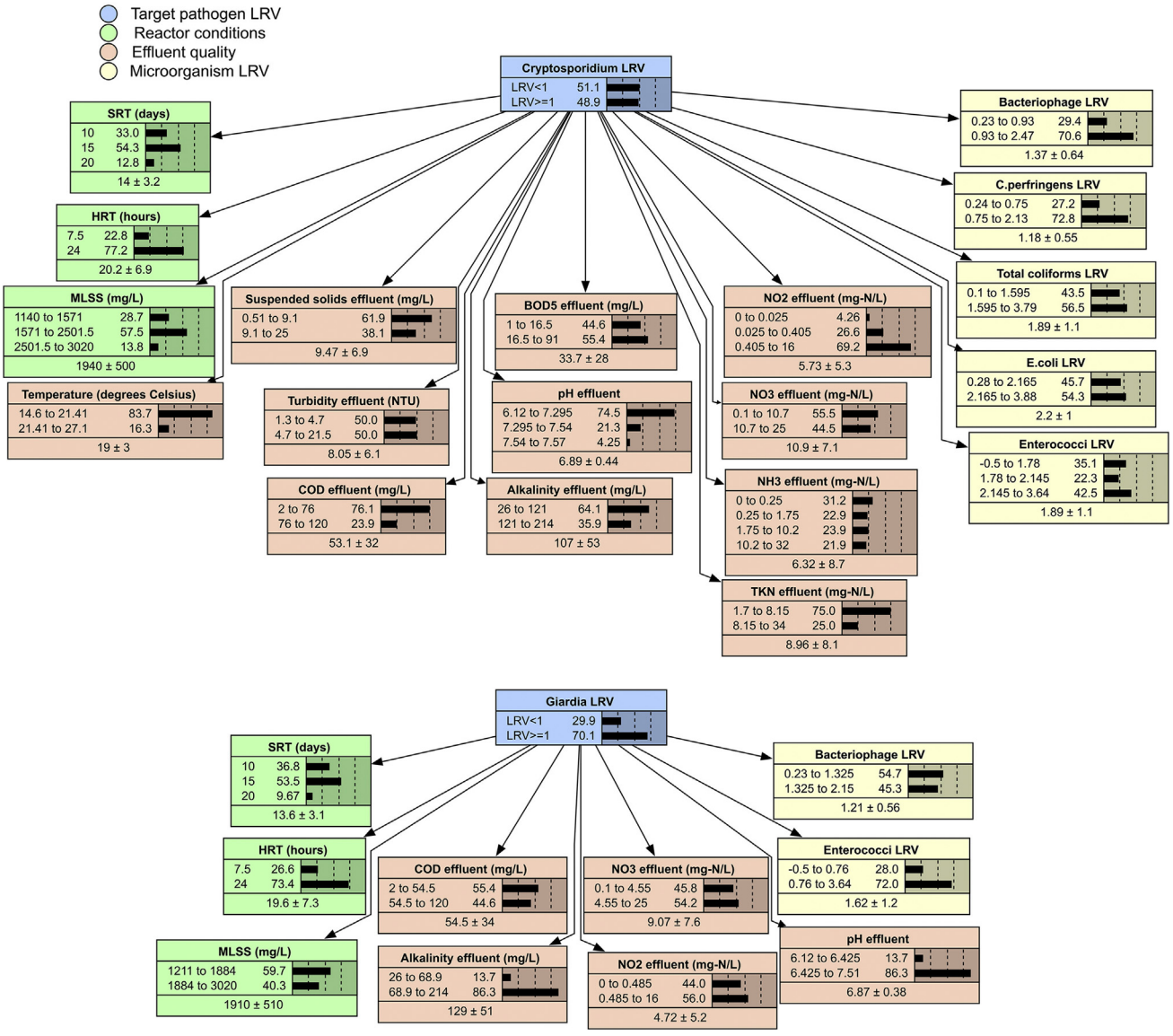


Fig. 3. Naïve Bayes models for (a) *C. parvum* LRV and (b) *G. lamblia* LRV showing discretization ranges.

for BBNs were used to compare the NB, SNB and ZeroR models: i) 3 model prediction performance metrics – prediction accuracy (PA), Kappa statistic (KS), area under the curve (AUC) for the receiver operating characteristic curve; ii) one goodness of fit metric – log-likelihood score (LL); and iii) four error matrix metrics – true positive rate (TPR), false positive rate (FPR), true negative rate (TNR) and false negative rate (FNR) (Korb and Nicholson, 2011; Marcot, 2012; Witten and Frank, 2005). Ten-fold cross validation was performed ten times to assess the variation in the metric estimates from different data randomizations. A one-way sensitivity to findings analysis was also performed. This analysis consisted of assessing the effect that each variable had on a target variable and is presented in Fig. A. 4 of the Supplementary information A.3.

A statistical analysis of the performance metrics' results was performed by nonparametric methods in Minitab 16 (Minitab, 2010). The Kruskal–Wallis test was conducted to determine whether there was a significant difference among the five models performance metrics medians. When the Kruskal–Wallis test indicated significant differences, Dunn's Test was used for the multiple comparisons ($n = 10$) among the individual groups with a family alpha probability of 0.1 equivalent to an individual pairwise comparison alpha probability of 0.01 (type I error).

3. Results and discussion

3.1. Identifying the best models

A comparison of eight performance metrics is presented in Table 2 (for acronym description see Table 1). These were calculated using the 10-fold cross validation procedure for the four candidate models and the baseline ZeroR model. All four refined models performed significantly better ($P < 0.01$) than the baseline ZeroR in almost all metrics with the exception of TNR and FPR for *C. parvum*, and TPR and FNR for *G. lamblia*. Because a threshold of 1 LRV splits the data in 51.1% (LRV < 1) and 48.9% (LRV ≥ 1) for *C. parvum*, ZeroR always predicted the removal to be LRV < 1. This meant that the testing cases where LRV < 1 were always predicted correctly (TNR = 1) and no false positives were obtained (FPR = 0) (predicting LRV ≥ 1 when testing cases LRV < 1) but the true positive rate was 0%. An equivalent analysis was undertaken for *G. lamblia*.

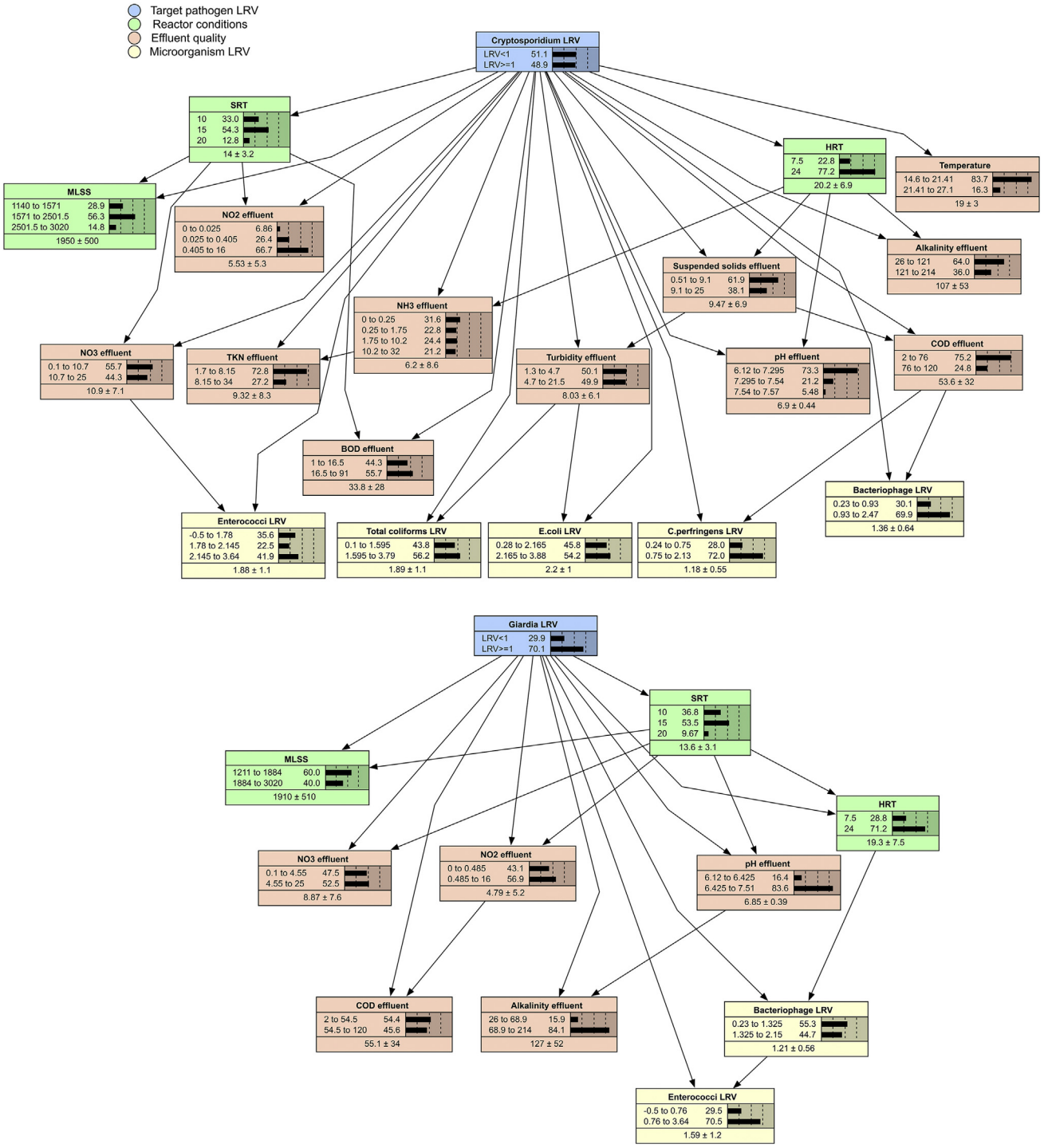


Fig. 4. Optimum semi-naïve Bayes net for *C. parvum* LRV (TAN (2)) and *G. lamblia* (TAN (2)).

Excluding the LL score, no significant differences ($P > 0.01$) in performance were observed between the metrics for the *C. parvum* NB and three SNB models.

The AUC metric for both pathogens indicated good classification and adequate prediction performances with a high ratio of true positive to false positive results. The fourth metric, log likelihood (LL), indicated the NB networks were significantly inferior to the semi-naive models ($P < 0.01$). SNB models for both pathogens showed no significant differences in the LL score ($P < 0.01$), indicating similar fit to the data.

Overall, these metrics indicated a significant improvement in the prediction results in the NB and SNB models over the ZeroR model for both pathogens and the SNBs over their NB equivalents based on the LL scores. However, the *G. lamblia* model metrics showed ≈2 times greater variability (standard deviation) compared to the *C. parvum* models.

The final 4 metrics measured True negative rate (TNR), False Positive Rate (FPR) (Type I error), True Positive Rate (TPR) and False Negative Rate (FNR) (Type II error). All *C. parvum* NB and SNB models and the *G. lamblia* BAN model especially predicted

Table 2
Arithmetic mean ± standard deviation of performance measures from the 10-fold cross validation for *C. parvum* and *G. lamblia* for the naïve and semi-naïve models.

Pathogen	Performance measure	NB	TAN (1)	TAN (2)	BAN	ZeroR
C. parvum	PA	93.2 ± 7.90[a]	88.6 ± 8.79	91.1 ± 8.26	91.5 ± 8.24	51.1 ± 4.87
	KS	0.86 ± 0.16	0.77 ± 0.18	0.82 ± 0.17	0.83 ± 0.16	0.00 ± 0.00
	AUC[b]	0.96 ± 0.07	0.95 ± 0.08	0.95 ± 0.09	0.95 ± 0.08	0.52 ± 0.15
	LL	−755 ± 42	−564 ± 43	−583 ± 44	−582 ± 43	−1131 ± 52
	TPR[b]	0.98 ± 0.07	0.90 ± 0.13	0.95 ± 0.10	0.96 ± 0.10	0.00 ± 0.00
	FNR[b]	0.02 ± 0.07	0.09 ± 0.13	0.05 ± 0.10	0.04 ± 0.10	1.00 ± 0.00
	TNR[b]	0.89 ± 0.14	0.87 ± 0.15	0.88 ± 0.15	0.88 ± 0.14	1.00 ± 0.00
	FPR[b]	0.11 ± 0.14	0.13 ± 0.15	0.12 ± 0.15	0.12 ± 0.14	0.00 ± 0.00
G. lamblia	PA	81.0 ± 12.6	83.2 ± 13.5	82.6 ± 12.7	84.4 ± 12.9	70.7 ± 4.41
	KS	0.54 ± 0.33	0.59 ± 0.33	0.57 ± 0.32	0.60 ± 0.33	0.00 ± 0.00
	AUC	0.87 ± 0.17	0.86 ± 0.18	0.87 ± 0.17	0.86 ± 0.17	0.35 ± 0.19
	LL	−413 ± 60	−345 ± 37	−350 ± 39	−347 ± 38	−491 ± 76
	TPR	0.85 ± 0.14	0.88 ± 0.16	0.88 ± 0.15	0.91 ± 0.14	1.00 ± 0.00
	FNR	0.15 ± 0.14	0.12 ± 0.16	0.12 ± 0.15	0.09 ± 0.14	0.00 ± 0.00
	TNR	0.72 ± 0.31	0.72 ± 0.32	0.70 ± 0.32	0.70 ± 0.32	0.00 ± 0.00
	FPR	0.28 ± 0.31	0.28 ± 0.32	0.30 ± 0.32	0.30 ± 0.32	1.00 ± 0.00

[a] Standard deviations were calculated from the results of 10-fold cross validation repeated 10 times.
[b] AUC and rates were computed considering LRV ≥ 1 as the target range.

reductions very well when they occurred (TPR metric). However, the FPR metric (crediting a plant with a LRV ≥ 1 when the opposite occurred) was nearly 3 times greater in the case of *G. lamblia* (FPR = 0.28–0.30).

Another criterion for comparing SNBs was whether the network structures were causally valid and logical. Though the metrics were comparable, WEKA created a *C. parvum* TAN (1) model which included illogical arcs which were absent from the TAN (2) and BAN models, for example ammonia controlling HRT and bacteriophage LRV controlling temperature (Fig. A. 2). Similarly there was an illogical arc from SRT to HRT in the three *G. lamblia* models (Fig. A. 3). Supporting the conclusion they provided comparable descriptions, the TAN (2) and BAN models had similar structures for both pathogens (Fig. A. 2 and Fig. A. 3), possibly due to them using the same search algorithm and score function (the AIC).

Increased uncertainty was observed in the attribute nodes' probability distributions when these were connected to each other in the SNBs. This uncertainty was reflected in the attribute nodes conditional probability tables as uniform distributions for combinations of parent node states not found in the data. This was expected as the dataset was limited in size but was not seen as significant problem since these networks were primarily designed to estimate classification LRVs and not estimate other parameters given these LRVs.

WEKA also allowed us to generate 'learning curves' for the two NB models by sequentially adding or removing 10% of the data during model construction and testing. During BBN learning, the average KS (prediction agreement) metric for *C. parvum* remained stable at 0.86 once more than 70 percent of the training data was incorporated and was 0.8 even when only 20 percent of the data had been incorporated. Similarly for *G. lamblia*, the KS statistic plateaued at 0.54 once 80 percent of the data had been incorporated. These stable plateaus indicated the data sets were sufficiently large for predicting the correct LRV range and obtaining models which were as accurate as possible given the data available.

The results of the model development and evaluation, using five different imputed datasets, indicated that the variation in the model performance was negligible (<1% difference) for both pathogens. This meant that the missing value imputation method did not significantly affect the measured performance of the models. The acceptable proportion of missing values in a dataset will depend on the specific context, including the degree of correlation between the variables. To quantify the influence of different proportions of missing values on model performance, the dataset was split into training (80%) and testing (20%) datasets and then a percentage of values was randomly removed from the training dataset. The effect on performance was assessed using AUC scores. The naïve Bayes model for *C. parvum* returned an AUC score of 0.95 or higher provided less than 30% of values were missing. We concluded that the number of missing values in the actual data sets was insufficient to substantially influence the final model performance.

Overall we concluded that for *C. parvum*, all NB and SNB models performed similarly for most of the metrics in Table 2. Consequently, LL and qualitative model assessment (logical structure) was used in this case to discriminate between the models. The *C. parvum* TAN (2) model (Fig. 4) achieved similar LL to the other SNBs (Table 2), but unlike TAN (1) its structure provided more insights into the system's behaviour (Fig. A. 2). On the other hand, it is possible that BAN model may have been overfitted by the number of permitted connections. Thus we selected TAN (2) as best for predicting *C. parvum* LRVs. For *G. lamblia*, the same results as for

Table 3
Individual attributes evaluation through AUC score (mean ± standard deviation) for *C. parvum* and *G. lamblia*.

Predictor	*C. parvum*	*G. lamblia*
Baseline[b]	0.52 ± 0.15[a]	0.35 ± 0.19[a]
SRT	**0.89 ± 0.09**	0.76 ± 0.17
HRT	0.71 ± 0.13	0.55 ± 0.15
MLSS	0.75 ± 0.13	0.67 ± 0.17
Temperature	0.61 ± 0.11	–[c]
SS	0.85 ± 0.10	–
Turbidity	**0.90 ± 0.10**	–
COD	0.72 ± 0.13	0.78 ± 0.13
BOD_5	0.81 ± 0.13	–
pH	0.58 ± 0.11	0.71 ± 0.16
Alkalinity	0.84 ± 0.11	0.70 ± 0.16
NO_2	0.70 ± 0.13	0.81 ± 0.14
NO_3	**0.89 ± 0.09**	0.66 ± 0.13
NH_4^+	0.81 ± 0.15	–
TKN	0.71 ± 0.13	–
Bacteriophage LRV	0.71 ± 0.13	0.66 ± 0.14
C. perfringens LRV	0.70 ± 0.13	–
Total coliforms LRV	**0.91 ± 0.09**	–
E. coli LRV	**0.91 ± 0.10**	–
Enterococci LRV	**0.90 ± 0.10**	0.66 ± 0.10

Bolded values indicate good predictor nodes.
[a] Mean AUC based on 10-fold cross validation repeated 10 times.
[b] "No evidence" no attributes are considered in the evaluation.
[c] Attribute was not included in the NB or SNBs.

C. parvum were obtained when comparing the models. The TAN (2) (Fig. 4) model was also selected as the best model for *G. lamblia*.

3.2. Operational control and monitoring parameters as predictors of protozoan LRVs

As well as informing overall model performance, the AUC score can be used to assess the predictive capacity of individual nodes alone (Table 3).

In the case of *C. parvum* there were several instances where AUC scores were comparable to those obtained using all nodes (Table 2). Turbidity, enterococci, *E. coli* LRV and total coliforms LRV returned high AUC scores ($\geq$0.9). SRT, SS and nitrate also generated a high (0.85–0.9) scores. SRT association with LRV indicated manipulating this variable might be used to maximise LRVs.

For *G. lamblia*, none of the attributes obtained an AUC > 0.9. The two highest scores were achieved with nitrite (AUC = 0.81), COD (AUC = 0.78), and SRT (AUC = 0.76). These results were consistent with the poorer overall *G. lamblia* model performance.

The predictive potential of groups of reactor and physicochemical parameters, and microbial indicator LRVs was assessed by selecting only the variables in such groups (coloured groups in Fig. 3) during the model construction phase. For *C. parvum*, the three reactor settings, SRT, HRT and MLSS together achieved an AUC of 0.93 ± 0.07. Similar predictive power was obtained using a combination of all five microbial indicators (AUC = 0.95 ± 0.07) and the eleven physicochemical water quality parameters together (AUC = 0.94 ± 0.10).

Disappointingly, for *G. lamblia* the three reactor parameters together were much less accurate in their prediction potential (AUC = 0.67 ± 0.17) while the two microbial indicators combined yielded only a slightly better AUC of 0.71 ± 0.13. All five physicochemical parameters (Fig. 3) provided a prediction comparable to the complete model (AUC = 0.84 ± 0.18). This performance evaluation showed that it is possible to predict the removal of *G. lamblia* for a threshold of 1 LRV under the given set of operating conditions. However, the performance metrics were not as good as the ones obtained for *C. parvum*. Moreover, it was not possible to obtain single operating or water quality parameters with an average AUC higher than 0.80, except from nitrite (AUC = 0.81). We concluded that operating parameters included in this study were not good indicators of the *G. lamblia* removal mechanisms in the activated sludge system. This weak relationship could also be observed in the scatterplots which do not evidence trends or clustering as in the case of *C. parvum* and indigenous microbial indicators (Fig. 5). A possible reason is high *G. lamblia* input concentrations variability compared to *C. parvum*. This is discussed further below.

The absence of useful microbial indicator parameters for predicting of *G. lamblia* LRVs reflected the weak association between *G. lamblia* and indicator LRVs generally (Fig. 5). The correlation between the variables can be more clearly seen through the locally weighted Kernel smoothers included on each scatterplot (Fig. 5). The randomness of the scatter in all *G. lamblia* LRV plots is consistent with low model prediction power, for example, between bacteriophage and enterococci LRVs. Conversely, the clear correlation between *C. parvum* and *E. coli* LRVs (as well as enterococci and total coliforms – not shown) is also evident. This is consistent with the high predictive power of the full *C. parvum* models (Table 2).

3.3. Estimating log_{10} reduction credits of protozoan pathogens for activated sludge

Both the optimized TAN (2) models and our modelling approach have potential applications and implications for setting AS operational and monitoring parameters and predicting protozoan pathogen reductions for the selected ranges. The *C. parvum* TAN (2) model quantified i) how operational, physicochemical microbial indicators related to removal and process settings for maintaining a removal range given by the model, and ii) what log_{10} credit might be assigned where AS is optimised for BOD_5 and nitrogen removal instead.

The model can be used to determine the conditions for which the LRV $\geq$ 1. These conditions are not necessarily the maximum pathogen reductions the system could achieve, but the performance that the model is able to reliably predict based on the available data.

By contrast the *G. lamblia* model and metrics indicated that its reduction is less well understood, and operating and monitoring parameters cannot as yet be tuned to optimize *G. lamblia* removal. That being said the average LRV was $\geq$1 indicating log credit assignment is still possible for AS even though indirect monitoring and removal optimization is not yet possible.

The *C. parvum* AUC score also suggested that a high degree of LRV prediction was possible using only one monitoring parameter including total coliform LRV, *E. coli* LRV, enterococci LRV, and turbidity. Most usefully this list includes a real time predictor, turbidity.

The achievement of ≈ 1 log_{10} removal for *G. lamblia* or *C. parvum* was not as striking as the >3 LRV reductions achieved with purpose designed disinfection agents. However, the result was robust. Not only did the model show the reduction was real but the prediction metrics confirmed the model reflected real trends in monitoring variables and were not the result of overfitting. This indicates that in the future, AS systems may be further optimised for improved pathogen removal, and the techniques described in this paper will be suitable for demonstrating any increased pathogen reduction. Separately our paper demonstrates how robust treatment targets can be robustly estimated for other novel or unconventional disinfection and contaminant treatment processes.

The optimum discretization thresholds and sizes of datasets to obtain high levels of model performance (e.g. prediction accuracy >0.9) will depend on the specific model, process, characteristics of the process, the number of nodes and dataset. Such data set characterization is undertaken as part of the initial data mining by statistically experimenting with the data set and subsets on a case by case basis. The option for estimating minimum data set size we adopted was the use of 'learning curves' (Frank et al., 2000) where prediction performance metrics are determined for increasing sample sizes until a stable plateau is reached. We found that >61 data records were required to obtain stable performance for both the simplest *C. parvum* and *G. lamblia* (NB) models. Other machine learning and data mining techniques (e.g. decision trees) can also inform on whether datasets are sufficiently large for models to be robust (Witten and Frank, 2005).

Node state discretization threshold, though was necessarily defined by us, the modellers. We needed to consider the influence of between state boundaries on prediction accuracy, while ensuring there were sufficient records corresponding to each state to permit us to estimate the performance credibly. Because of this, state thresholds currently need to be defined empirically and selecting a threshold where almost all observations are allocated to one state must be avoided if possible. In our study we also assessed the impact of the protozoan thresholds being 1.0, 1.5 and 2.0 LRV. The AUC scores indicated that all models maintained good performance. However, PA was not adequate for LRVs of 1.5 and 2.0, as these returned negligible improvements over their equivalent ZeroR models. Put another way, we could construct models able to predict higher LRVs using the higher thresholds, however, the states to be predicted became too 'unbalanced', and the models tended to over-predict the majority class.

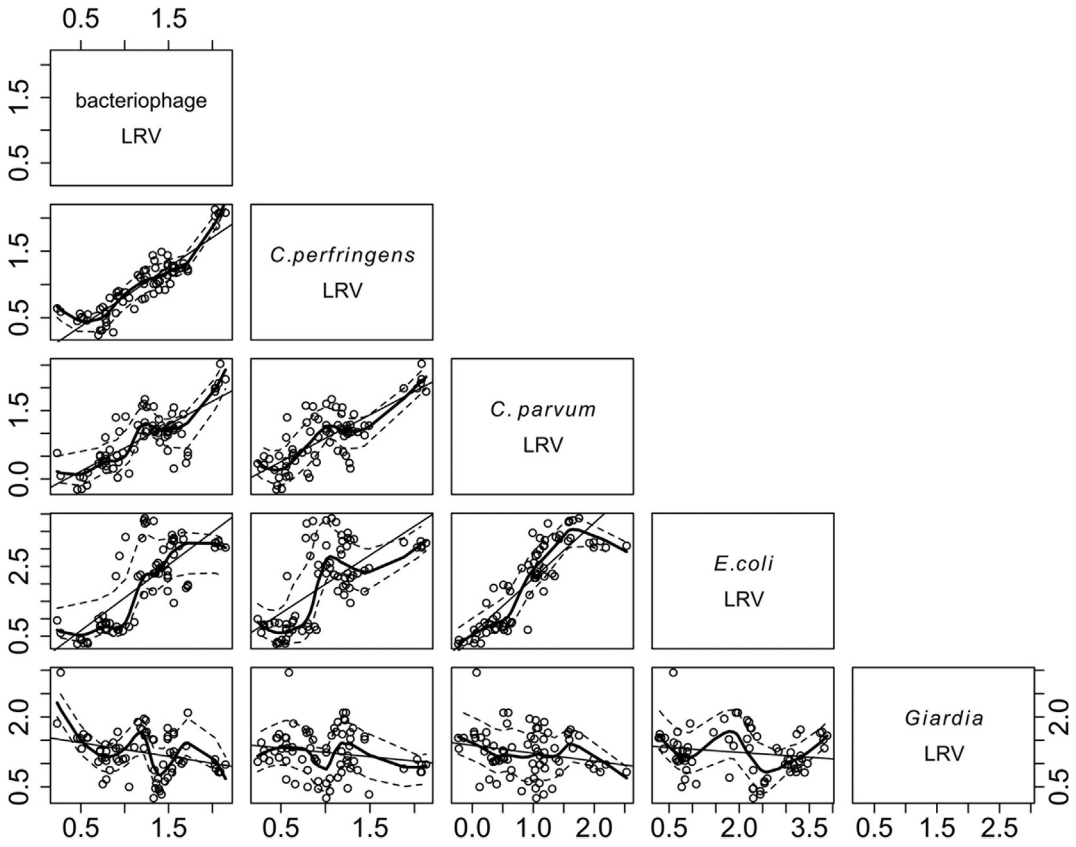


Fig. 5. Scatterplot matrix for bacteriophage LRV, *C. perfringens* LRV, *E. coli* LRV, *C. parvum* LRV and *G. lamblia* LRV. Graph includes a linear regression fit, smoothers and their 95% confidence intervals.

Using the 'Enter findings' Netica™ function we identified the most likely conditions (posterior probabilities) when LRVs were ≥1. This analysis was visualised in a tornado chart (Supplementary material Fig. A. 4a for *C. parvum* and Fig. A. 4b for *G. lamblia*). Consistent with our expectations, for *C. parvum* LRVs ≥1 were obtained when turbidity, SS, COD, TKN and alkalinity were in their lower ranges ($P > 0.9$) and microbial indicator LRVs (*E. coli*, enterococci, FRNA bacteriophage, *C. perfringens* and total coliforms) were in their higher ranges ($P > 0.9$). The optimal reactor parameter settings (SRT, MLSS and HRT), when *C. parvum* LRV was ≥1 and BOD_5 and ammonia were each in their minimum ranges, could be identified. These were the lowest range for MLSS (1140–1571 mg/L), lowest SRT (10 days), and highest HRT (24 h).

Recognizing the limitations of the *G. lamblia* model, we also noted that the conditions most associated with *G. lamblia* LRV ≥ 1 were high range pH ($P > 0.95$) and alkalinity ($P > 0.96$) and low range MLSS ($P = 0.73$). But counterintuitively these conditions were also associated with lower bacteriophage and enterococci LRVs and higher range COD. In light of these puzzling results and the poor model performance we concluded *G. lamblia* reduction needs further investigation.

Few studies have investigated the association between protozoa LRVs by AS and operational and microbial indicator variables. However, comparison with a recent literature review indicated that the BBN model pathogen and indicator LRVs were consistent with other research and protozoan removal is inversely correlated with SRT, effluent organic carbon and effluent SS (Flapper et al., 2010). A constraint on our *C. parvum* model's value is that LRVs are more difficult to measure than concentrations as both influent and effluent data are required. Accordingly the use of effluent indicator concentrations in place of LRVs was also evaluated and similar trends were observed.

The reproducibility of the *C. parvum* data was likely enhanced by seeding of standardized *C. parvum* oocysts into the influent of the test reactor at 3.6 $\log_{10}$ oocysts/L. This contrasts with the highly variable oocyst numbers and biotypes that would normally be encountered in wastewater influent but was essential within the earlier study to ensure valid LRVs would be established. *C. parvum* numbers in real wastewater influent are much more variable ranging from 10^1–10^4 oocysts/100 L (Harwood et al., 2005). Conceivably this increase and stability of oocyst numbers could have improved the model by reducing this source of variance in the LRV estimates and accounted for the contrasting performance of the *C. parvum* and *G. lamblia* models. This said *G. lamblia* cysts are generally present at higher numbers than *C. parvum* (10^4–$10^{5.5}$ cyst per 100 L in influent) (Harwood et al., 2005) and vary less between seasons.

3.4. Semi-naïve v. causal Bayesian modelling of water treatment processes

The models presented in this paper were not primarily designed to reflect cause-effect relationships. That is they are not 'causal' BBN models. However, they were shown to provide stable credible prediction of *C. parvum* LRVs. Among the reasons we trialled the semi-naïve Bayes approach, rather than a causal network approach, was the limited size of the available dataset compared to the number of possible variables, ease of model construction, avoidance of human bias and limited prior knowledge of the likely associations among the variables.

Causal models have advantages and different potential uses. For example, they would allow the whole system to be represented by a

single network and conceptually provide insights to the system behaviour. In causal models, derived numerical probabilities can be considered as representations of the probabilities of occurrence of a particular event. However, a disadvantage of causal models is that where a system is not well understood mechanistically and there are many dependent and independent nodes variables, the number of plausible models multiplies rapidly making parsimony a concern.

Semi-naïve Bayes models, on the other hand, allow the strong assumption of node independency given the target variable to be relaxed. These models are an intermediate step between the naïve Bayes model and a causal model and empirical experience in other fields has shown they can be very reliable (Korb and Nicholson, 2011). This approach also allowed dispassionate model construction using various performance metrics in a stepwise fashion based on rules developed for BBNs generally. WEKA allowed us to assess whether there were sufficient data records to generate stable model structures and what were credible discretization thresholds. The performance metrics informed us which nodes were most likely to influence LRVs. The metrics also allowed comparison of i) different model options, ii) their respective predictive power, and iii) assessment of whether the best models were credible or provided no improvement over the ZeroR model.

Real world activated sludge plants will differ in many respects for the pilot AS including SRT ranges, HRT ranges, temperature ranges, and MLSS concentration ranges and different data sets. So although the LRVs estimated here are valuable *per se* it may be preferable to repeat our model development process in such systems rather than use the models themselves uncritically. Nevertheless the method used here can be adapted to this task of constructing new candidate models and determining which if any validly describe the system being characterised.

4. Conclusions

A conceptual alternative to directly measuring pathogen removal efficiency is to predict LRVs using cost effective microbial and physicochemical monitoring, control parameters system operating conditions. However, conventional parametric statistical analyses have not yielded sufficiently convenient tools which describe AS processes and relate variables. In this investigation, we developed and assessed the potential of naïve and semi-naive Bayes models to predict and manage pathogen reductions. We made use of a real world data set to evaluate and quantify significant relationships between operating and monitoring parameters and estimate removal of two pathogens. The methodology we developed is objective, systematic and applicable to analysing water treatment processes more generally. Our study also identified operational parameters potentially useful for the prediction of *C. parvum* removal efficiency. Conversely the lack of success in modelling *G. lamblia* suggested that its removal by AS is not sufficiently understood and cannot yet be quantified based on removal of microbial indicators, even though assignment of average reduction credits of $\geq$1 $\log_{10}$ is still reasonable judging by the raw LRV probability density function.

Key outcomes from this study included:

- Useful predictors for *C. parvum* reduction included turbidity, SS, total coliform bacteria LRV and enterococci LRV.
- SRT, COD and nitrite were potential predictors of *G. lamblia* LRV. However, their AUC score was less than or equal than 0.81 indicating more work is needed before they can be reliably applied to this task.
- No microorganisms alone were reliably correlated with, or good predictors, of *G. lamblia*. This result highlighted the need to better understand the relationship between the removal of *G. lamblia* and other AS microbes.
- Naïve and semi-naive Bayes modelling of a real AS plant could reduce the costs of direct pathogen monitoring and encourage the gathering of informative process data which would permit LRV credits to be linked to the system's operating conditions.
- NB and SNB models can be used to understand whether optimal LRVs can be achieved concurrently with satisfactory BOD_5 and nitrogen removal.

Although causal BBNs were not constructed, our non-causal models provide a reference and starting point for such modelling by identifying those variables most likely to be useful when constructing causal models with the minimum of nodes. The SNB models provide an objective way of estimating the maximum accuracy that is possible with a causal Bayes model. The models are relatively easy to understand which should assist uptake by non-experts in Bayesian networks. Finally, the method here can reduce disagreements between model developers about what form BBNs should take.

Acknowledgements

Mr Guido Carvajal gratefully acknowledges scholarship support of Becas Chile, CONICYT. This research was supported by the Australian Water Recycling Centre of Excellence (NatVal2.2. WaterRA Project Number 3018-12) through the Australian Government's National Urban Water and Desalination Plan.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2015.08.035.

References

Cheng, J., Greiner, R., 1999. Comparing Bayesian network classifiers. In: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, pp. 101–108.
Chong, H., Walley, W.J., 1996. Rule-based versus probabilistic approaches to the diagnosis of faults in wastewater treatment processes. Artif. Intell. Eng. 10 (3), 265–273.
Chow, C.K., Liu, C.N., 1968. Approximating discrete probability distributions with dependence trees. IEEE Trans. Inf. Theory 14 (3), 462–467.
Cooper, G.F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Mach. Learn. 9 (4), 309–347.
Cote, M., Grandjean, B., Lessard, P., Thibault, J., 1995. Dynamic modelling of the activated sludge process: improving prediction using neural networks. Water Res. 29 (4), 995–1004.
Fayyad, U.M., Irani, K.B., 1993. Multi-interval discretization of continuous-valued attributes for classification learning. In: Proc. IJCAI93, Chambery, France, pp. 1022–1027.
Flapper, T., Campbell, B., Deere, D., Blackbeard, J., Halliwell, D., 2010. Quantifying Pathogen Log Reduction in Australian Activated Sludge Plants. Water (Australian Water Association), pp. 56–62. February.
Flapper, T., Campbell, B., O'Connor, N., Keegan, A., 2012. Quantification of Pathogen Removal in Australian Activated Sludge Plants (Phase 1 and 2), Viewed 10 January 2015. Available from: http://clearwater.asn.au//user-data/research-projects/swf-files/72m–-7104-quantification-of-pathogen-removal-in-activated-sludge-plant_final_report_.pdf.
Frank, E., Trigg, L., Holmes, G., Witten, I.H., 2000. Technical note: naive Bayes for regression. Mach. Learn. 41 (1), 5–25.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explor. Newsl. 11 (1), 10–18.
Harwood, V.J., Levine, A.D., Scott, T.M., Chivukula, V., Lukasik, J., Farrah, S.R., Rose, J.B., 2005. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Appl. Environ. Microbiol. 71 (6), 3163–3170.
Henze, M., Gujer, W., Mino, T., Van Loosedrecht, M., 2006. Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. IWA Publishing, Padstow, Cornwall, UK.
Honaker, J., King, G., Blackwell, M., 2011. Amelia II: a program for missing data. J. Stat. Softw. 45 (7), 1–47.
IEC/ISO, 2009. Risk Management–Risk Assessment Techniques. IEC/ISO 31010:

2009–11 International Electrotechnical Commission and International Organization for Standardization.
Kjræulff, U.B., Madsen, A.L., 2008. Bayesian Networks and Influence Diagrams: a Guide to Construction and Analysis. Springer Science+Business Media, LLC, New York.
Koller, D., Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques. MIT Press, Massachusetts.
Korb, K.B., Nicholson, A.E., 2011. Bayesian Artificial Intelligence, second ed. Chapman & Hall/CRC Press, London.
Lee, H., Lee, K., Park, C., Park, Y., 2005. Application of remote monitoring and automatic control system using neural network for small wastewater treatment plants in Korea. Water Sci. Technol. 51 (10), 249–257.
Li, D., Yang, H.Z., Liang, X.F., 2013. Prediction analysis of a wastewater treatment system using a Bayesian network. Environ. Model. Softw. 40, 140–150.
Liu, K.F.-R., Lu, C.-F., Chen, C.-W., Shen, Y.-S., 2012. Applying Bayesian belief networks to health risk assessment. Stoch. Environ. Res. Risk Assess. 26 (3), 451–465.
Marcot, B.G., 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecol. Model. 230, 50–62.
Metcalf & Eddy Inc., Tchobanoglous, G., Stensel, D., Tsuchihashi, R., Burton, F., 2014. Wastewater Engineering: Treatment and Resource Recovery, fifth ed. McGraw-Hill, New York.
Minitab, 2010. Minitab 16 Statistical Software. Minitab Inc., State College, Pennsylvania, USA.
NHMRC, EPHC, AHMC, 2006. Australian guidelines for water recycling: managing health and environmental risks (Phase 1). In: Natural Resource Management Ministerial Council, Environment Protection and Heritage Council, Australian Health Ministers' Conference, Australia.
Norsys, 2015. Norsys Software Corp. – Bayes Net Software. Available at: https://www.norsys.com/.
Pittman, K., 2008. Comparison of Data Mining Techniques Used to Predict Student Retention (PhD. thesis). Nova Southeastern University, Broward County.
Robertson, L., Paton, C., Campbell, A., Smith, P., Jackson, M., Gilmour, R., Black, S., Stevenson, D., Smith, H., 2000. Giardia cysts and Cryptosporidium oocysts at sewage treatment works in Scotland, UK. Water Res. 34 (8), 2310–2322.
Sahely, B.S., Bagley, D.M., 2001. Diagnosing upsets in anaerobic wastewater treatment using Bayesian belief networks. J. Environ. Eng. 127 (4), 302–310.
Smid, J., Verloo, D., Barker, G., Havelaar, A., 2010. Strengths and weaknesses of Monte Carlo simulation models and Bayesian belief networks in microbial risk assessment. Int. J. Food Microbiol. 139, S57–S63.
Stadterman, K., Sninsky, A., Sykom, J., Jakubowski, W., 1995. Removal and inactivation of Cryptosporidium oocysts by activated sludge treatment and anaerobic digestion. Water Sci. Technol. 31 (5), 97–104.
Suwa, M., Suzuki, Y., 2001. Occurrence of Cryptosporidium in Japan and countermeasures in wastewater treatment plants. Water Sci. Technol. 43 (12), 183–186.
Wen, Q., Tutuka, C., Keegan, A., Jin, B., 2009. Fate of pathogenic microorganisms and indicators in secondary activated sludge wastewater treatment plants. J. Environ. Manag. 90 (3), 1442–1447.
Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, third ed. Morgan Kaufmann, Massachusetts.