

Bayesian Network and System Thinking modelling to manage water quality related health risks from extreme events

Bertone, E. ^a, Sahin, O. ^a, Richards, R. ^b and Roiko, A. ^c

^a *Griffith School of Engineering, Griffith University, Queensland, Australia*

^b *School of Agriculture and Food Sciences, University of Queensland, Brisbane, Australia*

^c *Griffith School of Medicine, Griffith University, Queensland, Australia*

Email: e.bertone@griffith.edu.au

Abstract: The occurrence of extreme events can challenge the capacity of water utilities to deliver potable water of sufficient quality with respect to minimising health risks to consumers. As a consequence, proactive risk-assessment and decision support tools are necessary to assist in managing and mitigating such critical events effectively. However, the utility of these tools can be limited due to the lack of comprehensive data and a high degree of epistemic and stochastic uncertainty.

We use a combination of Bayesian Network (BN), System Dynamics (SD) and participatory modelling to develop a risk assessment tool for managing water-related health risks associated with extreme events. The combination of BN and SD modelling offers a number of advantages over other environmental modelling techniques; the capacity for dealing with a high degree of uncertainty, the use of feedback loops (SD only) and the ability to elicit and integrate quantitative and qualitative data (including expert opinion).

The risk assessment tool developed is applied to the raw water delivery system supplying Prospect water filtration plant system (Sydney, Australia), which is the main source of potable water for the Sydney metropolitan region. Key-stakeholders were engaged in developing and populating the conceptual models that form the basis of developing the BN and SD models. Conceptual models were developed by the stakeholders around the key indicator parameters of turbidity, water colour and *Cryptosporidium sp.* levels. These three conceptual models were combined into a single risk model and used for developing separate BN and SD models. Additional stakeholder workshops were conducted to refine the models (structure and parameter values) and to provide validation of the model outputs.

Here we present the development of a BN model designed to understand the risk of extreme events on the ability to provide potable water of a specified quality. The model has undergone development and preliminary parameterization via two participatory workshops. However, its development is an ongoing process with the next stage involving supplementing the 'expert opinion' used to parameterize the model so far with 'hard' data.

The completed models will quantify the sensitivity of the Prospect raw water delivery system to different types and combinations of extreme events (both natural and anthropogenic). The BN model will provide a risk management tool for estimating the probability of (top-down modelling), and requirements for (bottom-up modelling), meeting water quality guidelines. The SD model will provide a means of testing the implementation of different management scenarios and the impact that this has on water quality for different time horizons.

Overall, these complementary modelling methodologies will assist water treatment operators, water managers and other stakeholders in developing evidence-based mitigation strategies leading an to enhanced resilience of the system.

Keywords: *Bayesian networks, system dynamics, extreme events, water quality*

1. INTRODUCTION

Recent history in Australia has been characterised by a range of extreme weather events (e.g. droughts, Brisbane floods, cyclone Yasi, Victorian bushfires). These events have impacted on the ability of water utilities to provide potable water of a required standard to consumers or other bulk water clients. At issue, are the short- and long-term impacts of extreme events on the water quality at both pre- and post-treatment (including distribution and end-point) points in the system. Extreme events are projected to change in magnitude and frequency over the next century (IPCC, 2014), further exacerbating the pressures on water quality management. However, there are large uncertainties associated with the timing and nature of specific future events and this uncertainty is a major contributor to the challenge of water management. The key health-related water quality parameters that mostly concern the water utility involved in this research project (i.e. WaterNSW) in case of extreme events are: water colour, turbidity, and *Cryptosporidium sp.*

Water colour is a key parameter in drinking water reservoirs as it can affect the physical and biological properties of the whole lake, as well as creating discolouration of the raw water redirected to the water treatment plant (WTP). If discoloured water is supplied to the consumers, this may trigger complaints directed towards the water utility, thus the water utilities must meet the guidelines levels set by the Australian Drinking Water Guidelines. Furthermore, if coloured water leaves the WTP, the dissolved organic matter present in this water can react with chlorine when it enters the potable water system leading to the formation of carcinogenic trihalomethanes (THM's), one of the over 600 disinfection by-products currently reported in drinking water (Hrudey, 2009). However, the biggest issue with water colour is increased coagulation demand, thus increasing in turn the risk of filter failure at the WTP and breakthrough, leading to a direct health risk.

Water turbidity refers to how clear the water is. The greater the amount of total suspended solids in the water, the murkier it appears and the higher the measured turbidity. These impurities may include clay, silt, inorganic/organic matter, soluble coloured organic compounds, plankton and other microscopic organisms (EPA, 1999). Turbidity is a result of suspended particles and these can provide food and shelter for pathogens, which if not effectively removed, can promote regrowth of pathogens in the distribution system, leading to waterborne disease outbreaks (EPA, 1999). Higher turbidity levels are therefore pathogen risk factors (Khan *et al.*, 2013), typically associated with higher levels of disease-causing microorganisms such as parasites, viruses and some bacteria, which can cause e.g. cramps, diarrhoea, headache and nausea (Sarai, 2006).

Cryptosporidium sp. oocysts are often excreted in large amounts with the faeces of infected humans and animals (Graczyk and Fried, 2007), which can enter surface waters directly or through effluents and runoff from fields that are polluted by sewage sludge or manure (Graczyk *et al.*, 2008; Mons *et al.*, 2009) resulting in pollution of receiving waters. Importantly, these (oo)cysts have the capacity to remain infective for months in environmental waters and are highly resistant to chlorinated disinfectants (Betancourt and Rose, 2004). Therefore, waterborne contamination is a growing concern for water suppliers, causing widespread outbreaks of these diseases (Putignani and Menichella, 2010). For example, a contamination of cryptosporidium, along with *Giardia*, occurred in the water supply system of Greater Metropolitan Sydney during the 1998 Sydney water crisis (McClennan, 1998).

For the Sydney area, which is the location of this study, it is predicted that due to climate change, the number of days of extreme rainfall (> 40mm/day), as well as the number of very hot days (>37°C) and continued dry spells (>15days) will increase considerably (SCA, 2010). Therefore, the risk of bushfires, extreme rainfall events or prolonged droughts is expected to increase, which will have detrimental effects on the water quality in reservoirs. Water quality management in this context requires a multi- and inter-disciplinary approach that is both holistic and probabilistic, to develop appropriate management strategies. Strong support and active participation from practitioners within the water industry, whose experiences with past occurrences of extreme events and detailed understanding of many facets of the system is invaluable. Both qualitative and quantitative information about the system is also required. Traditional modelling approaches often deal poorly with such requirements.

This paper describes the methodology used to develop an extreme events risk assessment tools, using a combination of Bayesian Network (BN) modelling and System Dynamics (SD) modelling. Section 2 illustrates the benefits of such methodology and the key-steps for the development of such tool. Section 3 describes the activities that led to the development of the conceptual model, while Section 4 explains how this was then translated into a BN. Section 5 ends the paper by summarising the current findings and the next steps, such as the development of the SD model.

2. METHODS

The Research Team is developing an extreme event risk assessment tool using Bayesian Network (BN) modelling and System Dynamics (SD) modelling. A combination of these two modelling frameworks is proposed because of the following attributes:

- Both provide a modelling framework that allows prediction of an outcome (e.g. decline in water quality) even when the determining conditions (e.g. an extreme event) are both variable and uncertain.
- Both are able to integrate data from different sources (e.g. model output, monitoring and expert opinion) and of different types (environmental, social and economic) into a single model.
- SD is able to analyse the behaviour of complex systems (e.g. water quality management) and their interacting components with many feedbacks and with changes over time.
- BN provides an ideal representation for combining prior knowledge with data, and it is particularly helpful when dealing with uncertainty (Nadkarni and Shenoy, 2004)

The modelling process comprised the following core steps:

- An initial expert workshop was held in order to define the scope of the project, select the case-study sites, the key water quality parameters to be modelled and related levels of service, and to populate the preliminary conceptual models.
- The conceptual model was converted into a BN by the Research Team. In order to fill the Conditional Probability Tables (CPTs) attached to each node of the BN, a second expert workshop, with water utility experts from different fields, was held.
- The BN architecture and findings, along with available historical data, will be used to develop the SD model.

3. CONCEPTUAL MODEL DEVELOPMENT

3.1. First expert workshop, Sydney

The following section describes the outcomes of the first expert workshop held in Sydney (Australia) in 2015. The workshop process can be separated into three distinct components. The first part of the first expert workshop was used to identify the scope (being the Prospect water filtration plant raw water supply system), as well as the key water quality parameters of concern and the respective critical levels of these parameters that would lead to the expected Level of Service to be ‘not guaranteed’. These parameters were identified as: turbidity, water colour and *Cryptosporidium sp.* and the agreed levels of service were, respectively: 40 NTU, 60 CU⁴⁰⁰ and 10 IFA/10L (from “immunofluorescence assay” method) adjusted for recovery. The second part of the first expert workshop consisted of “unstructured” interviews, where the experts were asked to identify the parameters affecting the key-variables being modelled. The third part of the workshop consisted of “structured” interviews, meaning that the experts were asked to modify a preliminary conceptual model built based on the outcomes of the unstructured interviews. An outcome from the workshop was the development of three separate models, one for each critical parameter. Importantly, the type of “extreme events” (including combinations of these e.g. drought followed by flood), were identified at this stage of the project. Thus, extreme events were defined as being related to (both individually and cumulatively) inflow events (rainfall), bushfire and/or drought. The availability of critical infrastructure during extreme events was also identified as an important element to be factored into this analysis.

The following provides more detailed information about the process that was undertaken. As a first step of the conceptual model development, the main parameters directly affecting turbidity, water colour or *Cryptosporidium sp.* levels were identified (Table 1). The causality is positive if an increase in the input value relates to an increase in turbidity, colour and/or *Cryptosporidium sp.* (i.e. decrease in the water quality). Definitions of key-parameters from Table 1 are:

- Avoidance capacity: this is linked to the presence of, for example, intake towers with multiple gates at the reservoir, which allow the selection of the optimal (with regards to water quality) intake depth. However, its usefulness is limited during lake circulation periods (e.g. winter turnovers) as the water quality is uniform throughout the water column.
- Spill: if the dam is spilling (due to the storage level exceeding the full capacity), then the water quality is expected to deteriorate as the avoidance capacity is reduced due to the water moving from the bottom to the top of the dam wall (assuming the inflow coming as an underflow); the main factor affecting a possible spill is the storage level.

- Use of alternative reservoirs: the presence of other reservoir(s) that can be used to deliver raw water to the Prospect WTP. This allows for drawing raw water from other sources to provide the most appropriate water quality to optimise and ensure effective treatment. Raw water supplying the Prospect WTP is typically drawn from Warragamba Reservoir, but the Upper Canal supply route (which includes Cataract, Cordeaux, Nepean and Avon Reservoirs) and Prospect Reservoir itself can be also used as a backup source of water. Factors affecting the use of alternative reservoirs were identified during the workshop as asset failure and contamination (e.g. bushfire damage).
- Ashes: originated from bushfires and subsequently washed into the reservoir via surface runoff resulting in increased colour and turbidity in the reservoir. The main input factors were identified as the presence of a fire in forested areas around the catchment and rainfall events following the fire.
- Runoff and Crypto Runoff: the runoff following a high rainfall event will result in significant sediment and organic matter loading, increasing the levels of turbidity and colour in the reservoir. In some cases the amount of *Cryptosporidium sp.* also will increase. It was decided during the workshop for modelling purposes, to create separate variables for ‘runoff’ and ‘crypto runoff’. The rationale for this approach is that the runoff affecting turbidity and colour is mainly influenced by the amount of rainfall (intensity and duration) and catchment size, but in order for the runoff to generate high *Cryptosporidium sp.* levels, other inputs (e.g. the presence of intensive livestock production, onsite sewage systems, grazing, and the possibility of an overflow of a Sewage Treatment Plant) can play an important role.
- Swamp runoff: another special case of runoff, which affects colour only. As for the creation of a “crypto runoff” variable separated from the main “runoff” variable, this specific fraction of the runoff was kept separate as it is affected by different inputs (e.g. swamp size).
- Landslip event: indirect effect of rainfall events; it would increase the turbidity levels in the reservoir.
- Storage level: typically, a higher storage level implies more water column stability, more dilution, and generally a better water quality. It increases the avoidance capacity (i.e. more gates of the intake tower under water, thus more choice), but increases the risk of spill. It is affected by mainly the runoff and direct rainfall.

Table 1. List of water quality predictors and their effect (polarity) on turbidity, colour and cryptosporidium.

PREDICTOR	CORRELATION WITH TURBIDITY	CORRELATION WITH COLOUR	CORRELATION WITH CRYPTOSPORIDIUM
Spill	+	+	+
Avoidance Capacity	-	-	-
Use of Alternative Reservoirs	-	-	-
Ashes	+	+	NA
Runoff	+	+	+
Crypto Runoff	NA	NA	+
Swamp runoff	NA	+	NA
Landslip events	+	NA	NA
Storage Level	-	-	-

Following these considerations, three separate conceptual models were built collectively by the workshop participants, where connections and nodes were defined in accordance with Table 1. A feature of these models is that the main factors affecting water quality (as selected by the participants) were not only environmental (e.g. rainfall, drought, fires) but also related to the attributes and facilities of the water utility (e.g. variables such as avoidance capacity, alternative reservoirs, asset failure), land use (e.g. agricultural areas, forested areas, farms, grazing, intensive livestock) and even extreme human actions (such as intentional contaminations). The diversity of the conceptual model variables supported the choice of using BN and SD modelling frameworks for the project. Both frameworks are integrative and deal competently with limited and/or multi-field knowledge.

3.2. Development of a single system conceptualisation model

Following the first workshop, the three separate models were merged into a single one (Fig. 1). Thicker connections indicate that a node directly affects at least one of the three key-parameters. Additionally, these connectors are blue when the input is a positive factor (i.e. an increase in the input value implies a decrease in

the target parameter) and red if the input is a negative factor (i.e. an increase in the input value implies an increase in the target parameter). All the main input parameters' names (i.e. those directly affecting turbidity, colour and/or cryptosporidium) are also presented in bold and purple. All the secondary connections are thinner and in dark red.

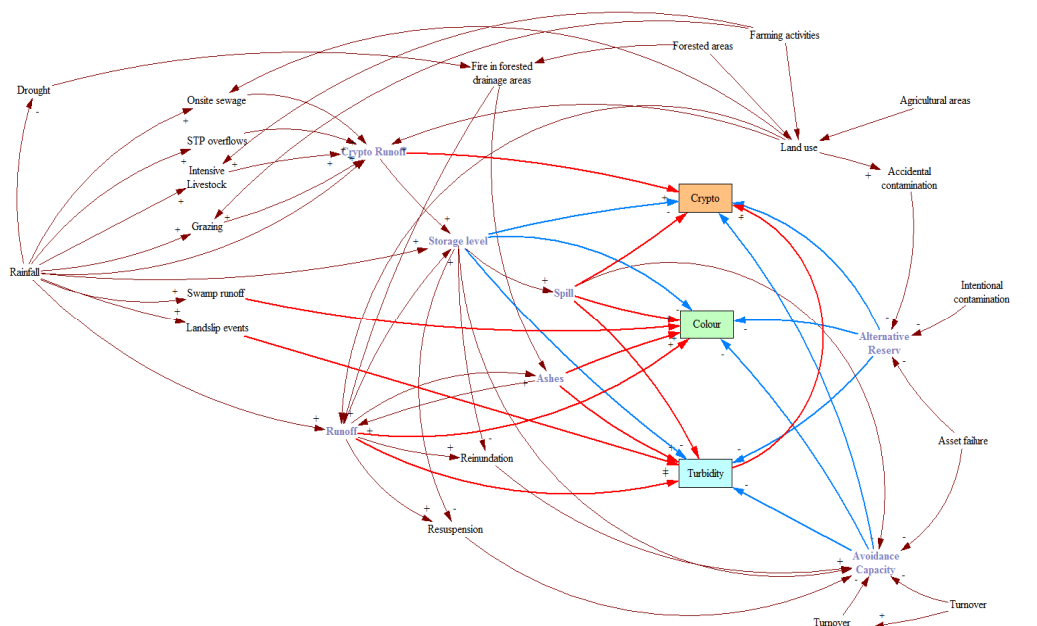


Figure 1. Final comprehensive conceptual model of factors affecting colour, turbidity and cryptosporidium.

4. BAYESIAN NETWORK DEVELOPMENT

The comprehensive conceptual model (Fig. 1) was used as the foundation for developing a Bayesian Network (BN) that would be used to assess the probability of delivering the required level of service under different conditions (scenarios). Much of BN development and application has emerged from Artificial Intelligence research (Korb and Nicholson, 2010) and they are an increasingly popular modelling technique, especially when the system being modelled presents a high degree of uncertainty and complexity, such as in ecosystems and environmental management (Korb and Nicholson, 2010). Each variable within a BN is presented as a node. A node that has direct input connections (arcs) from at least one other node (“parent”) is called a “child” node for that parent node. The strength of the connection (also known as conditional dependence) between a child node and its parent node(s) is quantified through probability distributions. There is one probability distribution for each combination of possible values of the parent node states. These conditional probabilities are defined in the Conditional Probability Tables (CPTs) attached to each node that has at least one parent node. Uncertainty is measured through probability by the BN, i.e. the higher the uncertainty, the wider the probability distribution; however, when more information/data becomes available and uncertainty decreases, the probability distribution usually becomes narrower and the knowledge of the true value of the node increases. Evidence is entered into the BN by substituting the a priori belief with observations (hard or soft evidence) or scenario-based values for a number of nodes (Chen and Pollino, 2012). Interactions between variables are displayed clearly and users can easily interrogate the reasoning behind the model output, thus providing a more transparent approach when compared to other “black-box” modelling techniques such as artificial neural networks (Chen and Pollino, 2012). In general, BNs are suitable for small or incomplete data sets: BNs can easily handle missing or little data, and typically can yield good prediction accuracy even with a small sample size, provided that the model structure is well defined (Uusitalo, 2007). Also, it is possible to combine different sources of data: that is, where ‘hard’ data (survey, model and/or monitoring data) are not available, probabilities can be entered manually through expert knowledge. Thus hybrid sources of data (historical data, expert knowledge) can be used to overcome historical data limitations (e.g. where historical trends are not good predictors of future events) or to enhance the model performance (Uusitalo, 2007). Overall, they provide a suitable support tool for decision makers, as costs and risks associated with different management strategies can be assessed easily; in addition, the model simulation is typically extremely fast compared to some process-based models (Uusitalo, 2007).

After the original conceptual model was built (Fig. 1), the Bayesian model structure was defined using the methodological framework of Chen and Pollino (2012). As model parsimony is essential (but balanced with

model accuracy), it is important to retain only influential variables (influential on the key nodes), and to reduce the number of states for each node to a minimum. This assists with producing conditional probability tables that can be reasonably populated by expert knowledge. An additional requirement for developing the BN structure is that feedback loops must be avoided. Thus, several minor modifications were performed by the Research Team after consultation with a water industry expert prior to conducting the second workshop (where the CPTs were populated by expert opinion). The final BN structure is illustrated in Fig. 2. Different colours represent different categories of variables (e.g. blue are environment-related, orange are anthropogenic, green are water utility-related, yellow are miscellaneous), thus clearly showing the capacity of BN of dealing with multi-field problems.

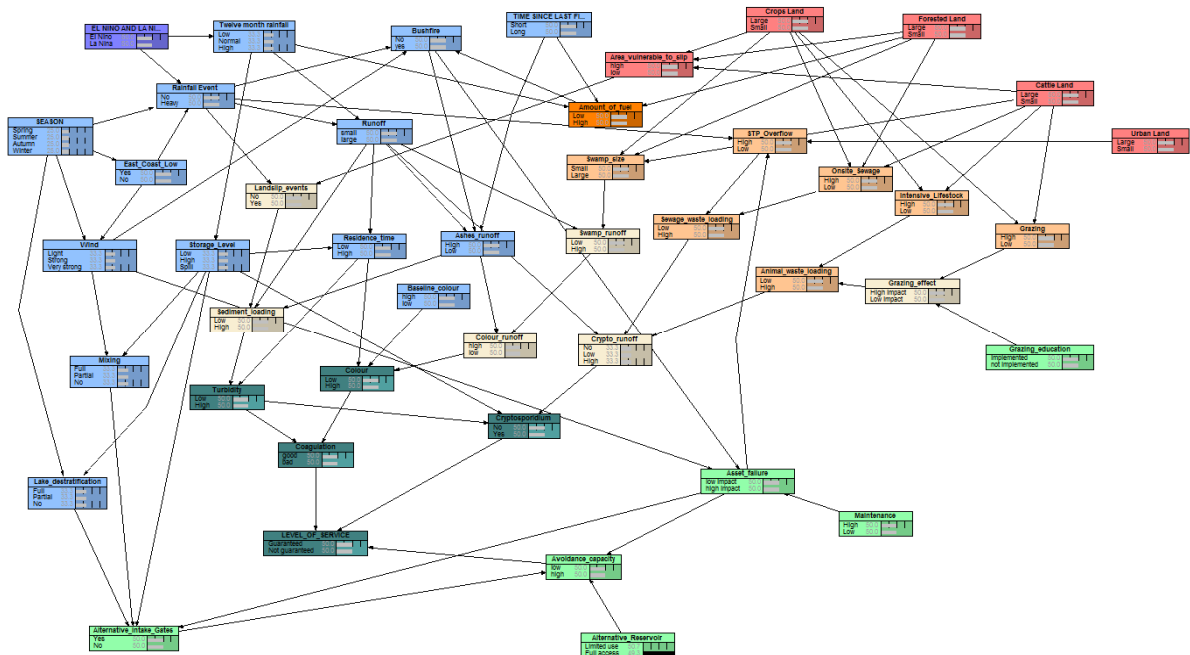


Figure 2. BN model of factors affecting the “Level of Service”

The structure of the conceptual model that led to the BN development was defined during and after the first participant workshop. This included preliminary definitions of the variables and connections. The second workshop was used as a mechanism to obtain feedback on the model (including slight modifications to the structure), populate the CPTs of the BN with expert evidence and identify where alternative data might be available (e.g. Bureau of Meteorology data regarding rainfall, wind, or El Niño/La Niña states). The workshop was held in the Water NSW’s main building in Penrith, Sydney (May 2015). Ten experts from different fields (e.g. water quality, water treatment, microbial risk, system configuration, risk management, operations management) attended the workshop and each was invited to populate all of the CPTs that sat behind the BN structure. This activity took about 3 hours for each expert using a ‘pen and paper’ approach (i.e. expert is given a blank CPT and asked to assign probabilities for each of the CPT scenarios). Some important issues (e.g. nodes definition, model structure, length of the CPT population activity, contrasting opinions between stakeholders) emerged and were addressed by the Research Team during the workshop. For instance, it was decided to collect the expert opinions independently rather than having an open discussion on each node. This made the process faster and avoided the potential impact of stronger personalities influencing the professional judgement of others and to prevail and eventually dictate his belief; it also provides more meaningful data, as in case of group-level engagement, the variability in the assigned probabilities would be lost as they will be the result of a compromise between the stakeholders. To account for the different probabilities assigned by the participants, an auxiliary node was included (see Richards *et al.*, 2013). This auxiliary node representing the stakeholder beliefs was connected to all the nodes whose associated CPT was filled by the experts. By doing so, it is possible to run the BN and assess the levels of risk based on probabilities assigned by people with different areas of expertise (e.g. the expert opinions of microbiologists versus operators versus bushfire experts). In this way it was also possible to assess how sensitive the model was to the different experts’ opinions and how these manifest on the probability of delivering water of the required standard (i.e. the focus of the BN). This is a strength of BN models when applied to a problem requiring the integration of multiple types of expertise (e.g. reservoir dynamics, bushfire dynamics, land-use expertise, microbial risks expertise, water treatment, etc.). By using a BN and populating

it with expert opinion, these diverse areas of expertise can be integrated robustly (using Bayes theorem, which provides the mathematical basis for BNs) within a single model. In case of strong diverging opinions around some nodes, it is also possible to engage the different stakeholders again in order to understand the different points of view.

5. DISCUSSION AND CONCLUSION

The research presented is part of an ongoing project and further activities and results are expected. Firstly, the developed BN can now be used by stakeholders to assess the risk of unacceptable levels of turbidity, water colour and *Cryptosporidium sp.* following one, or a combination of, extreme events. Importantly, although the generic negative effects of such extreme events were already known, numerical outcomes are now provided by the BN, so that it is possible to list those events from the most to those with least impact on the water quality, for this particular catchment. Also, BNs can be used, not only through a “top-down”, but also a “bottom-up” approach: that is, the worst input scenarios can be assessed by assuming unacceptable levels of one/multiple water quality parameters to be expected. Also, the effect of different management intervention options (e.g. increased maintenance) can be easily assessed, thus enabling the water managers to identify those providing the higher benefit compared to the cost.

Future activities will focus on the use of System Dynamics (SD) and the combination of both approaches. The available historical data will be collected and used to assess the expected future temporal trends in the water quality assuming a number of extreme events, consistent with historical data, will occur. The expected model will be complementary to the BN, thus bringing consistency and credibility to the overall modelling approach adopted.

ACKNOWLEDGMENTS

The Research Team is grateful to Water Research Australia and Water NSW for providing technical and financial support to this collaborative project.

REFERENCES

- Betancourt, W.Q., Rose, J.B. (2004). Drinking water treatment processes for removal of *Cryptosporidium* and *Giardia*. *Veterinary Parasitology*, 126 (1–2), 219–234.
- Chen, S.H., Pollino, C.A. (2012). Good practise in Bayesian Network modelling. *Environmental Modelling and Software*, 37, 134-145.
- Environmental Protection Agency of the United States. (1999). Guidance manual for compliance with the interim enhanced surface water treatment rule: turbidity provisions. Office of Water, April 1999.
- Graczyk, T.K., Fried, B. (2007). Human waterborne trematode and protozoan infections. *Advances in Parasitology*, 64, 111–160.
- Graczyk, T.K., Kacprzak, M., Neczaj, E., Tamang, L., Graczyk, H., Lucy, F.E., Girouard, A.S. (2008). Occurrence of *Cryptosporidium* and *Giardia* in sewage sludge and solid waste landfill leachate and quantitative comparative analysis of sanitization treatments on pathogen inactivation. *Environmental Research*, 106 (1), 27–33.
- Hrudey, S.E. (2009). Chlorination disinfection by-products, public health risk tradeoffs and me. *Water Research*, 43, 2057-2092.
- IPCC. (2014). Climate Change 2014: Synthesis Report. *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.*
- Khan, F. A., Ali, J., Ullah, R. and Ayaz, S. (2013). Bacteriological quality assessment of drinking water available at the flood affected areas of Peshawar. *Toxicological and Environmental Chemistry*, 95(8), 1448-1454.
- Korb, K. B., Nicholson, A. E. (2010). *Bayesian artificial intelligence*. CRC press.
- McClennan, P. (1998). Sydney Water Inquiry: Final Report. Volume 2 (Fifth Report ed.).
- Mons, C., Dumetre, A., Gosselin, S., Galliot, C., Moulin, L. (2009). Monitoring of *Cryptosporidium* and *Giardia* river contamination in Paris area. *Water Research*, 43 (1), 211–217.
- Nadkarni, S., Shenoy, P.P. (2004). A causal mapping approach to constructing Bayesian networks. *Decision Support Systems*, 38, 259-281.
- Putignani, L., Menichella, D. (2010). Global distribution, public health and clinical impact of the protozoan pathogen *Cryptosporidium*. *Interdisciplinary Perspectives on Infectious Diseases*, 2010.
- Richards R., Sano, M., Roiko, A., Carter, R.W., Bussey, M., Matthews, J., Smith, T.F.(2013). Bayesian belief modeling of climate change impacts for informing regional adaptation options.*Environmental Modelling & Software*,44,113-121.
- Sarai, D.S. (2006). *Water treatment made simple for operators*, pp. 263. John Wiley and Sons, Inc, Hoboken, New Jersey
- Sydney Catchment Authority (2010). Climate change impact assessment 2010.
- Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological modelling*, 203, 312-318.