Citation: Bertone, E.; Sahin, O.; Richards, R.; Roiko, A. (2015). Bayesian Network and System Thinking modelling to manage water-related health risks from extreme events. IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 6-9 December 2015

Bayesian Network and System Thinking Modelling to Manage Water-Related Health Risks from Extreme Events

E. Bertone¹, O. Sahin¹, R. Richards², R. A. Roiko³

¹ Griffith School of Engineering, Griffith University, Queensland, Australia

² School of Agriculture and Food Sciences, University of Queensland, Brisbane, Australia

³ Griffith School of Medicine, Griffith University, Queensland, Australia

Email: e.bertone@griffith.edu.au

Abstract - 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 has been developed. The risk assessment tool is applied to the Prospect water filtration plant system, main source of potable water for the Sydney metropolitan region. Conceptual models were developed by the stakeholders around the key indicator parameters of turbidity, water colour and cryptosporidium. These three conceptual models were used for developing separate BN and SD models. Here we present the development of a BN designed to understand the risk of extreme events on the ability to provide drinking water of a desired 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.

Keywords - Bayesian Networks, Extreme Events, Water Quality

I. INTRODUCTION

Recent history in Australia has been characterized 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 drinking water of a required standard to consumers. At issue, are the short- and long-term impacts of extreme events on the water quality at both the pre- and post-treatment (including distribution and end-point). Extreme events are projected to change in magnitude and frequency over the next century [1], 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. Water NSW) in case of extreme events are: water colour, turbidity, and cryptosporidium.

Water colour is a key parameter in drinking water reservoirs as it can affect 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 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 carcinogen trihalomethanes (THM's), one of the over 600 disinfection by-products currently reported in drinking water [2].

Water turbidity is another very important parameter that is monitored by water authorities. Turbidity refers to how clear the water is. It is a result of suspended particles that can provide food and shelter for pathogens, and if not effectively removed, can promote regrowth of pathogens in the distribution system, leading to waterborne disease outbreaks [3;4], which can cause e.g. cramps, diarrhea, headache and nausea [5].

The third key parameter of this project, cryptosporidium, is an intestinal protozoan pathogen that infects humans, domestic animals and wildlife worldwide. It can be found in the faeces of infected humans and animals [6], which can enter surface waters directly or through effluents and runoff from fields that are polluted by sewage sludge or [7;8] resulting in pollution of receiving waters. Importantly, the cryptosporidium (oo)cysts have the capacity to remain infective for months in environmental waters and are highly resistant to chlorinated [9]. Therefore, waterborne contamination is a growing concern for water suppliers, causing widespread outbreaks of these diseases [10]. 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[11].

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 [12] bringing detrimental effects for the water quality of reservoirs. Water quality management in this context requires a multi-disciplinary approach, both holistic and probabilistic, to develop appropriate management strategies. Strong support and active participation from the water industry itself, whose experiences with past occurrences of extreme events are invaluable sources of qualitative and quantitative information, is also required.

II. METHODOLOGY

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

- They 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.

- They 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 changing over time.

- BN provides an ideal representation for combining prior knowledge with data, and it is particularly helpful when dealing with uncertainty [13].

The modelling process comprised the following core steps:

- A first expert workshop was held in order to define 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 (CTPs) attached to each node of the BN, a second expert workshop, with water utility experts from different fields, was organized.

- The BN architecture and findings, along with collected historical data, will be used to develop the SD model.

III. CONCEPTUAL MODEL DEVELOPMENT

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 supply system), as well as the key water quality parameters of concern and the respective critical levels that would imply the expected Level of Service to be not guaranteed. These parameters were identified as: turbidity, water colour and cryptosporidium and the agreed levels of service were, respectively: 40 NTU, 60 CU400 and 10 IFA/10L 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 to be 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. 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. It was identified that the availability of critical infrastructure during

extreme events is an element to be factored into this analysis.

As a first step of the conceptual model development, the main parameters directly affecting turbidity, water colour or cryptosporidium levels were identified as being:

- 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. These structures reduce the risk of delivering raw water with very poor water quality features to the water treatment plant (e.g. after an extreme rainfall event). 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 spills (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 factors affecting a possible spill is the storage level and inflow.

- Use of alternative reservoirs: the presence of other reservoir(s) that can be used to deliver raw water to the Prospect Water Treatment Plant. This allows for drawing raw water from other sources (than the default reservoir where water is usually drawn from) if the water quality in the default reservoir is ‘poor’. Raw water reaching the Prospect Water Treatment Plant is typically drawn from Warragamba Reservoir, but Prospect Reservoir and the Upper Canal supply route (which includes Cataract, Cordeaux, Nepean and Avon Reservoirs) 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 whether accidental (e.g. bushfire damage) or intentional (e.g. terrorist attack).

- Ashes: originated from bushfires and subsequently washed into the reservoir via the surface runoff resulting in increased colour and turbidity in the reservoir, and affected by the presence of a fire in forested areas around the catchment and rainfall events following the fire.

- Runoff and Crypto Runoff: following a high rainfall event, the associated runoff will result in sediment and organic matter loading, increasing the levels of turbidity and colour in the reservoir. In some cases the amount of cryptosporidium 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 levels, other inputs (e.g. the presence of intensive livestock, onsite sewage, grazing, and the possibility of an overflow of an onsite sewage treatment plant) are important.

- Swamp runoff: another special fraction of runoff, affecting colour only and related to other inputs.

- Landslip event: another indirect effect of rainfall events, which 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.

Following these considerations, three separate conceptual models were built by the workshop participants. 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 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. Following the first workshop, the three separate models were merged into a single one (Fig. 1). Thicker connections indicate when 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. affecting directly turbidity, colour and/or cryptosporidium, are also in bold and purple. All the secondary connections are thinner and in dark red.

IV. 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 [14] 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 [14]. 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 termed a "child" node for that parent node. The strength of 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 per each combination of possible values of the parent node states. These conditional probabilities are defined in the Conditional Probability Tables (CTPs) attached to each node that has at least one parent node. The uncertainty is measured through probability by the BN, i.e. the higher

the uncertainty, the wider the probability distribution; however, when more information/data is available and uncertainty decreases, usually the probability distribution 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 scenarios values for a number of nodes [15]. Interactions between variables are clearly displayed 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 [15]. In general, BNs are suitable for small or incomplete data sets: BN 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 [16]. Also, it is possible to combine different sources of data: that is, where 'hard' data (survey, model and/or monitoring data) is not available, probabilities can manually be entered 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 [16]. Overall, they provide a suitable support tool for decision makers, as costs and risks associated to different management strategies can be easily assessed; additionally, the model simulation is typically extremely fast compared to some process-based models [16]. After the original conceptual model was built (Fig. 1), the Bayesian model structure was defined using the methodological framework of [15]. As model parsimony is essential (but balanced against 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 CTPs that are relatively small and therefore more easily populated by expert knowledge. Additionally, 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 CTP tables 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. 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. A second workshop was used as a mechanism to obtain feedback on the model (including slight modifications to the structure), populate the CTPs of the BN with expert evidence and identify where alternative "hard" data might be available. The workshop was held in the Water NSW main building in Penrith, Sydney (May 2015). Ten experts of different fields (e.g. water quality, water treatment, microbial risk, system configuration, risk management, operations

Fig. 1. Final comprehensive conceptual mode

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 CTP population activity, contrasting opinions between stakeholders) emerged and were addressed by the Research Team during the workshop. For instance, it was decided to independently collect the expert opinions rather than having an open discussion on each node. This makes the process faster and avoids strong personalities to prevail and eventually dictate his belief. To account for the different probabilities assigned by the participants, an auxiliary node was included (see [17]). This auxiliary node, representing the stakeholder beliefs, was connected to all the nodes whose associated CPT was filled by the experts, in order to run the BN and assess the risk based on different areas of expertise (e.g. the expert opinions of microbiologists versus operators versus bushfire experts). In this way it is also possible to assess how sensitive is the model 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.): BN allows their robust integration 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.

V. DISCUSSION AND CONCLUSION

The presented research is 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 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 the least impactful for water quality, for the study location. For instance, preliminary results show how a bushfire alone would still guarantee safe water with a probability of 99.1%, while if the fire is followed by heavy rain, that decreases to 67.6% due to the ash-rich runoff. Also, BN 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. Lastly, 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). The available historical data will be collected and used to assess the expected future temporal evolution of 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 adopted overall modelling approach. Despite being specifically designed for this case study, the proposed methodology can be reapplied to other systems to be modeled and it is especially relevant in cases where high uncertainty and lack of data is involved.

ACKNOWLEDGMENT

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

Fig. 2. BN structure

REFERENCES

- [1] 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.
- [2] Hrudey, S.E. (2009). Chlorination disinfection by-products, public health risk tradeoffs and me. *Water Research*, 43, 2057-2092.
- [3] 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..
- [4] 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.
- [5] Sarai, D.S. (2006). *Water treatment made simple for operators*. John Wiley and Sons, Inc.
- [6] Graczyk, T.K., Fried, B. (2007). Human waterborne trematode and protozoan infections. *Advances in Parasitology*, 64, 111-160. [7] J. E. Monzon, "The cultural approach to telemedicine in Latin American homes (Published Conference Proceedings style)," in *Proc. 3rd Conf. Information Technology Applications in Biomedicine, ITAB'00*, Arlington, VA, pp. 50-53.
- [7] 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.
- [8] 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.
- [9] Betancourt, W.Q., Rose, J.B. (2004). Drinking water treatment processes for removal of *Cryptosporidium* and *Giardia*. *Veterinary Parasitology*, 126 (1-2), 219-234.
- [10] Putignani, L., Menichella, D. (2010). Global distribution, public health and clinical impact of the protozoan pathogen *Cryptosporidium*. *Interdisciplinary Perspectives on Infectious Diseases*, 2010.
- [11] McClelland, P. (1998). *Sydney Water Inquiry: Final Report. Volume 2 (Fifth Report ed.)*.
- [12] Sydney Catchment Authority (2010). *Climate change impact assessment 2010*.
- [13] Nadkarni, S., Shenoy, P.P. (2004). A causal mapping approach to constructing Bayesian networks. *Decision Support Systems*, 38, 259-281.
- [14] Korb, K. B., Nicholson, A. E. (2010). *Bayesian artificial intelligence*. CRC press.
- [15] Chen, S.H., Pollino, C.A. (2012). Good practise in Bayesian Network modelling. *Environmental Modelling and Software*, 37, 134-145.
- [16] Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological modelling*, 203, 312-318.
- [17] 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.