

The value of stochastic modelling for the characterisation of maturation pond performance and health risk assessment of effluent reuse options

Kozak S^a, Xie G^{b c}, Leusch, F^b and Roiko A^{a b}

^a School of Medicine, Griffith University

^b Smart Water Research Centre, Griffith University

^c University of the Sunshine Coast

Email: a.roiko@griffith.edu.au

Abstract: Treated wastewater from Waste Stabilisation Ponds (WSPs) is a potential resource, especially in regional and remote areas facing water shortages. As the ultimate goal of reusing treated wastewater is ensuring that it is fit-for-purpose, the comprehensive assessment, management and communication of potential health risks are a necessity. Although not mandated by policy in Australia, Quantitative Microbial Risk Assessment (QMRA) is recommended in the Australian Guidelines for Water Recycling (AGWR) as a methodology for estimating the potential levels of health risk associated with exposures to microbial pathogens. With the rapid uptake of QMRA modelling in the water sector in Australia, it is necessary to explore the sensitivity of QMRA models to underlying assumptions.

This stochastic modelling study investigated the impacts of uncertainty in key input parameters underpinning QMRA and evaluated the sensitivity of a QMRA model to a range of underlying assumptions. This was conducted using the @Risk software program within the Palisade Decision Suite and operational monitoring data from a pond-based, Waste Water Treatment Plant (WWTP) in regional New South Wales, Australia.

This study was conducted in two phases. The first addressed two research questions regarding: 1: The impact of assumed sampling regimes (weekly, fortnightly or monthly), and 2: The impact of seasonal variability in characterising pond performance as measured by Log Reduction Values (LRVs). The fortnightly and monthly datasets were compiled from the weekly monitoring data collected from both the inlet and outlet of the 3 maturation-pond system within the overall treatment system. The weekly data were also stratified by season for investigation of potential seasonal patterns. The distributions of these LRVs were generated through Monte Carlo simulations and impacts of temporal variation investigated statistically. The second phase focused on estimating potential health risks associated with the hypothetical scenario of irrigating lettuce with pond effluent using QMRA. The results from the first phase were used as input data in the QMRA stochastic modelling. The QMRA incorporated the four steps: hazard identification, dose-response, exposure assessment and risk characterisation. Monte Carlo simulation was used to generate a probabilistic distribution of health risk estimates, and the built-in sensitivity analysis functions in @Risk were used to rank the input parameters by their effect on the estimated risk levels.

The results from the first phase revealed no significant difference between how weekly, fortnightly or monthly datasets characterised the microbial water quality from the maturation pond system in terms of LRVs. This suggests that the frequency of monitoring at the WWTP could be reduced without compromising the information value of the dataset. This would, in turn, reduce expenses. Seasonality, however, does appear to have a significant impact on pond performance as measured by LRVs. A pair-wise comparison of the weekly data by season revealed statistically significant differences between all seasons, except for winter and autumn. Summer provided the best performance and spring provided the worst. The results from the health risk assessment (second phase) suggested the microbial quality of the pond effluent would not be suitable for the proposed reuse scenario of irrigating lettuce eaten raw; however alternative, less risky reuse scenarios would be within the scope of the state guidelines. This phase of the research also opened up a broader range of issues such as how guidelines are interpreted and how different modelling approaches (e.g., deterministic versus stochastic) can yield disparate results. The stochastic methods provide more conservative estimates and therefore engender greater assurance in both regulators and the community that safety standards are being met. Overall, this research highlights that stochastic modelling is a powerful tool for risk communication and risk management for the water industry, the community and regulators.

Keywords: *Stochastic modelling, pond effluent reuse, pathogen log reduction values, quantitative microbial risk assessment*

1. INTRODUCTION

The field of Quantitative Microbial Risk Assessment (QMRA) is relatively young, but is growing (Haas, Rose & Gerba, 2014). It is being promoted as a rigorous methodology for modelling the human health risks associated with microbial hazards from exposure to water in a variety of contexts, such as for drinking water, recreational water and recycled water from storm and waste water. Early attempts to apply the traditional chemical health risk assessment paradigm (incorporating the four steps: hazard identification, dose-response, exposure assessment and risk characterisation) to microbial hazards faced many challenges; it was hampered by insufficient data, and knowledge of how to analyse existing data within each of these steps (enHealth, 2012; Haas *et al.*, 2014). In recent years, however, there has been a substantial growth in the knowledge of, and scope for, QMRA. Specifically, there has been better identification and enumeration of microorganisms, development of dose-response models and assessment of human health effects and greater access to risk modelling software platforms. This ongoing expansion of the current knowledge base equips health risk assessors with the capacity to demonstrate the value of stochastic modelling to an increasing range of real-world problems.

In the context of recycled water, the ultimate goal is safe, fit-for-purpose reuse. The comprehensive assessment, management and communication of potential health risks are therefore essential, particularly for microbial hazards which can result in immediate health threats (NRMCC, 2006; Toze, 2006). To this end, key guidance documents are being produced to facilitate its broader uptake, including the *Guidelines for Water Reuse* (US EPA, 2012) and guidelines being developed by the World Health Organisation (WHO) on QMRA.

Each of the four steps in the QMRA framework (Figure 1.) can be underpinned by either deterministic or stochastic analyses. Deterministic approaches utilise point estimates of the model parameters whereas stochastic approaches make use of probabilistic distributions in place of point estimates. The latter better accounts for the uncertainty and variability that is inherent in the input parameter estimates, such as pathogen concentrations, body weight and environmental decay rates of pathogens (Haas *et al.*, 2014; NRMCC, 2006). A comprehensive review of the how uncertainty is accounted for within these two approaches can be found in Haas *et al.* (2014). In essence, deterministic approaches have the advantage of being simple to compute and the information produced is relatively easy to communicate to stakeholders however they do not directly address variability and uncertainty and can convey a false sense of certainty in the computed estimates of risk (Haas *et al.*, 2014). Stochastic approaches, on the other hand, are able to describe both uncertain and variable inputs mathematically using probabilistic distributions and are more informative and inherently more representative than point estimates (NRMCC, 2006) when applied appropriately. They are, however, computationally more difficult and require supporting information that may not be readily available (Haas *et al.*, 2014; NRMCC, 2006).

Meaningful probabilistic distributions can only be generated when there are adequate data to describe the stochastic features of the input variables being used to estimate risk (Arcella & Leclercq 2004). If inadequate data are used, the resulting probabilistic distributions may be so uncertain that it can lead to inaccurate conclusions being drawn (Cullen & Frey, 1999). With regard to computational difficulty, this has been addressed to some extent through the wider accessibility to computer software programs such as @Risk (Palisade Corporation, 2014) and R (R Development Core Team, 2013).

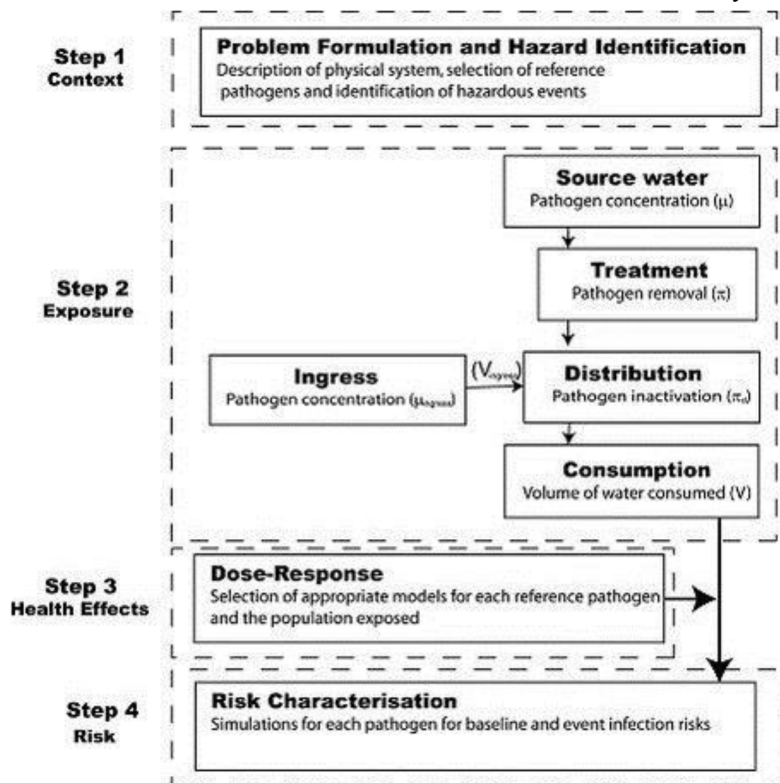


Figure 1. Overview of the quantitative microbial risk assessment process

programs such as @Risk (Palisade Corporation, 2014) and R (R Development Core Team, 2013).

Among stochastic approaches to QMRA there are several ways to quantify the uncertainty and variability of input parameters. These include likelihood ratio methods, interval methods and Bayesian and Monte Carlo methods, with the latter two allowing for an assessment of combined uncertainty from multiple inputs (Haas *et al.*, 2014). Monte Carlo simulation is becoming increasingly common in microbial risk assessments, particularly those concerned with recycled water reuse (Mok *et al.*, 2014; Hamilton *et al.*, 2006; Westrell *et al.*, 2004). One of the main impediments to the development and use of QMRA is the lack of complete information and data as input to this process. Identifying knowledge data gaps will facilitate the uptake of this methodology and benefit industry, community and regulators.

2. CASE STUDY SITE DESCRIPTION

A suitable WWTP with potential for treated wastewater reuse was found in regional New South Wales (NSW), Australia. Influent includes a mixture of both domestic and industrial wastewater. The hydraulic retention time of these three maturation ponds is approximately 25 days. After maturation pond treatment, the effluent is pumped via the effluent transfer pumping station to the storage and reuse farm.

Recycled water needs a market. Approvals for reuse schemes in NSW are governed by the *Environmental Guidelines: Use of effluent by irrigation* (2004) (Department of Environment and Conservation, 2004). The *Australian Guidelines for Water Recycling* (AGWR) (NRMCC, 2006) are also used to inform NSW regulatory and legislative requirements. With regards to treated wastewater reuse, the NSW *Guidelines* state that NSW Health should be consulted in regard to the level of treatment of effluent to be achieved when public health could be at risk through contact with irrigated effluent or products that have been produced with irrigated effluent (Department of Environment and Conservation, 2004).

3. METHODS

3.1. Treatment of raw data: Approximately three years of monitoring data were available from the case study WWTP. The raw data acquired included (roughly) weekly sampling for thermotolerant coliforms (TCs) along with fourteen other water quality parameters (e.g. pH, TSS, ammonia) between May 2011 and September 2014. This study utilised the following raw data from the case study WWTP dataset: sampling date and TC concentrations (CFU/100ml) from the Inlet Works (n=175), the inlet of Maturation pond 1 (n=179), the outlet of Maturation pond 3 (n=182) and the Farm Dam outlet (n=173). All TC concentrations were measured using a single 24 hour composite sample. Censored data in the maturation pond 3 dataset were treated by the substitution method whereby data below the limit of reporting (LOR) were substituted with values one half of the LOR, e.g., a data point is given a value of 1 for a LOR of <2. While other more sophisticated methods exist, this simple method is considered appropriate for microbial data where censored data constitute less than 25% of the dataset (NHMRC, 2008). Censored data constituted a significant proportion of the dataset (>25%) in the farm dam outlet dataset and therefore a more conservative method of substituting censored data with the LOR was used, e.g., a data point was given a value of 2 for a LOR of <2.

The TC concentration data were converted into pathogenic *Escherichia coli* (*E. coli*) concentration data using the following conversion: approximately 95% of TCs are *E. coli* (Mara & Horan, 2003), and 8% of *E. coli* are pathogenic (D. Rosser, personal communication, 18 October, 2014; Haas, 1999). The percentages of TC's that are *E. coli* have been reported to be much lower in other studies (between 47-71 %) (Elmund *et al.*, 1999). For the purpose of protecting public health, this study utilised the most conservative figure available, which is that 95% of TC are *E. coli*. Microorganism populations are known to follow the lognormal distribution in nature (McBride, 2009). The pathogenic *E. coli* concentration data were therefore transformed using the log₁₀ scale before being fitted with a normal distribution.

The research proceeded in two phases. The first phase of the research focused specifically on the impact of assumed sampling regimes (weekly, fortnightly or monthly) and seasonal variability in characterising pond performance as measured by effluent concentrations and LRVs. The second phase focused on estimating potential health risks associated with the hypothetical scenario of irrigating lettuce with treated wastewater using QMRA using the results from the first phase of the research as input.

3.2. Phase 1: The specific research questions developed for this initial phase were as follows:

1. What impact does sampling design have on characterising maturation pond performance, as measured by pathogen log reduction values?
2. To what extent does the treatment efficacy of maturation ponds vary between seasons, with regard to pathogen log reduction values?

The following steps were undertaken to explore the impact of different sampling designs and seasonal variability in turn

1. Normal distributions were fitted using @Risk (v 6.3.1; Palisade Corporation)
2. Distributions of LRV's were generated by subtracting outlet distributions from inlet distributions and running Monte Carlo simulations.
3. Analysis of variance (ANOVA) was then conducted to test whether any of the mean LRVs were significantly different. Tukey's Honestly Significant Difference (HSD) multiple comparison of means post-hoc tests were then performed where applicable

3.3. Phase 2: The following steps were used to model estimated levels of health risk

Exposure assessment: The model input parameters were used in conjunction with the recommended equation (Equation 1) to calculate the daily dose of pathogenic *E. coli* as a result of consuming the irrigated lettuce.

$$\lambda_k = M_i M_{\text{body}} c_{\text{iw}} V_{\text{prod}} e^{(-kt)} \quad (1)$$

where:

λ_k = *E. coli* dose ingested/person/day (CFU.mL⁻¹)

M_i = daily consumption per person per kg of body mass [g (kg/person/day)⁻¹];

M_{body} = human body mass (kg);

c_{iw} = concentration of pathogenic *E. coli* in the irrigation water (CFU mL⁻¹);

V_{prod} = volume of irrigation water caught by product (mL g⁻¹);

k = *E. coli* O157:H7 kinetic decay constant (day⁻¹); and

t = time between last reclaimed-water irrigation event and harvest/consumption/storage (*i.e.*, length of environmental exposure) (days).

Dose-response modelling: Dose-response modelling was performed using a 'piece-wise' dose-response approximation of the Beta Poisson model (Equations 2 and 3).

$$P_i(d) = 1 - \left(1 + \frac{d}{1.281}\right)^{-0.06699} \quad \text{for } d \leq 7; \quad (2)$$

$$P_i(d) = 1 - \left(1 + \frac{d}{0.6112}\right)^{-0.0502} \quad \text{for } d > 7, \quad (3)$$

where:

$P_i(d)$ = the probability of infection given a specific dose;

Note: d did not exceed 7 for the given scenario therefore only Equation 2 was needed

Risk Characterisation: The probability of infection per person per year (pppy) was calculated (Equation 4)

$$P_{\text{annual}} = 1 - (1 - P_i(d))^n \quad (4)$$

Where:

P_{annual} = the probability of infection pppy; $P_i(d)$ = the probability of infection from a single exposure;

d = the mean dose level; n = the frequency of consumption per year (WHO, 2006).

Monte Carlo simulation was used to generate probabilistic distributions for each of the steps above (10, 000 iterations per simulation). Mean, median and 95th percentile values of the health risk estimates were calculated. The built-in sensitivity analysis function in @Risk was used to rank the input parameters by their effect on the estimated health risk levels (*i.e.*, the probability of infection pppy).

Table 1. QMRA model parameters for estimating health risks associated with the consumption of lettuce irrigated with treated wastewater.

Model Parameter	Units	Distribution type (parameter) [mean]	References
M_{body} = Body mass	Kg-person	Lognormal (78, 13.5) [78.00]	(enHealth, 2012; Hamilton et al., 2006)

Model Parameter	Units	Distribution type (parameter) [mean]	References
M_i = Daily Consumption	g/kg/c/d	Lognormal (0.3994, 1.429) [0.40]	(AusVeg, 2011; Mok et al., 2014)
C_{iw} = Concentration of pathogenic <i>E. coli</i> in recycled wastewater	CFU.mL ⁻¹	Lognormal (normal fitted to log ₁₀ transformed data) Summer: RiskNormal (0.64352, 0.41912) Spring: RiskNormal (-0.4752, 0.7927)	Estimated from empirical Thermotolerant coliform data
V_{prod} = Volume of irrigation water caught by lettuce	mL. g ⁻¹	Lognormal3 (0.01049, 0.00559, 0.006) [1.05 x 10 ⁻²]	(Mok & Hamilton, 2014)
k = In-field <i>E. coli</i> decay constant	Day ⁻¹	Uniform (0.22, 2.61)	(Sjolander, 2012)
t = time between last irrigation event and consumption	Days	Uniform (0,2) [1.00]	(Mok et al., 2014)
<i>E. coli</i> dose-response parameters	$P_i(\lambda)$	$\alpha = 0.06699, \beta = 1.281$ (as $d \leq 7$)	

4. RESULTS

Distributions of LRVs modelled for datasets representing weekly, fortnightly and monthly sampling are presented in Figure 2. There were no statistically significant differences between any of the means of these distributions (ANOVA, $F_{2,281}=0.152, p=0.858$). The high p-value ($>>0.05$) implied insufficient evidence to reject the null hypothesis that all mean pathogen LRVs estimated by different data subsets were equal, and therefore the null hypothesis was accepted and no post-hoc tests were conducted.

Clear seasonal differences in pathogenic *E. coli* concentrations between the influent and effluent can be seen in Figure 3. ANOVA results revealed statistically significant differences ($F_{3,162}=23.32, p<0.001$) and pair-wise post-hoc comparisons of the seasonal means showed significant differences between all pairs ($p<0.05$) except for winter and autumn ($p=0.996$).

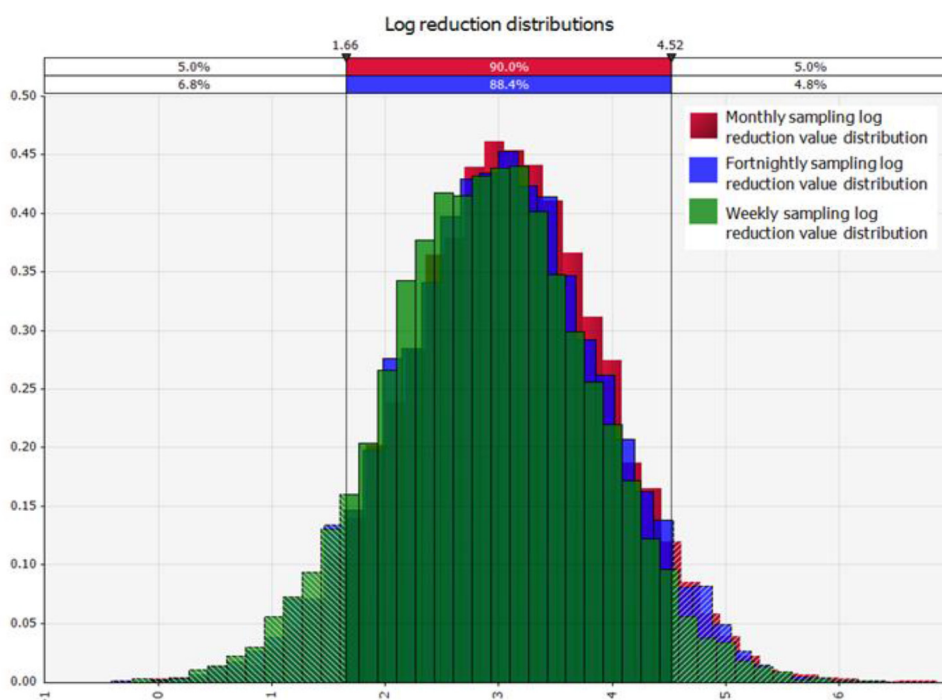


Figure 2. Comparison of estimated pathogen log reduction value distributions based on weekly, fortnightly and monthly sampling of maturation ponds.

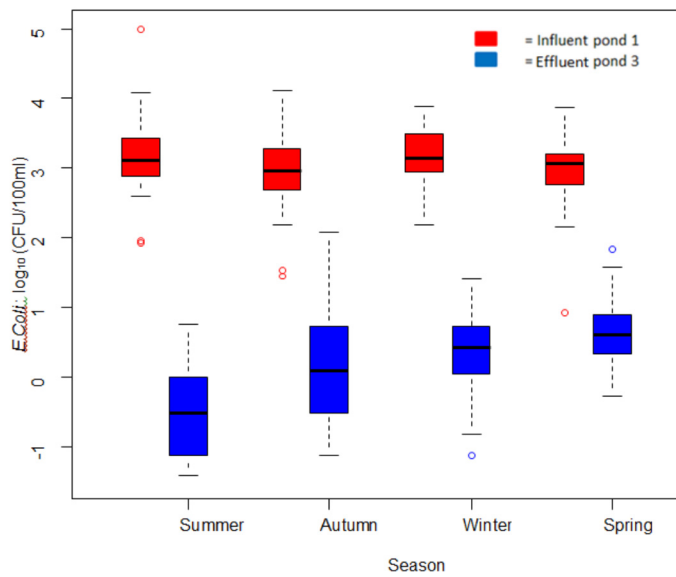


Figure 3. Seasonal variability of pathogenic *E. coli* concentrations in the case study maturation pond system.

Based on the distributions of the estimated probabilities of infection pppy, summer and spring represented the best and worst case scenarios respectively. The 95th percentile value for the summer probability of infection pppy was 0.06; much lower than the corresponding value for spring (0.31 infections pppy).

The tornado graphs shown in Figures 4 and 5 illustrate the differential sensitivity of the seasonal models predicting 95th percentile probabilities of infection pppy to the input parameters. There are two key differences. First, the spring model showed greater sensitivity to all input parameters compared with the summer model. Second, the summer model was most sensitive to variations in estimates of effluent pathogen concentrations while the spring model was most sensitive to estimates of daily consumption of lettuce.

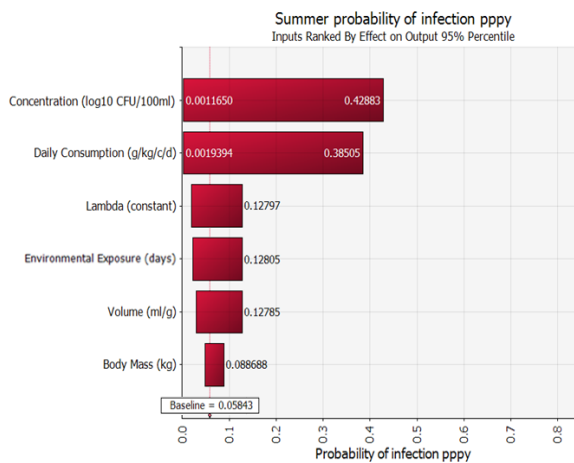


Figure 4. Model sensitivity to the input parameters for summer (predicting 95th percentile probability of infection per person per year).

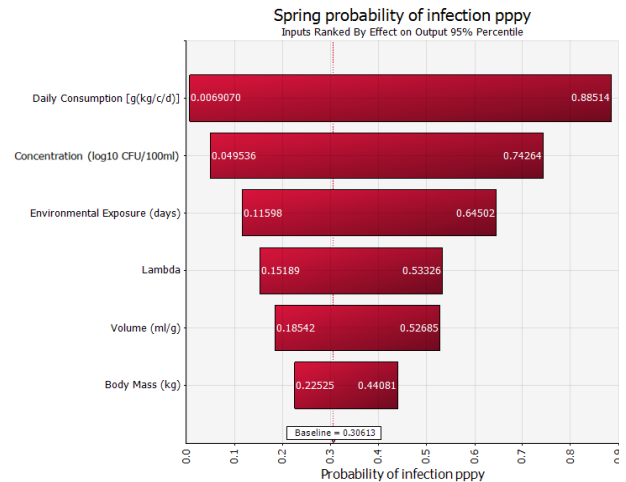


Figure 5. Model sensitivity to the input parameters for spring (predicting 95th percentile probability of infection per person per year).

5. DISCUSSION AND CONCLUSION

Treated wastewater is a potential resource, especially in areas such as regional and remote Australia where water scarcity is becoming an increasing challenge. WSPs provide a treatment option that is suitable for regional and remote areas due to their low cost and low maintenance despite their large land requirements. Treated wastewater from WSPs, however, cannot be reused until issues such as health risks have been considered adequately. Pathogens are of primary interest with respect to assessing the health risks of recycled water. It is known that microbial concentrations are highly variable both temporally and spatially in wastewater, yet the lack of attention to this variability and its impact on the characterisation of microbial water quality is a source of frustration to regulators evaluating QMRAs (D. Cunliffe, personal communication, October 29, 2014).

This study showed distinct benefits of stochastic modelling (through Monte Carlo simulation) for those involved in managing wastewater treated through maturation pond systems and its reuse. Essentially, these benefits can be grouped under two broad categories: those that ensue from being able to characterise the

temporal variability of the microbial water quality and those related to more transparent health risk modelling in relation to treated effluent reuse.

1. Allows WWTP operators to explore the extent to which monitoring programs can be rationalised without compromising their rigour (with respect to characterising microbial water quality) while still meeting legislative requirements. For example, it may be of strategic value to reduce the frequency of sampling during those seasons when variability is low to enable more intensive sampling, perhaps with replicates, during those periods exhibiting higher variability.
2. Extends the exploration of the impact of observed temporal variability of microbial water quality to estimations of health risks, ensuring more rigour.
3. It is useful for exploring the sensitivity of health risk estimates to all parameters utilised in the QMRA thereby providing guidance as to where resources should be focused.
4. The simulation capacity in @Risk facilitates a range of “what if” scenarios of interest to be explored (*e.g.*, different combinations of pathogens, reuse options and target populations)
5. The graphical outputs from modelling platforms such as @Risk facilitate more effective communication with stakeholders. For example, the tornado graphs clearly illustrate the parameters that have the most impact on the health risk estimates.

In summary, this study has shown that stochastic modelling and simulation enhances the investigative capacity of those involved in evaluating the performance of highly variable ecological systems such as maturation ponds. The same modelling techniques could also be applied to evaluate a broader range of exposure scenarios than the single scenario of irrigating lettuce considered here, as well as additional onsite treatment steps which may be needed to facilitate desired reuse options. Through explicating sources and implications of uncertainty and variability in this type of health risk modelling, this study cautions against relying too much on QMRA results where uncertainties have not been made explicit.

REFERENCES

- Cullen, A.C. & Frey, C.H. (1999). Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. New York, NY: Plenum Press.
- Elmund, G.K., Allen, M.J., Rice, E.W. (1999). Comparison of Escherichia coli, Total Coliform, and Fecal Coliform Populations as Indicators of Wastewater Treatment Efficiency. *Water Environment Research*, 71(3), 332-339.
- enHealth. (2012). Australian Exposure Factor Guide: Guidelines for assessing human health risks from environmental hazards. Canberra, Australia: Commonwealth of Australia.
- Haas, C. N., Rose, J. B., & Gerba, C. P. (2014). Quantitative Microbial Risk Assessment (2nd Ed.). New York: John Wiley & Sons, INC.
- Hamilton, A. J., Stagnitti, F., Premier, R., Boland, A. M., & Hale, G. (2006). Quantitative microbial risk assessment models for consumption of raw vegetables irrigated with reclaimed water. *Appl Environ Microbiol*, 72(5), 3284-3290. doi: 10.1128/AEM.72.5.3284-3290.2006
- McBride, G.B. (2009). Using Statistical Methods for Water Quality Management: Issues, Problems and Solutions. Hoboken, New Jersey: John Wiley & Sons Inc.
- Mok, H. F., Barker, S. F., & Hamilton, A. J. (2014). A probabilistic quantitative microbial risk assessment model of norovirus disease burden from wastewater irrigation of vegetables in Shepparton, Australia. *Water Res*, 54, 347-362. doi: 10.1016/j.watres.2014.01.060
- NRMMC, EPHC & AHMC. (2006). Australian Guidelines for Water Recycling: Managing Health And Environmental Risks (Phase1). Natural Resource Management Ministerial Council, Environmental Protection and Heritage Council and the Australian Health Ministers' Conference, Canberra, Australia.
- Toze, S. (2006). Reuse of effluent water – benefits and risks. *Agricultural Water Management*, 80, 147-159. doi:10.1016/j.agwat.2005.07.010
- U.S. Environmental Protection Agency (US EPA). (2012). Guidelines for Water Reuse. U.S. Agency for International Development, Washington: U.S.A.
- Westrell, T., Schönning, C., Stenström, T.-A., & Ashbolt, N. (2004). QMRA (quantitative microbial risk assessment) and HACCP (hazard analysis and critical control points) for management of pathogens in wastewater and sewage sludge treatment and reuse. *Water Science & Technology*, 50(2), 23-30.