WATER CLARITY-BASED PREDICTION OF E.COLI CONCENTRATIONS IN NEW ZEALAND RIVERS

Christopher Dada – Streamlined Environmental Ltd, PO Box 7003, Hamilton East, 3247 - chris@streamlined.co.nz

ABSTRACT

Regulatory authorities in New Zealand have communication strategies as part of their monitoring programmes to ensure the public is informed of a health risk at a swimming site, for example through the release of swimming advisories when *E.coli* levels exceed single sample bathing water criterions. By the time the swimming advisories are released, however, they are typically at least a day or two late, due to the 24-48h turn-around time before results from culture-based microbiological analysis are available. Methods are thus needed that improve the timeliness and accuracy of recreational water quality risk assessments. An important strategy until reliable continuous monitoring is in place, is to combine existing monitoring programmes with predictive models. While predictive faecal indicator bacteria (FIB) models have been used to estimate bacteriological water quality at some swimming sites, these models are largely 'top-down' in their approach to safeguarding public health. Beyond being simply 'advised when to avoid swimming', there is an increasing awareness amongst the general public regarding the role they can play in water quality monitoring. This presents novel opportunities for citizen participation in predictive FIB monitoring and modeling. This study reports on the possibility of developing intuitive, public-friendly models that are based on the physical appearance of water (clarity) as a predictive variable in estimating the *E.coli* concentrations in rivers, and to assess if water is safe to swim in. The goal of this study was to evaluate the possibility of water clarity-based *E.coli* models for now-cast prediction, at local and national scales in New Zealand.

Using an easily measured parameter (water clarity), this study calibrated and validated models that could be used by anyone, without the need for specialist technical knowledge. The models allow the user to assess whether or not it is safe to swim in their local waterways. At a national scale, the applicability of water clarity as a surrogate for *E.coli* concentration was also assessed using a total of 8103 *E. coli* datasets that have been routinely collected over the past two decades by regional authorities, for most New Zealand rivers and tributaries. Our results show that if swimmers were to avoid river waters with <1.1 m black disc visibility during autumn and summer or river waters with <0.5 during spring and winter, they would also avoid microbial hazards that are associated with exceedances of the 540 CFU/100 mL single sample bathing water standard. Regardless of the climatic season modelled, the clarity-based *E.coli* models performed well as they presented with sensitivity, specificity and accuracy values of at least 73%. The developed models offer the benefit of providing a faster method for estimating *E. coli* concentration, potentially engaging the public in water monitoring, and allowing them to make informed decisions on whether it is safe to swim at their favourite swimming spot.

KEYWORDS

Faecal indicator bacteria, water clarity, predictive model, water quality, public health

PRESENTER PROFILE

Chris Dada specializes in the fate, transport, detection, and control of pathogens in environmental media. He previously completed a Masters degree in Water Policy at Oxford University's Center for the Environment. His PhD research focused on faecal indicator bacteria and antibiotic resistant pathogens in aquatic environments.

1 INTRODUCTION

Regulatory authorities in New Zealand have communication strategies as part of their monitoring programmes to ensure the public is informed of a health risk at a beach or river; for example, through the release of swimming advisories when *E.coli* levels exceed single sample bathing water criterions. By the time the swimming advisories are released, however, they are typically at least a day or two late, due to the 24-48h turn-around time before results from culture-based microbiological analysis are available (Dada and Hamilton, 2016). Methods are thus needed that improve the timeliness and accuracy of recreational water quality risk assessments. An important strategy until reliable continuous monitoring is in place, is to combine existing monitoring programmes with predictive models.

Predictive FIB models provide a rapid estimation of the bacteriological condition, potentially assisting local beach managers in the decision process related to swimming advisories issuance. In recent years, many beach managers have increasingly adopted predictive tools, of which the most widely applied are models developed through multi-variable linear regression (e.g., Olyphant, 2005; Nevers and Whitman, 2005, Feng et al 2015). Process-based models, which couple hydrodynamic models with a microbe transport-fate model involving microbial loading, transport, and fate processes, have also been demonstrated to make predictions (e.g., Hipsey et al., 2008; Feng et al., 2015).

While these predictive faecal indicator bacteria (FIB) models have been used to estimate bacteriological water-quality, they have largely adopted a 'top-down' in their approach to safeguarding public health (i.e. models are by science staff of regulatory bodies who simply advise the public when its safe or not to swim). Beyond being simply 'advised when to avoid swimming', there is an increasing awareness amongst the general public regarding the role they can play in water quality monitoring. This presents novel opportunities for citizen participation in predictive FIB modeling. This study presents a classic example of developing intuitive, 'public friendly' and 'public-usable' models, using the physical appearance of water (as measured by water clarity) as a way of estimating *E.coli* concentrations in surface water, to assess if water is safe to swim in. The goal of this study was to evaluate the possibility of using clarity models for *E.coli* nowcast prediction, at both a local and national scale throughout New Zealand. This will constitute a milestone in efforts geared towards developing and deploying site-specific river clarity based *E.coli* models at local scales for nowcast prediction of *E.coli* concentrations at popular recreational sites in New Zealand.

2 METHODS

2.1 STUDY SITES

A total of 145,040 water quality datasets which has been routinely collected by regional authorities from as early as the late 1980s for most New Zealand rivers and tributaries (https://data.mfe.govt.nz/), was used in the analysis. This dataset contained measured values for several parameters including ammoniacal nitrogen, total nitrogen, nitrate-nitrogen, dissolved reactive phosphorus, total phosphorus, and *E.coli*. All *E.coli* datasets were extracted (n=8170). Among these, a total of 8103 *E. coli* datasets which had corresponding discharge data were subsequently used for the analysis. *E.coli* data used thus spanned the period 2005 to 2013 at a total of 77 freshwater swimming sites representing 49 rivers and tributaries throughout New Zealand (Figure 1).

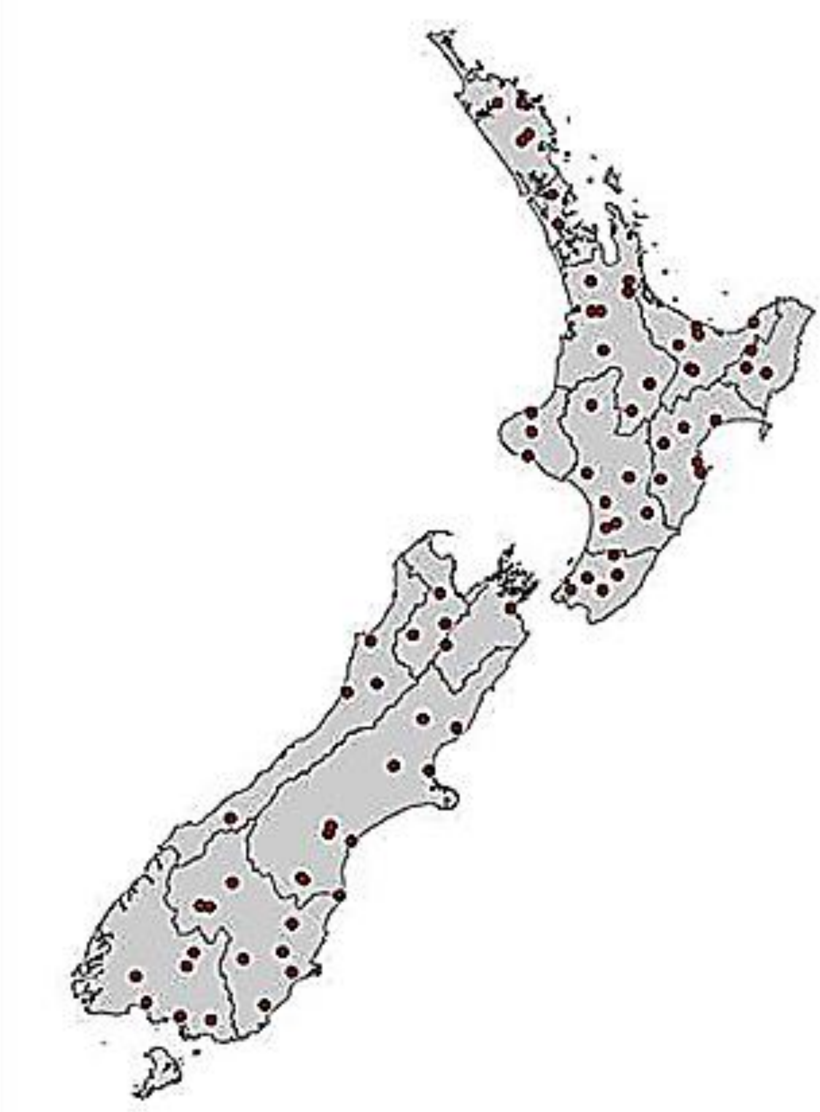

Figure 1: 77 New Zealand freshwater swimming sites (49 rivers and tributaries) in the *E.coli* predictive modeling using water clarity as a predictor variable.

2.2 DATA MANAGEMENT, STATISTICAL ANALYSIS, AND MODELING

To fit the 8103 *E.coli* datasets based on their clarity, a gradient approach was used in which incremental 'trigger' values or water clarity 'thresholds' were applied. Incremental 'trigger' values or water clarity 'thresholds' (i.e. from lowest to highest) were applied as 'thresholds' to predict exceedances and non-exceedances of the national bathing water standard. These triggers or thresholds are water clarity values that would warrant additional site-based investigation, as they are indicative of conditions of elevated faecal indicator bacteria levels higher than the national bathing water standard of 540 CFU/100mL.

2.3 MODEL PERFORMANCE AND SWIMMING ADVISORY ASSESSMENT

Exceedances of bathing water thresholds applied in this study were compared against national guidelines. An exceedance (or a positive model outcome) was recorded when sampled or predicted *E. coli* levels exceeded the single sample bathing water standard (BWS) of 540 CFU/100 mL (as stipulated in the New Zealand National Policy Statement for Freshwater Management) (NPS, 2014). A type I error (or a false positive outcome) was identified when the modeled *E. coli* level was above the thresholds, but the observed *E. coli level* was below the thresholds. When the modeled and observed *E. coli* levels were both above the thresholds, this was considered a true positive. On the other hand, a false negative result (type II error) was inferred when the modeled *E. coli* level was less than the thresholds but the observed *E. coli* level was higher. In such a case, potential microbial contamination would be undetected by the model and no swimming advisory would be issued. When the modeled and observed *E. coli* levels are both below the thresholds, this is identified as a true negative. Model accuracy is the percentage of correct advisory predictions. Sensitivity and specificity are defined as the rates of correctly predicted exceedances and non-exceedances, respectively. Specificity, sensitivity, and accuracy of the model were determined using the following equations:

$$Model\ Sensitivity\ (\%) = \frac{\text{Number of True Positives} * 100}{(\text{Number of True Positives} + \text{Number of False Negatives})}$$

$$Model\ Specificity\ (\%) = \frac{\text{Number of True Negatives} * 100}{(\text{Number of True Negatives} + \text{Number of False Positives})}$$

$$Model\ Accuracy\ (\%) = \frac{(\text{Number of True Negatives} + \text{Number of True Positives}) * 100}{(\text{Total number of observations})}$$

Receiver operating characteristic (ROC) curves, i.e. super-imposed plots of sensitivity, specificity and accuracy were used to determine the crossover (i.e. optimum trigger water clarity) values. This crossover value is the optimum decision threshold where the maximum number of exceedances are correctly identified and is a reasonable trade-off between sensitivity, specificity, and accuracy (Arad, Housh, Perelman, & Ostfeld, 2013).

3 RESULTS AND DISCUSSION

A correlational analysis of 8103 nation-wide *E.coli* and clarity datasets indicate that river *E.coli* concentrations was inversely proportional to river water clarity, with a simply fitted spline accounting for more than 60% of the variability in the national *E.coli* dataset (Figure 2).

Incremental water clarity values (i.e. from lowest to highest) were applied as 'thresholds' to predict exceedances and non-exceedances of the national bathing water standard. Based on this approach, Table 1 summarizes the model performance data obtained for different scenarios of climatic season. With increasing water clarity 'trigger value', sensitivity of the model increases, i.e. increase in the proportion of correctly predicted true exceedances but a concomitant reduction in the specificity of the model, i.e. decreases in the proportion of correctly predicted BWS non-exceedances. Regardless of the season, the *E.coli* clarity models performed well as they presented with at least 73% sensitivity, specificity and accuracy at the crossover value (see Table 1 and Figure 3).

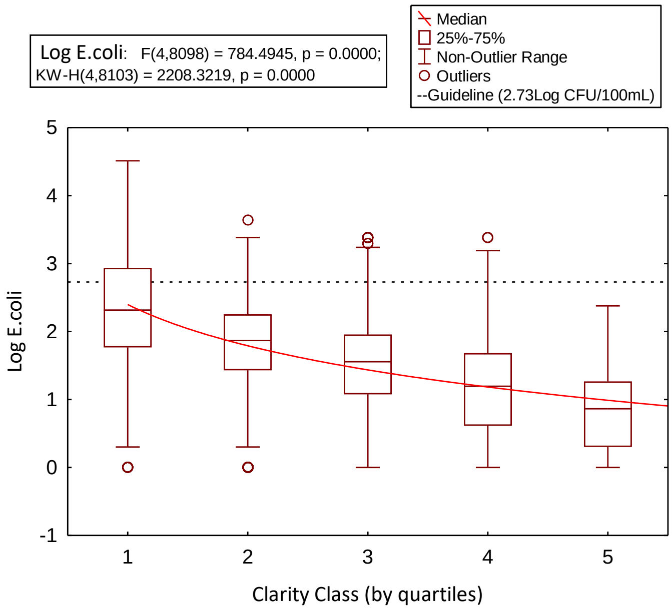

Figure 2: Plots of river E.coli concentrations versus water clarity.

Table 1 Model performance of seasonal water clarity trigger values used to fit exceedances and non-exceedances of 8103 nationwide E.coli data.

Parameter	Summer	Autumn	Winter	Spring	All seasons
Trigger Value (m)	1.1	1.1	0.5	0.6	0.8
True Exceedances of BWS	133	191	195	188	698
Total number of exceedances	181	242	231	234	887
Total number of observations	1952	2064	2043	2044	8103
Sensitivity (%)	73.48	78.93	84.42	80.34	78.69
Specificity (%)	76.76	79.36	79.08	80.22	77.63
Accuracy (%)	76.46	79.31	79.69	80.23	77.75

Trigger values were obtained from cross over plots for each scenario of climatic season. Performance of the classification scheme was assessed against a single sample BWS of 540 CFU/100mL as in the New Zealand National Policy Statement on Freshwater Management (NPS 2014).

During summer and autumn, the crossover (trigger) value was observed to be 1.1m (Table 1) i.e. on a nation-wide scale, there is a high likelihood that elevated levels of faecal indicator bacteria, above the *E.coli* BWS, would be present in rivers when the stream or river water clarity was lower than 1.1m. At this trigger value, 324 out of the 423 total *E.coli* BWS exceedances observed in the summers and autumn of the 9-year period were correctly predicted (i.e. 133 +191, see Table 1).

During winter and spring, the crossover (trigger) value was observed to be lower, 0.5m and 0.6m respectively (Table 1), i.e. on a nation-wide scale, during these seasons, there is a high likelihood that elevated levels of faecal indicator bacteria, above the *E.coli* BWS, would be present in rivers when the stream or river water clarity was lower than 0.6m. At this trigger value, at least 383 out of the 465 exceedances observed in the winters and springs over the 9-year period were correctly predicted (i.e. 195+198, see Table 1). This trigger value also correctly predicted a high proportion of the non-exceedances observed in the winters and springs of the 9-year period, with a minimum specificity of 79% (Table

1). Regardless of the climatic season considered in this study, the *E.coli* clarity models performed well as they presented with sensitivity, specificity and accuracy that ranged between 72.82% and 100% (Table 1).

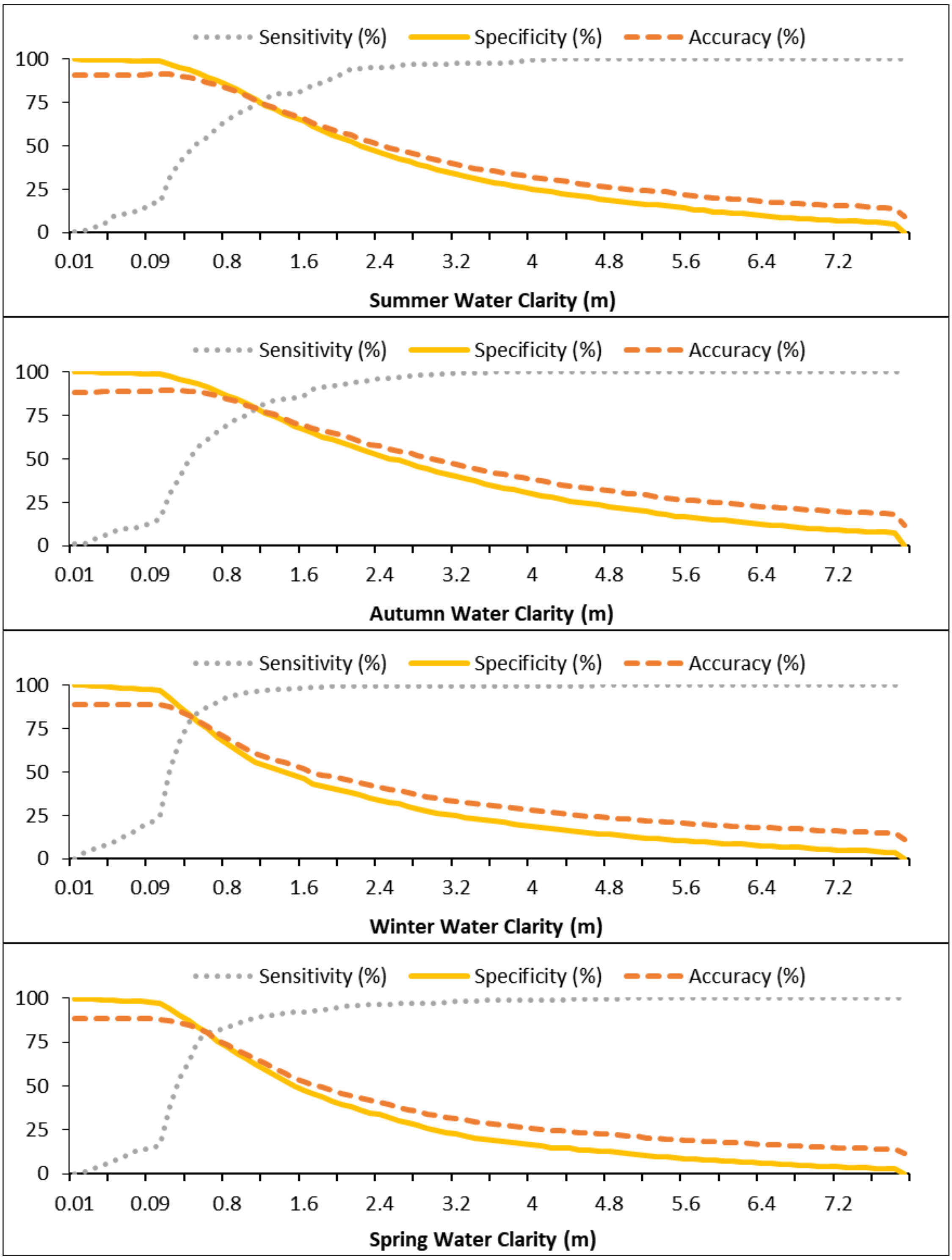

Figure 3: Receiver operating characteristic (ROC) plots of sensitivity, specificity, and accuracy versus incremental water clarity 'trigger' values that predicts E.coli BWS exceedances at 77 freshwater sites nationwide during different seasons. Performance of the classification scheme was assessed against a BWS of 540CFU/100mL as in the New Zealand National Policy Statement on Freshwater Management (2014).

Our results show that if swimmers were to avoid river waters with <1.1 m black disc visibility during autumn and summer or river waters with <0.5 during spring and winter, they would also avoid microbial hazards that are associated with exceedances of the 540 CFU/100 mL single sample bathing water standard. These water clarity thresholds could then be used by relevant authorities to build an early warning system which could be communicated to the public. This could result in warnings like, 'if you cannot see your feet in an ankle-deep water, don't bother swimming'. It is important to note that the 1.6m default water clarity guideline currently applied in New Zealand as indicative of safe swimming conditions, does not translate into safety, from a microbiological hazard perspective. Instead, it is more around water safety (seeing the bottom and obstructions) rather than a human health indicator. For instance, based on Figure 3, applying a 1.6m threshold will predict that a stream is safe for recreation when it actually contains *E.coli* at concentrations above the bathing water standard (almost half of the predicted times, based on the comparatively lower model specificity and accuracy associated with this threshold.

Although previous articles (Collins, 2003; Davies-Colley et al, 2018) reported on the use of water clarity to estimate *E.coli* concentrations, there was no consideration for the differences in the ability of water clarity to predict *E.coli* concentrations during different seasons, land use, geology, and stream order classifications. Our study thus advances the hypothesis reported in the Davies-Colley et al (2018) study by showing that water clarity thresholds applied to predict *E.coli* concentrations differ across different climatic seasons. Our study also advances the hypothesis reported in the Davies-Colley et al (2018) study by applying receiver operating characteristic (ROC) curves to optimize the determination of water clarity 'thresholds' that predict exceedances of the bathing water standard (see Figure 3). This ROC approach guides a selection of an optimum decision threshold where the maximum number of exceedances are correctly identified and is a reasonable trade-off between sensitivity, specificity, and accuracy (Arad, Housh, Perelman, & Ostfeld, 2013).

We note however that there are potential limitations to the modelling approach in our study. For instance, while our focus was to model the relationship between water clarity and faecal indicator bacteria, it does not differentiate between contributions of faecal bacteria from sediment bed and from the watershed. There was also no delineation between free and particle-associated faecal bacteria. In the future, sediment deposition and resuspension fluxes of faecal bacteria across the sediment bed–water interface at river-specific levels could be incorporated into the model. Based on this, it would be possible to apply the model to hypothetical scenarios that can potentially evaluate the impact of varying catchment management conditions as well as settling and resuspension conditions on E.coli concentrations observable in the water column.

Typically, bacteriological water quality at designated sampling sites are thought to be representative of that particular water body. However, considerable spatial variability has been documented over scales of 10 m and more (Schang et al, 2018; Boehm et al 2009). It is thus hoped that future studies in New Zealand will combine *E.coli* and water clarity data with geographic information systems in a way that dynamically captures both spatial and temporal dimensions. A similar approach was adopted by Money et al. (2009) in a study that combined *E.coli* and turbidity data in a river-based space/time geostatistical framework for basin-wide assessment of faecal contamination. This can harness the power of aerial photography and satellite-based remote sensing to provide real-time aerial estimates of water clarity-based prediction of *E.coli* conditions.

4 CONCLUSION

In this study, direct negative correlation between water clarity and FIB concentrations was observed for most major rivers and tributaries in New Zealand. This relationship was used to develop predictive models that can produce estimates of *E.coli* concentrations before waiting for the 24-48-hour reporting time that conventional monitoring procedures require. Results show that if swimmers were to avoid river waters with <1.1 m black disc visibility during autumn and summer or river waters with <0.5 during spring and winter, they would also avoid microbial hazards that are associated with exceedances of the 540 CFU/100 mL single sample bathing water standard. Water clarity trigger values defined by the model can be used by authorities to alert recreational users of possible high faecal bacteria values. The developed models can provide a faster estimation of *E. coli* concentrations, allowing the public to engage in water quality monitoring, and also to make informed decisions on whether it is safe to swim at their favourite swimming spot.

ACKNOWLEDGMENTS

Thanks to Ngaire Phillips for assisting with manuscript review.

REFERENCES

Arad, J., Housh, M., Perelman, L. and Ostfeld, A. (2013). A dynamic thresholds scheme for contaminant event detection in water distribution systems. *Water Research*, 47, 5, 1899–1908.

Boehm, A. B., Ashbolt, N. J., Colford, J. M., Dunbar, L. E., Fleming, L. E., Gold, M. A., ... & Soller, J. A. (2009). A sea change ahead for recreational water quality criteria. *Journal of Water and Health*, 7(1), 9-20.

Collins, R. (2003). Relationships between streamwater *E.coli* concentrations and environmental factors in New Zealand. In *Diffuse Pollution, 3:170-180.*.

Dada, A.C. and Hamilton, D.P. (2016). Predictive Models for Determination of *E.coli* Concentrations at Inland Recreational Beaches. *Water, Air and Soil Pollution*, 227, 9, 347.

Davies-Colley, R., Valois, A., & Milne, J. (2018). Faecal pollution and visual clarity in New zealand rivers: Correlation of key variables affecting swimming suitability. *Journal of Water and Health*, wh2018214.

Feng, Z., Reniers, A., Haus, B. K., Solo-Gabriele, H. M.,Wang, J. D., & Fleming, L. E. (2015). A predictive model for microbial counts on beaches where intertidal sand is the primary source. *Marine Pollution Bulletin*, 94(1), 37–47.

Hipsey, M. R., Brookes, J. D., Regel, R. H., Antenucci, J. P., & Burch, M. D. (2006). In situ evidence for the association of total coliforms and Escherichia coli with suspended inorganic particles in an Australian reservoir. *Water, Air, and Soil Pollution*, 170(1–4), 191–209.

Money, E. S., Carter, G. P., & Serre, M. L. (2009). Modern space/time geostatistics using river distances: data integration of turbidity and E. coli measurements to assess fecal contamination along the Raritan River in New Jersey. *Environmental science & technology*, 43(10), 3736-3742.

Nevers, M. B., & Whitman, R. L. (2005). Nowcast modeling of Escherichia coli concentrations at multiple urban beaches of southern Lake Michigan. *Water Research,* 39(20), 5250–5260.

NPS. (2014). National Policy Statement for Freshwater Management 2014. Retrieved from www.mfe.govt.nz/publications/fresh-water/national-policy-statement-freshwater-management-2014

Olyphant, G. A. (2005). Statistical basis for predicting the need for bacterially induced beach closures: emergence of a paradigm? *Water Research*, 39(20), 4953–4960.

Schang, C., Lintern, A., Cook, P. L. M., Rooney, G., Coleman, R., Murphy, H. M., ... & McCarthy, D. (2018) Escherichia coli survival and transfer in estuarine bed sediments. *River Research and Applications*. https://doi.org/10.1002/rra.3281