Building Tomorrow's Society
Bâtir la Société de Demain

Fredericton, Canada
June 13 – June 16, 2018/ *Juin 13 – Juin 16, 2018*

PREDICTIVE MODELLING OF RECREATIONAL WATER QUALITY AT TWO BEACHES IN WINDSOR ESSEX REGION, ONTARIO, CANADA

Madani, Mohammad[1], Seth, Rajesh[1,2]
[1] Department of Civil and Environmental Engineering, University of Windsor, Windsor, ON, Canada
[2] rseth@uwindsor.ca

1 INTRODUCTION

Swimming in contaminated waters may result in gastrointestinal and respiratory diseases. To protect bathers from swimming in such circumstances, microbial water quality is typically monitored for fecal indicator bacteria (FIB). Furthermore, in addition to the health effects, beach closure may have deep financial impacts as well. Annually this ranges from \$11.3M to \$117M in lost value for the Great Lake recreational swimmers for those days when swimming is banned (Shaikh 2006; Rabinovici et al. 2004). Although E. coli is considered the most suitable indicator for recreational water monitoring according to Guidelines for Canadian Recreational Water Quality, its measurement takes 18 to 24 h before results are available. However, water quality in near shore regions can change over a matter of hours (Myers et al., 1998; Boehm et al. 2002) so concentrations may change between the time of sampling and the reporting of results. Unsafe conditions are frequently announced late due to latencies in the E. coli measurement process. This process results in issuing closures based on previous day data rather than current water conditions. Developing rapid analytical methods which average two hours of laboratory analytical time is a possible solution to address this problem. Although these methods may be operationally available, but they need higher analytical cost than slower culture-based methods (Setty, 2012). Statistical models can be another solution which has been shown to also be an accurate approach (Nevers and Whitman 2005; Feng et al. 2015; Dada and Hamilton, 2016). In the late 1990s the first empirical models developed through statistical techniques such as Multiple Linear Regression (MLR), were used to reduce the risk of infection to users of recreational waters (USEPA, 1999). In Ohio, one-year data at three Lake Erie beaches was used to explore the effectiveness of predictive models (Francy and Darner, 1998). The simplest type of model, namely Rainfall-based alerts, have been used by several communities for a number of years (Francy D., 2009). Later, application of predictive models considering other water quality and hydro-metrological parameters on other beaches became more prevalent and many studies were conducted to develop such models for inland recreational lakes (Dada and Hamilton 2016; Olyphant and Whitman 2004; Nevers and Whitman 2005, 2011; Francy et al. 2013), coastal beaches (Thoe et al., 2012; Boehm et al., 2007) and reservoirs (Francy et al. 2013). Although some deterministic models were developed to predict the water quality and FIB concentration, these models are inferior to statistical models in most situations as they require the determination of FIB sources (Francy D., 2009). In the present study, collected water quality and weather data and their various transformations are examined to develop a multiple linear regression model using the United States Environmental Protection Agency (USEPA) Virtual Beach (VB) toolbox.

2 MATERIAL AND METHOD

Samples were collected on 30 consecutive days between the hours 8:00 and 12:00 (10 August 2010 to 8 September 2010) at Sandpoint beach (Lake St. Clair) and Holiday beach (Lake Erie) in the Windsor-Essex Region. Detailed descriptions of the analysis, field measured parameters and E. coli enumeration techniques were reported in previous work (McPhedran, 2013). Weather data was obtained from the nearest Environment Canada station for each beach. Data included air temperature (℃), daily rainfall (mm) and daily 10:00 am averages of hourly measurements of wind speed (m/s) and wind direction (degree). Minitab and XLSTAT were used to analyse the results. Briefly, to improve linear relations between E. coli concentrations and explanatory variables and also to take care of wide range of expected values, concentrations of E. coli were log10-transformed before any statistical testing and modeling (Francy et al., 2013). The candidate explanatory variables included: 24h and 48h antecedent cumulative rainfall, wind direction, wind speed, turbidity, conductivity, air and water temperature, dissolved oxygen, pH and conductivity. Model performance was examined by determination of metrics such as Root Mean Square Error (RMSE), Accuracy, Sensitivity and Specificity which are defined as:

[1] $RMSE = \sqrt{\sum_{i=1}^{N} \frac{(\log_{10} P_i - \log_{10} O_i)^2}{N}}$, $Sensitivity = \frac{TP}{(TP+FN)}$, $Specificity = \frac{TN}{(TN+FP)}$, $Accuracy = \frac{(TP+TN)}{N}$

Where N is number of observations, $\log_{10} P_i$ is the log10-transformed predicted model value, $\log_{10} O_i$ is the log10-transformed observation, and TP, TN, FP, and FN are numbers of true positives, true negatives, false positives (Type I error), and false negatives (Type II error) respectively.

3 RESULTS AND DISCUSSION

Time series illustration of observed and modeled E. coli count for the two beaches are shown in Figure 1 (a-b). Also, Figure 1 (c and d) show observed vs. model values of the advisory threshold for Sandpoint beach and Holiday beach respectively. During the study time, for Holiday beach, there were five EC exceedances (>100 CFU/100mL) while at Sandpoint beach, there were three EC exceedances. The non-parametric Mann-Kendall trend test reveals that there is no trend in the time series data (P-value of 0.28 and 0.97 for Sandpoint beach and Holiday beach respectively, $\alpha = 0.05$). Also, using the Durbin–Watson statistic ($\alpha = 0.05$), the selected variables were not autocorrelated ($D = 2.7$ and 2.3 for Sandpoint beach and Holiday beach respectively which are greater than $D_u = 1.93$). This check is important to make sure serial-correlation conditions which lead to improper estimation of the model are satisfactory to conduct MLR analysis (Ge and Frick, 2007; WYMER, 2007). The obtained RMSEs of 0.27 and 0.26 logCFU/100mL in Sandpoint beach and Holiday beach respectively are lower than the common range for MLR models reported in previous studies which were in the range of 0.4-0.5 logCFU/100 mL (Thoe et al., 2014). Virtual beach provides the top 10 best fit correlations, which based on cross-validation, the optimal parameters are those that minimize the mean squared error of prediction (MSEP).

Table 1: Summary of the model parameters

Beach Name	**Sandpoint Beach**	**Holiday Beach**
R2	0.78	0.89
RMSE	0.27	0.26
Accuracy (%)	96	96
Sensitivity	0.75	1.0
Specificity	1.0	0.96
Cross Validation MSEP	0.26-0.33	0.24-0.35
Correct Exceedance	3	5
Type I Error rate (%)	0	3.3
Type II Error rate (%)	3.3	0
Total Error rate (%)	3.3%	3.3%

Results show that although turbidity was the most important variable in VB models for both beaches, different combinations of other variables were found to be significant for each of the two beaches. While polynomial transformed turbidity, 24h rainfall and wind direction along shoreline were significant parameters in Sandpoint beach VB-Model, LOG(turbidity), wind direction, and INVERSE(wind_direction) were the most important variables for Holiday beach. Note that although EC levels at the study beaches can be impacted by a variety of other variables (humidity, cloud cover, animal and human sources etc.), these selected variables were reasonably satisfactory for building regression equations. Good correlation between the fitted VB models and observations were observed for both beaches (Figure 1). For the Sandpoint beach VB Model, the R^2 value was observed to be 0.78, sensitivity of 0.75, specificity of 1.0 and accuracy of 96%. Holiday beach VB model had higher R^2 of 0.89. Sensitivity, specificity and accuracy of model for Holiday beach are observed to be 1.0, 0.96 and 96% respectively. A summary of model evaluations for Sandpoint beach and Holiday beach, is shown in Table 1. This model resulted in 8 (22%) Type I and 5 (14%) Type II errors out of 30 testable outcomes at Sandpoint beach, and 3 (6%) Type I and 3 (6%) Type II errors out of 30 testable outcomes at Holiday Beach over the same time periods. 25% of the sample size (~ 8 days) and 1000 trails was set for cross-validation purpose. The best fit equation with the lowest MSEP is selected between top 10 results that provided by VB. Results of MSEP range for selected model are shown in the Table 1 for both beaches.

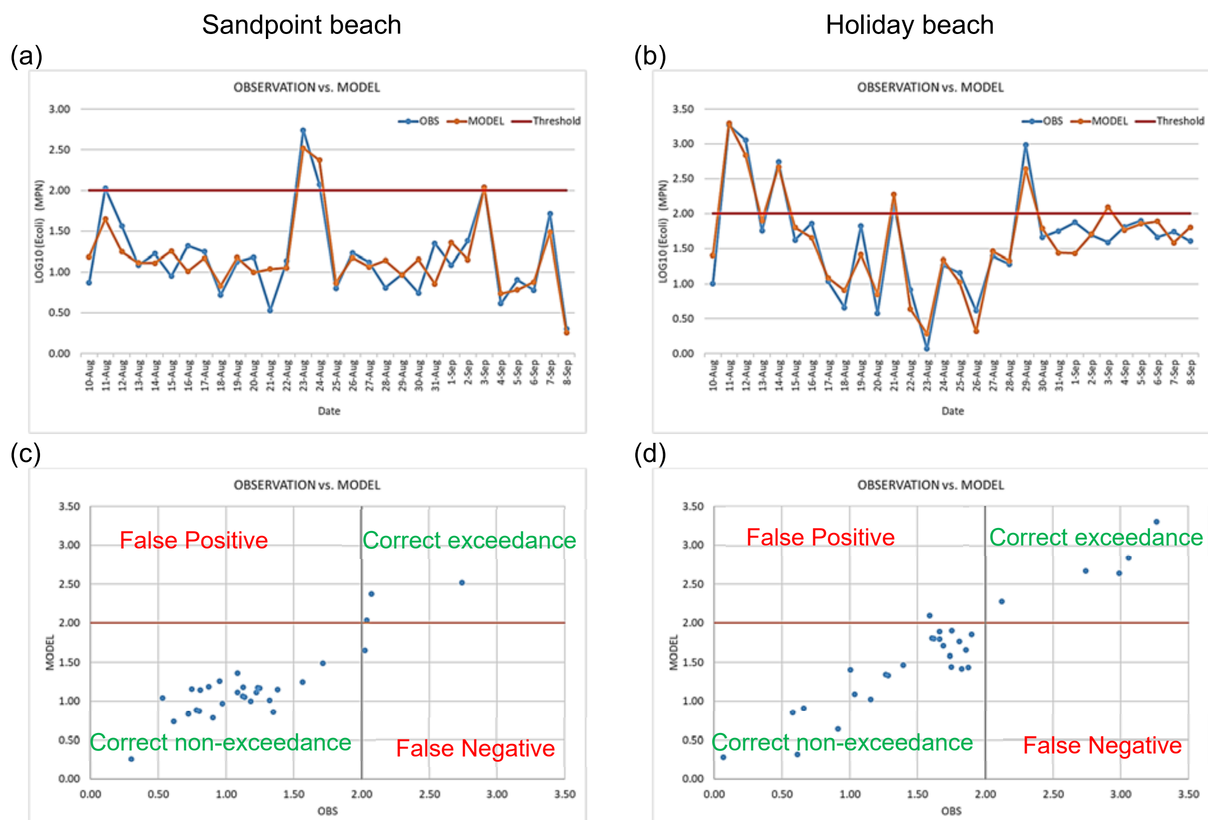

Figure 1: Actual vs predicted Log10-E. coli concentrations for Sandpoint beach and Holiday beach.

4 CONCLUSION

As a brief model interpretation, the structure of this model implies that approximately 78% of the variation of the E. coli concentration is accounted for by the variations of 24h rainfall, turbidity and wind direction for Sandpoint Beach. In the case of Holiday Beach, 96% of the variation is covered by explanatory variables such as turbidity, wind direction and water temperature. Since the variables are transformed, their effects are all nonlinear. Rainfall, which affects the transport of microbial pollution and spikes in turbid conditions, has an independent impact, while turbidity may be associated with the microbial build-up processes as a result of sediment resuspension that could influence E. coli concentrations into the water column. Storm events that usually coincide with high wind speed results in sediment resuspension as a result of wave/ current motions. In the case of Holiday Beach, rainfall is found to not be as effective as usually expected for these particular variables, which might be due to the beach geographical condition or sampling locations which need to be explored in more detail in future studies.

References

Boehm, A.B. 2007. Enterococci concentrations in diverse coastal environments exhibit extreme variability. *Environ. Sci. Technol.* **41**(24): 8227–8232.

Boehm, A. B.; Grant, S. B.; Kim, J. H.; Mowbray, S. L.; McGee, C. D.; Clark, C. D.; Foley, D. M.; Wellman, D. E. 2002. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. *Environ. Sci. Technol*, **36**: 3885–3892.

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

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**(1–2): 37–47.

Francy, D. S. 2009. Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches. *Aquatic Ecosystem Health and Management*, **12**(2): 177–182.

Francy, D.S.; Brady, A.M.G.; Carvin, R.B.; Corsi, S.R.; Fuller, L.M.; Harrison, J.H.; Hayhurst, B.A.; Lant, J.; Nevers, M.B.; Terrio, P.J.; and Zimmerman, T.M. 2013. Developing and implementing predictive models for estimating recreational water quality at Great Lakes beaches: *USGS Scientific Investigations Report* 2013–5166, 68

Francy, D.S.; Darner, R.A., 1998. Factors affecting Escherichia coli concentrations at Lake Erie public bathing beaches: *U.S. Geological Survey Water-Resources Investigations Report* 98-4241, 41

Ge, Z., Frick, W.E. 2007. Some statistical issues related to multiple linear regression modeling of beach bacteria concentrations. *Environ. Res*. **103**(3): 358–364.

McPhedran, K.; Seth, R.; Bejankiwar, R. 2013. Occurrence and predictive correlations of Escherichia coli and Enterococci at Sandpoint beach (Lake St Clair), Windsor, Ontario and Holiday beach (Lake Erie), Amherstburg, Ontario. *Water Quality Research Journal of Canada*, **48**(1): 99.

Myers, D.N., Koltun, G.F., and Francy, D.S. 1998. Effects of hydrologic, biologic, and environmental processes on sources and concentrations of fecal bacteria in the Cuyahoga River, with implications for management of recreation waters in Summit and Cuyahoga Counties, Ohio: *USGS Water-Resources Investigations Report* 98: 4089

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

Olyphant, G. A. and Whitman, R. L. 2004. Elements of a predictive model for determining beach closures on a real time basis: The case of 63rd Street Beach Chicago. *Environmental Monitoring and Assessment*, **98**(1–3): 175–190.

Rabinovici, S.J.M.; Bernknopf, R.L.; Wein, A.M.; Coursey, D.L.; Whitman, R.L. 2004. Economic and health risk trade-offs of swim closures at a Lake Michigan beach. *Environ. Sci. Technol*. **38**(10):2737–2745.

Setty, K. 2012. Adapting and adopting rapid molecular methods for beach water quality monitoring. *Urban Coast* **3**:48- 53.

Shaikh, S.L. 2006. Fact Sheet: The value of Chicago beaches, University of Chicago, Chicago, IL.

Thoe, W.; Gold, M.; Griesbach, A.; Grimmer, M.; Taggart, M. L. L.; Boehm, A. B. B. 2014. Predicting water quality at Santa Monica Beach: Evaluation of five different models for public notification of unsafe swimming conditions. *Water Research*, 67C, 105–117.

USEPA. 1999. Review of Potential Modeling Tools and Approaches to Support the BEACH Program. EPA-823-R-99-002. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.
WYMER, L. J. 2007. Statistical Framework for Recreational Water Quality Criteria and Monitoring. Presented at 7th Annual Great Lakes Beach Association Meeting, Cincinnati, OH, October 03 - 05,