



EPiC Series in Engineering

Volume 3, 2018, Pages 549–554

HIC 2018. 13th International
Conference on Hydroinformatics



Predicting Bacterial Levels in Recreational Beach Waters along U.S. Gulf Coast

Zhiqiang Deng^{1*} and Zaihong Zhang^{1†}

¹Louisiana State University, Baton Rouge, LA 70803-6405, U.S.A.
zdeng@lsu.edu, zaihongzhang1@gmail.com

Abstract

A Bayesian model was proposed for daily predictions of the probability of water quality standard violation of enterococci (ENT) levels in recreational beach waters. The Bayesian model consisted of a prior distribution and a likelihood function, which were constructed using seven years of environmental and bacteriological data collected at six recreational beach sites along the U.S. Gulf coast. The likelihood function followed best a normal distribution while the prior distribution was found to be best fitted with a Nakagami distribution. Modelling results showed that the Bayesian model was capable of explaining 86.13% of recreational water quality advisories issued by the U.S. Louisiana Beach Monitoring Program with a false positive rate of 7.24% and a false negative rate of 6.23%, indicating an excellent performance of the Bayesian model. The Bayesian model, presented in this paper, is unique and novel in terms of (1) the integration of a deterministic model and a probabilistic model for the prediction of recreation beach water quality; (2) the identification of important hydrodynamic processes controlling the source and transport of bacteria in coastal beach waters; and (3) the identification of key sinks of bacteria in coastal beach waters. It was found that the tidal washing process plays the most important role in causing the violation of ENT water quality standard for coastal beach waters, followed by the wash-off process of up to four-day antecedent rainfall.

1 Introduction

Recreational waters are regularly monitored to either confirm the compliance of bacterial levels in beach waters with recreational water quality standards or issue swimming advisories (Deng et al., 2014; Zhang et al., 2012). A major issue with the current beach monitoring programs is that decisions on beach water quality are made on a weekly, biweekly, or even monthly basis while beach use commonly takes place on a daily basis particularly in the swimming season, as shown in Figure 1. In

* Masterminded the development of the Bayesian model and created the final version of this document

† Created the first draft of this document under the direction of Zhiqiang Deng

addition, bacterial levels in recreational beach waters generally vary in a wide range and they are commonly measured using the most probable number (MPN). Mathematically, MPN could be best described with probabilistic models like Bayesian models.



Figure 1: A typical beach holiday without swimming advisory

The primary objective of this paper is to develop a predictive model that is capable of predicting the chance or probability of water quality standard violation at recreational beaches in a fashion similar to daily weather forecasts. To that end, a Bayesian model is constructed and demonstrated using seven years of environmental and bacteriological data collected at six recreational beach sites along the Gulf Coast of Louisiana, U.S.A.

2 Material and Methods

2.1 Data Collection

The study beach is Holly Beach that is located in the southwest Louisiana next to the Calcasieu River Outlet at the Gulf of Mexico, USA, as shown in in Figure 2. Holly Beach is one of the most visited beaches along the northern Gulf of Mexico coast. The bacteriological data on enterococci (ENT, the primary indicator of coastal beach water quality) levels in the beach water and swimming advisories along with environmental data for water temperature, salinity, rainfall, wind, cloud cover, solar radiation and tide type were collected weekly by the Beach Monitoring Program of Louisiana Department of Health and Hospitals, U.S.A. Water level data were downloaded from the NOAA website (<http://tidesandcurrents.noaa.gov/>) for Calcasieu Pass, Louisiana, U.S.A. The weekly data collected at six Holly beach sites in the swimming season (May 1 – October 31) in 2005 – 2012 were employed in this paper for the development and demonstration of a Bayesian model.



Figure 2: Location of study beach sites

2.2 Development of Bayesian Model

The water quality of recreational beaches is generally determined based on the level of Fecal Indicator Bacteria such as *enterococci* (ENT). Due to the high variability the ENT level in recreational beach waters is commonly measured using the most probable number (MPN), implying that beach water quality could be best described using the probability of water quality standard violations. To that end, a Bayesian approach is proposed in this paper to determine the probability of water quality standard violation. A Bayesian model primarily consists of a prior distribution and a likelihood function (Patil & Deng, 2011). The posterior probability of a model prediction M given data D , $P(M|D)$, can be defined as:

$$P(M|D) = \frac{P(D|M) \times P(M)}{P(D)} \quad (1)$$

where the variables D and M represent, respectively, observed ENT levels (data) and the ENT levels predicted with the Artificial Neural Network (ANN) model developed by the authors (Zhang et al., 2012); $P(D|M)$ is the likelihood function while $P(M)$ is the prior distribution function; and $P(D)$ is a normalization factor ($= 0.4482$). The prior distribution was produced by fitting various probability distributions to the frequency distribution histogram of the seven years of ENT concentration data, collected from 2006 – 2012, using the Easyfit software. It was found that the prior distribution can be best fitted with a Nakagami distribution. In order to determine the likelihood function, the residuals (errors) between the observed and the ANN model predicted ENT levels were first calculated. Then, the likelihood function was obtained by fitting various probability distributions to the frequency distribution histogram of the residuals using the Easyfit software. It was found that the likelihood

function can be best approximated with a normal distribution. By combining the Nakagami distribution-based prior distribution and the normal distribution-based likelihood function, the probability of beach water quality standard violation for an ANN model-predicted ENT level can be computed from Eq. (1) as the posterior probability. The detailed steps involved in the Bayesian modelling are described using the flow chart shown in Figure 3.

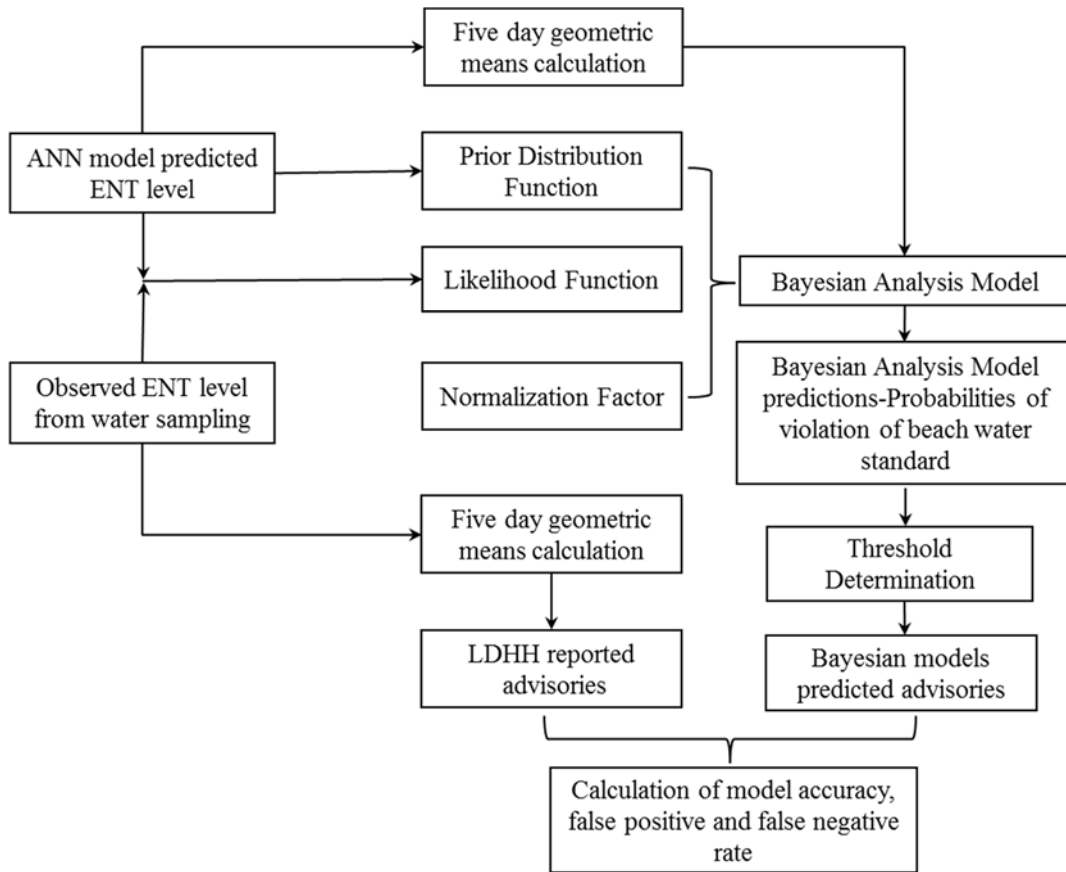


Figure 3: Steps for Bayesian model-based predictions of beach water quality standard violation

In order to understand the importance of individual hydrodynamic processes to the concentration of bacteria in recreational beach waters, a local (one-at-a-time) sensitivity analysis was conducted for the ANN model involved in the Bayesian model by varying one parameter at a time while all others are held constant. Specifically, all input variables involved in the ANN model were changed individually by $\pm 5\%$, $\pm 10\%$, $\pm 20\%$, $\pm 30\%$, and $\pm 40\%$, and the corresponding percent changes in the model output were calculated. The sensitivity of the model-predicted ENT levels to individual input variables can then be ranked from the highest (with the highest percent change in the model output) to the lowest (with the lowest percent change in the model output).

3 Results and Discussions

Combining the Nakagami distribution-based prior distribution and the normal distribution-based likelihood function yields the following Bayesian model for determining the posterior probability $P(M|D)$ of beach water quality standard violation.

$$P(M|D) = \int_0^{ANN} \frac{\exp\left[-\frac{1}{2}\left(\frac{x-0}{0.04655}\right)^2\right] \times \exp\left[-\frac{1}{2}\left(\frac{x-2.6299}{21.20}\right)^2\right]}{2.7777} dx \quad (2)$$

where the variable x represents the ENT level in beach waters and the upper limit ANN of the integral is the ANN model-predicted ENT level. Theoretically, the upper limit could be positive infinity. However, it was found in our actual calculations that there is an upper limit of 7.85786 for the Nakagami Distribution used in this paper. This means that the integration of Posterior Probability Density from the lower-limit of 0 to the upper-limit (7.85786) is close to the integration from the lower-limit to positive-infinity.

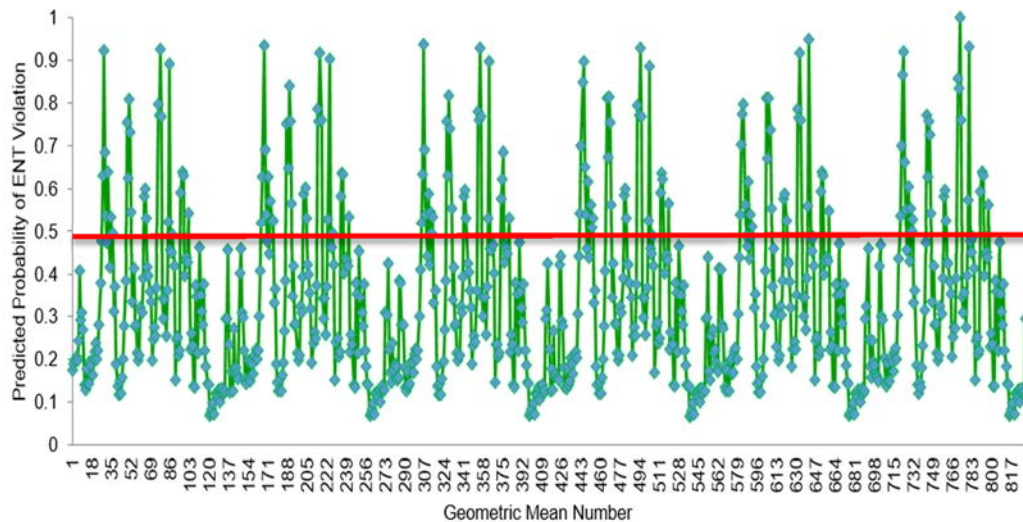


Figure 4: The probability of ENT level exceeding 35 MPN/100 mL predicted by the Bayesian model for each open beach holiday. The red line represents the probability threshold of 0.4875.

A beach water quality advisory is commonly issued by beach monitoring programs if the 5-day geometric mean of ENT level exceeds 35 MPN/100 mL. Accordingly, a beach water quality advisory should be issued if the probability of ENT level exceeding 35 MPN/100 mL, predicted by the Bayesian model, is higher than 0.4875. Figure 4 shows the probability of beach water quality standard violation for each geometric mean of ENT level predicted with the Bayesian model.

It is found, by comparing the Bayesian model predictions with the swimming advisories issued by the Louisiana Beach Monitoring Program, that the Bayesian model is capable of explaining 86.13% of the swimming advisories with a false positive rate of 7.24% and a false negative rate of 6.23%, indicating an excellent performance of the Bayesian model. The prediction of beach water quality violation using the probability is similar to weather forecasts and thus easy to understand and use by the general public. The Bayesian model shown in Equation (2) is in fact a hybrid model as the upper

limit (ANN) is based on the deterministic prediction of the ANN model. The integration of a deterministic model and a probabilistic model for prediction of recreation beach water quality is an innovative feature of this paper.

The results of sensitivity analysis can be summarized in terms of ENT sources and sinks. It was found that the model predicted ENT level and thereby the probability of water quality violation is most sensitive to the tidal washing process characterized by tide types (such as falling tide and raising tide) and tide level, followed by the runoff wash-off process produced by the up to four-day antecedent rainfall. This finding indicates that the primary source of ENT in the Holly Beach waters is the distributed nonpoint source that can be washed off and transported into the beach waters by the tidal washing process and the rainfall runoff following a heavy storm event, providing new insights into the source and transport of fecal indicator bacteria in recreational beach waters as bacterial sources of many recreational beaches and particularly the Holly Beach were considered to be unknown (Zhang et al., 2012). This is another innovative feature of this paper.

In terms of ENT sinks, the results of sensitivity analysis indicate that the increase in solar radiation and the decrease in rainfall are the two most important environmental predictors for the low concentration of ENT in beach waters. Basically, the increase in solar radiation (particularly the cumulative effect of antecedent solar radiation) enhances the inactivation of ENT while the decrease in rainfall means a low transport rate of ENT to beach waters, causing the reduction of ENT level in beach waters and maintaining good beach water quality. This finding is the third innovative feature of this paper.

4 Conclusions

The Bayesian model, presented in this paper, is capable to explain 86.13% of recreational water quality advisories issued by the Louisiana Beach Monitoring Program with a false positive rate of 7.24% and a false negative rate of 6.23%, indicating an excellent performance of the Bayesian model. The Bayesian model is innovative in terms of (1) the integration of a deterministic ANN model and a probabilistic model for the prediction of recreation beach water quality; (2) the identification of hydrodynamic processes controlling the source and transport of bacteria in coastal beach waters; and (3) the identification of key sinks of bacteria in coastal beach waters. It is found that the tidal washing process plays the most important role in triggering the violation of bacteriological water quality standard for recreational beach waters, followed by the runoff wash-off process of up to four-day antecedent rainfall. Strong solar radiation inactivates bacteria and thus reduces ENT levels in beach waters. While the Bayesian model is specifically developed for the U.S. Gulf Coast beaches, the methods (particularly the combination of the deterministic ANN model and the probabilistic Bayesian model) used in this paper are generally applicable to other coastal beaches.

References

- Deng, Z., Namwamba, F. and Zhang, Z. (2014). Development of decision support system for managing and using recreational beaches. *Journal of Hydroinformatics*. **16**(2), 447–457
- Zhang, Z., Deng, Z. & Rusch, K. A. (2012) Development of predictive models for determining Enterococci levels at Gulf coast beaches. *Water Research*. **46**(2), 465 – 474
- Patil, A. & Deng, Z. (2011). Bayesian approach to estimating margin of safety for total maximum daily load development. *Journal of Environmental Management*. **92**(3), 910–918