WATER RESEARCH 46 (2012) 465–474

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/watres

Development of predictive models for determining *enterococci* levels at Gulf Coast beaches

Zaihong Zhang, Zhiqiang Deng, Kelly A. Rusch*

Department of Civil & Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803-6405, USA

ARTICLE INFO

Article history:
Received 18 June 2011
Received in revised form
17 September 2011
Accepted 5 November 2011
Available online 19 November 2011

Keywords:
Coastal beaches
Enterococci levels
ANN model
Virtual Beach model

ABSTRACT

The US EPA BEACH Act requires beach managers to issue swimming advisories when water quality standards are exceeded. While a number of methods/models have been proposed to meet the BEACH Act requirement, no systematic comparisons of different methods against the same data series are available in terms of relative performance of existing methods. This study presents and compares three models for nowcasting and forecasting *enterococci* levels at Gulf Coast beaches in Louisiana, USA. One was developed using the artificial neural network (ANN) in MATLAB Toolbox and the other two were based on the US EPA Virtual Beach (VB) Program. A total of 944 sets of environmental and bacteriological data were utilized. The data were collected and analyzed weekly during the swimming season (May–October) at six sites of the Holly Beach by Louisiana Beach Monitoring Program in the six year period of May 2005–October 2010. The ANN model includes 15 readily available environmental variables such as salinity, water temperature, wind speed and direction, tide level and type, weather type, and various combinations of antecedent rainfalls. The ANN model was trained, validated, and tested using 308, 103, and 103 data sets (collected in 2007, 2008, and 2009) with an average linear correlation coefficient (LCC) of 0.857 and a Root Mean Square Error (RMSE) of 0.336. The two VB models, including a linear transformation-based model and a nonlinear transformation-based model, were constructed using the same data sets. The linear VB model with 6 input variables achieved an LCC of 0.230 and an RMSE of 1.302 while the nonlinear VB model with 5 input variables produced an LCC of 0.337 and an RMSE of 1.205. In order to assess the predictive performance of the ANN and VB models, hindcasting was conducted using a total of 430 sets of independent environmental and bacteriological data collected at six Holly Beach sites in 2005, 2006, and 2010. The hindcasting results show that the ANN model is capable of predicting *enterococci* levels at the Holly Beach sites with an adjusted RMSE of 0.803 and LCC of 0.320 while the adjusted RMSE and LCC values are 1.815 and 0.354 for the linear VB model and 1.961and 0.521 for the nonlinear VB model. The results indicate that the ANN model with 15 parameters performs better than the VB models with 6 or 5 parameters in terms of RMSE while VB models perform better than the ANN model in terms of LCC. The predictive models (especially the ANN and the nonlinear VB models) developed in this study in combination with readily available real-time environmental and weather forecast data can be utilized to nowcast and forecast beach water quality, greatly reducing the potential risk of contaminated beach waters to human health and improving beach management. While the models were developed specifically for the Holly Beach, Louisiana, the methods used in this paper are generally applicable to other coastal beaches.

* *Corresponding author*. Tel.: +1 225 5786850.
E-mail addresses: zzhan15@tigers.lsu.edu (Z. Zhang), zdeng@lsu.edu (Z. Deng), krusch@lsu.edu (K.A. Rusch).
0043-1354/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.watres.2011.11.027

1. Introduction

The US federal BEACH (Beaches Environmental Assessment and Coastal Health) Act requires beach managers to monitor bacterial water quality and issue swimming advisories when water quality criteria are violated. The main problem with the current beach monitoring program is that the level of water quality indicators like *enterococci* (ENT) may change between the time of sampling and reporting of results because current analysis methods commonly require an incubation step of 24–48 h. This incubation step makes protective actions such as preemptive beach closures impossible. This time lag of 24–48 h can lead to beach advisories and closures that cause unwarranted loss of valuable recreation access or to permit swimming when conditions present an unacceptable level of risk. People swimming during the time between sample collection and test results may be unnecessarily exposed to microbial pollutants at peak contamination times. Essentially, the efficacy of current beach monitoring procedure, called persistence model, generally depends on steadiness of bacterial concentration in beach water while the steady condition rarely occurs in coastal beach waters (Kim et al., 2004; Boehm et al., 2005; Hou et al., 2006; Heberger et al., 2008). Therefore, an alternative model is needed to meet the BEACH Act requirements.

Extensive efforts have been made to develop predictive models for nowcasting and forecasting the level of fecal indicator bacteria in beach waters (Heberger et al., 2008; Lin et al., 2008; Sanders et al., 2005; Kelsey et al., 2004; Kay et al., 1994). Hou et al. (2006) presented a Dynamic Partial Least Square Regression (DPLSR) model for predicting ENT levels at the Huntington State Beach (HSB) and the Huntington City Beach (HCB), California. Parameters involved in the DPLSR model included storm water discharge, rainfall, sea surface temperature, upwelling index, wind velocity, wave height and direction, visitor number, atmospheric pressure, solar insolation, sampling time, tide level and range, and rainfall. A total of 703 sets of ENT and environmental data from October 1999 through December 2000 (one swimming season) were used. Results showed that the DPLSR model performed better than the persistence model. Frick et al. (2008) proposed a multiple linear regression (MLR) model, called Virtual Beach (VB) model. Variables used in the VB model included air temperature, dewpoint temperature, cloud cover, precipitation potential, wave height, wind direction, wind speed, alongshore wind component, and cross-shore wind component. The adjusted coefficient of determination of the VB model for *Escherichia coli* nowcasting for a 42 day swimming period in the summer 2006 was about 0.40. He and He (2008) introduced artificial neural network models for predicting fecal indicator bacterial concentrations at the Torrey Pines State Beach and the San Elijo State Beach in southern California. Model variables involved temperature, conductivity, pH, turbidity, river flow, rainfall, and time lapse after a rainstorm. A total of 184 data sets collected from March 15 to April 14, 2003 were used to train, validate, and test the models. Results showed that the linear correlation coefficients for training, validation, and testing of the models were 0.883, 0.878, and 0.789, respectively. No independent data were used to test the ANN models. In addition, typical coastal water parameters (such as tide level and type) were not included in the ANN models. It is generally recognized that statistical models should be based on long-term data (Nevers and Whitman, 2005; Francy and Darner, 2006). However, most existing beach water quality models were trained and tested using the same data collected in a single/short swimming period. Few existing models were tested against beach water quality data observed over two swimming seasons, making it difficult for beach monitoring programs to find a reliable predictive model especially for coastal beaches with unknown bacterial sources. Therefore, more modeling efforts are needed to improve beach monitoring programs and to implement the BEACH Act.

The overall goal of this study was to enhance beach monitoring programs by finding a more reliable model for predicting the level of fecal indicator bacteria in coastal beach waters impaired frequently by unknown bacterial sources. The US Environmental Protection Agency (US EPA) recommended that the *enterococci* (ENT) be used as a bacteriological water quality indicator for marine waters (USEPA, 1986). Therefore, specific objectives of this study were (1) to develop an ANN model for nowcasting and forecasting the ENT level at the Holly Beach, Louisiana, USA, (2) to construct a VB model using the linear transformation and another VB model using the nonlinear transformation for the same purpose, and (3) to compare the three models against the same data series and thereby to find and recommend a more reliable model to meet the BEACH Act requirements under different environmental conditions. While exhibiting typical features of coastal beaches, the Holly Beach also has some unique characteristics. Unique features of the models developed in this study include: (1) they are developed using a relatively long time series of data observed over three swimming seasons; (2) they are further tested through a hindcast procedure using additional three years of independent data which are not used in the model development; (3) bacterial sources to the Holly Beach are unknown; and (4) annual average bacterial level at the Holly Beach keeps increasing or displays a dynamic behavior. Due to the unique features and highly variable and dynamic bacterial levels, no existing models are applicable to the Holly Beach.

2. Study area and data collection

2.1. Study area

The Holly Beach is located in the Cameron Parish along the southwest Louisiana shoreline, as shown in Fig. 1. The Holly Beach stretches along the Gulf of Mexico from the Calcasieu River Outlet in the East toward the Sabine River Outlet in the West in the Calcasieu River Basin, Louisiana, US. There are three national wildlife refuges (NWR) close to the Holly Beach in the Calcasieu River Basin. The use of Holly Beach is very high during the swimming season from May–October, with approximately 150 people using the beach on a typical weekday, 1000 people on a typical weekend, and 6000 people on a typical holiday. Six water quality sampling sites at the Holly Beach are shown in Fig. 1 and their coordinates are listed in Table 1.