Development of decision support system for managing and using recreational beaches
Zhiqiang Deng, Fulbert Namwamba and Zaihong Zhang

ABSTRACT
The paper presents a decision support system (DSS) for managing and using recreational beaches. The DSS consists of (1) a sensor-assisted water quality monitoring system, (2) a multiple linear regression (MLR) model, developed with the Virtual Beach (VB) program, for predicting enterococci levels, and (3) a web-enabled Geographic Information System (GIS) platform for displaying beach water quality conditions. The MLR model was tested using a total of 945 sets of environmental and bacteriological data collected over 6 swimming seasons. It is found that the MLR model is capable of correctly predicting over 88% of beach advisories and over 80% of no advisory events. The web-enabled GIS platform has nowcasting and forecasting functions. The nowcasting function reduces the decision-making time from current 2–3 days to near real-time while the forecasting function changes the decision-making time from current 2–3 days behind to 1–3 days ahead of actual use of beaches, greatly improving the decision-making in beach management and reducing potential health risks of fecal pollution to beachgoers. While the DSS was specifically developed for the Holly Beach, USA, the methods used in this paper are generally applicable to other coastal beaches.

Key words | decision support system, enterococci, forecasting, nowcasting, recreational beaches

INTRODUCTION
To prevent or reduce the burden of water-borne illness and injury related to recreational water use at public beaches, beach managers are generally required to monitor beach water quality and to issue swimming advisories when water quality standards are violated. The main problem with the current beach management practice is that levels 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 hours. This incubation step makes protective actions such as preemptive beach closures impossible. This time lag of 24–48 hours 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 (Zhang et al. 2012). People swimming during the time between sample collection and test results may be unnecessarily exposed to microbial pollutants at peak contamination times.

In order to address the problem in current beach monitoring programs and to manage and use beaches in a more effective and efficient way, increasing efforts have been made in developing decision support systems (DSSs) for beach water quality management with emphasis on the development of predictive models for enterococci (Abebe & Price 2005; Boehm et al. 2005; Nevers & Whitman 2005; Gregersen et al. 2007; Lin et al. 2008; Deng et al. 2012; Zhang et al. 2012). The US Environmental Protection Agency (EPA) recommends enterococci as the best indicator of health risk in salt water used for recreation and as a useful indicator in fresh water as well (http://water.epa.gov/type/rsl/monitoring/vms511.cfm). Frick et al. (2008) presented a basic framework of the US EPA Virtual Beach (VB) program using a multiple linear regression (MLR) model for predicting fecal indicator bacteria (including enterococci). He & He (2008) proposed an artificial neural network model for
predicting fecal indicator bacteria concentrations at the Torrey Pines State Beach and the San Elijo State Beach in southern California. Grant & Sanders (2010) developed a boundary layer-based mathematical framework for modeling fecal indicator bacterial concentrations in coastal beach waters. Recent investigations focused on the development of process-based models for predicting bacterial levels in beach waters (Zhu et al. 2011; Ge et al. 2012; López et al. 2013). In spite of the efforts, there is a gap between the existing models and the practical need for managing and using recreational beaches due to the lack of a simple yet effective decision support tool for recreational beaches.

The primary objective of this study is to fill the gap by developing a web-enabled and user-friendly DSS, called Holly Beach Watch, for beachgoers and beach managers. The DSS consists of (1) a near real-time water quality monitoring and data management system for coastal beaches, (2) a MLR model for predicting enterococci counts, and (3) a web-enabled Geographic Information System (GIS) platform for displaying beach water quality conditions, as shown in Figure 1.

**MATERIALS AND METHODS**

**Study area**

Holly Beach is one of the two most attractive beaches in Louisiana, USA and is the best among the west Louisiana beaches (Deng et al. 2012; Zhang et al. 2012). Holly Beach is located along the Gulf of Mexico, east of the Sabine River and west of the Calcasieu River Outlet in the Calcasieu River Basin within Cameron Parish, Louisiana, as shown in Figure 2. Use of this 5.55 km beach is very high, with approximately 150 people using the beach on a typical weekday, 1,000 people on a typical weekend, and 6,000 people on a typical holiday during the swimming season. Peak use occurs during May through Labor Day. Heaviest use during the peak period occurs during the 4th of July, Memorial Day, and Labor Day, in that order. The importance of Holly Beach to the local economy is high. While the DSS was specifically developed for the Holly Beach, Louisiana, USA, the methods used in this paper are generally applicable to other coastal beaches.

**Real-time water quality monitoring system**

To obtain near real-time fecal-indicator bacteria concentrations in beach waters, near real-time data are needed for water quality variables including salinity, wind speed and direction, antecedent rainfall, and tidal water level at Holly Beach. Site-specific meteorologic data, such as rainfall and wind (direction and speed), are collected from the weather station in the Cameron Parish. Radar rainfall data for the contributing watershed can be obtained from the US National Weather Service (NWS). The radar rainfall data are available for 4-km grids for each hour of the day from NWS.

Near real-time remote sensing data for our study area can be obtained from Louisiana State University (LSU) Earth Scan Laboratory (ESL) (http://www.esl.lsu.edu/). The LSU ESL uses SeaSpace Corporation hardware and a combination of software from SeaSpace and in-house software for image processing and atmospheric correction. It is a fully automated system that acquires MODIS direct broadcast data via a 4.4-m X-band antenna. The MODIS high resolution bands 1, 3, and 4 are used to produce ‘true color’ image quick looks, which are provided in quasi-real-time to the web site (http://www.esl.lsu.edu/). Hourly data on cloud cover, wind, salinity, and tide are available online for the
NOAA tide station #8768094 at Calcasieu Pass (Figure 2), LA, USA in the study area.

**Predictive model for beach water quality**

Predictive models have been increasingly used to nowcast and forecast beach water quality from environmental predictors and to make accurate and defensible preemptive advisory issuance decisions (Boehm et al. 2005; Choi et al. 2005; Nevers & Whitman 2005; Frick et al. 2008; He & He 2008; Lin et al. 2008; Grant & Sanders 2010; Zhang et al. 2012). A regression model is developed in this study using the US EPA VB program and three years of beach water quality observations. The VB program is a software package designed to construct site-specific MLR models for the prediction of pathogen indicator levels at recreational beaches (http://www.epa.gov/ceampubl/swater/vb2/).

The VB program requires a collection of fecal indicator bacteria data for dependent variables (such as enterococci and fecal coliforms) and concurrently collected data for independent variables that quantified environmental conditions at the beach sites. The user can provide data for any number of potential model independent variables. The VB program allows the user to import data from Microsoft Excel spreadsheets. The VB program uses a multilinear regression technique to relate a dependent variable to independent variables such as wind, tide, and salinity. If a relationship between the dependent variable and an independent variable is nonlinear, the independent variables can be transformed to make the relationship linear. The VB program provides five types of independent variable transformations, including polynomial, square root, inverse, natural log, and log to the base 10.

The VB program uses two methods to develop possible MLR models. If the number of independent variables is not large, an exhaustive search algorithm is used. In the exhaustive algorithm, all possible combinations of the independent variables are used and tested. More specifically, the exhaustive algorithm searches systematically all potential combinations of the independent variables to find the best combination (solution) or model. The advantage of this algorithm is that
every single possible combination will be checked and the best combination identified. The major steps, involved in the exhaustive search algorithm, include: (1) generating a list of all potential models by going through all possible combinations, (2) examining the potential models one by one by removing infeasible ones and ranking the feasible ones, and (3) listing the top 10 best models after searching and comparing all potential models meeting the objective or fitness function. The steps can be described diagrammatically in Figure 3(a).

As the number of independent variables increases, possible MLR models increase exponentially, resulting in trillions of possible models from a modest number (12–13) of independent variables. Then, the exhaustive search algorithm is infeasible and the VB program implements a Genetic Algorithm (GA) that effectively and efficiently searches for the best possible MLR models based on a user-selected evaluation criterion, such as $R^2$, Adjusted $R^2$, Akaike Information Criterion (AIC), Corrected AIC, Predicted Error Sum of Squares (PRESS), Bayes Information Criterion (BIC), Accuracy, Sensitivity, Specificity, or the model’s Root Mean Square Error (RMSE). The GA is an optimization and search technique based on the Darwinian principle of ‘natural selection’ on string structures to build unique searches with elements of structure and randomness (Haupt & Haupt 2004). The major steps (Figure 5(b)), involved in GA, include: (1) creating an initial population of strings as chromosomes, (2) defining the fitness function and selecting a proportion of the existing population to breed a new generation, (3) generating a second generation population of solutions from those selected through the genetic operators (including crossover reproduction and mutation), (4) decoding the strings before the fitness function is evaluated, (5) calculating the fitness level, and (6) repeating the generational process until one of the convergence conditions has been reached: (a) a solution is found that satisfies minimum criteria; (b) a fixed number of generations is reached; (c) the highest ranking solution’s fitness is reaching or has reached a plateau such that successive iterations no longer produce better results. Formulae for encoding and decoding the strings involved in steps (1) and (4) can be found in Haupt & Haupt (2004, page 34). Regardless of which criterion is chosen, the software ranks and records the 10 best models in terms of that criterion, as shown in Figure 4.

The selected MLR model for the Holly Beach involves seven environmental parameters or predictors, including salinity (SAL), wind speed type (WSTy), tide type (TdHNLOrd), wind direction (WDTyOnShore), rainfall 3 days before (RFlag3), rainfall in last 48 hours (RF48), and tidal water level (OWL), as shown in Equation (1):

$$
\begin{align*}
\text{Ln}(\text{ENT}) &= 4.3428 - 0.0340 \times [\text{SAL}] + 0.1541 \times [\text{WSTy}] \\
&\quad - 0.0745 \times [\text{TdHNLOrd}] + 0.0889 \\
&\quad \times [\text{WDTyOnShore}] + 0.4149 \times [\text{RFlag3}] \\
&\quad - 0.5086 \times [\text{RF48}] + 0.2185 \times [\text{OWL}]
\end{align*}
$$

where Ln(ENT) = natural log-transformed enterococci level.

The selected MLR model along with other models is constructed using a total of 514 sets of environmental and bacteriological data collected weekly in 2007, 2008, and 2009 by the Louisiana Beach Monitoring Program at six sites of the Holly Beach, Louisiana, USA. The MLR model is further tested using an additional three years of independent data which were collected in 2005, 2006, and 2010 and not utilized in the model development.
GIS platform for beach water quality management and use

A web-enabled GIS platform, called Holly Beach Watch, is constructed using ArcView GIS and ArcIMS. The web GIS system is designed for beach managers to issue advisories (Advisory in red color or No Advisory in green color) via the ‘Model’ page, beach visitors to check whether the current beach water quality is good for swimming on the ‘Nowcast’ page, and beachgoers to better plan for their beach holidays based on the forecasted beach water quality on the ‘Forecasting’ page, as shown in Figure 5.

The ‘Nowcast’ page is intended to show whether a beach site is open (no advisory shown in green color) or closed (advisory shown in red color) on the current day. Beachgoers may directly access the ‘Nowcast’ page on the Holly Beach Watch website to find current or nowcasting information on beach water quality (Advisory or No Advisory). Figure 6 illustrates the ‘closed’ or ‘advisory’ scenario on the ‘Nowcast’ page while Figure 7 demonstrates the ‘open’ or ‘no advisory’ scenario on the ‘Nowcast’ page. The nowcasting function reduces the time needed for decision-making on whether a beach should be closed or not from current 2–3 days to near real-time, significantly improving the decision-making in beach management and use and making the real-time decision-making on beach management and use possible.

The ‘Model’ page (Figure 8) is designed for beach managers to update beach water quality information and issue advisories (Advisory or No Advisory) for both the nowcasting webpage (Figure 6 or 7) and the forecasting webpage (Figure 9). Four steps are needed for updating the ‘Model’ page. First, the updated values for the seven input parameters should be entered in the corresponding fields, as shown in Figure 8. Secondly, the
date (10/22/2012 in Figure 8) should be selected by clicking the field beside the ‘Submit’ button. Clicking the date field will pop up a calendar where the beach manager can easily select the date. Thirdly, the ‘Calculate’ button should be clicked to run the model (Equation (1)) and see the updated value (56.4624 in Figure 8) for the log-transformed enterococci level Ln(ENT). Lastly, the ‘Submit’ button should be clicked to update the ‘Forecasting’ page (Figure 9) by showing the newly forecasted beach water quality information in terms of ‘ADVISORY’ (red color) or ‘NO ADVISORY’ (green color) on the calendar. By using the ‘Model’ function, beach managers are able to make decisions about the potential risk to swimmers following rainfall and other events, and can reliably predict where and when exposure to bathers is significant.

The ‘Forecasting’ page shows forecasted beach water quality information (Advisory or No Advisory) for the next 1–3 days, as shown in Figure 9. The user can easily find the Advisory or No Advisory information on the calendar for the next 1–3 days. Therefore, the forecasting function changes the decision-making time from current 2–3 days behind to 1–3 days ahead of actual use of beaches. Use of the forecasting function will also allow beachgoers to be able to better plan for their beach holidays and beach management agencies to reduce the number of sampling sites and staff time, while expanding coverage of their public health protection program.

RESULTS AND DISCUSSION

Figure 10 shows a comparison between the Ln(ENT) levels predicted using the selected MLR model and observed in the swimming seasons 2007–2009. It can be seen from the figure that the MLR model is able to follow the overall variation trend in the observed ENT.
levels. It should be pointed out that the best model from the model development is not necessarily the best model in terms of prediction of the independent data, as shown in Table 1. In fact, Equation (1) is ranked as the third best model in terms of its performance in the model development, described by the RMSE ($=1.3024$), as shown in Figure 4. Equation (1) rather than the top ranked model (RMSE = 1.3022) is finally selected because of the best performance in predicting ENT levels for both the model training data and the independent data collected in 2005, 2006, and 2010, as shown in Table 1. It should be pointed out that no calibration is needed in the VB program due to the availability of the 10 best models for final selection according to their performances in both the model development and the model predictions.

In terms of model prediction or application, the selected MLR model can be utilized to make two types of predictions of fecal indicator bacteria concentration: nowcast based on current and recently observed values of independent variables and forecast based on forecasted values of independent variables. Figure 11 shows a comparison between the Ln(ENT) levels predicted using Equation (1) and observed in the swimming seasons 2005, 2006, and 2010. The figure indicates that the selected MLR model is able to predict the average variation trend in ENT level while it is unable to capture the high variability in observed ENT levels. It should be pointed out that there are essentially two types of decisions to make on beach water quality management: (1) advisory (indicated with red color) and (2) no advisory (indicated with green color), as shown in
The single sample maximum criterion, recommended by US EPA, for the ENT is 104 cfu/100 ml. In addition to the single sample maximum criterion, the geometric mean (geometric mean of five consecutive weekly samples) steady state maximum criterion of 35 cfu/100 ml is also employed by Louisiana Beach Monitoring Program for the ENT. Regardless of which criterion is used, the ‘advisory’ is issued by the beach monitoring program if that criterion is violated. Otherwise, the ‘no advisory’ is issued. In terms of the ‘advisory’ (positive) or ‘no advisory’ (negative), the overall false positive rate and false negative rate of model predictions are 11.53% (=109/945) and 19.68% (=186/945), respectively. It means that the MLR model is capable of correctly predicting over 88% of advisories and over 80% of no advisory events, indicating very good performance of the MLR model in supporting decision-making on beach water quality.

More efforts will be made in future to further improve the DSS. More specifically, the ‘Model Calculator’ in Figure 8 will be revised to include the calculation of geometric mean steady state maximum criterion that is commonly used in current beach management. In addition, the MLR model may be replaced with some other models of better performance, such as Artificial Neural Network models and process-based models.
CONCLUSIONS

The paper presents a DSS for beach water quality management. The DSS consists of (1) a sensor-assisted water quality monitoring system, (2) a MLR model, developed with VB program, for predicting enterococci levels, and (3) a web-enabled GIS platform for displaying beach water quality conditions. The MLR model was tested using a total of 945 sets of environmental and bacteriological data collected over 6 swimming seasons. It is found that the MLR model developed in this study is able to predict the overall variation trend in observed enterococci levels in coastal beach waters. The model is also capable of correctly predicting over 88% of beach advisories and over 80% of no advisory events. The web-enabled GIS platform has nowcasting and forecasting functions. The nowcasting function reduces the decision-making time from current 2–3 days to near real-time, significantly improving the decision-making in beach management and use and making the real-time decision-making on beach management and use possible. The forecasting function changes the decision-making time from current 2–3 days behind to 1–3 days ahead of actual use of beaches. By using the forecasting function, beach managers are able to make decisions about the potential risk to swimmers following rainfall and other events, and can reliably predict where and when exposure to bathers is significant. Use of the forecasting function will also allow beachgoers to be able to better plan for their beach holidays. While the DSS was specifically developed for the Holly Beach, USA, the methods used in this paper are generally applicable to other coastal beaches.

Figure 8  |  The “Model” interface showing model calculator and input data layer for the Holly Beach. This page is designed for the beach monitoring program coordinator and manager to update the nowcasting and forecasting pages daily. This page can be accessed by clicking the ‘Model’ on the Holly Beach Watch homepage.
Figure 9 | The ‘Forecasting’ interface showing whether the forecasted water quality at the Holly Beach is safe (No Advisory) or not (Advisory) for the next 1–3 days. For instance, the calendar shows that there is No Advisory for the current day (October 26, 2012) but Advisory is issued or forecasted for the next 3 days (October 27, 28, and 29). The beachgoers can obtain the beach water quality information at their fingertips by simply clicking the ‘Forecasting’ on the Holly Beach Watch homepage.

Figure 10 | Comparison between the Ln(ENT) levels predicted using the selected MLR model and observed in swimming seasons 2007–2009 (1–197 are data from 2007 swimming season; 198–354 from 2008; and 355–514 from 2009).

Figure 11 | Comparison between the Ln(ENT) levels predicted using the selected MLR model and observed in swimming seasons 2005, 2006 and 2010 (1–157 are data from 2005 swimming season; 158–256 from 2006; and 257–430 from 2010).
**Table 1** | Comparison of prediction performance between the selected MLR Model 3 and seven other best MLR models using the Root Mean Square Error (RMSE) and the linear correlation coefficient (LCC)

<table>
<thead>
<tr>
<th>Model development data</th>
<th>Independent data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
</tr>
<tr>
<td>Model 1</td>
<td>1.3022</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.3023</td>
</tr>
<tr>
<td><strong>Model 3</strong></td>
<td>1.3024</td>
</tr>
<tr>
<td>Model 4</td>
<td>1.3025</td>
</tr>
<tr>
<td>Model 5</td>
<td>1.3025</td>
</tr>
<tr>
<td>Model 6</td>
<td>1.3025</td>
</tr>
<tr>
<td>Model 7</td>
<td>1.3025</td>
</tr>
<tr>
<td>Model 8</td>
<td>1.3025</td>
</tr>
</tbody>
</table>

**ACKNOWLEDGEMENTS**

The material is based upon work supported by US NASA (National Aeronautics and Space Administration) under award No NNX09AR62G and Louisiana Department of Health and Hospitals.

**REFERENCES**


First received 30 October 2012; accepted in revised form 4 March 2013. Available online 8 April 2013.