

Predicting culturable enterococci exceedances at Escambron Beach, San Juan, Puerto Rico using satellite remote sensing and artificial neural networks

Abdiel E. Laureano-Rosario, Andrew P. Duncan, Erin M. Symonds, Dragan A. Savic and Frank E. Muller-Karger

ABSTRACT

Predicting recreational water quality is key to protecting public health from exposure to wastewater-associated pathogens. It is not feasible to monitor recreational waters for all pathogens; therefore, monitoring programs use fecal indicator bacteria (FIB), such as enterococci, to identify wastewater pollution. Artificial neural networks (ANNs) were used to predict when culturable enterococci concentrations exceeded the U.S. Environmental Protection Agency (U.S. EPA) Recreational Water Quality Criteria (RWQC) at Escambron Beach, San Juan, Puerto Rico. Ten years of culturable enterococci data were analyzed together with satellite-derived sea surface temperature (SST), direct normal irradiance (DNI), turbidity, and dew point, along with local observations of precipitation and mean sea level (MSL). The factors identified as the most relevant for enterococci exceedance predictions based on the U.S. EPA RWQC were DNI, turbidity, cumulative 48 h precipitation, MSL, and SST; they predicted culturable enterococci exceedances with an accuracy of 75% and power greater than 60% based on the Receiving Operating Characteristic curve and F-Measure metrics. Results show the applicability of satellite-derived data and ANNs to predict recreational water quality at Escambron Beach. Future work should incorporate local sanitary survey data to predict risky recreational water conditions and protect human health.

Key words | bathing water quality, enterococci, machine learning, remote sensing, water quality prediction

Abdiel E. Laureano-Rosario (corresponding author)
Erin M. Symonds
Frank E. Muller-Karger
 College of Marine Science,
 University of South Florida,
 140 7th Avenue South, Saint Petersburg, FL 33701,
 USA
 E-mail: elias3@mail.usf.edu

Andrew P. Duncan
Dragan A. Savic
 Centre for Water Systems,
 University of Exeter,
 Harrison Building, North Park Road,
 Exeter EX4 4QF,
 UK

INTRODUCTION

Recreational water quality monitoring programs exist worldwide to protect humans from potential exposure to pathogens and are based upon fecal indicator bacteria (FIB; Colford *et al.* 2007). The FIB monitored varies across latitudes and water types, where *Escherichia coli*, fecal coliforms, and *Enterococcus* spp. are most common (Colford *et al.* 2007; U.S. EPA 2012), and can be correlated with illness in areas with known fecal contamination sources at temperate latitudes (e.g. ~33.4°N–37.8°N; Colford *et al.* 2012; Boehm & Sassoubre 2014). In the USA, the Environmental Protection Agency (U.S. EPA) monitors *Enterococcus* spp. in recreational marine waters. Based on the 2012

Recreational Water Quality Criteria (RWQC), enterococci cannot exceed the geometric mean of 35 colony forming units (CFU) per 100 mL, which represents 36 illnesses per 1,000 primary contact recreators (U.S. EPA 2012). This value was then modified to 70 CFU/100 mL based on the Beach Action Value (BAV), recommended by the U.S. EPA National Beach Guidance and Required Performance Criteria for Grants (U.S. EPA 2014). These guidelines were adopted by the Puerto Rico Environmental Quality Board (PREQB), where they monitor recreational water quality around the island of Puerto Rico biweekly in accordance with the 2000 U.S. Beaches Environmental Assessment and

Coastal Health Act (U.S. EPA 2000; Cordero *et al.* 2012) and other water quality standards of Puerto Rico (PREQB 2010).

Escambron Beach is located in San Juan, Puerto Rico and is one of the most visited beaches in the region. Escambron Beach is within the Rio Piedras watershed (Diaz 2007; Lugo *et al.* 2011) and has a stormwater drainage outfall (18.46 °N, 66.09 °W), which discharges rainwater, agricultural runoff, and other greywaters (Diaz 2007). This FIB point source and the Bayamon Regional and Puerto Nuevo Regional wastewater treatment plant (WWTP; primary wastewater treatment) ocean outfall, located 5 km offshore, are the most prominent point sources of fecal pollution at the beach (Ortiz-Zayas *et al.* 2006). The nearby Rio Grande de Loiza river mouth also discharges human and non-human fecal pollution to the coastline (Quiñones 2012). In addition to the impact of known fecal pollution sources, culturable enterococci concentrations are influenced by environmental factors (Sanchez-Nazario *et al.* 2014; Laureano-Rosario *et al.* 2017). Such factors include: precipitation through increased runoff (Cordero *et al.* 2012); solar radiance bacterial inactivation (Maraccini *et al.* 2012, 2016); turbidity being a source of FIB or protecting them from ultraviolet (UV) light (He & He 2008; Shibata *et al.* 2010); and the resuspension of FIB in sediment reservoirs through increased winds and waves (Byappanahalli *et al.* 2012; Feng *et al.* 2013).

Predicting when FIB exceed water quality criteria has been a management goal, and researchers have approached this using a variety of mathematical methods (e.g. linear and nonlinear statistical modeling). Some studies have applied linear models to understand FIB relationships with environmental factors; however, these complex interactions may not be adequately characterized by linear models, which typically describe less than 50% of the variability (Gonzalez & Noble 2014; Laureano-Rosario *et al.* 2017). Furthermore, previous modeling efforts lacked infrastructure and human activities data (i.e. land use); consequently, they did not accurately predict FIB concentrations (Rochelle-Newall *et al.* 2015). FIB vary depending on location, sources, and environmental factors; thus, a nonlinear approach is more appropriate due to FIB complexity and their relationship with multiple parameters. Thus, nonlinear modeling is essential to understanding the complex relationships between environmental variations and FIB.

Studies using nonlinear methods, mostly based on machine learning, focused on relationships between FIB and environmental factors to predict recreational water quality (He & He 2008; Thoe *et al.* 2014; Avila *et al.* 2018; Zhang *et al.* 2018). These studies used different methods, such as artificial neural networks (ANNs), Bayesian models, decision trees, and Monte Carlo approaches to predict recreational water quality in both marine and freshwaters (Jiang *et al.* 2013). These models take into account non-continuous relationships by creating a nonlinear combination of predictors to assess their relationship with FIB. For example, Choi & Bae (2018) applied ANNs and predicted total coliform concentrations in California based on rainfall and streamflow. Similarly, Ostad-Ali-Askari *et al.* (2017) applied ANNs and modeled nitrate pollution as the main water quality indicator in Iran. Furthermore, the ANNs models used in our study were previously compared to decision trees, where it was found that the ANNs approach allowed the specification of the relative importance of false positives and false negatives, whereas the decision tree methodology only provided a fixed operating point (Duncan *et al.* 2013a, 2013b). Therefore, our study helps fill research gaps in the Caribbean for recreational water quality predictions using ANNs in the context of environmental variability.

Since ANNs are self-driven data-adaptive methods, they can identify nonlinear, functional relationships between FIB and environmental factors. Forecasting recreational water quality can greatly improve the management of recreational waters as managers can overcome the time-lag associated with routine beach water quality monitoring (Enns *et al.* 2012; Thoe *et al.* 2014).

Even though ANNs have been widely applied to predict bathing water quality throughout the world, our study expands on this by using long-term satellite-derived data together with *in situ* bacterial sampling in Puerto Rico. Satellite-derived data has been used to study both land and the ocean for management purposes (McCarthy *et al.* 2017). These have provided data since the 1980s, allowing long-term studies regarding environmental variability. When it comes to beach water quality, these datasets have not been fully applied to predict bathing water quality, where satellites can provide long-term environmental data that can be combined with monitoring programs that have been in place for more than 10–15 years. There have been a few studies

looking at specific satellite-derived data (e.g. sea surface temperature, turbidity) in temperate and tropical areas (Kim *et al.* 2014; McCarthy *et al.* 2017; Zheng & DiGiacomo 2017). These have helped show the applicability of satellites; however, there are still some research gaps regarding the combination of satellite-derived data with predictive models in tropical areas. Therefore, this study implemented an ANNs approach, based upon 10 years of culturable enterococci concentration data together with *in situ* and satellite-derived environmental data, to predict recreational water quality at Escambron Beach, San Juan, Puerto Rico. More specifically, the model was developed using satellite-derived direct normal irradiance (DNI), turbidity, sea surface temperature (SST), and dew point with local observations of mean sea level (MSL), and cumulative precipitation from 24 h up to 120 h. The objectives of this study were: 1) to identify the most relevant environmental factors to predict culturable enterococci RWQC exceedances at Escambron Beach from 2005–2014; 2) to show the applicability of nonlinear modeling for an early warning system based on ANNs; and 3) show the benefit of incorporating remotely sensed data.

The results of this study can help understand the applicability of satellite-derived data in early warning systems and predictive models and the complex relationship between environmental factors and FIB in the Caribbean, with the aim of predicting exceedances and helping with management and mitigation of recreational water quality standards.

MATERIALS AND METHODS

Escambron Beach, San Juan, Puerto Rico

This study took place at Escambron Beach (Figure 1), one of the most popular beaches of San Juan, Puerto Rico (18.47°N, 66.08°W). This beach has a year-long swimming season. The municipality of San Juan (17.92°N–18.52°N, 65.62°W–67.28°W) has a tropical climate. Escambron Beach is classified as a low wave action beach, with mixed semidiurnal tides. Currents around the study area are generally westward; however, the Caribbean Coastal Observing System (CariCOOS; www.caricoos.org) buoy's current data shows very weak south-southeast semi-diurnal tidal currents on Puerto Rico's northern coast between 2 and 30 m

depth. In San Juan, the annual average precipitation is ~1,800 mm, and average air surface temperatures range between 24 and 29 °C. The study area is potentially influenced by the following sources of fecal pollution: stormwater outfall (Diaz 2007), the Rio Grande de Loiza river (Ortiz-Zayas *et al.* 2006; PREQB 2007), San Juan Bay Estuary (Perez-Villalona *et al.* 2015), and the Bayamon and Puerto Nuevo Regional WWTP ocean outfall (Ortiz-Zayas *et al.* 2006; PREQB 2007, 2011).

Culturable enterococci data

Culturable enterococci data for Escambron Beach were downloaded from the U.S. National Water Quality Monitoring Council from 2005 to 2012 (NWQC 2017). Data were for two sites separated by a distance of ~100 m; these were pooled due to their proximity and satellite data resolution. This dataset was extended from 2012 to 2014 with data provided by PREQB; thus, a total of 10 years of data were used ($n = 273$ observations for both sites combined). The culturable enterococci data were generated by the PREQB using U.S. EPA method 1,600 and had a detection limit of 4 CFU/100 mL. All enterococci concentrations described as below the limit of detection were substituted by the next highest concentration (e.g. 3 CFU/100 mL; Laureano-Rosario *et al.* 2017). Bacterial sampling was biweekly (i.e. every other week) and the combined geometric means from both sampling sites, commonly used due to bacterial variability, were then calculated. These geometric means were used in all further analyses.

Satellite-derived and *in situ* environmental data

Daily precipitation data (*in situ*) were obtained from the U.S. National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Information from 2005 to 2014. DNI and dew point were obtained from the satellite-derived U.S. National Solar Radiation Database (2005–2014; 30-min temporal resolution and 4 km spatial resolution). Daily MSL was obtained from the University of Hawaii Sea Level Center from 2005 to 2014. These datasets are obtained from a tide gauge located ~2 km from our study site. Day- and night-time SST were obtained from the U.S. NOAA Advanced Very High Resolution Radiometer

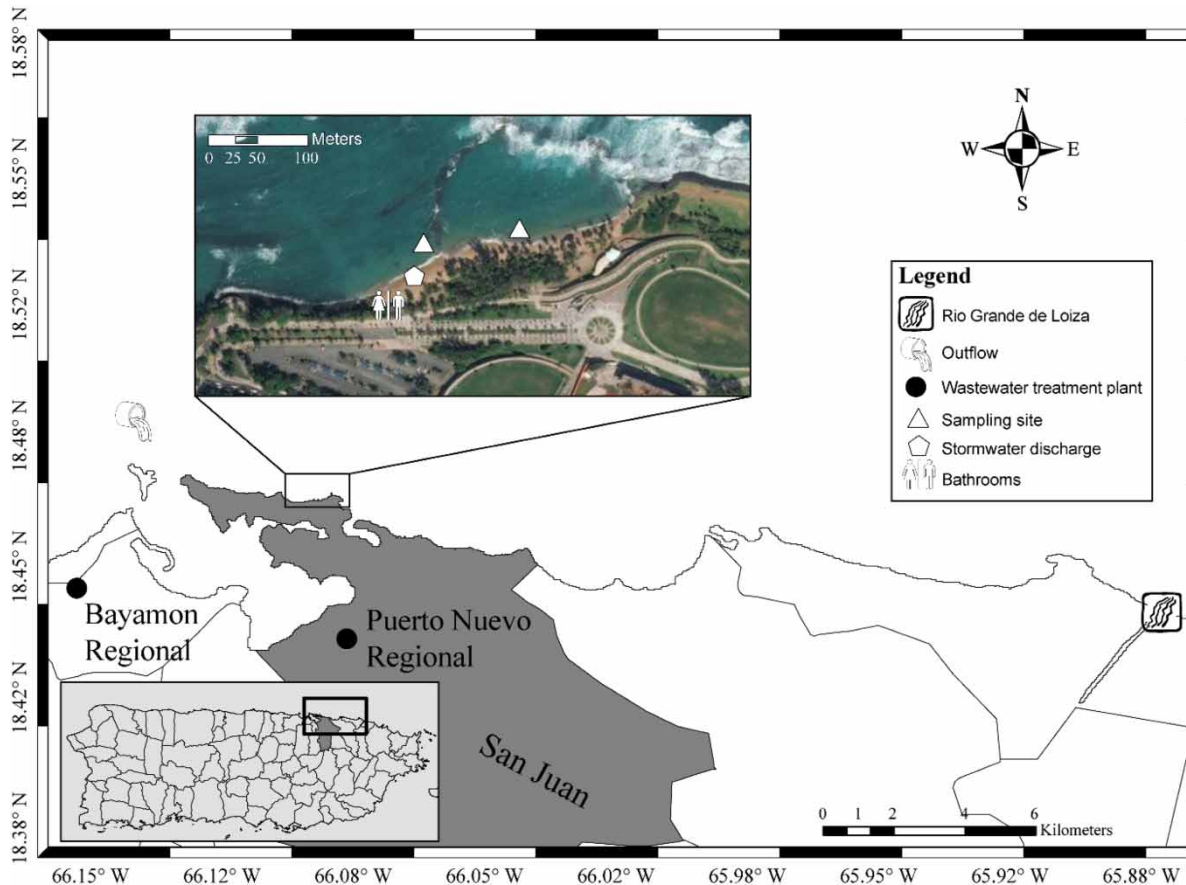


Figure 1 | Escambron Beach is located in San Juan, Puerto Rico. The inset map shows a stormwater outfall (pentagon), which is located on the beach, bathrooms (bathroom symbol), and sampling sites (triangles). The Rio Grande de Loiza river mouth (river symbol) discharges 22 km east of the beach. The combined ocean outfall, discharging primary-treated domestic wastewater (outfall symbol), is 5 km from the study site and discharges at a depth of 40 m.

(1 km spatial resolution) from 2005 to 2014. Data were extracted using the average of three 3×3 -pixel boxes, for the north coast of San Juan, Puerto Rico. Interactive Data Language (IDL; v. 7.2) was used to extract data. Remote sensing reflectance at 645 nm (R_{rs} 645; [Chen *et al.* 2007](#)) was used as a proxy for turbidity from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS-Terra; 250 m spatial resolution). Data were extracted using MATLAB (v. 2014b; The MathWorks Inc., Natick, MA, 2000); the average of two 3×3 -pixel boxes was used for turbidity for this coastal region. The environmental variables included in the model to predict culturable enterococci exceedances were: MSL, cumulative precipitation for 24, 48, 72, 96, and 120 h, SST, DNI, dew point, and turbidity. All satellite-derived data images corresponded to 1-day point samples. These images were compared with *in situ* collected data, where they

followed similar patterns. All input variables were log-transformed for predictive purposes.

ARTIFICIAL NEURAL NETWORK MODEL SETUP

Training, validation, and testing

The Radar Pluvial flooding Identification for Drainage System (RAPIDS) approach was implemented, which is based on ANNs ([Duncan *et al.* 2011, 2013a, 2013b](#)), to train, validate, and test the Escambron Beach recreational water quality model. This model was also used to predict bathing water quality in previous studies in the UK ([Duncan 2014](#)). The ANNs model used the non-dominated sorting genetic algorithm II (NSGA-II; [Deb *et al.* 2002](#)) for optimizing and

training. Data were subsequently validated using the Leave-One-Out-Cross-Validation (LOOCV) approach. The model included a weighting factor (a), which minimized the number of incorrectly predicted passes (i.e. values below thresholds; false positive ratios (FPR)) (Stidson *et al.* 2012). During the testing stage, a series of weighting factors (a) were tested, where the best and the one used for culturable enterococci exceedance predictions at Escambron Beach was $a = 3$, weighing health risks three times more important. Accuracy was determined by the accuracy band, which was calculated using the percentages of true positive rates (TPR) and true negative rates (TNR) compared to FPR and FNR. We used the F-measure (FM; Equation (1)) as one of the power metrics (i.e. how well the model predicted the outcome), providing the importance of false positives (FP) over false negatives (FN) through the weighting factor (a). In this case, bacterial exceedances (fails) were classified as negatives. By emphasizing the importance of FP, we reduced model misclassifications that could lead to potential health risks (i.e. saying it is safe to swim when bacterial concentrations are indeed higher). The second power metric used was the area under the Receiving Operating Characteristic (ROC) curve or AuC; this curve was based on the ratios of true positives (TP) and true negatives (TN) (Duncan *et al.* 2013a, 2013b; Duncan 2014). The AuC helped establish the ideal trade-off between FPR and FNR (i.e. $FPR = 1 - TPR$):

$$FM = \frac{(1 + a)TN}{(1 + a)TN + aFP + FN} \quad (1)$$

Culturable enterococci concentration was used as the target variable. Environmental variables (inputs) were weighted in to identify those that were the most relevant for predicting culturable enterococci exceedances. A total of 10 years (i.e. epochs) were included in the model. Data from 2012 and 2013 were pooled due to missing dates. The Neural Pathway Strength Feature Selection (NPSFS) method helped identify the most relevant inputs, through an ensemble of ANNs and comparing the similarities of the weight results (i.e. pathway strength) for the model inputs (i.e. environmental factors). Inputs with the most similarity of pathway strengths for the whole ensemble of ANNs were selected as the most relevant (Duncan *et al.* 2013a, 2013b; Duncan 2014). The strength of the relationship

as well as their relevancy (i.e. excitatory or inhibitory) to predict enterococci exceedances were identified.

ANNs calculate weights and biases to understand strengths and relationships between inputs and outputs (Basheer & Hajmeer 2000; Duncan *et al.* 2013a, 2013b). ANN weights were calculated for the hidden layers (W_1) and the output layer (W_2 ; Figure 2). The final weights (W_0) were calculated through matrix math of the ANNs hidden layer weights matrix and ANNs output layer weights vector (i.e. $W_1 \cdot W_2 = W_0$; Duncan *et al.* 2013a, 2013b; Duncan 2014). These final weights values were used to identify the most relevant parameters through NPSFS to predict culturable enterococci concentration exceedances. Members of the ensemble were trained on a similar but different subset of the full training dataset. Therefore, weights obtained in each ANN had different values. For a single output ANN, the result was a vector that specified the combined pathway strength of each input on the

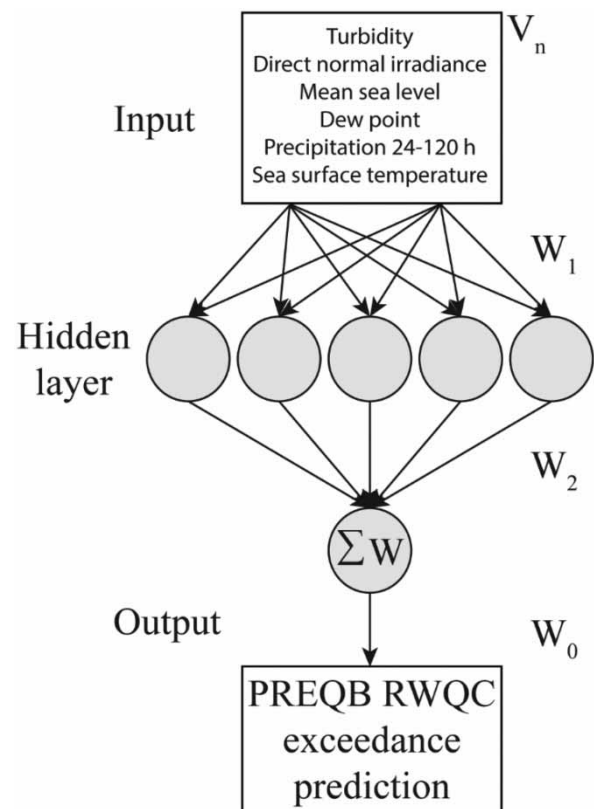


Figure 2 | The artificial neural networks model schematic for predicting Puerto Rico Environmental Quality Board Recreational Water Quality Criteria (PREQB RWQC) exceedances at Escambron Beach, San Juan, Puerto Rico.

output. Combined Neural Pathway Strength Analysis (CNPSA) was used to identify if the relationships were excitatory or inhibitory (Basheer & Hajmeer 2000). In that case, the input is considered relevant to predict enterococci exceedances. The model was run as categorical, where it identified either a pass or fail. The model used binary coding to classify passes (1) and fails (0). Therefore, when classified as excitatory it meant bacteria would pass (i.e. below threshold) and when classified as inhibitory it meant that bacteria would fail (i.e. above threshold).

Crossover and mutation rates, incorporated by NSGA-II during the training period, were used to optimize weights. These crossover and mutation rate factors differentiated new weights generations from the parent generation (Duncan 2014). Different crossover and mutation rate input values were tested; however, the crossover rate used was 0.2, and the mutation rate was 0.1. These input rates were used as they provided the best optimization results during the predictions of exceedances. Model's cost function used false positive rates and false negative rates. We used the minimum Euclidean distance to an ideal true positive ratio equal to one; these distances were derived from the ROC and used for optimization by NSGA-II to assess the quality of solutions (Duncan *et al.* 2013a, 2013b; Duncan 2014). Data were divided into ten epochs to ensure that the data used for training and validation were different than the data used for testing predictions as follows: epochs 1–3, 5, and 7 for training ($n = 152$); epochs 4, 6 and 8 for validation ($n = 66$); and epochs 9 and 10 for testing ($n = 55$).

Threshold selection for culturable enterococci exceedance predictions

To predict when enterococci exceeded the PREQB RWQC for safe recreation, the threshold selected for this study was the geometric mean concentration of 70 CFU/100 mL (PREQB 2016). This concentration is the BAV recommended by the U.S. EPA to ensure no more than 36 illnesses per 1,000 recreators and was adopted by the PREQB in 2015 (U.S. EPA 2014; PREQB 2016). The model compared the observed and predicted enterococci concentrations to this BAV threshold and identified them as 'safe for swimming' (i.e. below threshold) and 'potentially unsafe for swimming' (i.e. above BAV threshold). Results showed the influence, and magnitude, of

inputs to predict enterococci exceedances based on the specific thresholds mentioned above. These are shown as inputs having an inhibitory or excitatory influence on outputs crossing the set thresholds. Based on the 70 CFU/100 mL, we had a total of 238 passes, and 35 fails in the original data.

RESULTS

ANN model evaluation for accuracy and predictive power

The model predicted culturable enterococci exceedances with an accuracy band of 76% for Escambron Beach during 2005–2014 based on all ensemble models. This accuracy represented how many correct versus incorrect predictions were obtained compared to observed values. Overall, the model accurately predicted culturable enterococci exceedances based on the PREQB RWQC for safe recreation, with a power greater than 60%, where the FM was 0.61, and the AuC was 0.74 (Figure 3). Both FM and AuC provided the specificity and sensitivity of the model as these were based on the a -factor discussed above and the optimum Euclidean distance from the ideal point (i.e. False Positive Rate = 0 and True Positive Rate = 1; Duncan 2014).

Relevant environmental factors for culturable enterococci concentration predictions

The most relevant parameters to predict culturable enterococci concentrations at Escambron Beach from 2005 to 2014 were DNI, turbidity, 48 h cumulative precipitation, MSL, and SST (Figure 4). Only MSL and DNI showed an excitatory relationship; whereas turbidity, 48 h cumulative precipitation, and SST showed an inhibitory relationship. The most relevant variables were DNI and turbidity, where DNI showed a smaller spread of weights compared to turbidity. These environmental factors were identified as showing either an excitatory (positive weights) or inhibitory (negative weights) influence based on the binary coding. For example, DNI had an excitatory influence on predicting enterococci exceedances (Figure 4), which represented an overall stimulus of DNI on culturable enterococci concentrations not to cross the BAV threshold, meaning the results would be a pass (or

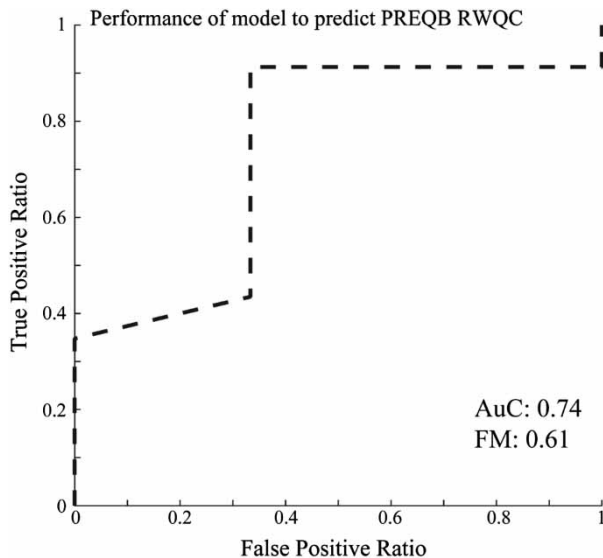


Figure 3 | Receiving Operating Characteristic curve for Puerto Rico Environmental Quality Board Recreational Water Quality Criteria (PREQ RWQC) exceedance predictions at Escambron Beach, San Juan, Puerto Rico depicts the performance of the model. The dashed line depicts the calculated true positive ratios (TPR) and false positive ratios (FPR). The area under the curve (AuC) was calculated based on the TPR/FPR ratios. The F-measure (FM) describes the power of model and the importance of false positives over false negatives (Stidson *et al.* 2012).

binary code 1). This represents a negative correlation between DNI and culturable enterococci, where it inhibits the bacterial cell to cross the BAV value. On the other hand, turbidity showed an inhibitory influence, meaning that it influences bacterial concentration to cross the BAV threshold, representing a fail (or binary code 0). This represents a positive correlation between turbidity and culturable enterococci, where it promoted higher concentrations that crossed the set threshold. Weight distributions were used as indicators of relevancy (Duncan 2014), where it was found that six variables (i.e. cumulative 24 h precipitation, cumulative 96 h precipitation, date, cumulative 120 h precipitation, and dew point), which crossed the zero line of the box and whisker plots, were not considered relevant to predict culturable enterococci exceedances at Escambron Beach surface waters for this time period.

DISCUSSION

This study investigated the use of satellite-derived data and a nonlinear model to predict exceedance of the PREQB

RWQC for safe recreation at Escambron Beach, San Juan, Puerto Rico. The most relevant variables in this model were DNI, turbidity, cumulative 48 h precipitation, MSL, and SST. These results showed that accurately predicting culturable enterococci exceedances, based on the 2014 BAV value, at Escambron Beach can be achieved using the aforementioned environmental variables. Notwithstanding, this model could make improved predictions by including a larger dataset and geo-referenced sanitation infrastructure data.

ANN model success for predicting exceedance of the PREQB RWQC

The ANN modeling described in this study showed the importance of identifying how environmental conditions can influence culturable enterococci concentration, as well as the complexity of these relationships between FIB and environmental factors. The use of ANNs to model culturable enterococci concentrations at Escambron Beach provided an accuracy band of 76% for exceedances, with greater than 60% model power, which is higher than previous models using linear approaches (e.g. Laureano-Rosario *et al.* 2017), and similar to those using ANNs for FIB predictions (e.g. He & He 2008; Chebud *et al.* 2012). Modeling enterococci exceedance at Escambron Beach was achieved by using the U.S. EPA and PREQB BAV (70 CFU/100 mL) as the model threshold concentration. By using this threshold, the model identified 35 occasions in which enterococci concentrations exceeded the BAV (i.e. model fails) in the original data and these events were then used for predictive purposes. AuC and FM provided the model's power and accounted for the ratios of true positives and true negatives. The accuracy band accounted for predicted values individually compared to the original values. These percentages may be affected by the number of passes ($n = 238$) and fails ($n = 35$) in the original observations.

These results also showed the applicability of combining satellite-derived data with nonlinear modeling. Sampling of monitoring programs usually is biweekly, which can greatly affect predictive power due to lower data resolution. Satellite remote sensing data can provide data once or twice a day depending on the sensor. For example, AVHRR

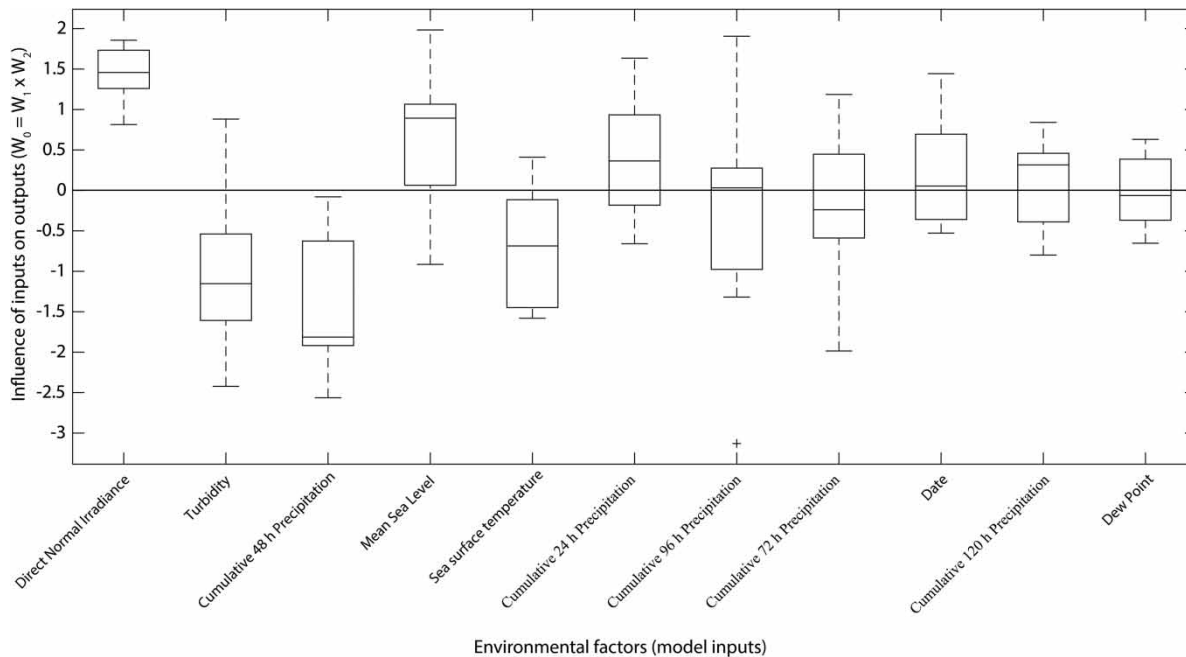


Figure 4 | Distribution of environmental variable weights in the ANNs model to predict Recreational Water Quality Criteria exceedances at Escambron Beach. The box and whisker plots show the distribution of weights from the ANNs model for each environmental factor: Direct normal irradiance, turbidity, cumulative 48 h precipitation, mean sea level, sea surface temperature, cumulative 24 h precipitation, cumulative 96 h precipitation, cumulative 72 h precipitation, cumulative 120 h precipitation, and dew point. Boxes are distributions of weights, lines inside boxes are mean values of the weights. The zero line represents no relevance for predicting outputs. Relevance for predictions was based on weight values calculated for predictions (W_0), multiplying those weights from the hidden layer (W_1) by weights from output layer (W_2). These W_0 weights represent the strength of the influence of input on output. Environmental factors are in order of importance for predictions, such that the most relevant variables are listed on the left.

provides SST data in the morning and afternoon. MODIS sensors also provide data in the morning (~10:30 AM; Terra satellite) and the afternoon (~1:30 PM; Aqua satellite). These datasets are freely available and are an important addition to management strategies if combined with monitoring programs that have been in place (McCarthy *et al.* 2017). These monitoring programs also help validate satellite-derived data.

Parameterization of the model was achieved with satellite-derived data as these provided larger datasets to be used during training, testing, and validation on the ANNs. Data were divided into a series of ANNs ensembles, and sensitivity and specificity were achieved by using FM and AuC. These were also possible by the inclusion of satellite-derived data, especially for products such as DNI and turbidity as they are not collected during every sampling effort. Therefore, by including satellite-derived data, it shows how models can be improved with better data resolution and potentially reducing sampling efforts (McCarthy *et al.* 2017).

Model improvement

Despite the high model power observed, future studies could improve upon the model created in this study by considering FIB watershed sources and longshore currents sources. For example, it is likely that failing sanitation infrastructure (e.g. leaky sewer pipes and septic systems) influenced FIB at Escambron Beach (Naidoo & Olaniran 2014). Additionally, WWTP, as well as stormwater discharges, could be a potential source of FIB throughout the year at various levels, and future studies should take them into account. Lastly, climatic conditions vary annually, and this natural variability can affect enterococci predictions over time.

The presence of enterococci in beach sands and vegetation (e.g. seagrass, green alga; Whitman *et al.* 2003; Sanchez-Nazario *et al.* 2014; Halliday *et al.* 2015) should also be considered to understand how these non-fecal sources influence enterococci concentrations (Feng *et al.* 2012, 2013). Thus, predictive models can likely be improved by the inclusion of these data. Furthermore, there is also the need

to identify other factors that might be of importance (e.g. through microbial source tracking, different fecal indicators, infrastructure data), to better predict these exceedances, identify when those are related to human fecal contamination versus non-human fecal contamination, and protect public health. Results of this model can potentially be implemented in an early warning system, using those variables identified here as predictors of bathing water quality. These, in turn, can be combined with ANNs, satellite-derived data, and ongoing monitoring programs to build now-casting models. Since satellite-derived data has been available for the past 20–30 years, it can help identify specific indicators correlated with FIB and reduced sampling efforts. Nevertheless, these relationships shown here are specific for Escambron Beach and culturable enterococci, and future work should modify and assess other environmental indicators that are correlated with fecal contamination.

Most relevant environmental factors influencing Escambron Beach water quality

Culturable enterococci concentration variability in coastal areas is influenced by fecal pollution sources, secondary, extraintestinal reservoirs, as well as by environmental factors (Viau *et al.* 2011). The current study accounted for specific environmental factors, such as DNI, turbidity, precipitation, MSL, SST, and dew point. These environmental factors have been shown to influence culturable enterococci concentrations, and other FIB, in temperate and tropical environments as well as marine and freshwaters (Enns *et al.* 2012; Lamparelli *et al.* 2015; Aranda *et al.* 2016).

As regards to environmental variables, precipitation most often explains the majority of FIB variability observed (He & He 2008; Feng *et al.* 2013; Laureano-Rosario *et al.* 2017); however, this study identified DNI as the most relevant environmental variable (Maraccini *et al.* 2012, 2016). The three most influential variables predicting PREQB RWQC exceedance were DNI, turbidity, and 48 h cumulative precipitation. DNI was the most important environmental variable to consider for PREQB RWQC exceedance predictions, likely due to bacterial inactivation (Maraccini *et al.* 2012, 2016). Since Escambron Beach is located in a tropical setting, it is no surprise that sunlight is one of the most influential environmental factors (Rochelle-Newall *et al.* 2015). Exposure to UV

light results in bacterial inactivation, and consequently a decrease in bacterial concentrations (Byappanahalli *et al.* 2012; Walters *et al.* 2014). The next most influential predictive environmental variable was turbidity, which has been documented to protect bacteria from UV light exposure. Turbidity is also associated with increased FIB when precipitation facilitates runoff into coastal waters (Halliday *et al.* 2015; Aragonés *et al.* 2016). Thus, the combined turbidity and DNI effects on enterococci concentrations could be the reason why these were identified as the most relevant parameters to predict culturable enterococci concentration exceedances.

The third most relevant parameter that predicted culturable enterococci exceedance at Escambron Beach was 48 h cumulative precipitation. He & He (2008) also identified 24–48 h of cumulative precipitation as significantly correlated with FIB at Torrey Pines State Beach and San Elijo State Beach, San Diego County, California, USA. Rainfall is known to increase FIB concentrations due to runoff (Colford *et al.* 2012), inadequately treated wastewater effluents (e.g. septic seepage; Naidoo & Olaniran 2014), and combined sewer-stormwater systems (He & He 2008). Since a nonlinear modeling approach was used in this study, the previously identified holistic influence of the aforementioned environmental conditions was able to be incorporated into the model, and improved predictions were generated (Noble *et al.* 2004).

The least two relevant environmental variables associated with PREQB RWQC exceedance predictions at Escambron Beach were MSL and SST. Previously at other beaches in Florida and California, USA, increased MSL was associated with lower culturable enterococci concentrations due to dilution and decreased MSL was associated with higher concentrations due to backwashing of waves and increased discharge into the coastal areas (Maraccini *et al.* 2012; Feng *et al.* 2016). However, Escambron Beach is a low-wave action beach, with a minimal tidal range; thus, MSL is not expected to strongly influence enterococci concentrations. Regarding SST anomalies, warmer waters have been documented to increase bacterial replication (Byappanahalli *et al.* 2012), and consequently, SST warm-anomalies have been shown to be related to increased culturable enterococci concentrations in tropical settings (Pachepsky *et al.* 2014; Laureano-Rosario *et al.* 2017). Even though SST was not the most influential environmental

variable identified by the model, it still provided information to predict PREQB RWQC exceedances.

CONCLUSIONS

This work shows that nonlinear models help to predict water quality with relatively good accuracy (76%). Data availability is an important aspect, especially the information regarding coastal water quality and both anthropogenic and environmental factors, due to their influence on FIB variability and phenology. Thus, a collection of data and water quality monitoring programs are important to better understand FIB variability. Through modeling culturable enterococci concentration exceedances, this study found the following:

- The most relevant parameters to predict culturable enterococci surface water concentrations at Escambron Beach from 2005 to 2014 were DNI, turbidity, cumulative 48 h precipitation, MSL, and SST.
- ANNs were able to predict enterococci concentration exceedances at Escambron Beach with an accuracy of 76% and a power greater than 60%, which is higher than most statistical linear models.
- Among the environmental variables evaluated, DNI, turbidity, and 48 h cumulative precipitation showed the highest influence on predicting culturable enterococci concentrations at Escambron Beach, which represent their holistic influence on enterococci concentrations.
- Only DNI and MSL showed a positive influence, whereas turbidity, 48 h cumulative precipitation, and SST showed an inhibitory (negative) influence on predicting culturable enterococci concentrations at Escambron Beach.
- Model predictive power may be improved by including sanitary survey data (e.g. septic system density), as well as other data describing enterococci sources, such as algal and sea-grass coverage, and stormwater and river discharges.

ACKNOWLEDGEMENTS

A.E.L.R. was supported by the U.S. National Science Foundation (NSF) Partnerships for International Research (PIRE) under Grant No. 1243510 and by the U.S. National

Aeronautics and Space Administration (NASA) Headquarters under the NASA Earth and Science Fellowship Program Grant No. NNX15AN60H. A.E.L.R. was also funded by the USF College of Marine Science Linton Tibbetts Endowed Fellowship. F.M.K. was supported by the U.S. EPA Science to Achieve Results (STAR) grant No. 83519301. E.M.S. was supported by U.S. NSF grant OCE-1566562. We would like to thank the teams from the Universidad Autonoma of Yucatan, Puerto Rico Environmental Quality Board, and Centre for Water Systems for their help and input for this work. We would also like to thank the IMaRS team for their input and help in manuscript revisions.

REFERENCES

- Aragones, L., Lopez, I., Palazon, A., Lopez-Ubeda, R. & Garcia, C. 2016 Evaluation of the quality of coastal bathing waters in Spain through fecal bacteria *Escherichia coli* and *Enterococcus*. *Sci. Total Environ.* **566**, 288–297.
- Aranda, D., Lopez, J. V., Solo-Gabriele, H. M. & Fleisher, J. M. 2016 Using probabilities of enterococci exceedance and logistic regression to evaluate long term weekly beach monitoring data. *J. Water Health* **14** (1), 81–89.
- Avila, R., Horn, B., Moriarty, E., Hodson, R. & Moltchanova, E. 2018 Evaluating statistical model performance in water quality prediction. *J. Environ. Manage.* **206**, 910–919.
- Basheer, I. A. & Hajmeer, M. 2000 Artificial neural networks: fundamentals, computing, design, and application. *J. Microbiol. Methods* **43** (1), 3–31.
- Boehm, A. B. & Sassoubre, L. M. 2014 Enterococci as indicators of environmental fecal contamination. In: *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection* (M. S. Gilmore, D. B. Clewell, Y. Ike & N. Shankar, eds). Massachusetts Eye and Ear Infirmary, Boston, MA, USA, pp. 1–17.
- Byappanahalli, M. N., Nevers, M. B., Korajkic, A., Staley, Z. R. & Harwood, V. J. 2012 Enterococci in the environment. *Microbiol. Mol. Biol. Rev.* **76** (4), 685–706.
- Chebud, Y., Naja, G. M., Rivero, R. G. & Melesse, A. M. 2012 Water quality monitoring using remote sensing and an artificial neural network. *Water Air Soil Pollut.* **223** (8), 4875–4887.
- Chen, Z. Q., Hu, C. M. & Muller-Karger, F. 2007 Monitoring turbidity in Tampa Bay using MODIS/Aqua 250-m imagery. *Remote Sens. Environ.* **109** (2), 207–220.
- Choi, S. W. & Bae, H. K. 2018 Daily prediction of total coliform concentrations using artificial neural networks. *KSCE J. Civ. Eng.* **22** (2), 467–474.
- Colford, J. M., Wade, T. J., Schiff, K. C., Wright, C. C., Griffith, J. F., Sandhu, S. K., Burns, S., Sobsey, M., Lovelace, G. &

- Weisberg, S. B. 2007 Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. *Epidemiology* **18** (1), 27–35.
- Colford, J. M., Schiff, K. C., Griffith, J. F., Yau, V., Arnold, B. F., Wright, C. C., Gruber, J. S., Wade, T. J., Burns, S., Hayes, J., McGee, C., Gold, M., Cao, Y. P., Noble, R. T., Haugland, R. & Weisberg, S. B. 2012 Using rapid indicators for *Enterococcus* to assess the risk of illness after exposure to urban runoff contaminated marine water. *Water Res.* **46** (7), 2176–2186.
- Cordero, L., Norat, J., Mattei, H. & Nazario, C. 2012 Seasonal variations in the risk of gastrointestinal illness on a tropical recreational beach. *J. Water Health* **10** (4), 579–593.
- Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6** (2), 182–197.
- Diaz, M. N. 2007 *Evaluation of Rainfall Intensity and its Effect on the Presence of Faecal Bacteria on the Beaches of Northern Puerto Rico*. Masters Thesis, Universidad Metropolitana, San Juan, Puerto Rico.
- Duncan, A. P. 2014 *The Analysis and Application of Artificial Neural Networks for Early Warning Systems in Hydrology and the Environment*. PhD Thesis, University of Exeter, Exeter, UK.
- Duncan, A., Chen, A. S., Keedwell, E., Djordjevic, S. & Savic, D. 2011 Urban flood prediction in real-time from weather radar and rainfall data using artificial neural networks. In: *Proceedings of the Weather Radar and Hydrology International Symposium*, 18–21 April, Exeter, UK.
- Duncan, A., Chen, A. S., Keedwell, E., Djordjevic, S. & Savic, D. 2013a RAPIDS: early warning system for urban flooding and water quality hazards. In: *Proceedings of the Machine Learning in Water Systems Symposium: Part of AISB Annual Convention 2013*, Exeter, UK, 3–5 April.
- Duncan, A., Tyrrell, D., Smart, N., Keedwell, E., Djordjevic, S. & Savic, D. 2013b Comparison of machine learning classifier models for bathing water quality exceedances in UK. In: *Proceedings of the 35th IAHR World Congress*, Chengdu, China, 8–13 September.
- Enns, A. A., Vogel, L. J., Abdelzaher, A. M., Solo-Gabriele, H. M., Plano, L. R. W., Gidley, M. L., Phillips, M. C., Klaus, J. S., Piggot, A. M., Feng, Z. X., Reniers, A., Haus, B. K., Elmir, S. M., Zhang, Y. F., Jimenez, N. H., Abdel-Mottaleb, N., Schoor, M. E., Brown, A., Khan, S. Q., Dameron, A. S., Salazar, N. C. & Fleming, L. E. 2012 Spatial and temporal variation in indicator microbe sampling is influential in beach management decisions. *Water Res.* **46** (7), 2237–2246.
- Feng, Z., Reniers, A., Haus, B., Solo-Gabriele, H., Fiorentino, L., Olascoaga, M. & MacMahan, J. 2012 Modeling microbial water quality at a beach impacted by multiple non-point sources. *Coastal Eng. Proc.* **1** (33), 74.
- Feng, Z. X., Reniers, A., Haus, B. K. & Solo-Gabriele, H. M. 2013 Modeling sediment-related enterococci loading, transport, and inactivation at an embayed nonpoint source beach. *Water Resour. Res.* **49** (2), 693–712.
- Feng, Z. X., Reniers, A., Haus, B. K., Solo-Gabriele, H. M. & Kelly, E. A. 2016 Wave energy level and geographic setting correlate with Florida beach water quality. *Mar. Pollut. Bull.* **104** (1–2), 54–60.
- Gonzalez, R. A. & Noble, R. T. 2014 Comparisons of statistical models to predict fecal indicator bacteria concentrations enumerated by qPCR- and culture-based methods. *Water Res.* **48**, 296–305.
- Halliday, E., Ralston, D. K. & Gast, R. J. 2015 Contribution of sand-associated enterococci to dry weather water quality. *Environ. Sci. Technol.* **49** (1), 451–458.
- He, L. M. & He, Z. L. 2008 Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern California, USA. *Water Res.* **42** (10–11), 2563–2573.
- Jiang, Y. C., Nan, Z. R. & Yang, S. C. 2013 Risk assessment of water quality using Monte Carlo simulation and artificial neural network method. *J. Environ. Manage.* **122**, 130–136.
- Kim, Y. H., Im, J., Ha, H. K., Choi, J. K. & Ha, S. 2014 Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *GISci Remote Sens.* **51** (2), 158–174.
- Lamparelli, C. C., Pogreba-Brown, K., Verhougstraete, M., Sato, M. I. Z., de Castro Bruni, A., Wade, T. J. & Eisenberg, J. N. 2015 Are fecal indicator bacteria appropriate measures of recreational water risks in the tropics: a cohort study of beach goers in Brazil? *Water Res.* **87**, 59–68.
- Laureano-Rosario, A. E., Symonds, E. M., Rueda, D., Otis, D. & Muller-Karger, F. E. 2017 Environmental factors correlated with culturable enterococci concentrations in tropical recreational waters: a case study in Escambron Beach, San Juan, Puerto Rico. *Int. J. Environ. Res. Publ. Health* **14** (12), 1602.
- Lugo, A. E., González, O. M. R. & Pedraza, C. R. 2011 *The Rio Piedras Watershed and Its Surrounding Environment*. USDA Forest Service, Washington, DC, USA.
- Maraccini, P. A., Ferguson, D. M. & Boehm, A. B. 2012 Diurnal variation in *Enterococcus* species composition in polluted ocean water and a potential role for the enterococcal carotenoid in protection against photoinactivation. *Appl. Environ. Microbiol.* **78** (2), 305–310.
- Maraccini, P. A., Mattioli, M. C. M., Sassoubre, L. M., Cao, Y. P., Griffith, J. F., Ervin, J. S., Van De Werfhorst, L. C. & Boehm, A. B. 2016 Solar inactivation of enterococci and *Escherichia coli* in natural waters: effects of water absorbance and depth. *Environ. Sci. Technol.* **50** (10), 5068–5076.
- McCarthy, M. J., Colna, K. E., El-Mezayen, M. M., Laureano-Rosario, A. E., Mendez-Lazaro, P., Otis, D. B., Toro-Farmer, G., Vega-Rodriguez, M. & Muller-Karger, F. E. 2017 Satellite remote sensing for coastal management: a review of successful applications. *Environ. Manage.* **60** (2), 323–339.
- Naidoo, S. & Olaniran, A. O. 2014 Treated wastewater effluent as a source of microbial pollution of surface water resources. *Int. J. Environ. Res. Public Health* **11** (1), 249–270.
- National Water Quality Council (NWQC) 2017 Water Quality Portal. Available from: <https://www.waterqualitydata.us/> (accessed on 15 January 2017).

- Noble, R. T., Lee, I. M. & Schiff, K. C. 2004 Inactivation of indicator micro-organisms from various sources of faecal contamination in seawater and freshwater. *J. Appl. Microbiol.* **96** (3), 464–472.
- Ortiz-Zayas, J. R., Cuevas, E., Mayol-Bracero, O. L., Donoso, L., Trebs, I., Figueroa-Nieves, D. & McDowell, W. H. 2006 Urban influences on the nitrogen cycle in Puerto Rico. *Biogeochemistry* **79** (1–2), 109–133.
- Ostad-Ali-Askari, K., Shayannejad, M. & Ghorbanizadeh-Kharazi, H. 2017 Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-rood River, Isfahan. *Iran. KSCE J. Civ. Eng.* **21** (1), 134–140.
- Pachepsky, Y. A., Blaustein, R. A., Whelan, G. & Shelton, D. R. 2014 Comparing temperature effects on *Escherichia coli*, *Salmonella*, and *Enterococcus* survival in surface waters. *Lett. Appl. Microbiol.* **59** (3), 278–283.
- Perez-Villalona, H., Cornwell, J. C., Ortiz-Zayas, J. R. & Cuevas, E. 2015 Sediment denitrification and nutrient fluxes in the San José Lagoon, a tropical lagoon in the highly urbanized San Juan Bay Estuary. *Puerto Rico. Estuar. Coasts* **38** (6), 2259–2278.
- Puerto Rico Environmental Quality Board (PREQB) 2007 *Total Maximum Daily Loads (TMDL) Rio Grande de Loiza Watershed*. Junta de Calidad Ambiental, San Juan, Puerto Rico, p. 281.
- Puerto Rico Environmental Quality Board (PREQB) 2010 *Water Quality Standards Regulation of Puerto Rico*. Available from: www.gobierno.pr/NR/rdonlyres/5A9F2F2E-94AE-4C69-8453 CA08D616ED7D/0/Reg_Estdares_Calidad_Agua_2010.pdf (accessed on 30 September 2017).
- Puerto Rico Environmental Quality Board (PREQB) 2011 *Total Maximum Daily Loads (TMDL) of Fecal Coliform for Evaluation Units, Puerto Rico*. Division de Planes y Proyectos Especiales, Area de Evaluacion y Planificacion Estrategica. Junta de Calidad Ambiental, San Juan, Puerto Rico, p. 164.
- Puerto Rico Environmental Quality Board (PREQB) 2016 *Beach Monitoring and Public Notification Program-Performance Criteria 2016–2017*. Available from: www2.pr.gov/agencias/jca/Documents/Monitoreo%20de%20Playas/Performance%20Criteria/BEACH_PERFORMANCE_CRITERIA_PREQB_2016-2017_final-draft.pdf (accessed on 31 July 2017).
- Quiñones, F. 2012 Impacto ambiental de pozos sépticos en Puerto Rico y su diseño y control. *Dimensión. Revista del Colegio de Ingenieros y Agrimensores de Puerto Rico* **1**, 16–22.
- Rochelle-Newall, E., Nguyen, T. M. H., Le, T. P. Q., Sengtehuanghoung, O. & Ribolzi, O. 2015 A short review of fecal indicator bacteria in tropical aquatic ecosystems: knowledge gaps and future directions. *Front. Microbiol.* **6**, 308.
- Sanchez-Nazario, E. E., Santiago-Rodriguez, T. M. & Toranzos, G. A. 2014 Prospective epidemiological pilot study on the morbidity of bathers exposed to tropical recreational waters and sand. *J. Water Health* **12** (2), 220–229.
- Shibata, T., Solo-Gabriele, H. M., Sinigalliano, C. D., Gidley, M. L., Plano, L. R., Fleisher, J. M., Wang, J. D., Elmir, S. M., He, G. & Wright, M. E. 2010 Evaluation of conventional and alternative monitoring methods for a recreational marine beach with nonpoint source of fecal contamination. *Environ. Sci. Technol.* **44** (21), 8175–8181.
- Stidson, R. T., Gray, C. A. & McPhail, C. D. 2012 Development and use of modelling techniques for real-time bathing water quality predictions. *Water Environ. J.* **26** (1), 7–18.
- Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M. L. & Boehm, A. B. 2014 Predicting water quality at Santa Monica Beach: evaluation of five different models for public notification of unsafe swimming conditions. *Water Res.* **67**, 105–117.
- United States Environmental Protection Agency (U.S. EPA) 2000 *Beaches Environmental Assessment and Coastal Health Act of 2000*. Public Law, Washington, DC, USA, pp. 106–284.
- United States Environmental Protection Agency (U.S. EPA) 2012 *Recreational Water Quality Criteria*. US EPA, Washington, DC, USA. Available from: www.epa.gov/wqc/microbial-pathogenrecreational-water-qualitycriteria (accessed on 30 July 2017).
- United States Environmental Protection Agency (U.S. EPA) 2014 *National Beach Guidance and Required Performance Criteria for Grants*. Available from: www.epa.gov/sites/production/files/2014-07/documents/beach_guidance-final-2014.pdf (accessed 30 July 2017).
- Viau, E. J., Goodwin, K. D., Yamahara, K. M., Layton, B. A., Sassoubre, L. M., Burns, S. L., Tong, H. I., Wong, S. H. C., Lu, Y. A. & Boehm, A. B. 2011 Bacterial pathogens in Hawaiian coastal streams – associations with fecal indicators, land cover, and water quality. *Water Res.* **45** (11), 3279–3290.
- Walters, E., Graml, M., Behle, C., Muller, E. & Horn, H. 2014 Influence of particle association and suspended solids on UV inactivation of fecal indicator bacteria in an urban river. *Water Air Soil Pollut.* **225** (1), 1822.
- Whitman, R. L., Shively, D. A., Pawlik, H., Nevers, M. B. & Byappanahalli, M. N. 2003 Occurrence of *Escherichia coli* and enterococci in *Cladophora* (*Chlorophyta*) in nearshore water and beach sand of Lake Michigan. *Appl. Environ. Microbiol.* **69** (8), 4714–4719.
- Zhang, J., Qiu, H., Li, X., Niu, J., Nevers, M. B., Hu, X. & Phanikumar, M. S. 2018 Real-time nowcasting of microbiological water quality at recreational beaches: a wavelet and artificial neural network based hybrid modeling approach. *Environ. Sci. Technol.* **52** (15), 8446–8455.
- Zheng, G. & DiGiacomo, P. M. 2017 Uncertainties and applications of satellite-derived coastal water quality products. *Prog. Oceanogr.* **159**, 45–72.

First received 14 May 2018; accepted in revised form 22 September 2018. Available online 22 October 2018