Journal of Environmental Management 223 (2018) 633–643

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

Implementation of an automated beach water quality nowcast system at ten California oceanic beaches

Ryan T. Searcy[a], Mitzy Taggart[a], Mark Gold[b], Alexandria B. Boehm[c,*]

[a] Heal the Bay, 1444 9th Street, Santa Monica, CA 90401, USA
[b] UCLA, 2248 Murphy Hall, 410 Charles E. Young Drive East, Los Angeles, CA 90095, USA
[c] Department of Civil and Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA, 94305, USA

ARTICLE INFO

Keywords:
Fecal indicator bacteria
Beach water quality
Enterococci
Escherichia coli
Nowcast
Modeling

ABSTRACT

Fecal indicator bacteria like *Escherichia coli* and enterococci are monitored at beaches around the world to reduce incidence of recreational waterborne illness. Measurements are usually made weekly, but FIB concentrations can exhibit extreme variability, fluctuating at shorter periods. The result is that water quality has likely changed by the time data are provided to beachgoers. Here, we present an automated water quality prediction system (called the nowcast system) that is capable of providing daily predictions of water quality for numerous beaches. We created nowcast models for 10 California beaches using weather, oceanographic, and other environmental variables as input to tuned regression models to predict if FIB concentrations were above single sample water quality standards. Rainfall was used as a variable in nearly every model. The models were calibrated and validated using historical data. Subsequently, models were implemented during the 2017 swim season in collaboration with local beach managers. During the 2017 swim season, the median sensitivity of the nowcast models was 0.5 compared to 0 for the current method of using day-to-week old measurements to make beach posting decisions. Model specificity was also high (median of 0.87). During the implementation phase, nowcast models provided an average of 140 additional days per beach of updated water quality information to managers when water quality measurements were not made. The work presented herein emphasizes that a one-size-fits all approach to nowcast modeling, even when beaches are in close proximity, is infeasible. Flexibility in modeling approaches and adaptive responses to modeling and data challenges are required when implementing nowcast models for beach management.

1. Introduction

Beach water quality is measured around the world to protect beachgoers from exposure to waterborne pathogens. Total and fecal coliforms, *Escherichia coli* and enterococci are fecal indicator bacteria (FIB) that are typically used to assess water quality. Epidemiology studies show that exposure to recreational waters contaminated with FIB from wastewater and urban runoff correlates with risk of diarrheal illness, respiratory disease, and skin ailments (Arnold et al., 2016; Colford et al., 2007; Haile et al., 1999; Wade et al., 2003; Yau et al., 2009, 2014). In the United States, 3943 beaches are monitored for FIB each year (USEPA, 2018). If concentrations exceed regulatory guidelines, then the beaches are posted as unfit for swimming or closed. Poor beach water quality not only affects the health of beachgoers, it also has large economic costs to surrounding communities (Rabinovici et al., 2004). In Southern California, Given et al. (2006) estimate that there are up to 1.5 million illnesses each year attributed to poor water quality at beaches costing as much as $51 million. Nationally, DeFlorio-Barker et al. (2018) estimate 90 million illnesses and costs of $2.2- $3.7 billion annually.

Analytical methods for detecting FIB require growing bacteria using selective microbiological media (USEPA, 2006, 2002). The methods take approximately 24 h in order to allow the bacteria to grow. Therefore, there is at least a 1 day lag between the time a water quality sample is collected and the result is obtained. Beach management decisions (posting or closing a beach) and public notification of water quality is therefore based on at least a 1 day old measurement (Kim and Grant, 2004). At most beaches, water samples are collected approximately weekly so that management decisions and public notification are based on even older measurements. It is well understood that FIB concentrations vary at periods smaller than a week, and smaller than a day (Boehm et al., 2002). Day-to-day changes in FIB sources, wind, tides, solar intensity, and rain, for example, may affect beach FIB concentrations (Hou et al., 2006; Jennings et al., 2018; Jovanovic et al.,

* Corresponding author. 473 Via Ortega, Stanford, CA 94305, USA.
E-mail address: aboehm@stanford.edu (A.B. Boehm).

https://doi.org/10.1016/j.jenvman.2018.06.058
Received 13 March 2018; Received in revised form 12 June 2018; Accepted 17 June 2018

2017; Laureano-Rosario et al., 2017; Nevers and Whitman, 2005; Thoe et al., 2015). Studies at marine beaches along the California coast have shown that FIB concentrations can even vary from minute-to-minute due to mixing processes in the coastal ocean, and the resultant patchiness of FIB contamination (Boehm, 2007).

Molecular methods such as quantitative PCR (QPCR) can be used to measure FIB concentrations at beaches (Haugland et al., 2005; Shanks et al., 2012) and have been proposed as a means to overcome the 1 day lag associated with culture-based method. QPCR methods can take a few hours to obtain results, however, when water sample collection and transport from the field to the laboratory is included, the total time is longer. US ambient water quality criteria allow for measurements of FIB by QPCR to complement measurements made by culture-based methods (USEPA, 2012). Although QPCR may provide water quality data for the same day the water is sampled, its not clear whether the data would be available in time for public notification that day. Additionally, as water samples are typically collected weekly, QPCR cannot provide information on water quality on days when a sample is not collected. On the other hand, FIB models can be used to provide predictions of water quality, even on days when a water sample is not collected.

FIB models can augment FIB measurements at beaches to overcome some of the problems associated with the use of FIB to make management decisions (Boehm et al., 2007; Frick et al., 2008; Hou et al., 2006; Nevers and Whitman, 2005; Thoe et al., 2015). Although process-based models that consider advection, dispersion, and non-conservative processes associated with FIB fate and transport have been developed and tested (Liu et al., 2006; Russell et al., 2013), they usually cannot provide the level of accuracy required for use in day-to-day beach management (Boehm et al., 2007). This is partly due to the uncertainties associated with FIB sources, lack of understanding of FIB fate in the environment, and difficulties parameterizing non-point FIB sources and fate processes within a model (Nevers and Boehm, 2010). Statistical models which take advantage of the correlative relationships between FIB and environmental variables, on the other hand, have been successfully used to develop FIB models that can accurately predict water quality standard exceedances (Avila et al., 2018; Boehm et al., 2007; Brooks et al., 2016; Francy, 2009; Frick et al., 2008; Gonzalez et al., 2012; Nevers and Whitman, 2005; Park et al., 2018). The USEPA supports the use of predictive models to supplement FIB measurements at beaches for public notification of water quality (USEPA, 2012).

Statistical FIB models have been used for beach management in the US (Great Lake beaches in Michigan, Ohio, and New York) (Francy, 2009; Francy et al., 2013), the UK (Crowther et al., 2001), and Hong Kong (Thoe and Lee, 2014). Multiple linear regression (MLR) models are used in these programs. Our previous work explored the ability of statistical FIB models to accurately predict beach water quality at California beaches (Thoe et al., 2014, 2015). We previously tested a variety of model types for their ability to accurately predict exceedances of the California single-sample standards for total coliform, fecal coliform, and enterococci at 25 California beaches. The results of those studies showed that the classification tree and tuned binary logistic regression (BLR-T) models best predicted whether beach water quality exceeded state single sample standards (SSS), and that these models' predictions were more accurate than the “current method” of using day-old or older measurements to make management decisions (referred to as the “persistence method” by some authors (Brooks et al., 2016; Francy et al., 2013)). Our previous work showed that accurate models could be created for most of the 25 beaches but that for a few, we could not find models that could be validated. Possible reasons for this include inter-annual trends in FIB concentrations that might result from changes in beach-specific infrastructure (installation of a new runoff diversion system for example), non-linear changes in climatic variables (for example, prolonged periods of rainfall, punctuated by drought or vice versa), or the stochastic, intermittent nature of FIB sources (Thoe et al., 2015).

In the present study, we develop and test an optimized variation of the MLR model– the tuned multiple linear regression model (MLR-T). This particular model type was not considered in our previous work. We also introduce a range of methods to partition historical data for model calibration and validation to create the best performing model. After identifying the best performing models, we created a custom Python code to actually implement the models for beach management during the 2017 summer swimming season at 10 California beaches, in collaboration with local beach managers. Model concentration predictions were compared against the California single sample FIB standards as outlined in the California Ocean Plan (10,000 most probable number (MPN)/100 ml total coliform, 400 MPN/100 ml fecal coliform, and 104 MPN/100 ml enterococci) to determine if a standard was exceeded and if a beach should be posted as unfit for swimming. During the implementation phase of the project, results were provided to beach managers and posted online before 10:00 h, 7 days a week. The three phases of the work (calibration, validation, and implementation) carried out with beach-specific MLR-T models set the present study apart from previous studies. We document the accuracy of the models and outline the successes and challenges to implementing a coast-wide California beach nowcast system. Overall, this study emphasizes that a one-size-fits all approach for creating nowcast models is infeasible, even when beaches are located in close proximity, and also that flexibility and adaptability is needed when implementing the models for beach management.

2. Methods

2.1. Overview

Ten beaches were chosen for the study. Historical FIB and environmental data were obtained for the beaches and divided into calibration and validation data. Models were calibrated and then validated using data to which they had not previously been exposed. The best performing models were chosen for each FIB at each beach and then used for actual beach management during the 2017 swimming season (referred to as the implementation phase). Model performance was evaluated using sensitivity and specificity metrics and compared to performance of the current method for beach management that uses day-to-week old measurements to estimate beach water quality. The modeling process is outlined in Fig. S1.

2.2. Beach selection

The 10 beaches were selected based upon their popularity among beachgoers and support for participation in the program among local beach managers (Fig. 1; each beach is denoted by a two letter abbreviation as shown in the figure). We included a variety of different beach types including those that were typically open ocean beaches, beaches with piers, and beaches with a drainage outlet. Nine of the beaches are located in Southern California; CB is located in Northern California. The 10 beaches are managed by 6 distinct local beach managers.

2.3. Historical data

Historical data collected between 2008 and 2016 during the swimming season (April through October) was used to calibrate and validate models. FIB concentrations (total coliform, fecal coliform, and enterococci), as well as the time of day and date the sample was collected, for the 10 beaches were obtained from the Heal the Bay database. Samples were collected on different days of the week including weekends, at all beaches. The database is constructed from data provided by local beach managers. If sample time was not available for a particular sample, the average sampling time at the beach (as stated by the beach manager or computed) was used. FIB data were measured using State-approved methods including IDEXX Coliert-18 and

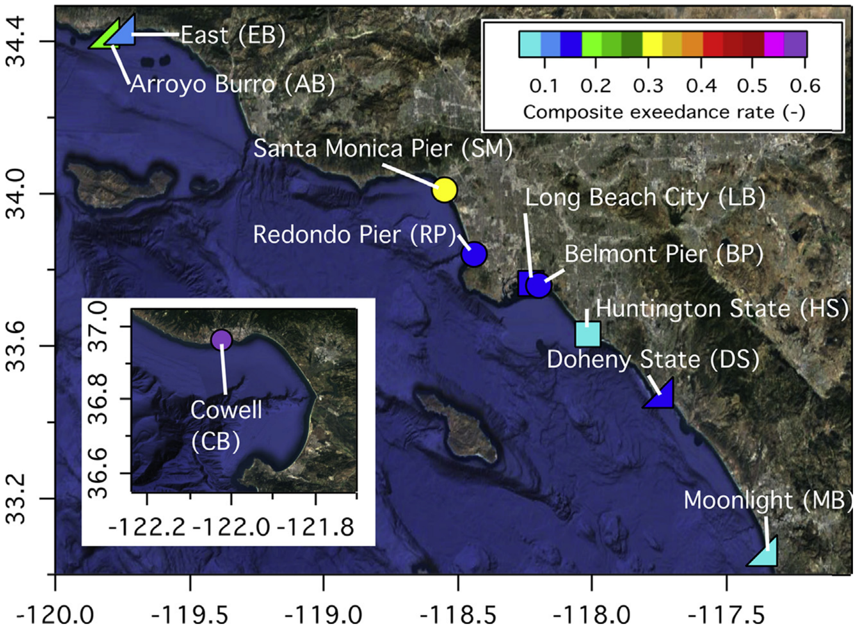

Fig. 1. Map of field sites along the California Coast. The inset shows Monterey Bay in Central California. The graph background is from Google Maps. The circles indicate pier beaches, the triangles are storm drain/river beaches, and the squares are open ocean beaches. The color corresponds to the composite exceedance rate for all three fecal indicator bacteria as also shown in Table S12. The positive latitude values indicate northern latitudes. The negative longitude values indicate western longitudes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Enterolert (Westbrook, ME) and standard methods (Table S1). For TC or FC concentrations below the limit of detection for the assay, the detection limit is substituted for the value. For ENT concentrations reported as below the limit of detection, the detection limit was substituted for two beaches in Orange County (HS and DS); for the remaining beaches, 1/2 the ENT detection limit was used. Although the different treatment of below detection limit values among FIB and beaches is not ideal, it reflects the manner in which historical data are provided by managers and archived and therefore, beyond our control. (We explored how the choice of below the detection limit (BDL) replacement value affected model performance for a subset of FIB-beach combinations and found that good performing models could be obtained regardless of the BDL replacement method, data not shown.) The FIB data were then $\log_{10}$-transformed (Thoe et al., 2015) for use as dependent variables in the models.

Raw environmental data were compiled for each beach including water temperature; wave period, height, and direction; river/stream flow rates; and weather conditions (e.g., rainfall, wind speed/direction, pressure, cloud cover, air temperature, dew point) from a variety of publicly available sources (regional airports, offshore buoys, local weather stations). Environmental data were obtained from sources as close to the beach as possible. Regional sources, including airport weather stations and offshore buoys, were used for all beaches. Meteorological data from sources local (within 12 km) to the beach sampling site (i.e. stations on a pier) were available for SM, LB, and BP. Water temperature data were measured at the Coastal Data Information Program (CDIP) offshore buoys and NOAA shore stations (when available). Tidal data were obtained from NOAA tide gauges. Data were checked for quality control purposes by manually examining data distributions and identifying extreme outliers and missing values. Because rainfall and streamflow data are skewed during the swimming season (that is, values for these parameters spike infrequently), the data were $\log_{10}$-transformed. Because of this parameterization, measurements of 0 inches of rainfall or 0 ft^3/s of streamflow were assigned at value of 0.005 inches or ft^3/s, respectively, before transformation (Thoe et al., 2015). Data sources are provided in Table S2.

Additional environmental variables were computed from the raw data (Tables S3–S7). Examples include offshore wave height three hours prior to sampling, tide level six hours prior to sampling, previous day's maximum tide, total rainfall over the previous three days, maximum tide level the day prior to sampling, cross-shore wind velocity, and littoral current direction. In all, over 100 different variables were considered for each beach (see SI for more details).

For each beach, individual predictive models for the three FIB regulated by the State of California (total coliform (TC), fecal coliform (FC), and enterococci (ENT)) were created. FIB concentrations were used as the models' predicted output and environmental parameters were considered as the models' independent variables.

2.4. Model calibration and validation: MLR-T models

We partitioned the historical data into calibration and validation subsets to create and validate predictive models, respectively. We used two methods to partition data: the 'chronological' method and the 'jackknife' method. The chronological method partitions all data but the most recent 1 or 2 summer seasons into a calibration set, leaving the last season or 2 for validation (hereafter C-1 and C-2, respectively). The jackknife method uses the *sample* function in the *pandas* Python package to randomly separate 60%, 70%, or 75% of the data points into a calibration set and the remaining 40%, 30%, or 25%, respectively, into validation set (hereafter JK-60/40, JK-70/30, and JK-75/25). C-1 was used as the default method of data partitioning. If resultant models were not sufficiently accurate (see criteria below), then the next method in a hierarchy (in the order: C-2, JK-75/25, JK-70/30, JK-60/40) was used to partition the variable dataset. This process was repeated until a sufficiently accurate model was found, or all of the data partitioning methods were exhausted (suggesting the FIB at the beach could not be modeled using our methods). The C-1 and C-2 methods were prioritized since they better reflect the way that models might be implemented by managers: created using all available data, and then implemented in the following season. However, if there are interannual trends or variations present in the FIB concentration time series (attributable to interannual variability in rainfall, for example), then the JK method might be better suited for creating accurate models.

We created tuned multiple linear regression (MLR-T) models for each of the three FIB at each of the 10 beaches. Whereas our previous work identified other model types that performed better than MLR (Thoe et al., 2015), beach managers had more familiarity with MLR models, in part because this is the type of model used in other nowcast programs around the world (Francy, 2009; Francy et al., 2013; Thoe and Lee, 2014) including USEPA's virtual beach (Cyterski et al., 2013). Adoption by local beach managers and efficient integration into their existing beach monitoring programs was an important goal of this project. Therefore, we sought to create and implement the California nowcast system using a MLR modeling approach, but improve upon the MLR model performance documented in our earlier work (Thoe et al., 2014, 2015) by tuning the MLR model to create MLR-T models.

2.4.1. Model calibration

MLR models were created from the calibration data set using SPSS Statistics (IBM, version 21). During model calibration, environmental variables (independent variables) are linearly combined to obtain predictions for FIB concentration (dependent variable) using a stepwise regression algorithm. The algorithm selects independent variables using entry and removal probability of 0.05 and 0.1, respectively. All independent variables (Tables S3–S7) were considered. We also included historical (previously measured) FIB concentrations as independent variables if we were unable to create a good-performing model excluding such variables (performance metrics below) (Brooks et al., 2016). Multicollinearity between independent variables was examined by calculating variance inflation factors (VIF). Variables with a VIF greater than 5 were highlighted, and from that set, the variable selected last by the stepwise algorithm was removed from the modeling dataset (Mason et al., 1989), and regression analysis was repeated.

To improve MLR model performance, we optimized MLR models by tuning them using a pre-multiplier (Francy et al., 2013; Francy and Darner, 2006) as described below. MLR models tuned using a pre-multiplier (PM) are referred to hereafter as MLR-T models. In brief, the MLR model prediction is multiplied by the PM to obtain the MLR-T prediction. The PM for each model was selected to maximize model sensitivity while still having a specificity of greater than 0.85 using the calibration data. Here, we define sensitivity as the fraction of correctly predicted single sample standard (SSS) exceedance days and specificity as the fraction of correctly predicted SSS compliance days. SSS exceedance and compliance days are those days when a measured FIB concentration was over and under or equal to its SSS, respectively.

Sensitivity = # of SSS exceedance days correctly predicted / # of actual measured SSS exceedance days

Specificity = # of SSS compliance days correctly predicted / # of actual measured SSS compliance days

A model was considered ‘calibrated’ if its sensitivity was greater than 0.30 and at least 0.10 greater than that of the ‘current method’ that used day to week old measurements to make management decisions, and the specificity was greater than 0.85 (Thoe et al., 2015). The justification and further description of these metrics are described by Thoe et al. (2014). See SM for additional description of the how the sensitivity and specificity of the current method are determined.

2.4.2. Model validation

The calibrated model was tested using validation period data. For validation, independent variable values are input into the model, and the prediction is compared against FIB measurements. A model was considered ‘validated’ if it passed the sensitivity and specificity criteria with the validation period data. If a model did not validate, a new method of data partitioning was used and the calibration and validation process was repeated. It is important to note that since the performance criteria depend on the performance of the current method in the validation and calibration periods, and these periods can change depending on the data partitioning method implemented, model creation is an iterative process.

Occasionally, we could not find a model for a specific beach and FIB combination that passed the sensitivity and specificity criteria for calibration and validation. In this case, we communicated with the local beach manger to determine if they would find models with lower sensitivity or specificity useful. If so, then the specificity and sensitivity criteria were adjusted.

2.5. Calibration and validation: BLR-T models

During the summer of 2017, in the midst of the implementation phase of the nowcast system, we noticed that four MLR-T models were not performing as well as we expected them to (as described in the results section), so we replaced those MLR-T models with BLR-T models, as described by Thoe et al. (2015). Historical data were split using the JK-75/25 method to create calibration and validation data sets. The dependent data were binary variables indicating the exceedance (1) or non-exceedance (0) of the single sample standards. Models were created from the calibration dataset using a custom Python script that employs the *LogisticRegression* class from the *scikit-learn* package. Before model fitting, multicollinearity between independent variables was examined by first identifying all pairs of predictor variables in the training dataset that have Pearson correlation coefficients (r_p) greater than 0.75. Within each pair, the variable with the lowest r_p with the $\log_{10}$-transformed FIB was dropped from the list of dependent variables. Variable selection was performed using a forward selection algorithm from the *mlxtend* Python package that iteratively adds variables to the model. In each step, the variable that maximizes the algorithm's scoring metric (sensitivity, accuracy, or area under the receiver operating characteristic curve (Brooks et al., 2016) is added to the model (Raschka, 2018). Initially, default values of the algorithm parameters - including the feature (variable coefficient) regularization, scoring metric and cross-validation parameters - were used, but were occasionally adjusted to increase model performance during calibration. For example, setting the cross-validation parameter to 3 folds, rather than the default 5 folds, yielded a better performing model for ENT at EB. Models were tuned by lowering the probability threshold from 0.5 to a level that maximized the model sensitivity while still meeting the specificity criteria (Thoe et al., 2015). Once a BLR-T model was calibrated (that is, it passed the sensitivity and specificity criteria), it was tested on the validation data set. Upon passing the performance criteria with the validation data, the BLR-T model was ready for implementation.

2.6. Model implementation

Validated models were programmed into a custom Python code (available on GitHub: https://github.com/rtsearcy/NowCast) that was implemented daily during the 2017 swimming season (approximately April–October depending on beach, Table 1). For each model, the code automatically harvests data from the internet by calling the application programming interface (API) of each data source. The code then calculates the model result for the day, and outputs the prediction as either a ‘Post’ or ‘No-Post’ beach management decision if the result was greater than, or less than or equal to the SSS, respectively.

For the first two weeks of implementation, model predictions from both the automated code and spreadsheets were generated by two independent technicians for QA/QC purposes. A technician collected input data from online sources and entered them into the spreadsheets that computed model predictions. The predictions generated using a spreadsheet were compared to those generated by the automated code to ensure that the automated code was working properly. After the two week QA/QC period, the code was run exclusively to generate model predictions. A technician always examined the outputs of the automated code before sending model predictions to beach managers. Results were sent to beach managers by 10:00 h each day and made available to the public via a Heal the Bay website (www.beachreportcard.org) and mobile app.

Most of the nowcast models rely on data collected by third parties (Table S2). If that third-party source was unavailable on a particular day, then the prediction for the beach was not provided to beach managers. Predictions for a beach were also not made if an acute event such as a sewage spill occurred, a rain advisory was in effect (issued at discretion of beach manager, usually when rainfall exceeds 0.01

Table 1
Years of historical data used in the calibration (cal) and validation (val) period, the start date of implementation (imp), and number of days in the imp period, the number of routine samples collected during imp, the number of days when no nowcast prediction was available, and the additional number of days of water quality (WQ) information (info) during the imp period provided by the nowcast models.

Beach	Years used for cal + val	Imp start date	# days in imp	# routine samples collected in imp	# days no Nowcast in imp	Additional days of WQ info during imp
CB	2010–2016	5/26/17	159	26	9	124
AB	2008–2016	4/4/17	211	29	6	176
EB	2008–2016	4/4/17	211	31	5	175
SM	2010–2016	4/4/17	211	140	17	54
RP	2010–2016	4/27/17	188	25	12	151
LB	2010–2016	4/27/17	188	28	6	154
BP	2010–2016	4/4/17	211	31	7	173
HS	2011–2016	5/26/17	159	54	4	101
DS	2010–2016	4/4/17	211	40	8	163
MB	2010–2016	5/18/17	167	39	5	123

inches), or the beach was closed for any other reason.

The FC model for Santa Monica Pier was the only model that used past FIB concentration (FC concentration) as an independent variable, as was required before in previous realizations of this model (Thoe et al., 2014). Data for this variable were provided by the beach manager to our modeling team as soon as they were available. However, if the beach manager did not provide the FIB data to us by 9:50 h, then we did not provide a prediction to the beach manager that day.

2.6.1. Model performance during implementation

During the 2017 implementation period, managers collected water samples for FIB analysis once per week at 9 of the 10 beaches. At SM, samples were collected 5 days each week. For the days on which samples were collected, nowcast predictions were compared to the measured FIB concentrations (particularly, if those concentrations were greater than, or below than or equal to the SSS) to calculate the sensitivity and specificity of the models. As a baseline, we also calculated the sensitivity and specificity of the 'current method' of using the most recent measurement to characterize whether the beach was in violation of the SSS. The sensitivity and specificities of the models and the current method were examined weekly to ensure model results were acceptable and meeting our sensitivity and specificity performance benchmarks.

3. Results

3.1. General overview of water quality at study beaches

The 10 beaches included in the study (abbreviations in Fig. 1) ranged from clean (MB) to polluted (CB) based on the composite FIB exceedance rates across the six to nine years of data used to create models (median = 0.15, range from 0.06 at MB to 0.63 at CB) (Fig. 1, Table S12). FC was the most exceeding indicator at 2 of the 10 beaches (SM and CB) while ENT was the most exceeding indicator at the remaining beaches (Table 2).

Based on the historical data used to calibrate and validate the model, most beaches had many samples with FIB concentrations at or below the limit of detection of the various assays used to measure FIB (Table 2). For example, 6 of 10 beaches had ~50% or more of the ENT measurements at or below the limit of detection. The co-occurrence of measurements at or below the detection limit with those above the single sample standard at single beaches is not unexpected given the high degree of spatial and temporal variability documented in coastal water quality datasets (Boehm, 2007).

The current method sensitivity ranged from 0 to 0.69 (median of 0.24) considering all FIB at the 10 beaches in the historical datasets used to calibrate models (Table 2 shows current method sensitivities for calibration and validation datasets, separately). This means that, based on the median, 24% of the beach postings were correctly identified using the current method. The Pearson's correlation coefficient between current method sensitivity and the FIB exceedance rate in the beach-specific historical datasets was 0.72 – that is, the 'dirtier' the beach, the better the current method performs in terms of sensitivity. This is a direct result of consecutive samples at dirtier beaches being more likely to exceed the SSS, than those at cleaner beaches.

3.2. Model attributes

We created 28 MLR-T models for implementation during the 2017 swimming season: TC, FC, and ENT models at 8 of 10 beaches, and FC and ENT models only at CB and MB. A sufficiently accurate TC model for CB could not be developed after all data partitioning methods were exhausted (likely because there were only 3 exceedances in the historical dataset). No TC model for MB was developed because MB did not have a TC exceedance in its historical dataset. Henceforth, we will refer to the models by the FIB modeled and the beach (for example, ENT-SM for enterococci model at Santa Monica).

The methods used to parse data into calibration and validation periods varied between beaches (Table 2). Two, 10, 12, 2, and 2 models were created using the C-1, C-2, JK-75/25, JK-70/30, and JK-60/40 parsing methods, respectively. The MLR-T models contained between 2 and 10 independent variables, depending on beach (Table 3, Tables S8, S9). The pre-multiplier (PM) varied between 1.008 and 1.990.

The variables used in the MLR-T models for ENT are provided in Table 3 (Tables S8 and S9 show similar results for TC and FC, respectively). 27 of 28 MLR-T models included a rainfall variable; with the majority including at least one rain variable with a positive association with $\log_{10}$-FIB. 22 of 28 MLR-T models included a tide variable. Tide variables were diverse (such as tide level at time of sampling, and change in tide level over the previous hour) and generally the association with $\log_{10}$-FIB and the tide variable was positive. 16 of 28 MLR-T models used a wave variable (such as wave height, period, or direction), and the association between wave variables and $\log_{10}$-FIB was mixed between positive and negative. 8 of 15 models for a beach that had available data on nearby stream flow included a stream flow variable in the model, and associations were generally positive between $\log_{10}$-FIB and streamflow. Only one model (FC-SM) included a previous FC measurement as an independent variable (logFC1, see Table S7).

Twenty-five of the 28 MLR-T models achieved the sensitivity and specificity criteria in calibration and validation (Fig. 2). The exceptions were FC-SM, FC-CB and ENT-CB which did not attain the specificity criterion of 0.85. For these three beaches, which tended to have relatively poor water quality (Table 2), we purposely reduced the specificity criterion to match or exceed that of the current method in both calibration and validation, after discussions with the local beach managers. They had calibration and validation specificities of 0.81 and 0.81 (FC-SM), 0.63 and 0.52 (FC-CB), and 0.76 and 0.83 (ENT-CB), respectively (Table 2). Table S10 provides the number of SSS exceedances predicted versus observed during calibration and validation.

Table 2
Attributes of the FIB data sets at each beach as well as performance of the current method (CM) during calibration (cal) and validation (val) periods in terms of sensitivity (sens) and specificity (spec). Data partitioning method is also shown. SSS = single sample standard, DL is lower detection limit.

Beach	FIB	N (cal + val)	fraction above SSS (cal + val)	fraction ≤ DL (cal + val)	CM sens (cal)	CM spec (cal)	CM sens (val)	CM spec (val)	Parsing method for MLR-T
CB	FC	254	0.571	0.02	0.67	0.628	0.758	0.516	JK-75/25
CB	ENT	252	0.25	0.238	0.326	0.755	0.353	0.826	JK-75/25
AB	TC	230	0.065	0.2	0.111	0.938	0.167	0.942	JK-60/40
AB	FC	230	0.087	0.478	0.176	0.901	0	0.949	C-2
AB	ENT	227	0.123	0.617	0.409	0.872	0	0.941	JK-75/25
EB	TC	201	0.065	0.179	0.273	0.951	0	0.957	C-2
EB	FC	201	0.08	0.423	0.286	0.928	0	0.957	JK-75/25
EB	ENT	217	0.129	0.636	0.316	0.886	0.111	0.895	JK-75/25
SM	TC	973	0.022	0.314	0.375	0.988	0	0.971	C-1
SM	FC	967	0.282	0.451	0.444	0.806	0.484	0.773	JK-70/30
SM	ENT	1033	0.1	0.487	0.304	0.924	0.167	0.915	JK-75/25
RP	TC	243	0.025	0.062	0	0.983	0	0.954	C-2
RP	FC	242	0.107	0.355	0.267	0.942	0.455	0.903	JK-70/30
RP	ENT	242	0.116	0.649	0.389	0.933	0.5	0.941	JK-75/25
LB	TC	196	0.051	0.071	0.2	0.963	0	0.9	C-2
LB	FC	200	0.08	0.245	0	0.893	0	0.931	JK-60/40
LB	ENT	198	0.091	0.566	0.077	0.89	0	0.932	JK-75/25
BP	TC	195	0.041	0.077	0	0.963	0	0.942	C-2
BP	FC	195	0.051	0.246	0	0.957	0.2	0.955	JK-75/25
BP	ENT	195	0.113	0.564	0.111	0.869	0.25	0.941	C-2
HS	TC	501	0.014	0.279	0	0.987	0.5	0.991	C-2
HS	FC	502	0.062	0.345	0.292	0.929	0.286	0.966	JK-75/25
HS	ENT	503	0.085	0.298	0.3	0.932	0	0.959	C-1
DS	TC	372	0.008	0.083	0	0.993	0	0.989	C-2
DS	FC	372	0.086	0.102	0.357	0.917	0.25	0.965	C-2
DS	ENT	372	0.159	0.054	0.581	0.932	0.75	0.922	JK-75/25
MB	FC	362	0.025	0.304	0.167	0.977	0.333	1	JK-75/25
MB	ENT	362	0.052	0.713	0.143	0.974	0.167	0.917	C-2

3.3. Model implementation

Nowcast predictions for the 10 beaches were produced for the 2017 summer swimming season during the implementation phase. Beach model predictions began on different days of the season (Table 1), with predictions for half the beaches beginning 4 April 2017; predictions for all 10 beaches were available starting 26 May 2017. Nowcast predictions ended on 31 October 2017 coinciding with the end of the California swimming season, as defined in the California Ocean Plan.

The number of days during the swimming season when both routine monitoring samples were collected to measure FIB and nowcast predictions were made varied among beaches from 25 at RP to 140 at SM (Table 1). Nowcast predictions were made on 150–206 days, depending on the beach, thereby providing an average of 140 additional days of updated water quality information compared to the current method for each beach (Table 1). There were a few days when nowcast predictions could not be made (between 4 and 17 days, depending on the beach). These included days when the FIB data used as independent variables were not provided to the nowcast team in time for a model prediction (before 10:00 h), and days when there were data server failures.

The fraction of days where FIB was observed during routine sampling to exceed the SSS varied among beaches. AB and MB experienced no exceedance of single sample standards while CB and SM had exceedances of the FC SSS in 42% and 41% of the measurements (Table 4).

On a weekly basis during the implementation phase, nowcast predictions were compared to actual FIB measurements available to date in order to assess model performance in terms of sensitivity and specificity. A few months into implementation, we determined that four models should be replaced. The MLR-T models for ENT-AB, ENT-EB, FC-EB, and ENT-MB were not performing as well as we expected them to, based on their performance during validation. Between 4 April and 12 August 2017, the ENT-AB, ENT-EB, and FC-EB MLR-T models had specificities of 71%, 78%, and 72% although the validated specificities

Table 3
The variable types used in ENT models. Results are for the MLR-T models unless otherwise noted (BLR indicates results for the BLR-T models). Numbers indicate the number of each variable type used in the models. The corresponding sign with the number indicates association with $\log_{10}$-ENT ("+" = positive; "-" = negative). For example, "2+/1-" indicates that within the variable type, there are two variables with positive association, and one variable with negative association, with $\log_{10}$ENT. "/" indicates that the variable type was not included in the possible variable list for that beach. A blank cell indicates that the variable type was available to the model, but not selected. "Met" is meteorological. "Local Met" includes nearshore water temperature.

Variable Type	CB	AB	AB (BLR)	EB	EB (BLR)	SM	RP	LB	BP	HS	DS	MB	MB (BLR)
Rainfall	1+	1+	2+/2-	2+	1-	3+	1+	3+	2+	1+	1+	2+	
Other Regional Met		2-	1+					1-		1+/1-			
Local Met		/	/	/	/		/		1-	/	/		
Wave	1-	2+	2-			1-	1+			1+/1-	1-	2+/1-	1-
Tide	1+		1+/2-	3+	1-	2+	1+/1-	1+	1+	2+	1+/1-		
Weekend								1+					1-
Labor Day							1+		1+		1+	1+	1-
Past FIB	/	/	/	/	/		/	/	/	/	/	/	/
Streamflow		/	/	1+		/	/	/	/		1+	/	/

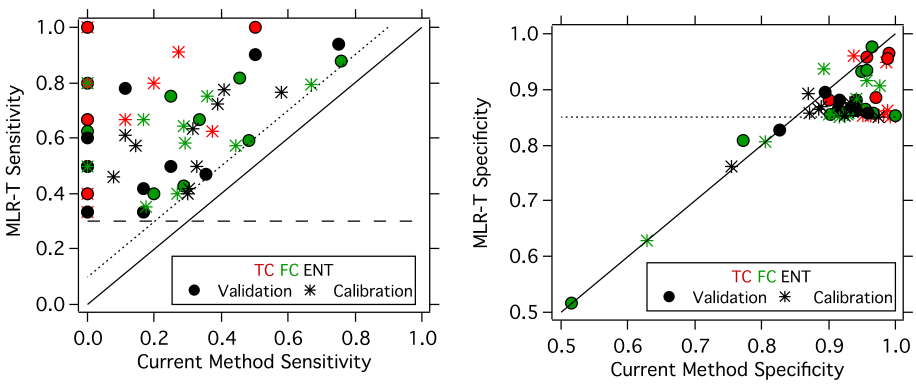

Fig. 2. The sensitivity (left) and specificity (right) of the MLR-T models versus the current method for each FIB at each of 10 beaches in the calibration and validation periods. TC is total coliform, FC is fecal coliform, and ENT is enterococci. The solid black line shows the one-to-one line. The diagonal dashed line in the sensitivity plot shows the line y = x + 0.1, and the horizontal dashed line shows y = 0.3 and illustrates that models had sensitivities greater than 0.3 and at least 0.1 higher than the current method. The dotted line in the specificity plot shows that all models had specificities higher than 0.85; the exceptions are described in the text. The data plotted here are also show in Table 2.

were nearly 90%. Similarly, the ENT-MB specificity between 18 May and 10 October 2017 was 82% although it also validated near 90%. Moreover, the four models were predicting SSS exceedances even though all measurements reported by the local beach managers for the beaches to date for the 2017 swim season were near or below the FIB assay lower detection limit. We developed new models for these beaches using the BLR-T method, because models developed using this method performed well previously at California beaches (Thoe et al., 2015). The four replacement BLR-T models contained similar variables as the MLR-T models (Table 3 and Table S9); and they met the sensitivity and specificity criteria in calibration and validation (Fig. 3). The new BLR-T models were implemented starting 13 August 2017 at AB and EB, and starting 11 October 2017 at MB. Thus, two distinct models (MLR-T and BLR-T) were used in the implementation period for ENT-AB, ENT-EB, FC-EB, and ENT-MB.

The median sensitivity and specificity of the nowcast models during the implementation period were 0.5 and 0.87, respectively (Fig. 4, Table 4). There were 11 models for which the sensitivity could not be evaluated because there were no measured single sample standard exceedances for the modeled FIB. During the same time period, the current method median sensitivity and specificity were 0 and 0.97, respectively (Fig. 4, Table 4). These values included results from AB, EB, and MB that were generated using both MLR-T and BLR-T models that were implemented in the beginning and at the end of the season, respectively.

Overall, nowcast sensitivities (that could be calculated) were greater than or approximately equal to the current method sensitivities (Fig. 4). The exception was TC-SM where the nowcast model had a sensitivity of 0 and the current method a sensitivity of 0.25. Nowcast specificities were generally greater than 0.85. Four models had specificities during implementation that were less than 0.8: FC-CB, ENT-AB, FC-SM, and ENT-HS. Table S10 provides the number of SSS exceedances predicted versus observed during implementation.

4. Discussion

The nowcast models provided water quality predictions every day, including weekends when most beach visits occur. This is an improvement over the current method in California where a majority of beaches are monitored once a week and new information is typically not provided on weekend days. Over the near 200 days in the 2017 swimming season, the nowcast system provided water quality

Table 4
Properties of data and model performance during implementation phase. Sens. = sensitivity, Spec. = specificity, CM = current method, N is the number of samples collected to measure the FIB type during the implementation phase, SSS = single sample standard.

Beach	FIB	N	fraction above SSS	nowcast sens	nowcast spec	CM sens	CM spec
CB	FC	26	0.42	0.73	0.67	0.64	0.73
CB	ENT	26	0.04	0	0.92	0	0.96
AB	TC	29	0	NA	0.9	NA	1
AB	FC	29	0	NA	0.9	NA	1
AB	ENT	29	0	NA	0.69	NA	1
EB	TC	31	0.06	1	0.93	0	0.93
EB	FC	31	0.03	0	0.83	0	0.97
EB	ENT	31	0.03	1	0.83	0	0.97
SM	TC	140	0.03	0	0.9	0.25	0.98
SM	FC	140	0.41	0.42	0.73	0.49	0.66
SM	ENT	140	0.03	0.25	0.88	0.25	0.99
RP	TC	25	0	NA	0.84	NA	1
RP	FC	25	0.08	0.5	0.87	0	0.91
RP	ENT	25	0	NA	0.92	NA	1
LB	TC	28	0.04	0	0.96	0	0.93
LB	FC	28	0.04	0	0.96	0	0.93
LB	ENT	28	0	NA	1	NA	0.96
BP	TC	31	0.03	1	0.87	0	0.93
BP	FC	31	0.06	0.5	0.83	0	0.93
BP	ENT	31	0.03	0	0.83	0	0.93
HS	TC	59	0	NA	1	NA	1
HS	FC	60	0.06	0.67	0.84	0	0.94
HS	ENT	60	0.06	0.67	0.76	0	0.94
DS	TC	40	0	NA	0.85	NA	1
DS	FC	40	0	NA	0.82	NA	0.98
DS	ENT	40	0.03	1	0.87	0	0.95
MB	FC	39	0	NA	0.87	NA	1
MB	ENT	39	0	NA	0.82	NA	1

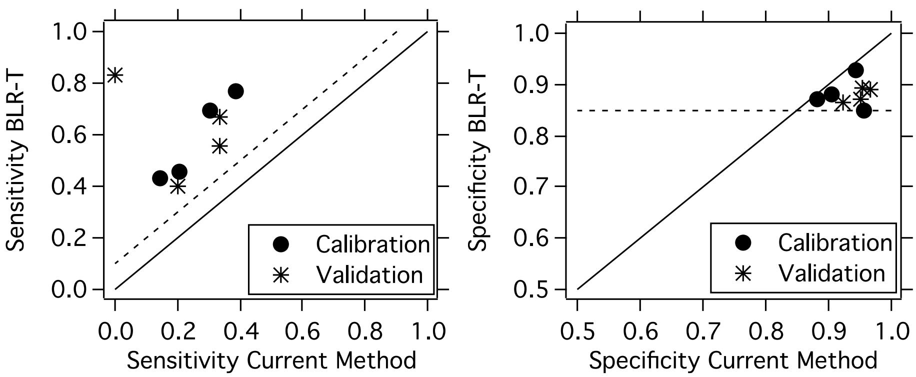

Fig. 3. The sensitivity (left) and specificity (right) of the 4 BLR-T models versus the current method during calibration and validation periods. The solid black line shows the one-to-one line. The dotted line in the sensitivity plot shows the line y = x + 0.1 and illustrates that models had sensitivities at least 0.1 higher than the current method. The dotted line in the specificity plot shows that all models had specificities higher than 0.85.

information on average 140 more days than the current method. For the weekly sampled beaches in our study, monitoring samples were collected (on average) 34 days during the season, while nowcast provided predictions on 182 days (on average) during the season. Local beach managers and Heal the Bay posted predictions on their websites, and local beach managers placed nowcast prediction signs at their beaches to ensure nowcast predictions were readily available to the public (Fig. S2). During the implementation phase, sensitivities of the nowcast models were on par with, or better than, the current method. Of the 17 models for which sensitivity could be calculated, 15 achieved a sensitivity equal to or higher than the current method. Although the performance of the models could not be verified explicitly on days when a water quality measurement was not made, it is logical to assume that model performance on these days is well approximated by model performance on days when a sample was collected.

Models can be evaluated using other metrics such as cross validation error rate (Avila et al., 2018), area under the receiver operator curve (Brooks et al., 2016), R^2 values (Brooks et al., 2016; Thoe et al., 2014, 2015; Thoe and Lee, 2014), and root mean square errors (Thoe et al., 2015). However, for the application of statistical models to public notification of SSS FIB exceedances at beaches, those metrics are only informative as a statistical measure. More importantly, beach managers are most interested in the sensitivity and specificity of the models; that is, managers need to know the fraction of beach postings that are correctly and incorrectly predicted by the model. However, we did examine R^2 values of the models during calibration and the values compared favorably with those we found previously for California beaches (Hou et al., 2006; Thoe et al., 2015), but were generally lower than those reported at Great Lake beaches (Francy et al., 2013) (R^2 varied from 0.15 (ENT-EB) to 0.51 (ENT-MB), with a median value of 0.26; data not shown). The abundance of samples with measurements at or below the lower detection limit of the assays, and the resultant left-censored data certainly affected model development and performance. However, the MLR model residuals were normally distributed (data not shown), thus use of the MLR model in conjunction with these left-censored data was deemed acceptable per assumptions of MLR.

Previous work has shown that rainfall (Boehm et al., 2002; Jennings et al., 2018), tides (Boehm and Weisberg, 2005), waves (Boehm et al., 2005), and water temperature (Boehm et al., 2004a), for example, affect FIB at oceanic beaches. The predictive models created for this nowcast system use these variables and others to predict FIB single sample exceedances. Rainfall was used as a variable in nearly every model. Rainfall rarely occurs in the summer season in coastal California, but when it does rain, urban runoff, which contains high levels of FIB (Reeves et al., 2004; Walters et al., 2011), can enter coastal waters via storm drains, streamflow, and over-beach flow. Tides were also included as variables in the majority of models. Tides can modulate the flow of groundwater (Boehm et al., 2004b; Yau et al., 2014) and surface waters (Grant et al., 2001) into coastal waters, and can wash FIB from foreshore sands (Feng et al., 2016; Yamahara et al., 2007) and beach wrack (Imamura et al., 2011; Russell et al., 2013) into the water column. Wave variables appeared in some models, potentially a result of waves impacting the speed and direction of the littoral current, or the formation of rip cells (Inman et al., 1971) and the potential for waves to suspend FIB from sands (Phillips et al., 2014). Although previous work has shown that salinity (He and He, 2008; Palani et al., 2008), turbidity (He and He, 2008; Paule-Mercado et al., 2016; Thoe and Lee, 2014), and solar radiation (Jennings et al., 2018; Yamahara et al., 2007), can be good predictors of FIB at beaches, reliable, daily data for these parameters were not available at the 10 beaches considered in our study. Variables describing flow of runoff into the ocean could be useful for nowcast models, but unfortunately, such variables are not readily available. In general, data availability is an important factor to consider when choosing environmental data to include as independent factors for nowcast models. Data must be available daily, in the morning, if they are to be used for generating daily nowcast predictions by 10:00 h. It is important to note that model predictions represent water quality status at the time of day when water samples are typically collected by

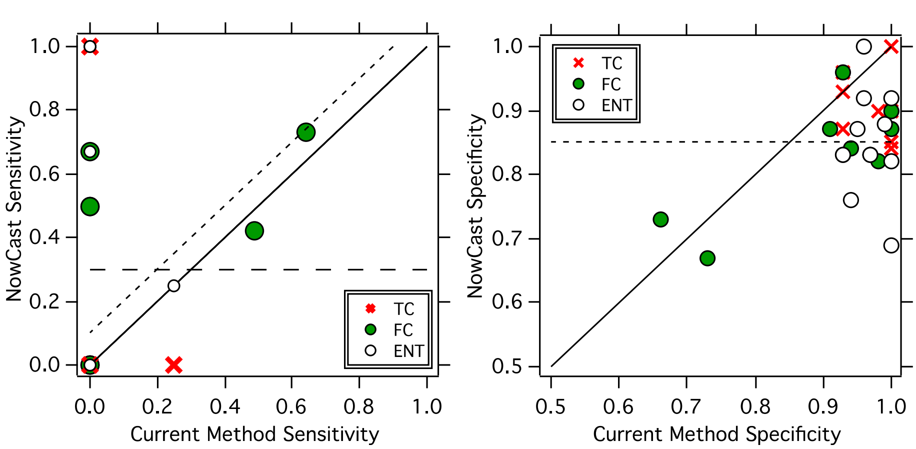

Fig. 4. The sensitivity (left) and specificity (right) of the nowcast models versus the current method during implementation. TC is total coliform, FC is fecal coliform, and ENT is enterococci. The solid black line shows the one-to-one line. The dotted line diagonal line in the sensitivity plot shows the line y = x + 0.1 and the horizontal dashed line shows y = 0.3. The horizontal dotted line in the specificity plot shows 0.85. Different symbols are used in this plot relative to Fig. 2 to aid in visualization as some of the symbols are in the same location in the sensitivity plot. The data plotted here are also provided in Table 4.

beach managers, even though the predictions may be available earlier in the day.

There is a trade off in terms of sensitivity and specificity when selecting and tuning models. In interviews with the local beach managers, we learned they valued sensitivity over specificity in order to prioritize public health protection. At most beaches in the nowcast system, specificities of the current method are high because the beach water quality exceeds SSS in < 10% of the measurements. Beach managers indicated that they were willing to accept specificities of 0.85 if doing so could maximize sensitivity. However, at beaches with poor water quality, where SSS exceedances are common (~50% of measurements are exceedances), the current method specificity can be relatively low (~0.5–0.75 at CB and SM, for example). At these beaches, local beach managers indicated they were willing to accept nowcast specificities less than 0.85 given the already low specificities they were experiencing with the current method. In general, nowcast models can be tuned by manipulating the pre-multiplier (for MLR-T) or the probability threshold (for the BLR-T) to achieve the sensitivities and specificities desired by beach managers. The method of tuning nowcast models has been previous applied in the Great Lakes (Francy et al., 2013; Francy and Darner, 2006).

An effective design element of our nowcast system is centralization. Beach specific models were developed, run, and maintained by Heal the Bay, and then results were communicated to the various health agencies responsible for public safety at the beaches throughout the state. This management structure is similar to the Great Lakes beaches where the USGS served as the centralized entity for predictive models for several years (Francy et al., 2013). We found this centralized structure had several advantages. A centralized system provides efficiency and cost benefits that come with scale. Modeling expertise was housed in one location, a centralized database was maintained, and the development of automation functions and a public notification system was applied to the beaches. Centralization also ensured model development and implementation are consistent across beaches. It also allowed greater flexibility and adaptability when confronted with poor performing models or unavailable data sources. Additionally, we found the centralized nowcast system integrated easily into local beach manager's existing monitoring and public notifications programs. Water quality predictions were available by 10:00 h, providing beach managers time to distribute the information before most beachgoers arrived at the beach. When surveyed, most beach managers found the daily availability of water quality information that came from the nowcast system to be beneficial. All beach managers indicated additional beaches for inclusion in the nowcast system in future seasons.

An essential part of our nowcast system was automation because it allows predictions to be made rapidly for multiple beaches. The automated code produced predictions for the 28 models in ~1 min. We implemented a rigorous quality assurance/quality control process whereby the automated model output was scored against a technician-run spreadsheet. In cases where the automated model and spreadsheet result disagreed, it was usually due to a technician input error in the spreadsheet. Thus, implementation of the automated Python code eliminated the potential for human error in the modeling process while increasing the speed and reliability of daily model predictions. Future iterations of the modeling system will run automatically at a pre-determined time (for example, 7:00 h) and will send results directly to beach managers to fully automate the process. Additional improvements to the automation can include identifying back-up data sources should third party websites fail. A similar system is used in the Great Lakes (https://ny.water.usgs.gov/maps/nowcast/).

A “one size fits all” approach to modeling FIB at California beaches was not possible. Although some models had similar independent variables (rainfall, tides, for example), FIB models were distinct for different beaches with diverse variables and coefficients. In addition, we considered different ways of splitting data between calibration and validation sets, inclusion of historical FIB data as independent variables, and found that different approaches worked at different beaches. We attempted to determine if there were patterns among the required approaches and parameters – for example if certain beach-types (defined by their typical water quality or their physiographic attributes) or FIB required certain variables or specific calibration/validation data splitting approaches, but we could not identify any patterns (data not shown). This emphasizes the importance of flexibility and adaptivity when creating nowcast models for beach management.

We found routine evaluation of model performance and adaptive improvement of models throughout the summer was necessary to improve prediction accuracy. We used multiple lines of evidence to inform the decision to replace a model during implementation including information on (1) model predictions, (2) model sensitivity and specificity, as well as (3) input from the local beach managers. Throughout the implementation period, we compared the models' sensitivities and specificities to their values during validation to assess if model performance was as expected. We replaced 4 MLR-T models for BLR-T models after noting that the sensitivities and specificities of the models were lower than we expected based on validation period performance, and very low measured FIB concentrations compared to occasional elevated predicted FIB concentrations. Replacing underperforming models with a different model type led to increased performance for the remainder of the season, a noted benefit of evaluating performance weekly. Had the 4 replacement BLR-T models been used during the entire implementation period, nowcast performance would have been even better than reported herein for the 4 models (Table S11). It is not clear why the ENT MLR-T models at AB, EB, and MB, and the FC MLR-T model at AB performed worse in implementation than validation, but a possible explanation is that water quality at the beaches, based on these bacterial indicators, was improved in implementation compared to validation (Thoe et al., 2015). Additional experience implementing models for management should provide additional insight into the conditions that result in a well calibrated and validated model underperforming. Future implementation of the California nowcast system may employ BLR-T models exclusively, particularly given the binary nature of the information provided to beach managers.

Communication with the public and beach managers was an important component of the nowcast system. The program communicated on a daily basis with local beach managers. The program operated a public notification program that includes a webpage, mobile apps, and beach signs; and a public awareness campaign. These communication components served to disseminate easy-to-understand water quality information to as many people as possible. Year-round communication with local beach managers who have intimate knowledge of their beaches (including historic data, beach usage, potential sources, and events in the watershed that could potentially affect water quality such as sewage spills) was crucial to the producing quality predictions.

Beach water quality is highly variable even within a day (Boehm, 2007; Whitman et al., 2004). The nowcast models presented herein provide predictions for water quality once per day, at the time of day when water samples are typically collected by beach managers. An implicit assumption in the modeling approach is that the historical FIB data were sampled over the range of climatic and oceanographic conditions that influence FIB as measured at the usual monitoring time. Nowcast models could be created to provide hourly predictions of water quality if hourly FIB data were available to calibrate and validate models. We know that sunlight affects FIB concentrations, so calibration and validation data would need to include samples collected during diverse sunlight conditions throughout the day (hourly data). Currently, such data are not routinely collected by beach managers. It is only in special circumstances that beach managers acquire such high frequency data (during a sanitary sewage spill, a source tracking study, or other special study, for example). Also, the question remains as to whether a beach manager would be interested in and able to providing hour-to-hour updates to beach users as to the water quality status.

5. Conclusions

We calibrated and validated nowcast water quality models for 10 beaches in California using historical data. We used a new approach that involved tuning multiple regression models to achieve manager-desired sensitivities and specificities. Validated models were then implemented daily for actual beach management during the 2017 swimming season using an automated code. Models performed better than or on par with the current method (in terms of sensitivity and specificity) of using day-to-week old measurements to inform beach management. The work underscores that beach-specific models are needed for California beaches, even when beaches are located in close proximity, and that flexibility and adaptability is needed to deal with unexpected model performance and unreliable data sources when implementing nowcast models for beach management. In the future, the nowcast system will expand to provide wet weather and dry weather predictions at more beaches across California. Another goal is to provide predictions earlier in the morning (~7:00 h) and explore the potential for forecasting water quality into the future.

Acknowledgements

Funding for this project has been provided through the California State Water Resources Control Board. Any opinions in this paper are those of the authors. The contents of this document do not necessarily reflect the views and policies of the State Water Resources Control Board, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. We thank Wai (Anthony) Thoe, Steve Weisberg, and Donna Francy for their input on the paper as well as our beach agency partners: County of Santa Cruz Environmental Health, Santa Barbara County Environmental Health Services, Los Angeles County Department of Public Health, Orange County Health Care Agency, and County of San Diego Department of Environmental Health.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jenvman.2018.06.058.

References

Arnold, B.F., Wade, T.J., Benjamin-Chung, J., Schiff, K.C., Griffith, J.F., Dufour, A.P., Weisberg, S.B., Colford, J.M., 2016. Acute gastroenteritis and recreational water: highest burden among young us children. Am. J. Publ. Health 106, 1690–1697. https://doi.org/10.2105/AJPH.2016.303279.

Avila, R., Horn, B., Moriarty, E., Hodson, R., Moltchanova, E., 2018. Evaluating statistical model performance in water quality prediction. J. Environ. Manag. 206, 910–919. https://doi.org/10.1016/j.jenvman.2017.11.049.

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

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

Boehm, A.B., Keymer, D.P., Shellenbarger, G.G., 2005. An analytical model of enterococci inactivation, grazing, and transport in the surf zone of a marine beach. Water Res. 39, 3565–3578.

Boehm, A.B., Lluch-Cota, D.B., Davis, K.A., Winant, C.D., Monismith, S.G., 2004a. Covariation of coastal water temperature and microbial pollution at interannual to tidal periods. Geophys. Res. Lett. 31. http://dx.doi.org/10.1029/2003GL019122.

Boehm, A.B., Shellenbarger, G.G., Paytan, A., 2004b. Groundwater discharge: a potential association with fecal indicator bacteria in the surf zone. Environ. Sci. Technol. 38, 3558–3566.

Boehm, A.B., Weisberg, S.B., 2005. Tidal forcing of enterococci at marine recreational beaches at fortnightly and semidiurnal frequencies. Environ. Sci. Technol. 39, 5575–5583.

Boehm, A.B., Whitman, R.L., Nevers, M.B., Hou, D., Weisberg, S.B., 2007. Now-casting recreational water quality. In: Wymer, L., Dufour, A. (Eds.), Statistical Framework for Water Quality Criteria and Monitoring.

Brooks, W., Corsi, S., Fienen, M., Carvin, R., 2016. Predicting recreational water quality advisories: a comparison of statistical methods. Environ. Model. Software 76, 81–94. https://doi.org/10.1016/j.envsoft.2015.10.012.

Colford Jr., 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, 27–35.

Crowther, J., Kay, D., Wyer, M.D., 2001. Relationships between microbial water quality and environmental conditions in coastal recreational waters: the fylde coast, UK. Water Res. 35, 4029.

Cyterski, M., Brooks, W., Galvin, M., Wolfe, K., Carvin, R., Roddick, T., Fienen, M., Corsi, S., 2013. Virtual beach 3: User's Guide. (No. EPA/600/R-13/311).

DeFlorio-Barker, S., Wing, C., Jones, R.M., Dorevitch, S., 2018. Estimate of incidence and cost of recreational waterborne illness on United States surface waters. Environ. Health 17, 3. https://doi.org/10.1186/s12940-017-0347-9.

Feng, Z., 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, 54–60. https://doi.org/10.1016/j.marpolbul.2016.02.011.

Francy, D., 2009. Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches. Aquat. Ecosys. Health Manag. 12, 177–182.

Francy, D.S., Brady, A.M.G., Carvin, R.B., Corsi, S.R., Fuller, L.M., Harrison, J.H., Hayhurst, B.A., Lant, J., Nevers, M.B., Terrio, P.J., Zimmerman, T.M., 2013. Developing and Implementing Predictive Models for Estimating Recreational Water Quality at Great Lakes Beaches.

Francy, D.S., Darner, R.A., 2006. Procedures for Developing Models to Predict Exceedances of Recreational Water-quality Standards at Coastal Beaches (Techniques and Methods No. 6–B5). U.S. Department of the Interior, U.S. Geological Survey.

Frick, W.E., Ge, Z.F., Zepp, R.G., 2008. Nowcasting and forecasting concentrations of biological contaminants at beaches: a feasibility and case study. Environ. Sci. Technol. 42, 4818–4824.

Given, S., Pendleton, L.H., Boehm, A.B., 2006. Regional public health cost estimates of contaminated coastal waters: a case study of gastroenteritis at Southern California beaches. Environ. Sci. Technol. 40, 4851–4858.

Gonzalez, R.A., Conn, K.E., Crosswell, J.R., Noble, R.T., 2012. Application of empirical predictive modeling using conventional and alternative fecal indicator bacteria in eastern North Carolina waters. Water Res. 46, 5871–5882.

Grant, S.B., Sanders, B.F., Boehm, A.B., Redman, J.A., Kim, J.H., Mrse, R.D., Chu, A.K., Gouldin, M., McGee, C.D., Gardiner, N.A., Jones, B.H., Svejkovsky, J., Leipzig, G.V., 2001. Generation of enterococci bacteria in a coastal saltwater marsh and its impact on surf zone water quality. Environ. Sci. Technol. 35, 2407–2416.

Haile, R.W., Witte, J.S., Gold, M., Cressey, R., McGee, C., Milikan, R.C., Glasser, A., Harawa, N., Ervin, C., Harmon, P., Harper, J., Dermand, J., Alamillo, J., Barrett, K., Nides, M., Wang, G.-Y., 1999. The health effects of swimming in ocean water contaminated by storm drain runoff. Epidemiology 10, 355–363.

Haugland, R.A., Siefring, S.C., Wymer, L.J., Brenner, K.P., Dufour, A.P., 2005. Comparison of Enterococcus measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. Water Res. 39, 559–568.

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, 2563–2573.

Hou, D., Rabinovici, S.J., Boehm, A.B., 2006. Enterococci predictions from a partial least squares regression model can improve the efficacy of beach management advisories. Environ. Sci. Technol. 40, 1737–1743.

Imamura, G.J., Thompson, R.S., Boehm, A.B., Jay, J.A., 2011. Wrack promotes the persistence of fecal indicator bacteria in marine sands and seawater. FEMS Microbiol. Ecol. 77, 40–49. https://doi.org/10.1111/j.1574-6941.2011.01082.x.

Inman, D.L., Tait, R.J., Nordstrom, C.E., 1971. Mixing in the surf zone. J. Geophys. Res. 76, 3493–3514.

Jennings, W.C., Chern, E.C., O'Donohue, D., Kellogg, M.G., Boehm, A.B., 2018. Frequent detection of a human fecal indicator in the urban ocean: environmental drivers and covariation with enterococci. Environ. Sci. Process. Impacts 20, 480–492. https://doi.org/10.1039/C7EM00594F.

Jovanovic, D., Coleman, R., Deletic, A., McCarthy, D.T., 2017. Tidal fluctuations influence E. coli concentrations in urban estuaries. Mar. Pollut. Bull. 119, 226–230. https://doi.org/10.1016/j.marpolbul.2017.04.004.

Kim, J.H., Grant, S.B., 2004. Environ. Sci. Technol. 38, 2497.

Laureano-Rosario, E.A., Symonds, M.E., Rueda-Roa, D., Otis, D., Muller-Karger, E.F., 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. https://doi.org/10.3390/ijerph14121602.

Liu, L., Phanikumar, M.S., Molloy, S.L., Whitman, R.L., Shively, D.A., Nevers, M.B., Schwab, D.J., Rose, J.B., 2006. Modeling the transport and inactivation of E. coli and enterococci in the near-shore region of Lake Michigan. Environ. Sci. Technol. 40, 5022–5028.

Mason, R.L., Gunst, R.F., Hess, J.L., 1989. Statistical Design and Analysis of Experiments: Applications to Engineering and Science.

Nevers, M.B., Whitman, R.L., 2005. Nowcast modeling of Escherichia coli concentrations at multiple urban beaches of southern Lake Michigan. Water Res. 39, 5250–5260.

Nevers, M.N., Boehm, A.B., 2010. Modeling fate and transport of fecal bacteria in surface water. In: Sadowsky, M.J., Whitman, R.L. (Eds.), The Fecal Bacteria. ASM Press, Washington DC.

Palani, S., Liong, S.-Y., Tkalich, P., 2008. An ANN application for water quality forecasting. Mar. Pollut. Bull. 56, 1586–1597. https://doi.org/10.1016/j.marpolbul.2008.05.021.

Park, Y., Kim, M., Pachepsky, Y., Choi, S.-H., Cho, J.-G., Jeon, J., Cho, K.H., 2018. Development of a nowcasting system using machine learning approaches to predict fecal contamination levels at recreational beaches in Korea. J. Environ. Qual. https://doi.org/10.2134/jeq2017.11.0425.

Paule-Mercado, M.A., Ventura, J.S., Memon, S.A., Jahng, D., Kang, J.-H., Lee, C.-H., 2016.

Monitoring and predicting the fecal indicator bacteria concentrations from agricultural, mixed land use and urban stormwater runoff. Sci. Total Environ. 550, 1171–1181. https://doi.org/10.1016/j.scitotenv.2016.01.026.
Phillips, M.C., Feng, Z., Vogel, L.J., Reniers, A.J.H.M., Haus, B.K., Enns, A.A., Zhang, Y., Hernandez, D.B., Solo-Gabriele, H.M., 2014. Microbial release from seeded beach sediments during wave conditions. Mar. Pollut. Bull. 79, 114–122. https://doi.org/10.1016/j.marpolbul.2013.12.029.
Rabinovici, S.J.M., Bernknopf, R.L., Wein, A.M., Coursey, D.L., Whitman, R.L., 2004. Economic and health risk trade-offs of swim closures at a Lake Michigan beach. Environ. Sci. Technol. 38, 2737–2745.
Raschka, S., 2018. MLXTEND. [WWW Document]. URL. https://rasbt.github.io/mlxtend/, Accessed date: 25 January 2018.
Reeves, R.L., Grant, S.B., Mrse, R.D., Oancea, C.M.C., Sanders, B.F., Boehm, A.B., 2004. Sources, management, and scaling of fecal indicator bacteria in runoff from a coastal urban watershed in Southern California. Environ. Sci. Technol. 38, 2637–2648.
Russell, T.L., Sassoubre, L.M., Wang, D., Masuda, S., Chen, H., Soetjipto, C., Hassaballah, A., Boehm, A.B., 2013. A coupled modeling and molecular biology approach to microbial source tracking at Cowell Beach, Santa Cruz, CA, USA. Environ. Sci. Technol. 47, 10231–10239.
Shanks, O.C., Sivaganesan, M., Peed, L., Kelty, C.A., Blackwood, A.D., Greene, M.R., Noble, R.T., Bushon, R.N., Stelzer, E.A., Kinzelman, J., Anan'eva, T., Sinigalliano, C., Wanless, D., Griffith, J., Cao, Y.P., Weisberg, S., Harwood, V.J., Staley, C., Oshima, K.H., Varma, M., Haugland, R.A., 2012. Interlaboratory comparison of real-time PCR protocols for quantification of general fecal indicator bacteria. Environ. Sci. Technol. 46, 945–953.
Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M.L., Boehm, A.B., 2015. Sunny with a chance of gastroenteritis: predicting swimmer risk at California beaches. Environ. Sci. Technol. 49, 423–431. https://doi.org/10.1021/es504701j.
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.
Thoe, W., Lee, J.H.W., 2014. Daily forecasting of Hong Kong beach water quality by multiple linear regression models. J Env. Eng – ASCE 140.
USEPA, 2018. BEACON - Beach Advisory and Closing On-line Notification. [WWW Document]. URL. https://ofmpub.epa.gov/apex/beacon2/f?p=137:8:::NO:::, Accessed date: 25 January 2018.
USEPA, 2012. Recreational Water Quality Criteria. (No. OFFICE OF WATER 820-F-12-058).
USEPA, 2006. Method 1600: Enterococci in Water by Membrane Filtration Using Membrane-Enterococcus Indoxyl-beta-d-glucoside Agar (MEI). Environmental Protection Agency: Office of Water, Washington D. C EPA 821-R-06-009 (No. EPA-821-R-06-009).
USEPA, 2002. Method 1603: Escherichia coli in Water by Membrane Filtration Using Modified Membrane-tolerant Escherichia coli agar (Modified M-TEC). (No. 821-NaN-02–023). Office of Water, U.S. Environmental Protection Agency., Washington, DC.
Wade, T.J., Pai, N., Eisenberg, J., Colford, J.M.J., 2003. Do US EPA water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ. Health Perspect. 111, 1102–1109.
Walters, S.P., Thebo, A.L., Boehm, A.B., 2011. Impact of urbanization and agriculture on the occurrence of bacterial pathogens and stx genes in coastal waterbodies of central California. Water Res. 45, 1752–1762. https://doi.org/10.1016/j.watres.2010.11.032.
Whitman, R.L., Nevers, M.B., Korinek, G.C., Byappanahalli, M.N., 2004. Solar and temporal effects on Escherichia coli concentration at a Lake Michigan swimming beach. Appl. Environ. Microbiol. 70, 4276–4285.
Yamahara, K.M., Layton, B.A., Santoro, A.E., Boehm, A.B., 2007. Beach sands along the California coast are diffuse sources of fecal bacteria to coastal waters. Environ. Sci. Technol. 41, 4515–4521.
Yau, V., Wade, T.J., de Wilde, C.K., Colford Jr., J.M., 2009. Skin-related symptoms following exposure to recreational water: a systematic review and meta-analysis. Water Qual Expo Health 1, 79–103.
Yau, V.M., Schiff, K., Arnold, B., Griffith, J., Gruber, J., Wright, C., Wade, T., Burns, S., Hayes, J., McGee, C., Gold, M., Cao, Y., Boehm, A.B., Weisberg, S.B., Colford, J.M., 2014. Effect of submarine groundwater discharge on bacterial indicators and swimmer health at Avalon Beach, CA, USA. Water Res. 59, 23–36.