Know before you go: Data-driven beach water quality forecasting

Ryan T. Searcy[1], Alexandria B. Boehm[1*]

1. Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA, USA, 94305

* Corresponding author: 473 Via Ortega, Room 189, Stanford, CA 94305, aboehm@stanford.edu

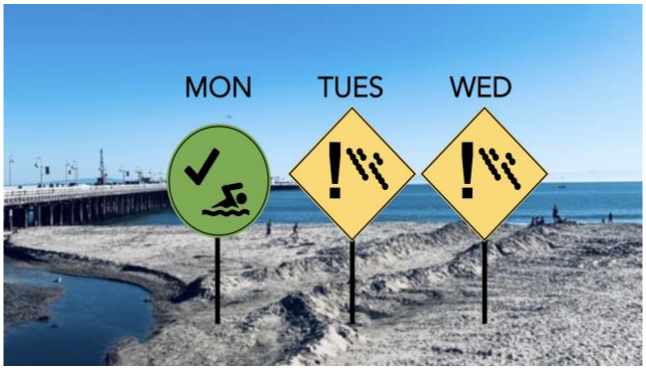
Abstract

Forecasting environmental hazards is critical in preventing or building resilience to their impacts on human communities and ecosystems. Environmental data science is an emerging field that can be harnessed for forecasting, yet more work is needed to develop methodologies that can leverage increasingly large and complex datasets for decision support. Here we design a data-driven framework that can, for the first time, forecast bacterial standard exceedances at marine beaches with three days lead time. Using historical datasets collected at two California sites, we train nearly 400 forecast models using statistical and machine learning techniques and test forecasts against predictions from both a naive 'persistence' model and a baseline nowcast model. Overall, forecast models are found to have similar sensitivities and specificities to the persistence model, but significantly higher areas under the ROC curve (a metric distinguishing a model's ability to effectively parse classes across decision thresholds), suggesting forecasts can provide enhanced information beyond past observations alone. Forecast model performance at all lead times was similar to that of nowcast models. Together, results suggest that integrating the forecasting framework developed in this study into beach management programs can enable better public notification and aid in proactive pollution and health risk management.

Keywords: data-driven models, water quality forecasting, machine learning

Synopsis: New environmental data science methodologies are needed to improve environmental hazard prediction. This study presents a framework to forecast regulatory exceedances of fecal indicator bacteria with three days lead time.

TOC Art:



Introduction

Environmental hazards including earthquakes, heat waves, wildfires, disease outbreaks, and acute water pollution threaten ecosystems and human communities around the planet. For example, it is estimated that weather- and climate-related disasters in the US have been responsible for more than 9,000 deaths and costs greater than $1.8 trillion since the year 2000.[1] Forecasting these events is important to better enable disaster mitigation and resilience, but is often difficult to do accurately due to the rare and complex nature of these events.[2,3]

One challenge of hazard forecasting relates to the heterogeneity of environmental systems. It is difficult to predict the frequency and strength of hazards because they are functions of the interaction of multiple natural phenomena; they often vary in space in different regions of the planet; and they can change over time (especially as climate change and anthropogenic impacts accelerate).[2,4] Perhaps resulting from this heterogeneity, many ways of issuing environmental forecasts have been developed ranging from completely non-technical (such as using animal behavior to forecast earthquakes[5]) to entirely computational (such as global climate models that require the use of supercomputers).

Another barrier to effective forecasting relates to the nature of the data representing environmental systems which are often available from multiple sources in multiple spatial and temporal resolutions; noisy and sparse due to the difficulty in data collection (particularly for biological parameters); and skewed due to the rarity of events. Data-driven techniques - which take advantage of statistical and algorithmical relationships amongst datasets and are the primary tools of the emerging field of *environmental data science* - can be effective in extracting meaning from complex data.[4] Particularly, environmental data science is well-equipped to leverage the increasingly large and more real-time environmental datasets that are available from sources such as remote sensing platforms,[6,7] ecosystem monitoring stations,[8,9] individual sampling campaigns, and model output.[10,11] As a result, data-driven hazard forecasting applications previously documented include predicting flooding,[12] air pollution,[13] foreign species invasion,[14] and harmful algal blooms.[15,16]

However, new and enhanced methodologies are still needed for data-driven environmental forecasts to have broad application. This is partially due to the unique complexity of environmental data which provides a barrier to using methods that have been successful in forecasting applications in other fields. For example, time series analysis techniques such as ARIMA have been successfully used for predicting financial outcomes but can break down when applied to environmental time series that are heavily-skewed (as is the case with rare event prediction) or unevenly spaced.[17] Developing new methods for specific environmental data science problems will improve the translation of complex datasets and subsequently improve decision support for environmental management.[4]

A specific opportunity to enhance environmental data science methodology is in the management of recreational beach water quality. At beaches around the planet, fecal indicator bacteria (FIB) - organisms that can be indicative of the presence of enteric pathogens - are monitored in water. However, FIB monitoring is often conducted infrequently (e.g. weekly or less often)[18] and relies on laboratory methods that delay result availability 24-48 hours; thus, beach management decisions and subsequent public notification often do not reflect current water quality conditions.[19] To augment monitoring, data-driven models have been used previously to predict FIB standard exceedances at beaches.[20–22] FIB models are often regression- or machine learning-based, and use environmental observations like tide, wave, and meteorological parameters as inputs to make predictions. Models can be more sensitive than the ‘persistence method’ of using a single day-old measurement to represent water quality[23,24] and can provide more frequent information to beach managers and beachgoers than sampling alone.[25]

Though often referred to as providing water quality ‘forecasts’,[26–28] most models are designed to issue FIB ‘nowcasts’; technically, model outputs are near real-time predictions of current water quality conditions rather than true forecasts which are predictions of future conditions. Nowcasts have limited use because in cases where water quality is predicted to be poor, same-day predictions give managers only a few hours to conduct adaptive sampling and make beach management decisions. Water quality forecasts – or predictions of water quality issued one or more days in advance – could alleviate these limitations, allowing for more effective beach management from local agencies as well as providing beachgoers more time to decide where to recreate based on water quality.

Few studies have investigated FIB forecasting. Frick et al.[29] converted a linear regression-based nowcast model developed for Lake Erie to provide forecasts for one day in the future. To do this, they substituted observed model variables with their forecast versions as model inputs. Zhang et al.[30] used wavelet analysis and neural networks to forecast FIB in Lake Michigan up to 24 hours in advance without using independent environmental variables. Both studies focus on lacustrine beaches with forecast lead times (the hours or days between when a prediction is issued and when it is valid) of one day or less. While there are examples of forecasting other beach parameters days in advance (such as tide level, harmful algal blooms, and air quality),[7,31,32] to our knowledge no work exists testing FIB forecasts at marine beaches.

The objective of the present study is to develop and test a data-driven framework that can effectively forecast FIB at marine beaches up to three days in advance. Our approach is based on methodology that is commonly applied in the development of nowcast models, including using a standard data science ‘pipeline’ (or workflow) to develop models and applying well-studied regression and machine learning model types. However, a key difference between nowcast methodology and our framework is that model inputs are intentionally limited to environmental observations made at least the number of days in advance of the time for which the FIB forecast is made. This means that as forecast lead time increases, the larger the gap in time between FIB and the observed model inputs. This approach is supported by knowledge that predictive information can be found in environmental observation temporally lagged from an FIB observation.[33,34]

Using forecasting datasets composed of historical observations of FIB and environmental data, we apply a custom data science workflow to develop a series of machine learning models for two California marine beaches. We evaluate the models' ability to forecast FIB standard exceedances for one, two, and three days lead time, and assess the most common environmental parameters for each time step. We compare forecast performance to that of both naive models and of baseline nowcast models. The results of this work will allow practitioners to extend beyond nowcasting applications to make FIB predictions days in advance. Further, this work serves as a foundation that can built upon with future work to improve forecasting of poor beach water quality as well other environmental hazards.

Methodology

Study Sites

Two California marine beaches were chosen for the study: Cowell Beach (CB, 36.962 N, 122.023 W) and Huntington State Beach (HSB, 33.633 N, 117.966 W) (Figure S1). Both sites have a Mediterranean climate in which most precipitation events occur between the months of November and March. Sites were selected based on popularity with beachgoers, historical data availability, and differences in geomorphology, climate, and water quality conditions; a detailed description of the sites including notes on recent infrastructure changes can be found in the Supporting Information.

FIB and Environmental Data

FIB Data

FIB monitoring is conducted year-round at both sites; a description of monitoring methodology can be found in the SI. *Escherichia coli* (EC) and Enterococcus (ENT) data were collated from a

State of California online database (http://ceden.org/index.shtml). Samples collected during the summer months (April-October) between 2007 and 2021 were used in this analysis. Beach sampling occurred approximately weekly at CB and twice weekly at HSB; no significant serial correlation was found in these datasets. Samples measured below the limit of quantification (LOQ) were flagged and assigned a value of ½ the LOQ. FIB samples with concentration in exceedance of the State of California's regulatory standard (104 and 400 CFU/100 ml for EC and ENT samples, respectively)[35] were also flagged.

Environmental Data

Historical environmental data were compiled for each beach from internet sources through manual download or API access; specific station details (including the sampling interval for each data type) are provided in Table S1. Oceanic data include tide, wave, and water quality parameters. Tide level predictions based on harmonic constants were available from the National Oceanic and Atmospheric Administration (NOAA). We used tidal predictions as opposed to observations because tides can be forecast accurately years in advance and NOAA maintains a historical archive of tide predictions (which is required for forecast model training, see below)[36]. Wave parameters including significant wave height, average wave period, and dominant wave period were measured by regional buoys maintained by the Coastal Data Information Program (CDIP). These buoys also collected water temperature data. Other water quality parameters such as chlorophyll, turbidity, dissolved oxygen, pH, salinity, and conductivity were measured by automated pier-based stations maintained by the Central and Northern California Ocean Observing System (CenCOOS).

Meteorological data include air and dew point temperatures; wind speed and directions; precipitation totals; and solar irradiance. These parameters were aggregated from multiple meteorological stations maintained by the National Climatic Data Center (NCDC), the California Irrigation Management Information System (CIMIS), and CenCOOS. Stream discharge data collected from automated USGS gauges located in the San Lorenzo River and Santa Ana River were used in CB and HSB analyses, respectively. Finally, alongshore and crosshore current velocities measured by high-frequency radar maintained by CenCOOS were used in HSB analyses. All data sources provide quality assurance information on their websites to verify data veracity.

Forecast Modeling

Overview

We developed a suite of models that can forecast up to three days in advance whether FIB concentrations at a given beach will exceed regulatory standards based on the input of environmental variables. Rather than a single model that could at once forecast FIB at all times in the future, FIB forecasts were made by three individual models each developed to predict FIB on a specific day in the future. The following terminology and notation will be used henceforth to describe forecast models. The time upon which a prediction is output by a model is the *forecast issue time*, while the time in the future for which a forecast is made is the *forecast validity time*. The difference between the forecast issue and validity times is the *forecast lead time*. In this study, the forecast validity time (i.e. the time for which a forecast is made) will be assigned a timestamp $t = 0$. It then follows that a forecast of lead time L is issued by the model on $t = -L$ where the units of L are days. This frame of reference enables consistency when organizing modeling datasets for models of varying lead times, and is equivalent to a frame of reference where forecast issuance (rather than validity) occurs at $t = 0$. For example, using variables observed prior to $t = -3$ to issue a forecast for $t = 0$ (i.e. a three day lead time) is equivalent to a forecast valid at $t = 3$ issued by a model with input variables observed prior to $t = 0$.

Model development included aggregating FIB and environmental variables into bulk datasets, pre-processing bulk modeling datasets by reducing dimensionality and removing missing values; partitioning data into training and test datasets; selecting the final model variables; and fitting and evaluating the predictive model (Figure 1). This process was followed for each model developed for a given beach, FIB type, data partition, forecast lead time, and model type. Model development was performed using the Python programming language and specifically using the *scikit-learn* machine learning package.[37]

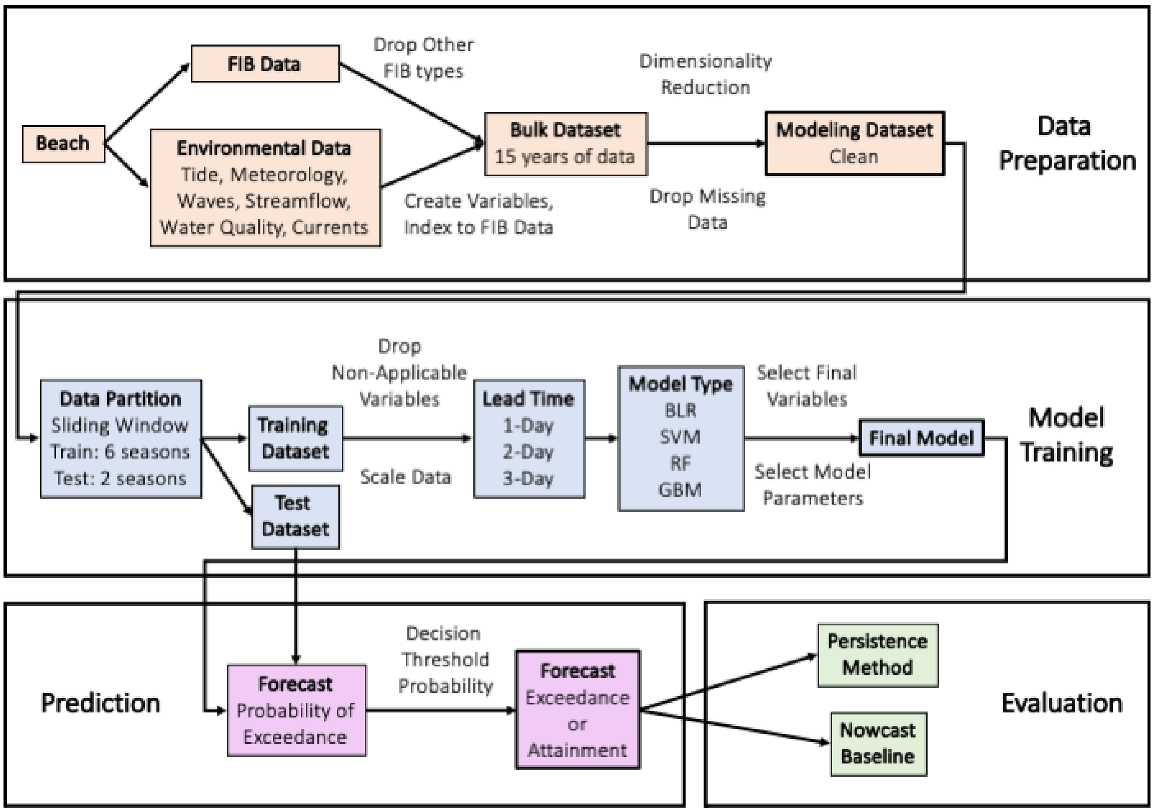


Figure 1: Flow diagram of the development of an individual forecast model. Procedure is repeated for each beach, FIB type, data partition, lead time, and model type.

Modeling Datasets

Bulk modeling datasets for a given beach were composed of historical FIB observations (dependent variable) and environmental (independent) variables. Environmental variables were created by time indexing raw data to each FIB observation at a given beach. Because the temporal resolution of the FIB datasets are on the order of days, raw environmental data with higher frequency sampling resolution were transformed into bulk daily statistics (e.g. daily maximum tide level, daily mean significant wave height, the sum of precipitation over the previous seven days). Categorical variables were created by grouping by environmental condition on a given day. Such categorizations include whether there was spring or neap tide; during periods of discrete wind and current directions (offshore or onshore and upshore or downshore); and when wind speed, significant wave height, dominant wave period, chlorophyll, and turbidity were above or below their historical 75th percentile (which served to indicate relatively elevated levels of the condition).

Temporally lagged variables were created by employing a temporal shift of up to 7 days between environmental variables and FIB observations. Examples include mean air temperature observed two days prior, dichotomized wind conditions observed four days prior, and total precipitation observed between three and seven days prior. While lagged variables have been shown to significantly improve the predictive performance of FIB nowcast models,[33] they are necessary in this work primarily because of operational constraints specific to forecasting; that is, only environmental observations made on or before the forecast issuance time (which lags from the forecast validity time) can be used. The exception to this were tidal parameters. The instantaneous tide level at the time of FIB sampling was also used as a model variable because this variable is available for operational FIB forecasting as tide data were forecasts themselves.

In addition to the aforementioned environmental categorizations, we created variables that indicated the month of sample collection and whether samples were collected on weekends. Precipitation, streamflow, chlorophyll, and turbidity parameters were $\log_{10}$+1 transformed (i.e. $\log_{10}$-transformed after adding 1) to reduce skew. Finally, we chose not to include autoregressive FIB variables (e.g. the most recent FIB measurement) as independent variables due to the unevenly spaced FIB time series available at the study beaches. A maximum of 266 environmental variables spanning seven types (meteorological, tide, wave, water quality, streamflow, current, and date) were included in bulk modeling datasets (Table

S2.1 - 2.7). It should again be emphasized that all environmental variables except for tide variables were composed of measured data as opposed to forecast (or modeled) data; this is primarily due to poor availability of forecast environmental data. All data used in this study are available in the Stanford Digital Repository and can be accessed at https://purl.stanford.edu/nw799cp6263.[38]

The bulk datasets represent an aggregation of data from multiple sources and a range of temporal lags, and thus often contained hundreds of modeling variables and included missing data points. Pre-processing was required prior to individual model development in order to reduce the likelihood of overfitting (which can occur when training models on a large number of variables) and to yield clean modeling datasets.

Dataset dimensionality was performed by first removing irrelevant variables from the bulk modeling datasets. This included zero-variance variables, variables related to observations of FIB of the other type (i.e. EC and ENT were modeled independently), and variables (except for tidal) representing data observed on the same day of an FIB observation as they are often not available for use operationally in a forecast model. To further reduce dataset dimensionality as well as multicollinearity, highly correlated environmental variables (Spearman Rank Correlation > 0.9) were identified and the variable of the pair with the lowest magnitude Spearman correlation with FIB concentrations were dropped from the dataset. Remaining variables with a number of missing data points greater than 15% of the total length of the dataset were dropped. A threshold of 15% was selected because we found it balanced dataset length (i.e. number of observations available for model training and testing) and maximizing the total number of the variables available for inclusion in models. Finally, if missing values remained, the entire record for those time points (i.e. the FIB observation and the environmental variables indexed to it) was omitted from the modeling data set. We chose not to employ imputation because there were often large gaps in the environmental variable datasets (i.e. spanning seasons) in which imputation could not be useful. Further, dropping missing values did not significantly reduce the number of data points available for model training and testing (i.e. at least 85% of datapoints maintained. These steps yielded a clean modeling dataset specific to a given beach and FIB type that enabled consistency when developing models of varying lead times.

Data Partitioning

Modeling datasets were partitioned into training and test datasets. Six consecutive summer seasons of data were included in the training datasets which were used to select final modeling variables and parameters and fit models. The subsequent 2 seasons of data composed the test datasets which were used to evaluate predictive performance of the forecasts.

Multiple data partitions were created using a sliding window with a step of one year over the range of the entire modeling dataset. For example, if one partition included a training dataset with data collected between 2007 - 2012 and a test dataset with data collected during 2013 and 2014, the following partition would entail a training dataset with data collected between 2008 - 2013 and a test dataset with data collected during 2014 and 2015. Thus, 8 total data partitions per modeling dataset were available.

Environmental variables were subsequently standardized in order to optimize model training. Variables were first centered by subtracting the variable’s mean value and subsequently scaled by dividing by the variable’s variance. The means and variances from the partition’s training dataset were used to standardize the data in both the training and test datasets.

Forecast Model Training

The data of each partition were subsequently used to develop models for all three forecast lead times as well as baseline nowcast models. Prior to training forecast models of a given lead time, a final removal of environmental variables (with the exception of tide variables which are technically available for all lead times because they are composed of forecast data) was conducted such that the models were fit only on data available for operational use. For example, total observed precipitation on t = -2 is available for use in nowcast (lead time of 0) and 1-day forecast models, but is operationally unavailable for 2- and 3-day forecast models. This final removal led to an increasingly smaller number of variables available for model training as lead time increased.

Models were trained as binary classifiers. Predictions were made in two-steps: first outputting the probability of whether FIB concentrations were in exceedance of the water quality regulatory standard, and then converting that probability to either the positive class (i.e. FIB standard exceedance) or negative class (i.e. attainment or 'non-exceedance' of FIB standard) based on if it was above or equal to or below the decision threshold probability, respectively. The default decision threshold probability used was 0.5.

We trained four models for each forecast lead time using the following statistical and machine learning model types: binary logistic regression, support vector machine, random forest, and gradient boosted machine. Each model type is available in the *scikit-learn* package in Python, and has been used previously to predict FIB or other water quality parameters.[39–42] Binary logistic regression (BLR) is a statistical model that has an interpretable relationship between environmental variables and predicted FIB. We used a BLR model with an 'elastic net' penalty. Support vector machine (SVM) is a model that employs kernel functions which nonlinearly map data such that the hyperplane that separates predicted classes is optimized. Random forest (RF) is an ensemble model where the prediction is an aggregate of the predictions of multiple decision 'trees'. Gradient boosted machine (GBM) is similar to RF in that multiple weak estimators are fit as an ensemble; however, each subsequent tree fit is 'boosted' by learning the error resulting from the previous. Default model parameters are listed in Table S3.

Models were trained using a process that used cross-validation to first select final model variables from the full training dataset and then optimize model parameters from the training datasets. Balanced accuracy (the mean of model sensitivity and specificity, defined below) was used as the cross-validation score metric in all steps of this process. Variable selection involved a two-step process. The first step further reduced the dimension of the training dataset by using the *permutation feature importance* algorithm in the *scikit-learn* package.[37] The permutation feature importance (PFI) of a given variable is indicative of how dependent a model is on that variable and is calculated as the change in model score metric upon training a model after randomly shuffling the variable's data. Each variable was shuffled five times, and the mean PFI resulting from the shuffling was calculated. Variables with mean PFI less than 1.5 times the grand mean PFI of the entire variable set were dropped from the training dataset. A value of 1.5 was selected because it was found to balance model training time and variety for subsequent variable selection.[43]

The second step in variable selection was a recursive feature elimination algorithm with cross-validation. The training dataset was first split into five 'folds' or subsets, four of which were used to fit submodels and the remaining used to cross-validate the submodel's score. In a stepwise manner, the algorithm dropped the one variable from the dataset such that the submodel score (i.e. balanced accuracy) on the validation data is maximized. Variable removal was repeated until the minimum number of submodel variables remained (for this study, we set this parameter to 3). The entire stepwise process was repeated five times using all combinations of data folds for submodel fitting and validation. Upon completion, the set of variables corresponding to the highest average cross-validation score was selected as the final modeling variable set.

A grid search algorithm with cross-validation was then used to optimize the model parameters specific to the given model type. Similarly to recursive feature elimination, grid search involved splitting the training dataset into five folds, fitting submodels with varying model parameters on four folds, and evaluating the model score on the fifth. The procedure was repeated until all combinations of model parameters were exhausted. The parameters used in the search for each model type are listed in Table S3. The entire parameter search was repeated a total of five times exhausting all combinations of data folds for submodel fitting and validation. The model with parameters that maximized the average cross-validation score averaged over the submodels was used as the final model.

Forecast Evaluation

The test datasets were used to evaluate the predictive performance of models. This was done by running models to output forecasts of whether FIB exceeded regulatory standards, and comparing predictions to FIB measurements. Three performance metrics were used to evaluate models: sensitivity, specificity, and area under the receiver operating characteristic curve. *Sensitivity* is defined as the proportion of positive values (i.e. FIB standard exceedances) correctly predicted by the model; models with higher sensitivity are more health protective. *Specificity* is defined as the proportion of negative values (i.e. FIB in attainment of the standard) correctly predicted. It is valuable to consider both sensitivity and specificity because beach managers typically desire models to be effective in predicting both days of FIB standard exceedances and attainment. *Area under the receiver operating characteristic curve (AUC)* is a bulk metric that integrates sensitivity and specificity across a range of potential decision threshold probabilities, and is less biased than those metrics when comparing performance between models tested on datasets with differing proportions of FIB standard exceedances.[21,44] AUC ranges from 0 to 1, with a value of greater than 0.5 indicating that a model’s predictive ability is greater than guessing at random and a value of 1 indicating perfect delineation of positive and negative classes. Aggregate performance was calculated across all developed models, by individual model type, and for all models developed for a given beach and FIB type.

Model performance was contextualized by comparing these metrics to those calculated for the ‘persistence method’, a naive model which assumes the forecast FIB condition at a beach is equivalent to that indicated by the most recently collected observation prior to forecast issuance. The persistence method is the means by which beach management in California is currently conducted, and thus serves as a practical baseline to evaluate model performance.

Finally, performance metrics for forecast models were also compared to those calculated for nowcast models (lead time of 0) developed for the same data partition and model type. This allowed for evaluation on how forecasts of increasing lead time compared to the nowcast baseline on the same data.

Results

FIB and Environmental Data

Bulk modeling datasets for each beach were composed of FIB and environmental data collected over 15 summer seasons (April through October from 2007 - 2021) (Table S4). HSB had a higher proportion of samples measured below the LOQ (66% and 61% of EC and ENT samples, respectively) than CB (8% and 43%). Of the two FIB types considered in this study, EC exceeded the regulatory standard most frequently at CB (19% of all samples) while ENT exceeded most frequently at HSB (8%).

Forecast Model Development

The proportion of FIB exceedances varied between training and test datasets within partitions and between individual partitions, reflecting changing FIB distributions over the years (Table S5). The average proportion of FIB exceedances at CB (13% and 11% for EC and ENT test datasets, respectively) was higher than that for HSB (4% and 8%).

Across all combinations of beach, FIB type, dataset partition, forecast lead time, and model type, we developed 384 total forecast models. An additional 128 nowcast models were developed for baseline comparison.

A total of 130 unique variables across all models trained were selected through the variable selection process. The most common variable types selected in fitting forecast models were meteorological (appearing in 97% of models), tide (91%), and wave (83%) variables, while currents and water quality variables were the least common (appearing in 0% and 17% of models, likely owing to the poor availability of their data). Instantaneous tide level was the most common individual variable selected, appearing in 43% of models. The most frequently selected variables for models developed for each beach and FIB are listed in Table S6.

The number of modeling variables selected differed by model type (Figure S2). RF and GBM tended to have fewer variables (mean of 7 and 9, respectively) than SVM and BLR (mean of 11 and 12). Across all forecast models, the mean number of selected variables in a model was 10 and the maximum was 38. The number of variables in individual forecast models tended to decrease with increasing model lead time, mostly due to the increasingly limited availability of variables to select from.

Forecast Model Performance

Overall Performance

FIB forecasts were made by running models on the test datasets and were compared to observations in order to evaluate predictive performance. Across all forecast models, median sensitivity was 0.00 (interquartile range (IQR) 0.00 - 0.31), median specificity was 0.92 (IQR 0.77 - 0.97), and median AUC was 0.58 (IQR 0.50 - 0.66). The majority of forecast models we developed (76%) had AUC values greater than 0.5 (i.e. more informative than random guessing). Comparatively, the persistence method's median sensitivity was 0.00 (IQR 0.00 - 0.15), median specificity was 0.95 (IQR 0.89 - 0.98), and median AUC was 0.50 (IQR 0.48 -

0.51). Further, the persistence method had 32% of instances with AUC greater than 0.5. At this aggregate level, forecast models do not appear to perform well in terms of sensitivity and specificity compared to the persistence model; however, we next disaggregate the results to investigate if specific models perform well.

Metrics grouped by individual model type showed that certain model types tended to outperform the persistence model (Figure 2, Table S7). Across all beaches, FIB types, data partitions, and lead time, the most sensitive model type was BLR (median of 0.40, IQR 0.22 - 0.56) while the least sensitive was SVM (0.0, IQR 0.0 - 0.0). Specificity was highest for SVM models (1.00, IQR 1.00- 1.00), because all predictions were for the negative class. The next most specific model type was GBM (0.95, IQR 0.92 - 0.97); the lowest specificities came from BLR models (0.69, IQR 0.62 - 0.76). AUC was highest for RF (0.60, IQR 0.54 - 0.68) and lowest for SVM (0.56, IQR 0.48 - 0.65). BLR and RF were the model types with AUC most frequently greater than 0.5 (80% and 79% of models, respectively), while SVM was the least frequent (68%). No single model type was superior in all of the evaluation metrics, yet depending on the performance criteria of the end user (i.e. required sensitivity or specificity), each can be an effective model type for operational forecasting.

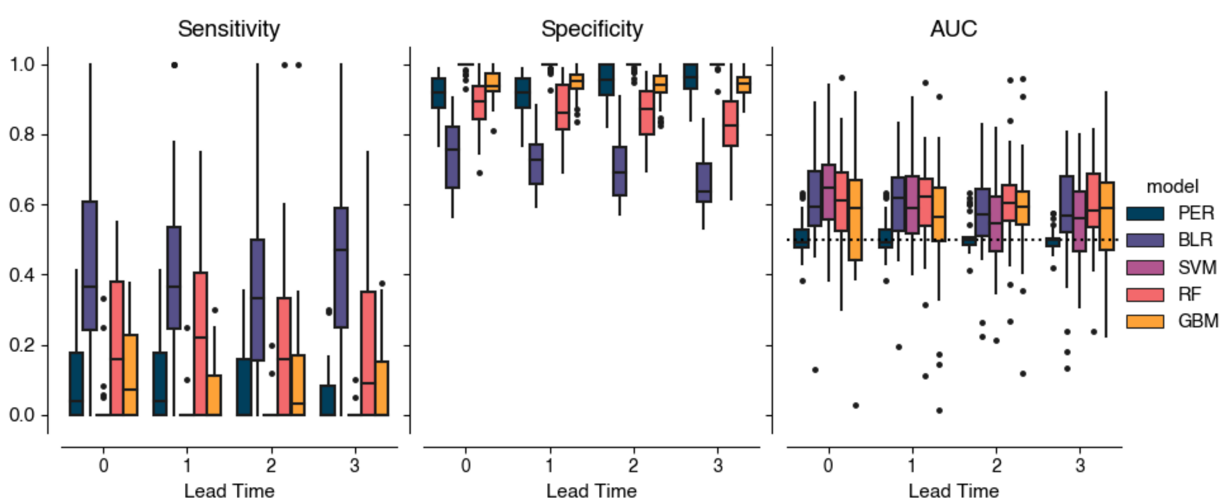


Figure 2: Predictive performance of forecast models on test datasets. Model sensitivities, specificities, and AUC values are plotted in the left, middle, and right subplots, respectively. The dotted line in the AUC

subplot is set at a value of 0.5 for reference. Performance metrics are categorized by forecast lead time (x-axis). The lead time zero boxplots represent the results from the baseline nowcast models. Color indicates model type; PER corresponds to the persistence method. The middle line in the boxplots represents the median; the upper and lower edges of the boxes represent the 75th and 25th quantiles, respectively. The whiskers extend to 1.5 times the interquartile range (75th quartile−25th quartile).

Forecast models are developed specific to a given beach and FIB type, so we next aggregate performance metrics as such and show that forecast models perform well relative to the persistence method for beach management. For brevity, we present results below for EC at CB and ENT at HSB, yet metrics for all beach and FIB type groupings are listed in Table S8. For EC at CB, median sensitivity was 0.12 (IQR 0.0 - 0.34) for forecast models compared to 0.11 (0.0 - 0.26) for the persistence method; median specificity was 0.91 (0.71 - 0.98) for forecast models and 0.90 (0.85 - 0.95) for the persistence method; and median AUC was 0.58 (0.48 - 0.65) for models and 0.51 (0.47-0.58) for the persistence method. For ENT at HSB, median sensitivity was 0.10 (IQR 0.0 - 0.41) compared to 0.0 (0.0 - 0.0) for the persistence method; median specificity was 0.92 (0.77 - 0.99) for models and 0.97 (0.94 - 1.0) for the persistence method; and median AUC was 0.62 (0.56 - 0.70) and 0.50 (0.48-0.50) for the persistence method.

Comparison to Nowcast Models

We developed 128 nowcast models using the same methodology described above for forecast models, but allowing for the inclusion of the most recently observed variables (i.e. from one day previous of prediction validity). Compared to persistence method predictions with a lead time of 0, nowcast models had higher median sensitivity, specificity, and AUC (Table S9).

Overall, average predictive performance of nowcast models was similar to those of forecast models of all three forecast lead times. We assessed how performance changed between individual models of a given lead time and their associated baseline nowcast model (N

= 128 model comparisons per lead time). Model sensitivity decreased in a minority of forecast models (24%, 26%, and 27% for 1, 2, and 3 days lead time, respectively) relative to the associated nowcast model; forecast model sensitivity remained unchanged relative to nowcast for approximately half of forecast models of each lead time (likely an artifact of many models having sensitivities of 0.0). Model specificity decreased in 40%, 45%, and 45% of forecast models relative to the associated nowcast model for 1, 2, and 3 days lead time, respectively, while it increased in approximately one third of forecast models of each lead time respectively. Thus the median difference in specificity between forecast and nowcast models was 0.0 for all three lead times. Finally, forecast model AUC decreased in 59%, 57%, and 61% relative to nowcast models for 1, 2, and 3 days lead time, respectively; the median change in AUC was small (0.02, 0.02, and 0.04 for one, two, and three days lead time, respectively). Cumulatively, these results suggest that a comparable quality of information can be provided by forecast models as from their baseline nowcast models.

Discussion

We established an automated framework that can be used to develop data-driven models that forecast FIB at marine beaches for up to three days lead time. Our methodology extends beyond that used to develop FIB nowcast models by leveraging temporally-lagged environmental observations and multiple statistical and machine learning model types. We used the framework to train nearly 400 forecast models for two California beaches and tested their predictions against FIB observations. To our knowledge, this is the first demonstration that FIB levels at marine beaches can be predicted days in advance.

Our results show that forecast models can provide enhanced beach water quality information beyond single past observations alone (i.e. the persistence method). FIB concentrations tend to exhibit ‘patchy’ time series and can be heavily skewed, making them difficult to predict with autoregressive techniques alone. Predictive models instead rely on observed environmental parameters which can be monitored more frequently than FIB and represent the physical, chemical and biological processes that control FIB fate and transport in the environment. Moreover, the statistical and machine learning models we considered in the framework have the potential to capture nonlinear relationships between FIB and environmental drivers.

An underlying design component specific to the forecasting framework is to utilize the knowledge that there is often a delayed effect between a given environmental driver (e.g. precipitation, wind-driven flow) and the expressed effect in FIB fate; this is represented in our framework through the use of lagged observed environmental variables. For example, precipitation observed six days prior was often included as a model variable, perhaps due to a six day gap between rainfall and contaminated runoff reaching the study site. Site-specific studies devoted to exploring FIB fate and transport mechanisms would be needed to verify this.

However, the effect of some environmental mechanisms on FIB fate can be more immediate (e.g. water temperature, solar irradiance).[45–47] Thus the presence of lagged environmental variables in models (which was more common as forecast lead time increased) may be more related to variable autocorrelation than a delayed effect on FIB. This may explain why solar irradiance observed four days prior was a frequently included parameter in models; if a parameter observed on or prior to forecast issuance is strongly correlated with its observation on the time of forecast validity $t = L$, then the lagged parameter acts more like a forecast variable.

It is interesting to note that instantaneous tide level - the single forecast variable used in this study - was the most often selected variable in FIB forecast models. Tide level is regarded as a parameter that can be forecast with high accuracy, and it has also been shown to be important in nowcast models developed previously at the study sites.[25,43] Future studies could examine the accuracy of forecasts of additional environmental parameters and the efficacy of their inclusion as variables in forecast models. We expect precipitation forecasts could be beneficial model variables that may lead to improved model performance, especially during periods when precipitation occurs between forecast issuance and validity.. In this study we chose not to use forecast variables (aside from tide level) because model accuracy is affected by the accuracy of the forecast variables (as is evidenced by Frick et al.)[29] and historical records of forecast environmental parameters are often unavailable from the data source. However, a concerted effort to archive a sufficient number of environmental forecasts for model training could be made, or forecasts could be created using time series techniques.[48]

Predictive metrics of forecast models were similar to those of the comparative baseline nowcast models, suggesting that the predictive efficacy of a forecast modeling system can be maintained for at least three days. This result is somewhat surprising considering forecast accuracy of most environmental parameters is generally expected to decline with increasing lead time. That predictive metrics were similar between forecasts and nowcasts may further validate the notion that information predictive of FIB can be found in environmental observations collected less recently than solely the closest days to prediction validity. Future work could extend our framework to lead times of one week (or longer) to determine the lead times at which predictive efficacy drops significantly.

Despite frequently having improved performance over the persistence method, the sensitivities of forecast models (developed using the default decision threshold probability of 0.5) were generally low (median of 0.0 overall, median of 0.4 for BLR models) and thus may be unacceptable for use for beach management where swimmer health protection is the priority. However, it should be noted that the status quo has even lower sensitivity than the forecast models. Low sensitivity could be due to a number of data limitations including site infrastructure changes which alter the FIB distributions between training and test sets, or lack of environmental data that truly explain FIB fate and transport at the site. An additional and likely limitation could be due to the imbalanced nature of the FIB datasets at the specific beaches we tested, where many more samples were in attainment than in exceedance of the regulatory standard; this is evidenced by the poor performance of the persistence method. The effect of class imbalance on model sensitivity has been observed previously by Thoe et al. (2015).[23] To account for low default sensitivities, their group optimized their models by 'tuning' the decision threshold probability at which a prediction is for a positive or negative class on the training data. This increased the sensitivity of their BLR models on test data from 19% to 50%. When we tuned the decision threshold probabilities of a random subset of our models, we saw an improvement in sensitivity with a drop in specificity that would likely be acceptable for beach management[25] (see the Supporting Information for more details). However, because there is a tradeoff between model sensitivity and specificity, models should be tuned based on the risk tolerance of the forecast user.

Comparison of the AUC of our models to those in the literature show that forecast models (agnostic of tuning) can perform on par to what is already expected in the water quality prediction field. Our models (particularly, RF and GBM models) were able to distinguish between positive and negative classes similarly to those developed by Brooks et al. .[21] Over the 15 nowcast model types they tested, the median AUC was 0.67 (IQR 0.63-0.71) which is similar to ours.

The framework presented here serves as a foundation for future beach water quality forecasting research and implementation. Additional potential research topics include assessing the effect of training dataset size (i.e. the number of observations) on model performance; exploring different model outputs like categorical or probabilistic predictions; testing deep learning model types such as long short-term memory neural networks (which was out of scope of the present study);[12,16] and incorporating more locally-collected environmental data as well as previous FIB observations as predictor variables (which may become operationally feasible with the onboarding of rapid[49] and autonomous[50] sampling methods that enable higher resolution datasets).

This framework can be employed by health agencies, community groups, or other beach water quality stakeholders who desire FIB forecasts up to three days in advance. With adequate support, the framework can also be integrated into existing nowcast modeling systems. Attention must be given to the time indexing of environmental data; specifically, we suggest organizing environmental variables such that forecast issuance occurs at $t = 0$. Many environmental data sources can be accessed via the internet, enabling the automation of modeling pipelines and the issuance of daily forecasts of all lead times simultaneously. Forecasts can be compared to routine monitoring data as they are collected, and models can be re-tuned or retrained based on what performance level is desired by operators. Further, forecasts could be used as a basis to conduct proactive sampling in the case that FIB exceedances are predicted. Thus, water quality forecasts would provide agencies and beach communities much more information than what sampling alone provides, leading to improved environmental awareness and management.

Though our framework was developed specifically for FIB prediction, we expect it may be useful for predicting other environmental hazards that have similarly structured and available data (such as harmful algal blooms or shellfish toxicity). We encourage the extension of this framework to other environmental prediction applications especially as the availability of unique environmental datasets and novel modeling types continues to grow. Developing new techniques or extending upon existing methodologies is an important step in advancing environmental data science and thus better understanding our interactions with the natural environment.

Data Availability Statement

The data that support the findings of this study are openly available at the Stanford Digital Repository: https://purl.stanford.edu/nw799cp6263. The Python code used to manipulate data and develop and test models can be found at: https://github.com/rtsearcy/wq-forecasting

Acknowledgements

This study was supported by the University of Southern California Sea Grant Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, under grant number NA18OAR4170075.

Supporting Information: Model tuning case study. List of environmental variables used for modeling and the stations from which data were acquired. Grid search parameters for each model type. Data partition summaries. Most frequent model variables. Predictive performance by model type and by beach.

References

(1) National Centers for Environmental Information (NCEI). *U.S. Billion-dollar Weather and Climate Disasters, 1980 - present*. https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:0209268 (accessed 2022-08-10).

(2) Sun, Y.; Wong, A. K. C.; Kamel, M. S. Classification of Imbalanced Data: A Review. *Int. J. Pattern Recognit. Artif. Intell.* **2009**, *23* (04), 687–719. https://doi.org/10.1142/S0218001409007326.

(3) Ho, C.-H.; Bhaduri, M. On a Novel Approach to Forecast Sparse Rare Events: Applications to Parkfield Earthquake Prediction. *Nat. Hazards* **2015**, *78* (1), 669–679. https://doi.org/10.1007/s11069-015-1739-1.

(4) Blair, G. S.; Henrys, P.; Leeson, A.; Watkins, J.; Eastoe, E.; Jarvis, S.; Young, P. J. Data Science of the Natural Environment: A Research Roadmap. *Front. Environ. Sci.* **2019**, *7* (121). https://doi.org/10.3389/fenvs.2019.00121.

(5) Wikelski, M.; Mueller, U.; Scocco, P.; Catorci, A.; Desinov, L. V.; Belyaev, M. Y.; Keim, D.; Pohlmeier, W.; Fechteler, G.; Martin Mai, P. Potential Short-Term Earthquake Forecasting by Farm Animal Monitoring. *Ethology* **2020**, *126* (9), 931–941. https://doi.org/10.1111/eth.13078.

(6) Xu, Y.; Vaughn, N. R.; Knapp, D. E.; Martin, R. E.; Balzotti, C.; Li, J.; Foo, S. A.; Asner, G. P. Coral Bleaching Detection in the Hawaiian Islands Using Spatio-Temporal Standardized Bottom Reflectance and Planet Dove Satellites. *Remote Sens.* **2020**, *12* (19), 3219. https://doi.org/10.3390/rs12193219.

(7) Izadi, M.; Sultan, M.; Kadiri, R. E.; Ghannadi, A.; Abdelmohsen, K. A Remote Sensing and Machine Learning-Based Approach to Forecast the Onset of Harmful Algal Bloom. *Remote Sens.* **2021**, *13* (19), 3863. https://doi.org/10.3390/rs13193863.

(8) Searcy, R. T.; Boehm, A. B.; Weinstock, C.; Preston, C. M.; Jensen, S.; Roman, B.; Birch, J. M.; Scholin, C. A.; Van Houtan, K. S.; Kiernan, J. D.; Yamahara, K. M. High-Frequency and Long-Term Observations of EDNA from Imperiled Salmonids in a Coastal Stream: Temporal Dynamics, Relationships with Environmental Factors, and Comparisons with Conventional Observations. *Environ. DNA* **2022**, *4* (4), 776–789. https://doi.org/10.1002/edn3.293.

(9) Truelove, N. K.; Patin, N. V.; Min, M.; Pitz, K. J.; Preston, C. M.; Yamahara, K. M.; Zhang, Y.; Raanan, B. Y.; Kieft, B.; Hobson, B.; Thompson, L. R.; Goodwin, K. D.; Chavez, F. P. Expanding the Temporal and Spatial Scales of Environmental DNA Research with Autonomous Sampling. *Environ. DNA* **2022**, *4* (4), 972–984.

https://doi.org/10.1002/edn3.299.

(10) Chan, S. N.; Thoe, W.; Lee, J. H. W. Real-Time Forecasting of Hong Kong Beach Water Quality by 3D Deterministic Model. *Water Res.* **2013**, *47* (4), 1631–1647. https://doi.org/10.1016/j.watres.2012.12.026.

(11) Panda, U. S.; Pradhan, U. K.; Kumar, S. S.; Naik, S.; Begum, M.; Mishra, P.; Murthy, M. V. R. Bathing Water Quality Forecast for Chennai Coastal Waters. *J. Coast. Res.* **2020**, *89* (sp1), 111–117. https://doi.org/10.2112/SI89-019.1.

(12) Le, X.-H.; Ho, H. V.; Lee, G.; Jung, S. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* **2019**, *11* (7), 1387. https://doi.org/10.3390/w11071387.

(13) Freeman, B. S.; Taylor, G.; Gharabaghi, B.; Thé, J. Forecasting Air Quality Time Series Using Deep Learning. *J. Air Waste Manag. Assoc.* **2018**, *68* (8), 866–886. https://doi.org/10.1080/10962247.2018.1459956.

(14) Srivastava, V.; Griess, V. C.; Padalia, H. Mapping Invasion Potential Using Ensemble Modelling. A Case Study on Yushania Maling in the Darjeeling Himalayas. *Ecol. Model.* **2018**, *385*, 35–44. https://doi.org/10.1016/j.ecolmodel.2018.07.001.

(15) Cruz, R. C.; Reis Costa, P.; Vinga, S.; Krippahl, L.; Lopes, M. B. A Review of Recent Machine Learning Advances for Forecasting Harmful Algal Blooms and Shellfish Contamination. *J. Mar. Sci. Eng.* **2021**, *9* (3), 283. https://doi.org/10.3390/jmse9030283.

(16) Shan, K.; Ouyang, T.; Wang, X.; Yang, H.; Zhou, B.; Wu, Z.; Shang, M. Temporal Prediction of Algal Parameters in Three Gorges Reservoir Based on Highly Time-Resolved Monitoring and Long Short-Term Memory Network. *J. Hydrol.* **2022**, *605*, 127304. https://doi.org/10.1016/j.jhydrol.2021.127304.

(17) Ömer Faruk, D. A Hybrid Neural Network and ARIMA Model for Water Quality Time Series Prediction. *Eng. Appl. Artif. Intell.* **2010**, *23* (4), 586–594. https://doi.org/10.1016/j.engappai.2009.09.015.

(18) Heal the Bay. *Annual Beach Report Card 2020*. https://healthebay.org/beachreportcard2020/ (accessed 2020-10-28).

(19) Kim, J. H.; Grant, S. B. Public Mis-Notification of Coastal Water Quality: A Probabilistic Evaluation of Posting Errors at Huntington Beach, California. *Environ. Sci. Technol.* **2004**, *38* (9), 2497–2504. https://doi.org/10.1021/es034382v.

(20) Avila, R.; Horn, B.; Moriarty, E.; Hodson, R.; Moltchanova, E. Evaluating Statistical Model Performance in Water Quality Prediction. *J. Environ. Manage.* **2018**, *206*, 910–919. https://doi.org/10.1016/j.jenvman.2017.11.049.

(21) Brooks, W.; Corsi, S.; Fienen, M.; Carvin, R. Predicting Recreational Water Quality Advisories: A Comparison of Statistical Methods. *Environ. Model. Softw.* **2016**, *76*, 81–94. https://doi.org/10.1016/j.envsoft.2015.10.012.

(22) Francy, D. Use of Predictive Models and Rapid Methods to Nowcast Bacteria Levels at Coastal Beaches. *Aquat. Ecosyst. Health Manag.* **2009**, *12* (2), 177–182.

(23) Thoe, W.; Gold, M.; Griesbach, A.; Grimmer, M.; Taggart, M. L.; Boehm, A. B. Sunny with a Chance of Gastroenteritis: Predicting Swimmer Risk at California Beaches. *Environ. Sci. Technol.* **2015**, *49* (1), 423–431. https://doi.org/10.1021/es504701j.

(24) Francy, D. S.; Brady, A. M. G.; Cicale, J. R.; Dalby, H. D.; Stelzer, E. A. Nowcasting Methods for Determining Microbiological Water Quality at Recreational Beaches and Drinking-Water Source Waters. *J. Microbiol. Methods* **2020**, 105970. https://doi.org/10.1016/j.mimet.2020.105970.

(25) Searcy, R. T.; Taggart, M.; Gold, M.; Boehm, A. B. Implementation of an Automated Beach Water Quality Nowcast System at Ten California Oceanic Beaches. *J. Environ. Manage.* **2018**, *223*, 633–643. https://doi.org/10.1016/j.jenvman.2018.06.058.

(26) Thoe, W.; Lee, J. H. W. Daily Forecasting of Hong Kong Beach Water Quality by Multiple Linear Regression Models. *J. Environ. Eng.* **2014**, *140* (2), 04013007. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000800.

(27) Vijayashanthar, V.; Qiao, J.; Zhu, Z.; Entwistle, P.; Yu, G. Modeling Fecal Indicator Bacteria in Urban Waterways Using Artificial Neural Networks. *J. Environ. Eng.* **2018**, *144* (6), 05018003. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001377.

(28) Palani, S.; Liong, S.-Y.; Tkalich, P. An ANN Application for Water Quality Forecasting. *Mar. Pollut. Bull.* **2008**, *56* (9), 1586–1597. https://doi.org/10.1016/j.marpolbul.2008.05.021.

(29) Frick, W. E.; Ge, Z. F.; Zepp, R. G. Nowcasting and Forecasting Concentrations of Biological Contaminants at Beaches: A Feasibility and Case Study. *Environ. Sci. Technol.* **2008**, *42* (13), 4818–4824. https://doi.org/10.1021/es703185p.

(30) Zhang, J.; Qiu, H.; Li, X.; Niu, J.; Nevers, M. B.; Hu, X.; Phanikumar, M. S. Real-Time Nowcasting of Microbiological Water Quality at Recreational Beaches: A Wavelet and Artificial Neural Network-Based Hybrid Modeling Approach. *Environ. Sci. Technol.* **2018**, *52* (15), 8446–8455. https://doi.org/10.1021/acs.est.8b01022.

(31) Kim, S.; Mehrotra, R.; Kim, S.; Sharma, A. Probabilistic Forecasting of Cyanobacterial Concentration in Riverine Systems Using Environmental Drivers. *J. Hydrol.* **2021**, *593*, 125626. https://doi.org/10.1016/j.jhydrol.2020.125626.

(32) Ma, J.; Cheng, J. C. P.; Lin, C.; Tan, Y.; Zhang, J. Improving Air Quality Prediction

Accuracy at Larger Temporal Resolutions Using Deep Learning and Transfer Learning Techniques. *Atmos. Environ.* **2019**, *214*, 116885. https://doi.org/10.1016/j.atmosenv.2019.116885.

(33) Cyterski, M.; Zhang, S.; White, E.; Molina, M.; Wolfe, K.; Parmar, R.; Zepp, R. Temporal Synchronization Analysis for Improving Regression Modeling of Fecal Indicator Bacteria Levels. *Water. Air. Soil Pollut.* **2012**, *223* (8), 4841–4851. https://doi.org/10.1007/s11270-012-1240-3.

(34) Boehm, A. B.; Grant, S. B.; Kim, J. H.; Mowbray, S. L.; McGee, C. D.; Clark, C. D.; Foley, D. M.; Wellman, D. E. Decadal and Shorter Period Variability of Surf Zone Water Quality at Huntington Beach, California. *Environ. Sci. Technol.* **2002**, *36* (18), 3885–3892. https://doi.org/10.1021/es020524u.

(35) Wayne. *CA Assembly Bill 411: Beach Sanitation: Posting.*; 1997.

(36) National Oceanographic and Atmospheric Administration (NOAA). *Tide Prediction Error for the United States Stations*. https://tidesandcurrents.noaa.gov/pdf/Tide_Prediction_Error_for_the_United_States_Coastline.pdf (accessed 2022-08-17).

(37) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825–2830.

(38) Searcy, R. T.; Boehm, A. B. *Data for FIB Forecasting Project*. Stanford Digital Repository. https://purl.stanford.edu/nw799cp6263 (accessed 2022-08-15).

(39) Thoe, W.; Gold, M.; Griesbach, A.; Grimmer, M.; Taggart, M. L.; Boehm, A. B. Predicting Water Quality at Santa Monica Beach: Evaluation of Five Different Models for Public Notification of Unsafe Swimming Conditions. *Water Res.* **2014**, *67*, 105–117. https://doi.org/10.1016/j.watres.2014.09.001.

(40) Naloufi, M.; Lucas, F. S.; Souihi, S.; Servais, P.; Janne, A.; Wanderley Matos De Abreu, T. Evaluating the Performance of Machine Learning Approaches to Predict the Microbial Quality of Surface Waters and to Optimize the Sampling Effort. *Water* **2021**, *13* (18), 2457. https://doi.org/10.3390/w13182457.

(41) Xu, T.; Coco, G.; Neale, M. A Predictive Model of Recreational Water Quality Based on Adaptive Synthetic Sampling Algorithms and Machine Learning. *Water Res.* **2020**, 115788. https://doi.org/10.1016/j.watres.2020.115788.

(42) Park, Y.; Kim, M.; Pachepsky, Y.; Choi, S.-H.; Cho, J.-G.; Jeon, J.; Cho, K. H.

Development of a Nowcasting System Using Machine Learning Approaches to Predict Fecal Contamination Levels at Recreational Beaches in Korea. *J. Environ. Qual.* **2018**. https://doi.org/10.2134/jeq2017.11.0425.

(43) Searcy, R. T.; Boehm, A. B. A Day at the Beach: Enabling Coastal Water Quality Prediction with High-Frequency Sampling and Data-Driven Models. *Environ. Sci. Technol.* **2021**, *55* (3), 1908–1918. https://doi.org/10.1021/acs.est.0c06742.

(44) Madani, M.; Seth, R. Evaluating Multiple Predictive Models for Beach Management at a Freshwater Beach in the Great Lakes Region. *J. Environ. Qual.* **2020**, *49* (4), 896–908. https://doi.org/10.1002/jeq2.20107.

(45) Boehm, A. B.; Lluch-Cota, D. B.; Davis, K. A.; Winant, C. D.; Monismith, S. G. Covariation of Coastal Water Temperature and Microbial Pollution at Interannual to Tidal Periods. *Geophys. Res. Lett.* **2004**, *31* (6). https://doi.org/10.1029/2003GL019122.

(46) Olyphant, G. A.; Whitman, R. L. Elements of a Predictive Model for Determining Beach Closures on a Real Time Basis: The Case of 63rd Street Beach Chicago. *Environ. Monit. Assess.* **2004**, *98* (1–3), 175–190. https://doi.org/10.1023/B:EMAS.0000038185.79137.b9.

(47) Desai, A. M.; Rifai, H. S. Escherichia Coli Concentrations in Urban Watersheds Exhibit Diurnal Sag: Implications for Water-Quality Monitoring and Assessment. *J. Am. Water Resour. Assoc.* **2013**, *49* (4), 766–779. https://doi.org/10.1111/jawr.12039.

(48) Qin, M.; Li, Z.; Du, Z. Red Tide Time Series Forecasting by Combining ARIMA and Deep Belief Network. *Knowl.-Based Syst.* **2017**, *125*, 39–52. https://doi.org/10.1016/j.knosys.2017.03.027.

(49) Lucius, N.; Rose, K.; Osborn, C.; Sweeney, M. E.; Chesak, R.; Beslow, S.; Schenk, T. Predicting E. Coli Concentrations Using Limited QPCR Deployments at Chicago Beaches. *Water Res. X* **2019**, *2*, 100016. https://doi.org/10.1016/j.wroa.2018.100016.

(50) Yamahara, K. M.; Demir-Hilton, E.; Preston, C. M.; Marin, R.; Pargett, D.; Roman, B.; Jensen, S.; Birch, J. M.; Boehm, A. B.; Scholin, C. A. Simultaneous Monitoring of Faecal Indicators and Harmful Algae Using an In-Situ Autonomous Sensor. *Lett. Appl. Microbiol.* **2015**, *61* (2), 130–138. https://doi.org/10.1111/lam.12432.