

Predicting *E. coli* concentrations using limited qPCR deployments at Chicago beaches

Nick Lucius, many more and Tom Schenk Jr.

November 07, 2017

Culture-based methods to measure *Escherichia coli* (*E. coli*) are used by beach administrators to inform whether bacteria levels represent an elevated risk to swimmers. However, these methods take up to 12 hours for processing. Statistical models are used to forecast bacteria levels in lieu of test results. Often, these results underestimate days with elevated *E. coli* levels. In Chicago, these models only accurately predicted elevated levels 6.5% percent of the time in 2015. EPA approval of enterococci quantitative PCR (qPCR) methods allows for testing that is available within 3 hours, eschewing the need for predictive models. Unfortunately, these “rapid testing” methods are 100% to 400% more expensive than culture-based methods. This paper presents a prediction model which uses limited deployments of “rapid testing” sites with inter-beach correlation to predict when bacteria will exceed acceptable thresholds. The model can be used to inform management decisions on when to warn or close beaches due to exposure to the bacteria. Using data from Chicago, we find that the approach would have issued accurate advisories for 90 poor water quality “beach days” compared to 15 advisories issued by the previous model. The accuracy of the model rose from 3.2% to 11.2%—a 250% increase.

1 Introduction

Managers of recreational beaches use culture-based methods to monitor bacteria levels at recreational beaches. Sampling is conducted early in the morning, but despite those efforts, results take upward of 12 hours (Kinzelman et al. 2003). Between the time of sampling and subsequent results, beach conditions will often change so the water sample is not relevant to today’s beach conditions (Whitman, Nevers, and Gerovac 1999; A. B. Boehm et al. 2002). To get around this, researchers have developed statistical models to estimate water quality conditions for the day (Francy et al. 2013). These models account for the past trend at the beaches and other factors such as precipitation, wind, and water conditions.

Yet, these models usually do not correctly identify days with elevated bacteria levels (known as Type II errors or “false negatives”) (Nevers and Whitman 2011; Rabinovici et al. 2004; A. B. Boehm et al. 2002). For instance in 2015, Chicago beaches had 200 events where bacteria levels were too high; however, only 13 (6.5%) of these events were forecasted by the predictive model. As we will discuss, these models do have good overall fit, but elevated bacteria levels are statistically rare events and models often fail to predict them.

Meanwhile, scientists have developed new methods which measure bacteria levels in water with substantially less delay. *Enterococci* quantitative polymerase chain reaction (qPCR) methods can determine bacteria levels within 3 to 4 hours and yield similar results as culture-based methods (Haugland et al. 2005; Kinzelman et al. 2003).

However, this approach has a drawback of cost and equipment availability. qPCR testing can cost between 2 to 5 times more than traditional culture-based methods (News and News 2017). Thus, managers are faced with a dilemma of choosing between expensive qPCR methods or choosing slower culture-based methods and using predictive models to produce swim advisories.

While culture-based and qPCR measure two completely different markers for bacteria, both of these tests have generally-acceptable thresholds for acceptable bacteria levels. Acceptable levels for culture-based methods should not exceed 235 CFU/100 ml while acceptable levels for qPCR testing should not exceed 1,000 cell equivalents (CE) (Byappanahalli et al. 2010). Within the context of beach management, these rules can

be leveraged to create a new predictive model which mixes the short turn-around of qPCR with statistical models to produce swim advisories.

This paper proposes a hybrid model which uses readily available results from qPCR testing at select beaches to predict whether bacteria levels will be high at other beaches. We exploit the historical correlation between beaches to estimate water quality readings based solely on qPCR results that are available on the same day. This approach allows limited deployment of qPCR equipment to reduce overall costs but provides a higher quality statistical model. As a result, these predictions can be used to “nowcast” water quality conditions.

This method eschews time-series based methods, which seem unreliable between summers and, as mentioned, fail to predict when bacteria levels are too high. This paper discusses attempts to formulate time-series models—including the introduction of ensemble models—but we conclude that time-series models often exhibit inconsistency between beach seasons and relatively low rates of true positives and high rates of false negatives.

We find that the hybrid model is more accurate and consistent between seasons. Using data from 20 beaches managed by the Chicago Park District, we use 10 years of historical measurements to create a model to forecast whether bacteria levels at a beach will be elevated. During the summer of 2017, a operationalized version of the model was able to increase the accuracy of prediction of beaches by 250 percent. For years which recorded data exist (2015 - 2016), the existing predictive model typically correctly suggested 15 advisories. In 2017, the piloted process, using selective rapid testing and predictive modeling, correctly suggested 90 advisories. Meanwhile, the model did not increase the rate of falsely identified days of high bacteria levels (known as Type I errors or “false positives”).

2 Background

Timing is a crucial factor for monitoring bacteria levels. We’ve adopted the typical shorthand to denote time periods by t to denote now; $t - 1$, $t - 2$, $t - n$ to denote last period, two periods ago, and *some* periods ago; and $t + 1$, $t + 2$, $t + n$ to denote the next period, two periods from now, and *some* period in the future. For the purposes of this paper, we often use t to denote *today*, $t - 1$ to denote *yesterday*, and $t + 1$ to denote *tomorrow*.

2.1 Prior-day Nowcast Model

Researchers have developed forecasting models using prior bacteria testing results at beaches and other available data such as precipitation (Ackerman and Weisberg 2003; Morrison et al. 2003), sunlight (Whitman et al. 2004), wind (P. Smith et al. 1999; Olyphant and Whitman 2004), wave and tidal levels (Le Fevre and Lewis 2003; Crowther, Kay, and Wyer 2001), lake levels (Francy, Bertke, and Darner 2009), turbidity (Olyphant et al. 2003), and density of humans and animals (Alexandria B. Boehm et al. 2003; Reeves et al. 2004).

These models rely on prior-day data since culture-based testing is not available until upward of 12 hours after the samples were collected. The prediction at beach x_i at time t is dependent on the prior periods culture-based results, x_i^{t-1} and other covariates, w_1, w_2, \dots, w_j . These models often take the form of

$$x_i^t = f(x_i^{t-1}, w_1, w_2, \dots, w_j)$$

where $f(\dots)$ is a some function or algorithm that inputs raw data and outputs a probability. For instance, the linear regression model is typically $x_i^t = \beta_1 x_i^{t-1} + \beta_2 w_1 + \beta_3 w_2 + \dots + \beta_j w_j$, where β_j are the coefficients that weight the importance of each input. But, models often use log transformations (Nevers and Whitman 2005), polynomial coefficients (Frick, Ge, and Zepp 2008), logistic regression, partial least squares (Hou, Rabinovici, and Boehm 2006), generalized boosted regression modeling (Cyterski et al. 2014), random forest, and genetic algorithm approaches (Brooks et al. 2016). This class of models often has the same structure of using *E. coli* levels from time $t - 1$ —the prior-day—and covariates from time t to predict beaches at time t .

Shively et al. (2016) implemented a model which automatically collected covariates on water and atmospheric conditions (e.g., water temperature, turbidity) through hydrometeorological sensors placed in the water and on shore. This technology allowed automated “nowcasts” of beaches to rely on prior-day bacteria levels and current water and atmospheric conditions. Moreover, it allowed the automation of forecasts based on available data.

3 Statistical Models

3.1 Existing Methodology

One of the first automated prior-day nowcast models was developed by Shively et al. (2016). Between 2011 and 2016, Chicago Park District teamed with scientists to use this model (hereafter to as “Chicago prior-day model”) to forecast elevated bacteria levels in Lake Michigan. It used air and water sensors to estimate the covariates, w_i , that indicate future levels of bacteria and culture-based measurements of bacteria for the previous day, x_{t-1} .

Buoys were installed at five Chicago beaches—Foster, Montrose, Oak, 63rd Street, and Calumet—to collect turbidity, wave height, wave period, water temperature, and depth of the sensor. Weather stations were installed at three beaches—Foster, Oak, and 63rd Street—and collected wind direction, wind speed, air temperature, rainfall, solar radiation, relative humidity, and barometric pressure. Data was collected from the hydrometeorological sensors between every 2 and 5 minutes through the summer and aggregated.

Similar to other models discussed, it used a multivariate model between covariates and the log of *E. coli* measured at each beach. The authors used a regression tree to determine the factors that portend elevated *E. coli* levels. Once predictions were obtained, they were published online for beach visitors at cpdbeaches.com and archived online at data.cityofchicago.org.

3.2 Ensemble Modeling

Typically, previous research has only explored one of the above models or compared individual models (Brooks et al. 2016). In this section, we explore various different prior-day nowcast models and use ensemble modeling techniques to combine various statistical and machine-learning models. Ensemble models allows two or more models to be combined to yield predictions. This allows weaknesses from any single model to be mollified by other statistical models. Often, this yields better and more stable predictions.¹

...

3.3 Hybrid Model

The proposed hybrid model relies on the observed correlation of bacteria level between beaches, as suggested by Whitman and Nevers (2008). Bacteria levels at Chicago beaches often fluctuate with each other on the same day where extreme highs and extreme lows were simultaneous for most beaches (figure 1). It is therefore possible to use selected sampling at each beach to infer bacteria levels at other beaches were sampling was not available.

Hybrid modeling uses selected observations (x_i^t) from today t at some (n) beaches, that is $X^t : x_1^t, x_2^t, \dots, x_n^t$, to forecast bacteria levels y_i^t at other (m) beaches: $Y^t : y_1^t, y_2^t, \dots, y_m^t$. In order to derive the predictions, we first determine clusters, $J(c_k)$ of beaches which have similar bacteria levels. Each cluster contains at least one beach to be sampled each day while the rest of the beaches are not. Then, we use machine learning techniques to derive predictions for each beach based on qPCR data.

¹By analogy, ensemble modeling is akin to asking a panel of experts instead of relying on a single expert.

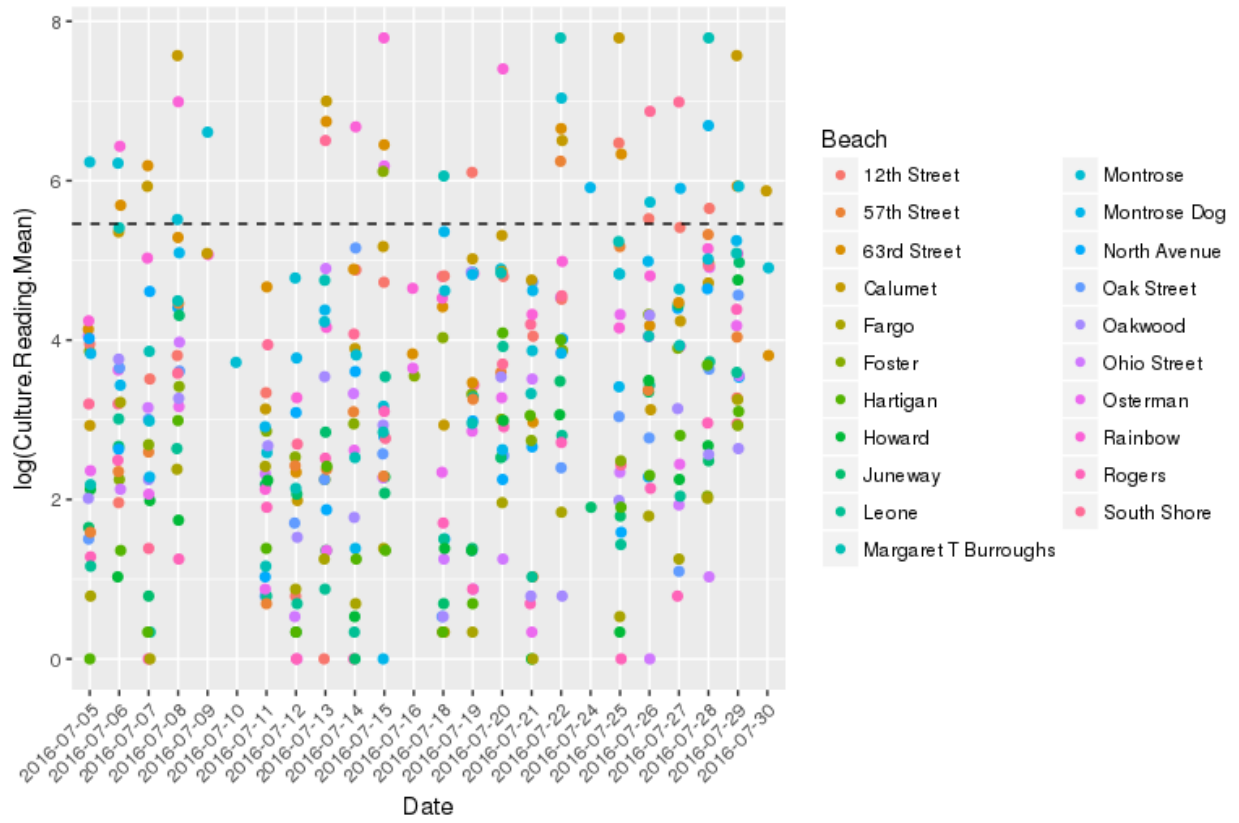


Figure 1: Chicago Beaches often fluctuate with each other. The dotted line represents the logarithm of the typical beach advisory decision threshold (235 CFU/100ml)

This hybrid model attempts to recast the research into a missing value problem. That is, from a research perspective, use information from the observed beaches to “fill-in” the data. Missing-value problems are common and often tackled in statistics in a variety of ways. For example, researchers can “fill-in” responses to surveys that were left partially blank. The general approach to missing-value problems is using observed correlations to fill-in missing values.

3.3.1 Identifying Beach Clusters

We used clustering techniques to determine groups of beaches that tend to move in tandem. Specifically, we identified beaches where bacteria levels and advisories are correlated with each other. We used K-means clustering algorithm to detect when beaches have similar movements.

Given a number of desired clusters (k), the objective is to group a beach (i) with a cluster of other beaches (B_k) together that minimize the sum of squares. Mathematically:

$$J(c_k) = \sum_{i=1}^k \sum_{x_i \in B_k} \|x_i - \mu_k\|^2$$

It is an algorithm which cannot be solved as a closed-form solution. Instead, the algorithm provides an initial guess for clusters and runs over multiple iterations to find the clusters which minimize the overall error. First, there is an initial, random guess for which beaches belong in clusters. The distance or error is then measured, $x_i - \mu_k$ for all clusters. Then, membership of clusters is slightly altered, the error is then measured and determined if the error is increasing or decreasing. This process is repeated until the lowest measured error is obtained.

We chose to limit to 5 clusters ($k = 5$) because it aligned to the number of qPCR testing sites by Chicago Parks District. K-means was applied to beaches based on latitude, longitude, total *E. coli* exceedances, and length of the longest breakwater. Each of these variables was scaled and centered by calculating z-scores prior to clustering.

Some beaches were removed from the clustering because they were historical outliers for bacteria levels or had distinct physical features. Namely, we removed beaches that have lengthy breakwaters which have an impact on bacteria levels at those beaches. We tested this hypothesis by measuring the distance of the southernmost part of the beach to the northeasternmost edge of the breakwater. The correlation between the breakwater and bacteria exceedances between 2006 and 2017 were positive ($R^2 = 0.54$). We removed 63rd Street, Rainbow, Montrose, and Ohio, which have long breakwaters or a similar feature. Likewise, our earlier analytical modeling showed that beaches with a high frequency of high bacteria levels often confounded the model. Calumet, which had high exceedances as well as a medium-sized breakwater, was also removed from the analytical model. These 5 removed beaches comprised two of the clusters from the K-means analysis.

The remaining three clusters were reanalyzed using 5-cluster k-means with the same variables.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Foster	North Avenue	Leone	31st	South Shore
Osterman	Oak Street	Juneway	12th	57th
Albion		Howard	39th	
		Rogers		
		Jarvis		

Table 1: Final results of K-means clustering.

Within each cluster, the beach with the most *E. coli* exceedances was selected to be rapid tested as model input. By rapid testing the beaches with the most frequent exceedances, we capture an added operational

benefit of maximizing the number of correct advisories. The remaining beaches were selected to be predicted by the model.

Feature Beaches (to be rapid tested)	Predicted Beaches
South Shore	57th
31st	39th
North Avenue	12th
Foster	Oak Street
Leone	Osterman
	Albion
	Juneway
	Howard
	Rogers
	Jarvis

Table 2: Hybrid model features (inputs), and beaches to be predicted.

3.3.2 Building the Predictive Model

A random forest regression model was trained with the following features: the predicted beach’s name and 10 years of culture-based test results for each of the feature beaches. The model was fit to predict the culture-based result for the predicted beach on the same day. The model was tuned and validated using leave-one-year-out cross validation, in which the validation set consisted of all the observations in one year, and the training set consisted of all other observations. The process was repeated until each year had been in the validation set. The year 2016 was left out as a final validation set.

Historically, the Chicago prior-day model from 2015 through 2016 had a false-positive rate of 1.8 percent. That is, the model falsely predicted elevated bacteria levels 1.8 percent of all beach days over the course of the summer—typically 30 beach-days per season. We chose to align the parameters of our model to this historical false-positive rate. As Figure 2 shows, we could increase the number of times we correctly predict elevated bacteria levels, but it comes at a cost of also increasing the number of days we provide unnecessary warnings to beach-goers.

Figure 2 shows the receiver operating characteristic (ROC), which shows the trade-off between increasing the number of days with true predictions versus days with falsely identify elevated levels. By constraining the model to only have a 1.8 percent false-positive rate, the true positive rate is forecasted to be 22 percent.

A threshold was chosen to transform the model prediction to a binary outcome. To keep the model’s false positive rate (FPR) near 1.8 percent, the threshold within each year that corresponded to a 1.8 percent was noted, and the mean threshold was then used to generate predictions for the holdout validation set.

In summer 2017, the model concept was piloted with some limitations imposed by available data. The feature beaches were Calumet, Rainbow, South Shore, 63rd Street, and Montrose, and the predicted beaches were the other 15 regularly tested beaches. The model was trained using qPCR test results for 2015 and 2016 and fit to predict the culture-based levels for the predicted beach on the same day. An application was developed and deployed which regularly checked Chicago’s public data portal for qPCR results for the feature beaches. Every day during the summer, once all the qPCR results were posted, the application automatically ran the model and uploaded the predictions on the public data portal within five minutes.

3.3.3 Validating the Algorithm - Summer 2017

The pilot analytical model was created using historical data from 2015 and 2016. However, to correctly assess the accuracy of the model, we need to use out-of-sample predictions. As can happen with machine learning,

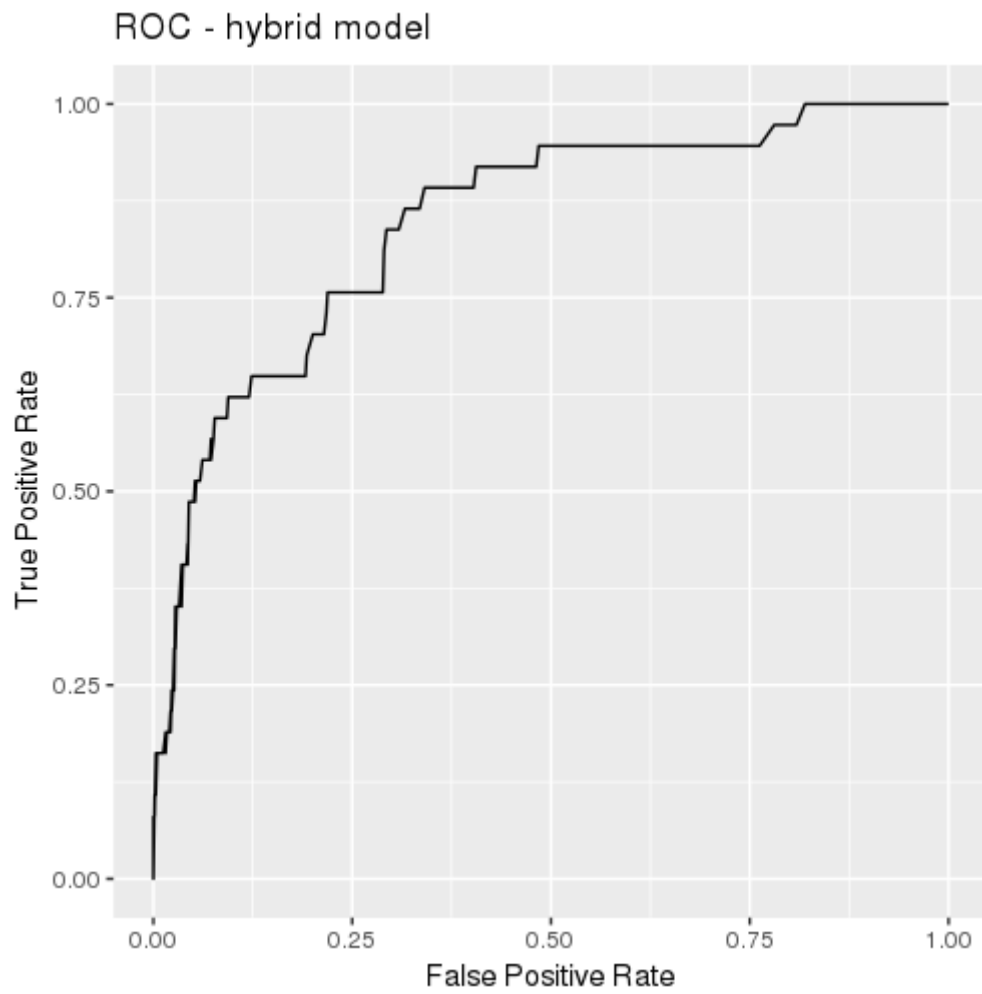


Figure 2: Receiver Operating Characteristic (ROC) Plot for hybrid model.

it is possible that the accuracy of the model was caused by the analysts tuning the model to fit existing data and accompanying noise. Fitting existing data too well can lead to the model failing to accurately predict when encountering new data.

Using the training data through 2016, we used the summer of 2017 to predict elevated levels and then measure the accuracy at the end of the season. This provided the same scenario as beach forecasting models: read the most recent data, formulate a prediction, collect water samples, and then compare the predictions to the collected samples.

qPCR measurements from the 5 beaches noted above were used to predict the qPCR levels at the other 15 regularly tested beaches. Beginning on Friday, May 26, water samples were collected by the Chicago Park District and tested by the University of Illinois at Chicago. Those results were then posted to the City of Chicago Data Portal when they were publicly available—typically around noon. About 5 minutes later, we posted the predictions based on our model to the Data Portal. The last predictions and samples were conducted on September 4th.

Both predictions and samples were collected each day during the week. This process was different from prior years where samples were usually collected only during weekdays and non-holidays, except when levels had been elevated on the previous Friday or it was suspected to be a busy weekend on the beaches.

To translate the predicted E. coli level to an advisory decision, the threshold level used to validate the original model was 381 CFU/100ml. Table 3 shows the confusion matrix for the results of the pilot. Each quadrant shows the relationship between the prediction (whether the model provided an advisory) versus the actual results from qPCR testing (whether there were levels above 381 CFU/100ml).

2017 Pilot	Predicted Elevated	Predicted Normal
Actual Elevated	9 (11.2%)	71 (88.8%)
Actual Normal	24 (2.0%)	1186 (98.0%)

Table 3: Hybrid model 2017 pilot results.

Over the summer, the analytical model had a true positive rate of 11.2 percent compared to a true positive rate of 3.2 percent in 2016. From 2015 through 2016, the historical average true positive rate for those beaches was 3.3 percent. That is, the hybrid model increased the true positive rate by 250 percent over the historical average. Meanwhile, the false positive rate was slightly higher at 2.0 percent during 2017. Our target rate was 1.8 percent to remain consistent with prior summers.

In terms of impact on beachgoers, the hybrid model issued 90 correct advisories while there were 71 incorrect advisories. While we cannot directly compare it to the Chicago prior-day models, in 2016 that model issued 12 correct advisories while issuing 184 incorrect advisories. Likewise in 2015, the prior-day model issued 14 correct advisories versus 184 incorrect advisories.

On the other hand, the hybrid model identified 24 days as having elevated levels, when the actual bacterial level was normal.

4 Summary

References

- Ackerman, Drew, and Stephen B. Weisberg. 2003. “Relationship Between Rainfall and Beach Bacterial Concentrations on Santa Monica Bay Beaches.” *Journal of Water and Health* 1 (2): 85–89.
- Boehm, A. B., S. B. Grant, J. H. Kim, S. L. Mowbray, C. D. McGee, C. D. Clark, D. M. Foley, and D. E. Wellman. 2002. “Decadal and Shorter Period Variability of Surf Zone Water Quality at Huntington Beach,

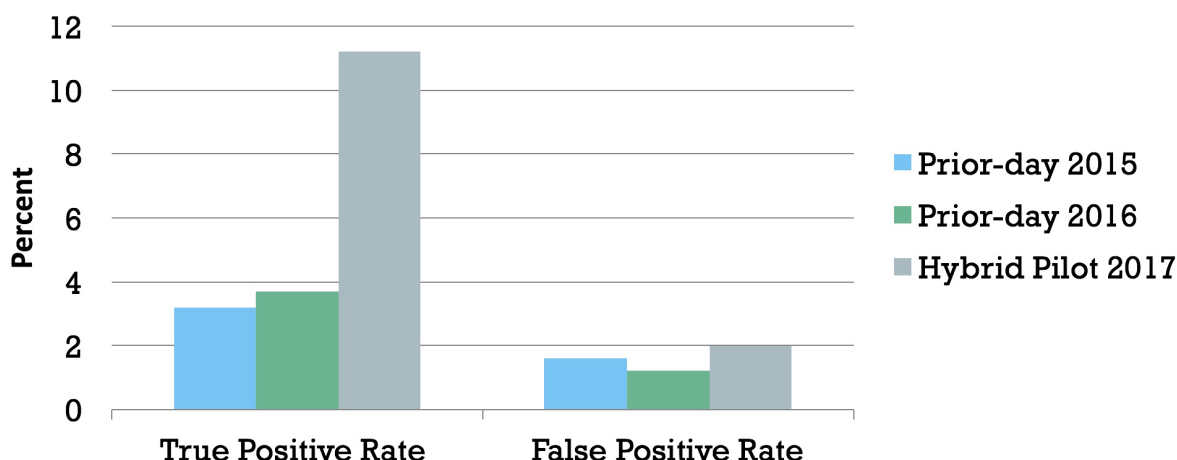


Figure 3: Historical TPR and FPR for Chicago prior-day models and the 2017 pilot of hybrid model.

California.” *Environmental Science & Technology* 36 (18): 3885–92. doi:10.1021/es020524u.

Boehm, Alexandria B., Jed A. Fuhrman, Robert D. Mrše, and Stanley B. Grant. 2003. “Tiered Approach for Identification of a Human Fecal Pollution Source at a Recreational Beach: Case Study at Avalon Bay, Catalina Island, California.” *Environmental Science & Technology* 37 (4): 673–80. doi:10.1021/es025934x.

Brooks, Wesley, Steven Corsi, Michael Fienen, and Rebecca Carvin. 2016. “Predicting Recreational Water Quality Advisories: A Comparison of Statistical Methods.” *Environmental Modelling & Software* 76 (February): 81–94. doi:10.1016/j.envsoft.2015.10.012.

Byappanahalli, Muruleedhara N., Richard L. Whitman, Dawn A. Shively, and Meredith B. Nevers. 2010. “Linking Non-Culturable (qPCR) and Culturable Enterococci Densities with Hydrometeorological Conditions.” *Science of The Total Environment* 408 (16): 3096–3101. doi:10.1016/j.scitotenv.2010.04.051.

Crowther, John, David Kay, and Mark D. Wyer. 2001. “Relationships Between Microbial Water Quality and Environmental Conditions in Coastal Recreational Waters: The Fylde Coast, UK.” *Water Research* 35 (17): 4029–38. doi:10.1016/S0043-1354(01)00123-3.

Cyterski, Mike, Wesley Brooks, Mike Galvin, Kurt Wolfe, Rebecca Carvin, Tonia Roddick, Mike Fienen, and Steve Corsi. 2014. “Virtual Beach 3.0.4: User’s Guide.” Athens, Georgia: Center for Exposure Assessment Models, US Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-02/documents/vb3_manual_3.0.4.pdf.

Francy, Donna S., Erin E. Bertke, and Robert A. Darner. 2009. “Testing and Refining the Ohio Nowcast at Two Lake Erie Beaches-2008.” USGS Open-File Report 2009–1066. Reston, VA: U.S. Geological Survey. <https://pubs.usgs.gov/of/2009/1066/>.

Francy, Donna S., Amie M.G. Brady, Rebecca B. Carvin, Steven R. Corsi, Lori M. Fuller, John H. Harrison, Brett A. Hayhurst, et al. 2013. “Developing and Implementing the Use of Predictive Models for Estimating Water Quality at Great Lakes Beaches.” USGS Numbered Series 2013-5166. Reston, VA: U.S. Geological Survey. <http://pubs.er.usgs.gov/publication/sir20135166>.

Frick, Walter E., Zhongfu Ge, and Richard G. Zepp. 2008. “Nowcasting and Forecasting Concentrations of Biological Contaminants at Beaches: A Feasibility and Case Study.” *Environmental Science & Technology* 42 (13): 4818–24. doi:10.1021/es703185p.

Haugland, Richard A., Shawn C. Siefring, Larry J. Wymer, Kristen P. Brenner, and Alfred P. Dufour. 2005.

“Comparison of Enterococcus Measurements in Freshwater at Two Recreational Beaches by Quantitative Polymerase Chain Reaction and Membrane Filter Culture Analysis.” *Water Research* 39 (4): 559–68. doi:10.1016/j.watres.2004.11.011.

Hou, Deyi, Sharyl J. M. Rabinovici, and Alexandria B. Boehm. 2006. “Enterococci Predictions from Partial Least Squares Regression Models in Conjunction with a Single-Sample Standard Improve the Efficacy of Beach Management Advisories.” *Environmental Science & Technology* 40 (6): 1737–43. doi:10.1021/es0515250.

Kinzelman, Julie, Clement Ng, Emma Jackson, Stephen Gradus, and Robert Bagley. 2003. “Enterococci as Indicators of Lake Michigan Recreational Water Quality: Comparison of Two Methodologies and Their Impacts on Public Health Regulatory Events.” *Applied and Environmental Microbiology* 69 (1): 92–96. doi:10.1128/AEM.69.1.92-96.2003.

Le Fevre, N. M., and G. D. Lewis. 2003. “The Role of Resuspension in Enterococci Distribution in Water at an Urban Beach.” *Water Science and Technology: A Journal of the International Association on Water Pollution Research* 47 (3): 205–10.

Morrison, Ann Michelle, Kelly Coughlin, James P. Shine, Brent A. Coull, and Andrea C. Rex. 2003. “Receiver Operating Characteristic Curve Analysis of Beach Water Quality Indicator Variables.” *Applied and Environmental Microbiology* 69 (11): 6405–11. doi:10.1128/AEM.69.11.6405-6411.2003.

Nevers, Meredith B., and Richard L. Whitman. 2005. “Nowcast Modeling of Escherichia Coli Concentrations at Multiple Urban Beaches of Southern Lake Michigan.” *Water Research* 39 (20): 5250–60. doi:10.1016/j.watres.2005.10.012.

———. 2011. “Efficacy of Monitoring and Empirical Predictive Modeling at Improving Public Health Protection at Chicago Beaches.” *Water Research* 45 (4): 1659–68. doi:10.1016/j.watres.2010.12.010.

News, Brian Bienkowski, Environmental Health, and Brian Bienkowski News Environmental Health. 2017. “DNA Tests Offer Quicker Results for Beach Bacteria.” *Scientific American*. Accessed September 21. <https://www.scientificamerican.com/article/dna-tests-offer-quicker-results-for-beach-bacteria/>.

Olyphant, Greg A., and Richard L. Whitman. 2004. “Elements of a Predictive Model for Determining Beach Closures on a Real Time Basis: The Case of 63rd Street Beach Chicago.” *Environmental Monitoring and Assessment* 98 (1-3): 175–90.

Olyphant, Greg A., Judith Thomas, Richard L. Whitman, and Denver Harper. 2003. “Characterization and Statistical Modeling of Bacterial (Escherichia Coli) Outflows from Watersheds That Discharge into Southern Lake Michigan.” *Environmental Monitoring and Assessment* 81 (1-3): 289–300.

Rabinovici, Sharyl J. M., Richard L. Bernknopf, Anne M. Wein, Don L. Coursey, and Richard L. Whitman. 2004. “Economic and Health Risk Trade-Offs of Swim Closures at a Lake Michigan Beach.” *Environmental Science & Technology* 38 (10): 2737–45. doi:10.1021/es034905z.

Reeves, Ryan L., Stanley B. Grant, Robert D. Mrse, Carmen M. Copil Oancea, Brett F. Sanders, and Alexandria B. Boehm. 2004. “Scaling and Management of Fecal Indicator Bacteria in Runoff from a Coastal Urban Watershed in Southern California.” *Environmental Science & Technology* 38 (9): 2637–48. doi:10.1021/es034797g.

Shively, Dawn A., Meredith B. Nevers, Cathy Breitenbach, Mantha S. Phanikumar, Kasia Przybyla-Kelly, Ashley M. Spoljaric, and Richard L. Whitman. 2016. “Prototypic Automated Continuous Recreational Water Quality Monitoring of Nine Chicago Beaches.” *Journal of Environmental Management* 166 (January): 285–93. doi:10.1016/j.jenvman.2015.10.011.

Smith, P., C. Carroll, B. Wilkins, P. Johnson, S. N. Gabhainn, and L. P. Smith. 1999. “The Effect of Wind Speed and Direction on the Distribution of Sewage-Associated Bacteria.” *Letters in Applied Microbiology* 28 (3): 184–88.

Whitman, Richard L., and Meredith B. Nevers. 2008. “Summer E. Coli Patterns and Responses Along 23

Chicago Beaches.” *Environmental Science & Technology* 42 (24): 9217–24.

Whitman, Richard L., Meredith B. Nevers, Ginger C. Korinek, and Muruleedhara N. Byappanahalli. 2004. “Solar and Temporal Effects on Escherichia Coli Concentration at a Lake Michigan Swimming Beach.” *Applied and Environmental Microbiology* 70 (7): 4276–85. doi:10.1128/AEM.70.7.4276-4285.2004.

Whitman, Richard L., Meredith Becker Nevers, and Paul J. Gerovac. 1999. “Interaction of Ambient Conditions and Fecal Coliform Bacteria in Southern Lake Michigan Beach Waters: Monitoring Program Implications.” *Natural Areas Journal* 19 (2): 5. <http://pubs.er.usgs.gov/publication/1000927>.