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

Nick Lucius[a,*], Kevin Rose[c], Callin Osborn[b], Matt E Sweeney[c], Renel Chesak[b], Daniel Y. Little[c], Scott Beslow[c], Tom Schenk Jr.[a]

[a] *City of Chicago, 333 S State St., Suite 420, Chicago, IL 60604*

[b] *DePaul University, Department of Computer Science, 243 S Wabash Ave., Chicago, IL 60604*

[c] *Chi Hack Night, 2543 N Spaulding Ave, Suite 2, Chicago, IL 60647*

Abstract

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. Since result take up to 12 hours, statistical models are used to forecast bacteria levels in lieu of test results; however they underestimate days with elevated fecal indicator bacteria levels. Quantitative polymerase chain reaction (qPCR) tests return results within 3 hours but 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 residents or close beaches due to exposure to the bacteria. Using pilot data from Chicago, the model proposed in this paper increased accuracy from 3.4% to 11.2%–a 230% increase. Thus, limited deployments of qPCR testing can be used to control costs, but are blended with statistical forecasting to gain an overall improvment in forecast accuracy.

Keywords: random forest, *Escherichia coli*, recreational water quality, fecal indicator bacteria, Chicago,

1. Introduction

Managers of recreational beaches use culture-based methods to measure levels of fecal indicator bacteria (FIB) levels at recreational beaches. Sampling is conducted early in the morning, but 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; Boehm et al. 2002). To get around this, researchers have developed statistical models–dubbed "nowcast" models–to estimate FIB for the day based on the previous day *E. coli* levels and factors such as precipitation, wind, and water conditions (Francy et al. 2013).

These models often incorrectly predict that beaches will not have elevated FIB levels–known as Type II errors or "false negatives" (Nevers and Whitman 2011; Rabinovici et al. 2004; Boehm et al. 2002). For instance in 2015, Chicago beaches had 200 events where bacteria levels were too high; however, only 14 (7.0%) of these "beach days" were forcasted by the predictive model. These models do have good overall fit, but elevated FIB levels are statistically rare events and models often fail to predict them–known as low "sensitivity".

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

*Corresponding author

Email addresses: nicholas.lucius2@cityofchicago.org (Nick Lucius), kevin@maypark.com (Kevin Rose), callin620@gmail.com (Callin Osborn), mesweene@uic.edu (Matt E Sweeney), rchesak1@gmail.com (Renel Chesak), daniel.y.little@gmail.com (Daniel Y. Little), scott.beslow@gmail.com (Scott Beslow), tom.schenk@cityofchicago.org (Tom Schenk Jr.)

3 to 4 hours and yield similar results as culture-based methods (Haugland et al. 2005; Kinzelman et al. 2003). The rapid turnaround lets managers determine warnings or closures based on tests within the same day instead of relying on statistical models.

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 (Bienkowski 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.

Both of these tests have generally-acceptable thresholds for acceptable FIB 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). 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.

We exploit the historical correlation between beaches to estimate FIB readings. We use limited qPCR results at specific beaches to predict elevated bacteria levels at other beaches using clustering algorithms and random forest regressions. This hybrid approach allows limited deployment of qPCR equipment to reduce overall costs and provides a higher quality statistical model.

We use 10 years of historical measurements from 20 beaches in Chicago to create a model to forecast whether bacteria levels at a beach will be elevated. During the summer of 2017, the model was able to increase the accuracy of prediction of beaches by 230 percent. 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. Material and Methods

2.1. Prior-day Nowcast Models

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*.

The defining characteristic of prior-day nowcast models is using FIB levels from the previous day ($t-1$) to predict bacteria levels for the current day (t). Models will also incorporate covariates or predictors to improve the accuracy of models, such as precipitation (Ackerman and Weisberg 2003; Morrison et al. 2003), sunlight (Whitman et al. 2004), wind (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 (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, $w_1, w_2, ..., w_j$. These models often take the form of

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

where $f(...)$ is a some function or algorithm that inputs raw data and outputs a probability. For instance, the linear regression model is typically $x_i^t = \beta_1 x_i^{t-1} + \beta_2 w_1 + \beta_3 w_2 + ... + \beta_j w_j$, where β_j are the coefficients that weight the importance of each input.

Various models are used to improve accuracy, such as log transformations (Nevers and Whitman 2005), polynomial coefficients (Frick, Ge, and Zepp 2008), logistic regression, partial least squares (Hou, Rabinovici, and Boehm 2006), generlized 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 FIB levels from time $t-1$–the prior-day–and covariates from time t to predict beaches at time t.

Yet, the reliance on prior-day bacteria levels are likely the source of inaccuracy. The cause of elevated levels is unlikely to persist between days. Covariates can help determine when conditions are optimal for bacteria growth, but are still imprecise. Importantly, most statistical models are tuned to maximize overall accuracy (e.g., high R^2), but elevated bacteria levels are often rare events. Thus, models frequently have low sensitivity; that is, the ratio of accurately estimated elevated bacteria levels over instances of actual exceedences.

2.2. Chicago Prior-day Nowcast Model

Chicago Park District measures water quality of 20 beaches along 26 miles against Lake Michigan on Chicago's eastern shore. There is no single source of bacteria in Lake Michigan; rather it is likely introduced by birds, beaches, and people. The Chicago River is contiguous to Lake Michigan but a lock limits water flowing into the lake. While the lock is sometimes opened, it is generally less than once per year, and beaches are immediately closed (Chicago Metropolitan Water Reclamation District 2016).

Beaches operate between Memorial Day weekend through Labor Day – approximately 122 days. As a result, there are approximately 2,440 "beach days," which each represent an observation in the model. An individual beach may have a warning when *E. coli* levels have exceeded acceptable thresholds.

Between 2011 and 2016, Chicago Park District placed hydrometeorological sensors to automatically collect covariates on water and atmospheric conditions. 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.

Water samples were collected for culture-based testing each morning and recorded, usually around noon. Sampling was done on weekdays; however, weekend and holiday sampling was conducted if the prior readings were elevated or if the weekend was particularly busy.

Shively et al. (2016) used this data to build a prior-day nowcast model which used the prior-day culture tests and same-day hydrometeorological data to predict bacteria levels at all beaches. The novel model leveraged the sensors to automate the management process by providing estimates to beach managers each weekday.

Predictions were obtained from a random forest model and the predictions were published online for beach visitors. In 2015 and 2016, the overall accuracy was 90 and 93%, respectively, and specificity (true negative rate) was 98 and 99%. However, the sensitivity (true positive rate) was only 7 and 11.9%.

Beginning in 2015, Chicago Park District began to use limited qPCR testing at five beaches. Data was collected but not incorporated into the predictive model. During the summer of 2017, qPCR testing was expanded to all 20 beaches and culture-based testing was paused.

2.3. Hybrid Nowcast Model

Whitman and Nevers (2008) observed that 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 shows the correlation of *E. coli* measurements between beaches from 2006 through 2017. The branches denote the "nearest neighbor" for each pair of beaches, a simple way to show similar beaches. Beaches are not uniformally correlated. Some beaches display strong clusters of correlation that exceed the correlation of readings within the same beach between days.

Beaches are placed into clusters where bacteria levels are highly correlated, but readings between clusters is low. A clustering algorithm, $J(\dots)$, can assign beaches, x_i, to a cluster k based on correlation between historical bacteria levels:

$$K = J\left(x_1^t, x_2^t, \dots, x_n^t\right)$$

We select a feature beach, $\hat{x}_i^t$, in each cluster to predict the bacteria levels for the remaining m beaches in the cluster, x_i^t. Thus, each cluster (k) will have the following membership: $K_i : \{\hat{x}_1^t, x_1^t, x_2^t, \ldots, x_m^t\}$.

To generate the predictions, we can formulate a model that limits predictions to each cluster:

$$x_{i\in k}^t = f\left(\hat{x}_{i\in k}^t\right)$$

so the feature beach $\hat{x}_{i\in k}^t$, the ith beach in cluster k uses data from time t to predict the remaining beaches ($x_{i\in k}^t$) in the same cluster in the same time period.

This model leverages observations from the same time t to predict bacteria levels at the other beaches on the same day. The rapid results from qPCR testing allows recreational beach managers to observe beach readings at selected beaches, $\hat{x}_{i\in k}^t$, to predict FIB levels at other beaches. The factors and conditions which led to elevated levels are captured and preserved in the model to render predictions.

2.4. Identifying Beach Clusters

We apply the above approach to Chicago's beach data. We used K-means clustering algorithm to detect when beaches have similar movements. First, there is an initial, random guess for which beaches belong in clusters. The distance or error is then measured for all clusters. Then, membership of clusters is slightly altered, the error is then measured and determined if it is increasing or decreasing. This process is repeated until the lowest measured error is obtained.

Mathematically:

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

We chose to limit to 5 clusters ($k = 5$) because it aligned to the number of qPCR testing sites by Chicago Park District starting in 2015. 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 since it has 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 final list of beaches and their respective clusters are listed in Table 1.

Table 1: Final results of K-means clustering. The ˆ denotes feature beaches used to predict remaining beaches in the cluster.

k	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
$\hat{x}$	Fosterˆ	North Avenueˆ	Leoneˆ	31stˆ	South Shoreˆ
$x_{1\in k}$	Osterman	Oak Street	Juneway	12th	57th
$x_{2\in k}$			Howard	39th	
$x_{3\in k}$			Rogers		
$x_{4\in k}$			Jarvis		

Within each cluster, the beach with the most *E. coli* exceedances was selected to be feature beach and rapid testing as model input. By rapid testing the beaches with the most frequest 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.

2.5. Building the Predictive Model

A random forest regression model was trained with the predicted beach's name and 10 years of culture-based test results for each of the feature beaches. Variable importance for the random forest model was assessed by calculating the percentage increase in mean squared error (MSE) as a result of permuting the values for each variable (Figure 2). The most important feature was the *E. coli* level at Leone Beach. In fact, the importance of each beach's *E. coli* level decreased as the beach location was further south. This is likely due to the fact that more beaches on the north side (10) are being predicted than on the south side (5).

The name of the beach being predicted had the lowest importance. Without this predictor, the predicted *E. coli* for a specific day would be the same for each beach. Indeed, because of the correlation that exists between certain beaches, it is not surprising that beach identity would rank last against the importance of E. coli levels along the whole lakefront. Still, permuting the beach name variable led to a 15.6% increase in the MSE.

The random forest regression model fit predicted *E. coli* levels in CFU/100ml units. These raw levels were transformed to a binary advisory decision. Due to left censoring at 1 CFU/100ml and an abundance of low E. coli readings, a limit can be observed at the lower left portion of the plot.

As the number of trees in the random forest model grew, model error decreased until around 400 trees. No performance benefit was observed beyond 400 trees.

2.6. Validation

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 3 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 3 shows the receiver operating characteristic (ROC), which shows the trade-off between increasing the number of days with true predictions versus days with falsely identified 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.

3. Results and Discussion

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, 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.

qPCR testing was piloted in 2015 and 2016; it was implemented at all 20 beaches during the summer of 2017. Culture-based methods were phased-out in 2017 and no longer collected. We used the modeling concept described in Figure 2.

In summer 2017, the model was piloted with some limitations imposed by available data. The model inputs were qPCR test results for Calumet, Rainbow, South Shore, 63rd Street, and Montrose, and the predicted beaches were the other 15 regularly tested beaches. The model was trained using 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.

Beginning on Friday, May 26, water samples were collected by the Chicago Park District and qPCR testing was completed by the University of Illinois at Chicago. Those results were then posted to the City of Chicago Data Portal–typically around noon (City of Chicago 2016b) . About 5 minutes later, we posted the predictions based on our model to the Data Portal (City of Chicago 2016a). The last predictions and samples were conducted on September 4th. While qPCR data was available for all beaches, we ignored any qPCR reading except for the 5 beaches used as model inputs.

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 S2 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). We compared whether our model predicted elevated bacteria levels compared to the results from the actual test results (City of Chicago 2016c).

3.1. Results

Over the summer, the analytical model had a sensitivity (true positive rate) of 11.2 percent compared to a sensitivity of 3.7 percent in 2016 (Table 2). From 2015 through 2016, the historical average true positive rate for the predicted beaches was 3.4 percent. That is, the hybrid model increased the true positive rate by 230 percent over the historical average. Additionally, these beaches have typically not been predicted as well as the other beaches – the same prior model that averaged 3.4 percent for the 15 predicted beaches averaged 9.5 percent for all 20 beaches.

The precision of the model–the portion of predicted high FIB that were accurate compared to all predicted FIB beach days–also grew from 17% to 27 percent–a 59% increase. Meanwhile, the false positive rate was slightly higher at 2.0 percent during 2017.

Accuracy of the hybrid nowcast model is in line with the previous performance of the Chicago prior-day nowcast model. Since most days do not have elevated *E. coli* levels, the increase in sensitivity (true positive rate) does not make a major impact on overall accuracy.

Table 2: Comparing specificity, sensitivity, and accuracy between Hybrid and Prior-day Nowcast models.

Model	Specificity	Sensitivity	Precision	Accuracy
2017 Hybrid	0.980	0.112	0.273	0.926
2016 Prior-day	0.988	0.037	0.167	0.930
2015 Prior-day	0.985	0.031	0.176	0.898

3.2. Discussion

Overall, the hybrid model correctly raised warnings by 230 percent. When a beach is predicted to have an elevated level, it is 58% less likey to have been a "false positive" (type I error). The increase in correct predictions, however, is not induced by simply raising more erroneous warnings. The "false negative" rate only slightly increases (type II errors) by two-tenths of a percent.

In terms of impact on beachgoers, the hybrid model issued 90 correct advisories while there were 71 incorrect advisories. In 2016, the Chicago prior-day nowcast 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.

The difference in performance is likely due to the use of bacteria levels on the same day. Chicago beaches rarely encounter consecutive days of elevated bacteria levels, so prior-day readings do not predict elevated levels on the next day. The factors that cause bacteria levels to grow rarely persist between days as few beaches have consecutive days of beach warnings. However, elevated bacteria levels portend ideal conditions for bacteria in other beaches.

Although Chicago was able to deploy qPCR testing at all 20 beaches, the cost and complexity of qPCR equipment limits widescale deployment. For instance, the lab conducting qPCR testing for Chicago has no remaining equipment to expand testing to other beaches or to recreational beaches along the Chicago river. The hybrid model presented in this paper allows beach managers to balance the constraints of qPCR equipment with beachgoer safety.

We removed four beaches from this model because of physical characteristics, such as long breakwaters, that prevented them from being clustered with other beaches; one beach was removed due to frequently high readings; and we used 5 beaches to predict FIB levels. As a result, 10 total beaches would need to undergo qPCR testing and ten beaches would be able to forgo any testing. Of the 10 undergoing qPCR testing, 5 would be tested to estimate FIB levels for 10 other beaches and the other 5 would need qPCR testing since they have distinct geographies, breakwaters, or other unique physical characteristics.

This model selected feature beaches that were prone to elevated bacteria levels; Foster, North Avenue, Leone, 31st, and South Shore. Moreover, we removed Calumet from the model since it frequently had elevated levels. qPCR testing can be deployed to the worse behaved beaches and rely on statistical models for the remaining beaches. This will naturally improve model performance since beaches which comprise the mode number of issues will be accurately measured, leading to more accurate beach warnings.

These results did not include any other covariates to help predict performance at other beaches. Initial attempts to include information—such as weather, air and water temperature, turbidity, etc.—did not improve overall model performance. This can be explored further to see if covariates or if other models besides random forest can improve model performance.

While this paper showed a 230% increase in sensitivity, the ability to predict elevated bacteria levels is still relatively low. Sensitivity can be significantly improved, but at a cost of increasing the rate of false positives which is an inconvenience for beachgoers. However, there is a lack of guidelines or studies which provide guidelines for acceptable levels of false positives. Our target rate of 1.8% was determined based on prior years, but it may prove worthwhile to incorrectly predict additional beach warnings if it deters exposure to bacteria.

4. Conclusions

- Hybrid nowcast modeling preserves information from the current day, instead of test data from the prior day.
- The modeling technique allows for limited deployments of the costly qPCR testing equipment.
- Accuracy of predicting elevated bacteria levels increased from 3.4% to 11.2%–a 230% increase–while reducing false positive rates by 58% and maintaining similar rate for false negatives.

5. Acknowledgements

We are indebted to Chicago Park District staff, especially Cathy Breitenbach and Carol Kim, who provided incredible guidance and feedback while taking time to work with the research team. Meredith Nevers graciously provided the research team with her considerable expertise. This was a unique project initially developed by a team of volunteers. We are grateful for Chi Hack Night which provided a forum for the volunteers to contribute to this project. In particular, we thank Forest Gregg who helped spur this project. We appreciate Jonathan Levy's work to publish the necessary data to make it available to the research team and the public and Sean Thornton's helpful comments and edits on early drafts. Finally, we wish to thank the attendees of State of Lake Michigan conference in 2017 for their comments and feedback.

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary Materials

Raw data from beach reading and historical forecasts generated by the hybrid nowcast model and past swim advisories are available on the City of Chicago Data Portal (City of Chicago 2016b; City of Chicago 2016c; City of Chicago 2016a) and updated during the beach season. The statistical model developed and described in this paper is also available online and is open source (Lucius et al. 2018). Finally, the code used to develop this paper is entirely reproducable, meaning all code and formula to generate figures and tables are reviewable.

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.

Bienkowski, Brian. 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/.

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, 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.

Chicago Metropolitan Water Reclamation District. 2016. "Reversals to Lake Michigan: 1985 to 2015." *Harvard Dataverse*. doi:10.7910/DVN/POEVFH.

City of Chicago. 2016a. "Beach E. Coli Predictions." *Beach E. Coli Predictions City of Chicago Data Portal*. https://data.cityofchicago.org/d/xvsz-3xcj.

———. 2016b. "Beach Lab Data." *Beach Lab Data City of Chicago Data Portal*. https://data.cityofchicago.org/d/2ivx-z93u.

———. 2016c. "Beach Swim Advisories." *Beach Swim Advisories City of Chicago Data Portal*. https://data.cityofchicago.org/d/t62e-8nvc.

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.

Lucius, Nick, Kevin Rose, Callin Osborn, Rebecca Jones, Matt E. Sweeney, Melissa McNeill, Chris Prokop, et al. 2018. "Chicago/Clear-Water: V1.1.0." *Zenodo*, January. doi:10.5281/zenodo.1145910.

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.

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.

List of Figures

A.1 Correlation heat map of Chicago beaches between 2006 and 2017. Each level of tree shows the nearest-neighbor correlation. 14

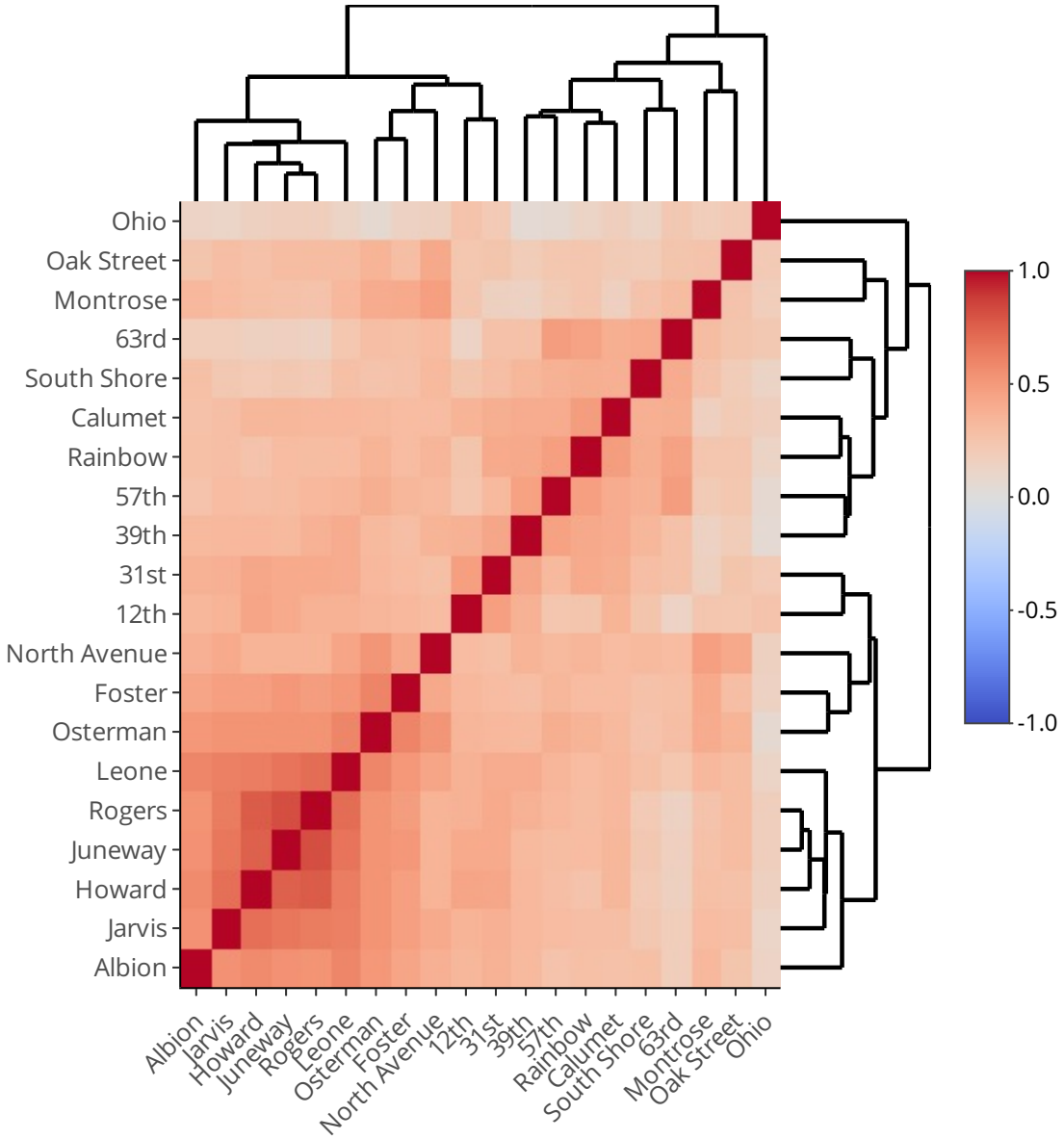
A.2 Variable importance of each factor when added to random forest model as measured by mean squared error 15

A.3 Receiver Operating Characteristic (ROC) for Hybrid Model and Prior-day Model. Dashed line shows historical false positive rate for prior-day model in Chicago. 16

A.4 Comparison of the 2017 hybrid pilot to existing prior-day model for "true positive rates" (sensitivity) and "false positive rates" (type I error) 17

A.5 Plot of the log of raw fitted values versus residuals from the random forest model. 18

A.6 Model performance measured by mean squared error (MSR) as the number of trees grow within the random forest model. 19


Ohio
Oak Street
Montrose
63rd
South Shore
Calumet
Rainbow
57th
39th
31st
12th
North Avenue
Foster
Osterman
Leone
Rogers
Juneway
Howard
Jarvis
Albion
1.0
0.5
0.0
-0.5
-1.0

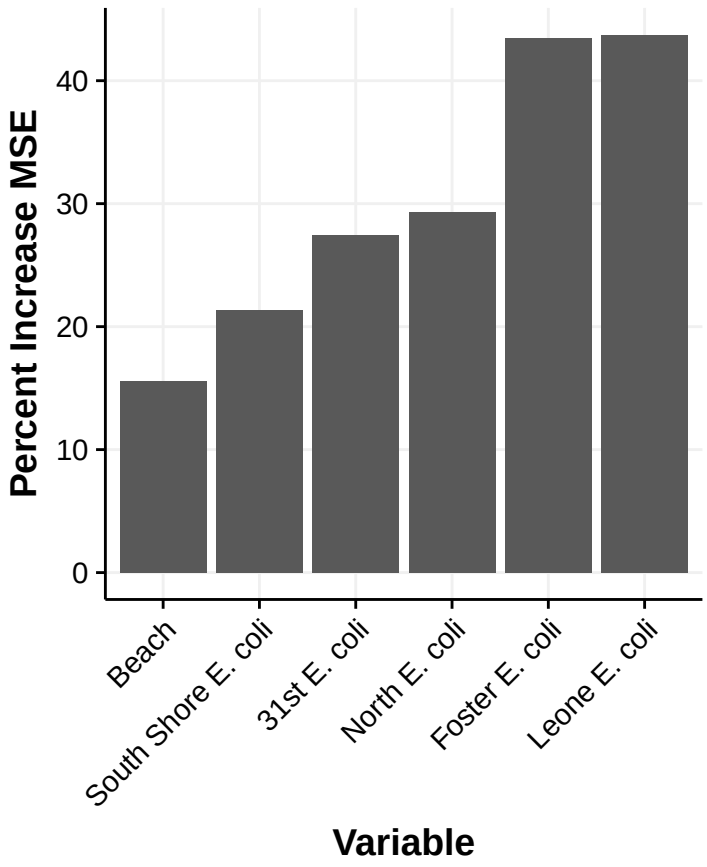


Figure A.2: Variable importance of each factor when added to random forest model as measured by mean squared error

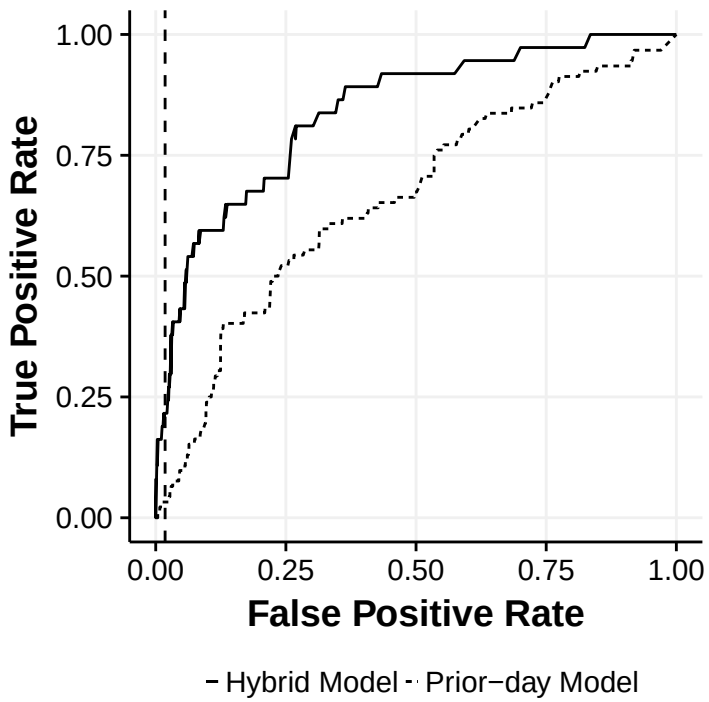


Figure A.3: Receiver Operating Characteristic (ROC) for Hybrid Model and Prior-day Model. Dashed line shows historical false positive rate for prior-day model in Chicago.

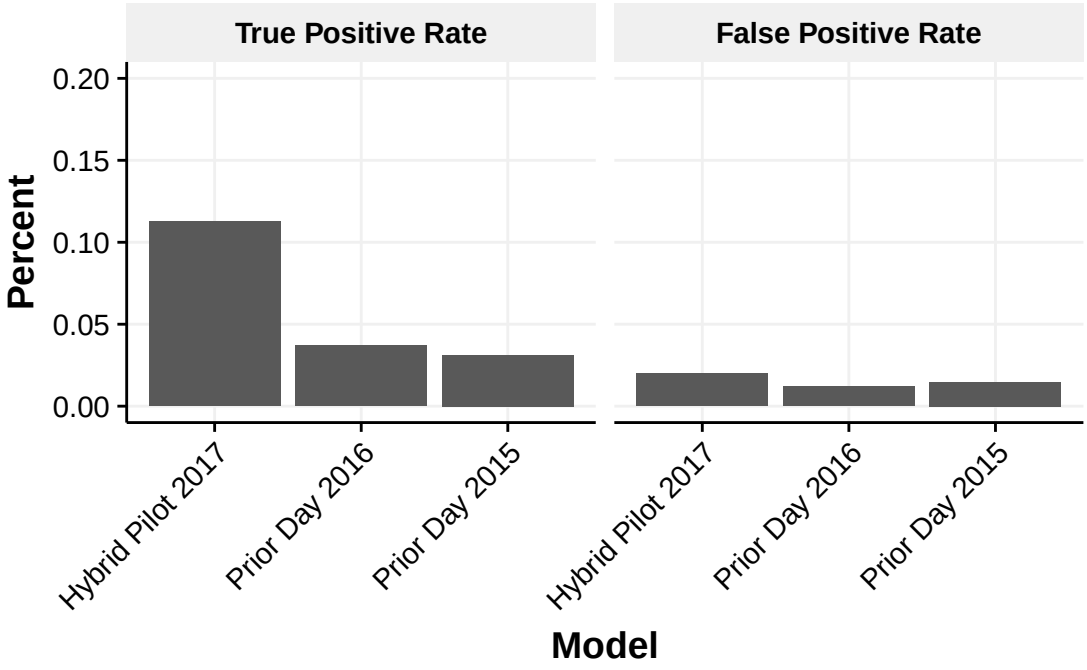


Figure A.4: Comparison of the 2017 hybrid pilot to existing prior-day model for "true positive rates" (sensitivity) and "false positive rates" (type I error)

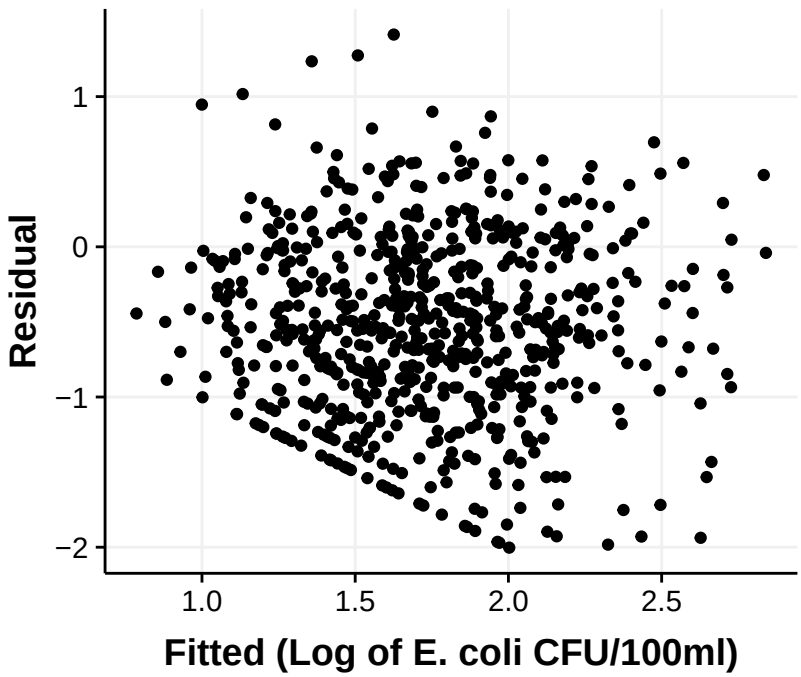


Figure A.5: Plot of the log of raw fitted values versus residuals from the random forest model.

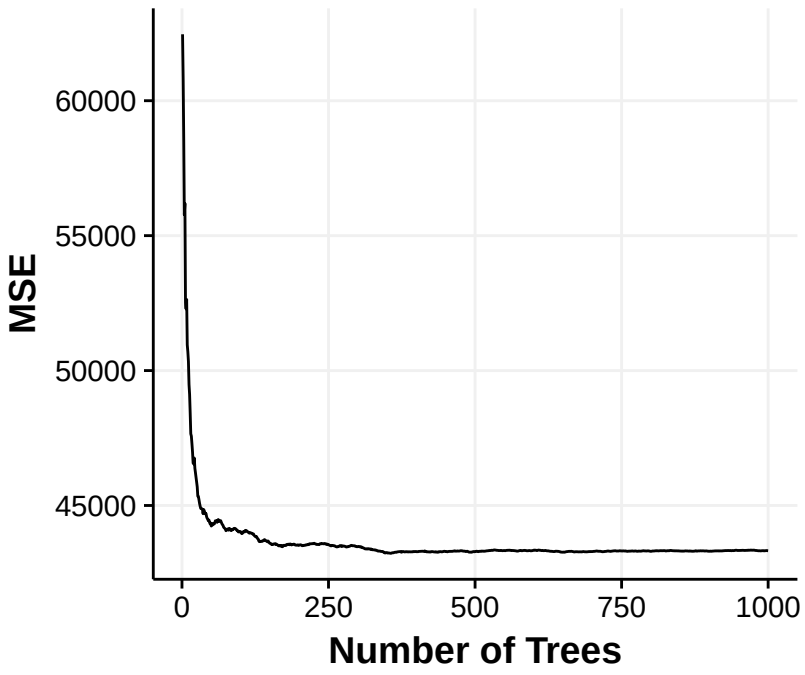


Figure A.6: Model performance measured by mean squared error (MSR) as the number of trees grow within the random forest model.