Comparing methods for predicting health advisories for beach water

Wesley Brooks, Rebecca Carvin, Steve Corsi, Mike Fienen

Abstract

Pithy, concise and informative. May bring the reader to tears due to the beauty of it.

Introduction

Fecal indicator bacteria (FIB) in beach water are often used to indicate contamination by harmful pathogens (Cabelli et al.; T. J. Wade et al.; Timothy J. Wade et al.; Fleisher et al.). The United States Environmental Protection Agency (USEPA) has established, through epidemiological studies, that FIB concentration is associated with human health outcomes (Cabelli; Dufour; USEPA 1986). Accordingly, the state of Wisconsin has established regulatory standards for beach water quality, stating that a beach should be posted with a swimmer's advisory when the concentration of the FIB *Escherichia coli* exceeds 235 colony forming units (CFU) / 100 mL (USEPA 2012; WDNR 2012). Traditional analysis methods for FIB concentration requires 18–24 hours for culturing a sample, so the decision to post an advisory is often made based on the previous day's FIB concent[...] [...] hod of beach management (USEPA). Recent research has show[...] [...] er can vary substantially during the 18-24 h analysis period, w[...] often w[...] [...] ost a warning (Whitman et al.; [...]) [...] ged by [...] the public is sometimes expose[...] recrea[...]

In order to have [...] models that "nowc[...] and running 24 h[...] have been used t[...] squares (OLS) (N[...]) [...] rtial least squares (PLS) (Hou, Rabinovici, and Boehm; Brooks, Fienen, and Corsi 2013), logistic regression (Waschbusch et al.; Jin and Englande Jr.), decision trees (Stidson, Gray, and McPhail 2012), random forests (Parkhurst et al., ???), and artificial neural networks (Kashefipour, Lin, and Falconer 2005; He and He). A thorough review of the regression techniques being used in nowcast models for FIB concentration is provided by Brauwere, Ouattara, and Servais (2014).

Ordinary least squares regression is the most commonly used regression technique in the nowcast models (Brauwere, Ouattara, and Servais 2014). However, OLS is well-known for drawbacks like overfitting, difficulty [...] of its linear modeling structure (Ge and Frick). The literature [...] have been successfully used for nowcast modeling, but due to [...] s, data handling, and performance validation, it is not possible to [...] owcast modeling by comparing different models at different sites. [...] s are evaluated in nowcast models at seven sites over four years of [...] ify the techniques that tend to produce more accurate predictions [...] osted.

[...] ollows: in the next section we discuss data collection a[nd] [...] ain how the comparisons were made. Next, we present [...] of comparing the methods by several metrics including: area under the ROC curve; predictive e[...] squares; and raw number of correct/incorrect predictions. Finally, we discuss what the comparis[...] about which are the best choices for a regression technique in a nowcast model.

Methods

The availability of large data sets for building regression models to predict the bacterial counts in beach water is both an opportunity and a challenge.

Data Sources

Possibly move this to the end of the section

Which sites

Where are they

What specific sources sources of data (plug EnDDAT)

Will include a map and tables

Definitions

At any site, den[...] xplanatory variables, by X, which is an $ns \times p$ matrix where n is the number of [...] e number of covariates. The vector of n observations of bacterial concentration is [...] regulatory standard δ and the decision threshold $\hat{\delta}$. The decision threshold $\hat{\delta}$–des[...]w–is used to balance the conclusions of exceedance or nonexceedance of δ. Structural [...]deling process results in bias. This bias can cause poor performance with respect to predicted actual concentration in cases with a continuous response, but adjusting the threshold to account for the bias can improve exceedance calcualtions.

Listing of specific statistical techniques

Fourteen different regression modeling techniques were considered. Each technique uses one of five modeling algorithms: gradient boosted models (GBM), the adaptive lasso, the genetic algorithm, partial least squares (PLS), or sparse PLS. Each technique is applied to either continuous or binary regression and to either variable selection and model estimation, or variable selection only.

Continuous vs. binary regression The goal of predicting exceednaces of the water quality standard is approached in two ways: one is to predict the bacterial concentration and then compare the prediction to a threshold, which is referred to as continuous modeling. The other is referred to as binary mo[...] we predict the state of the binary indicator z_i:

$$z_i = \left\{ \begin{array}{l} I\left(\hat{y}_i < \delta\right) = 0 \\ I\left(\hat{y}_i \geq \delta\right) = 1 \end{array} \right.$$

where $\hat{y}_i$ is the predicted concentration. The indicator is coded as zero when the concetration is below the regulatory standard and one when the concentration exceeds the standard. All of the binary modeling techniques herein use logistic regression (Hosmer Jr and Lemeshow 2004). Binary regression methods are indicated with a (b).

I tried to explain the meaning and use of decision threshold but may have fallen short. In any case, I think it's important that we make some explanation of it here.
We also need to explain whether y_i is referring to observed or modeled. Should we use y_hat to be modeled? I will adopt this throughout, but feel free to veto...OOPS! Later I found the tilde used for model prediction. I still moved that definition to be earlier in the text, but where it looks like hat in this draft, it is now tilde. Sorry for confusion.

you called GBM a gradient boosting machine but in the Statistical Learning book it's called a gredient boosted model. Just being precise....

I added the full names here.

see revised equation – optional to keep it. If "line ball" (e.g. equal to threshold) that's an exceedance, yeah?

2

Weighting of observations in binary regression   A weighting scheme was implemented for some of the binary regression techniques. In the weighting scheme, observations were given weights w_i where:

$$w_i = (y_i - \delta)/\hat{sd}(y)$$

$$\hat{sd}(y) = \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/n}$$

$$\bar{y} = \sum_{i=1}^{n} y_i/n$$

That is, the weights are equal to the number of standard deviations that the observed concentration lies from the regulatory threshold. Any technique that was implemented with this weighting scheme was separately implemented without any weighting of the observations. The methods that weight the observations are indicated with a (w).

Selection-only methods   The contest investigated whether certain modeling methods should to select covariates. Once the covariates were selected, the regression model using those estimated using ordinary least squares for the continuous methods, or ordinary logistic binary methods. Selection-only methods are indicated by an (s).

GBM

A gradient boosting machine (GBM) model is a so-called random forest model - a collection of trees (Friedman 2001). Prediction is done by averaging the outputs of the trees. Two GBM are explored - we refer to them as GBM-OOB and GBM-CV. The difference is in how the optimal number of trees is determined - GBM-CV selects the number of trees in a model using leave-one-out CV, while GBM-OOB uses the so-called out-of-bag error estimate. The CV method is much slower (it has to construct as many random forests as there are observations, while the OOB method only requires computing a single random forest) but GBM-CV should more accurately estimate the prediction error.

Adaptive Lasso

The least absolute shrinkage and selection operator (Lasso) is a penalized regression method that simultaneously selects relevant covariates and estimates their coefficients (Tibshirani 1996). The adaptive lasso (AL) is a refinement of the Lasso that possesses the so-called "oracle" properties of asymptotically selecting exactly the correct covariates and estimating them as accurately as would be possible if their identities were know in advance (Zou 2006). To use the AL for prediction requires selecting a tuning parameter. For the contest, the AL tuning parameter λ is selected to minimize the corrected Akaike Information Criterion (AICc) (Akaike 1973; Hurvich, Simonoff, and Tsai 1998).

Genetic algorithm

The genetic algorithm is a variable-selection method that works by analogy to natural selection, where so-called chromosomes represent regression models (Fogel 1998). A covariate is included in the model if the corresponding element of the chromosome is one, but not otherwise. Chromosomes are produced in successive generations, where the first generation is produced randomly and subsequent generations are produced by combining chromosomes from the current generation, with additional random drift. The chance that a chromosome in the current generation will produce offspring in the next generation is an increasing function of its fitness. The fitness of each chromosome is calculated by the AICc.

PLS

Partial least squares (PLS) regression is a tool for building regression models with many covariates (Wold, Sjostrom, and Eriksson 2001). PLS works by decomposing the covariates into mutually orthogonal components, with the components then used as the covariates in a regression model. This is similar to principal components regression (PCR), but the way PLS components are chosen ensures that they are aligned with the model output, whereas PCR is sometimes criticised for decomposing the covariates into components that are unrelated to the model's output. To use PLS, one must decide how many components to use in the model. This study follows the method described in (Brooks, Fienen, and Corsi 2013), using the PRESS statistic to select the number of components.

SPLS

Sparse PLS (SPLS) combines the orthogonal decompositions of PLS with the sparsity of lasso-type variable selection (Chun and Keles 2007). To do so, SPLS uses two tuning parameters: one that controls the number of orthogonal components and one that controls the lasso-type penalty. The optimal parameters are those that minimize the mean squared prediction error (MSEP) over a two-dimensional grid search. The MSEP is estimated by 10-fold cross-validation.

Data transformations for beach regression

The response for our continuous regression models is the base-10 logarithm of the *E. coli* concentration. For the binary regression models, the response is an indicator of whether the concentration exceeds the regulatory threshold $\delta = 235$ CFU/mL. Transformations were applied to some of the data during pre-processing: the beach water turbidity and the discharge of tributaries near each beach were log-transformed, and rainfall variables were all square root transformed.

Cross Validation

Our assessment of the modeling techniques is based on their performance in predicting exceedances of the regulatory standard. Two types of cross validation was used to measure the performance in prediction: leave-one-out (LOO) and leave-one-year-out (LOYO). In LOO CV, one observation is held out for validation while the rest of the data is used to train a model. The model is used to predict the concentration of that held out observation, and the process is repeated for each observation. Each cycle of LOYO CV holds out an entire year's worth of data for validation instead of a single observation. It is intended to approximate the performance of the modeling technique under a typical use case: a new model is estimated before the start of each annual beach season and then used for predicting exceedances during the season. The LOYO models in this study were estimated using all the available data except for the held out year, even that from future years. So for instance the 2012 models were estimated using the 2010-2011 and 2013 data.

Some methods also used cross-validation internally to select tuning parameters. In those cases the internal CV was done using only the model data, and never looking at the held-out observation(s). This process is separate from - and does not affect - the CV to assess predictive performance.

Comparing methods, and quantifying uncertainty in the ranks

Results were compiled into one table for each site, such as the one below which cotains the results of running the contest at Hika. Each observation corresponds to a row in the table. The results table has a column for the observed log *E. coli* concentration and, for each method, columns for the predicted concentration by LOO CV and by LOYO CV. From the table, we can calculate the predictive error sum of squares (PRESS) and the area under the receiver operating characteristic (ROC) curve (AUROC), which are the statistics we use to summarize performance of the modeling methods.

Row	Actual	PLS (LOO)	PLS (LOYO)	. . .	SPLS (LOO)	SPLS (LOYO)
1	2.54	2.35	2.22	. . .	2.29	2.55
2	2.59	1.87	1.79	. . .	1.91	1.23
⋮	⋮	⋮	⋮	. . .	⋮	⋮
166	1.57	1.93	2.06	. . .	1.83	2.07
167	3.38	1.84	2.01	. . .	1.80	1.71

In order to identify which modeling methods have the best performance across all sites, the summary statistics at each site were converted to ranks. We report the mean rank of each method across the sites. Uncertainty in the rankings is quantified by the bootstrap. Since PRESS and AUROC are functions of the results tables, the bootstrap procedure is carried out by resampling the rows of each results table and recalculating the ranks for each bootstrap sample. We used 1001 bootstrap samples of each results table in the analysis that follows.

Results

AUROC

predictions are sorted. The AUROC averages the model's ... A model which perfectly separates exceedances from ..., while a model that predicts exceedances no better than
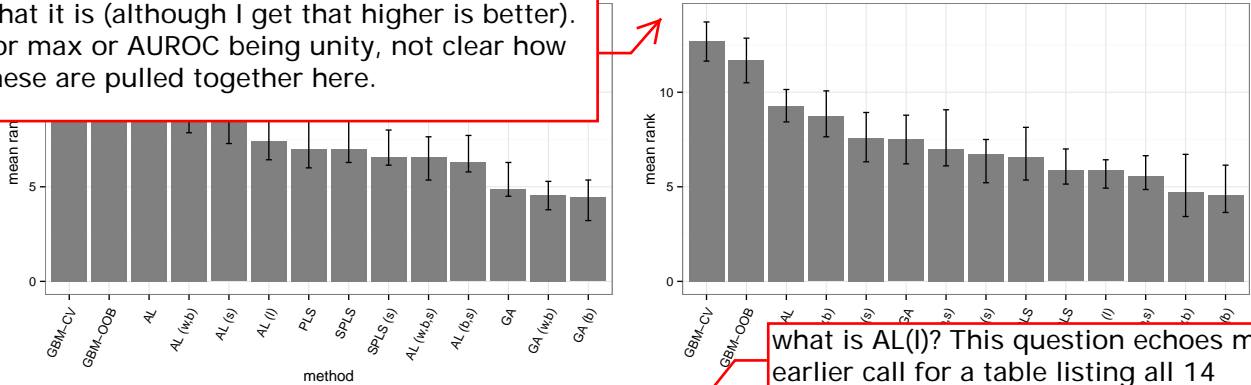
Figure 1: Mean ranking of the methods by area under the ROC curve (higher is better). The error bars are 90% confidence intervals computed by the bootstrap. At left are the AUROC rankings from the leave-one-year-out cross validation, at right are the AUROC rankings from the leave-one-out cross validation.

The mean LOO and LOYO ranks for all the methods are plotted in Figure [fig:auroc-boxplot]. The three top-ranked methods were GBM-CV, GBM-OOB, and the adaptive lasso. In order to facilitate a pairwise comparison between modeling methods, Tables [table:auroc.pairs.annual] (for the leave-one-year-out analysis) and [table:auroc.pairs] (for the leave-one-out analysis) show the frequency that the mean AUROC rank of GBM-OOB, GBM-CV, or the adaptive lasso exceeded each of the other modeling methods.

	GBM-OOB	AL	AL (w,b)	AL (s)	AL (l)	PLS	SPLS	SPLS (s)	AL (b,s)	AL (w,b,s)	GA
GBM-CV	0.82	0.82	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
GBM-OOB		0.73	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AL			0.73	0.91	1.00	0.91	1.00	1.00	1.00	1.00	1.00

Table 2: Under leave-one-year-out cross validation, how often the mean AUROC rank of GBM-OOB, GBM-CV, or AL (in the rows) exceeded that of the other methods (in the columns).

	GBM-OOB	AL	AL (w,b)	AL (s)	GA	AL (w,b,s)	SPLS	SPLS (s)	PLS	AL (l)	AL (b,s)
GBM-CV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
GBM-OOB		1.00	0.91	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00
AL			0.73	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00

Table 3: Under leave-one-out cross validation, how often the mean AUROC rank of GBM-OOB, GBM-CV, or the adaptive lasso (in the rows) exceeded that of the other methods (in the columns).

PRESS

While AUROC quantifies how well a model sorts exceedances and non-exceedances, PRESS measures how accurately a model's predictions match the observed bacterial concentration. The PRESS can only be computed for continuous regression methods. Let the model's predictions be denoted $\tilde{y}_i$ and letting the actual observed bacterial concentrations be denoted y_i for $i = 1, \ldots, n$ where n is the total number of predictions. Then PRESS computed as follows:

$$\text{PRESS} = \sum_{i=1}^{n} \left(\tilde{y}_i - y_i \right)^2.$$

The PRESS statistic is of interest because a good model should accurately predict the bacterial concentration, but AUROC is a more important as a metric of model performance than the PRESS because it directly measures the ability of a model to separate exceedances from non-exceedances. That said, we expect the two statistics to usually agree on which modeling methods are the best.

The rankings of the methods by PRESS are plotted in Figure [fig:press-boxplot]. The top three techniques under both LOO and LOYO analysis were GBM-CV, GBM-OOB, and the adaptive lasso. The pairwise comparison of modeling methods by PRESS are in Tables [table:auroc.pairs.annual] (for the leave-one-year-out analysis) and [table:auroc.pairs] (for the leave-one-out analysis).

	GBM-OOB	AL	SPLS	PLS	SPLS (s)	AL (s)	GA
GBM-CV	0.73	0.91	1.00	1.00	1.00	1.00	1.00
GBM-OOB		0.91	1.00	1.00	1.00	1.00	1.00
AL			0.73	0.91	1.00	1.00	1.00

Table 4: Under leave-one-year-out cross validation, how often the mean PRESS rank of GBM-OOB, GBM-CV, or AL (in the rows) exceeded that of the other methods (in the columns).

Narrowing the focus

By both AUROC and PRESS, and for both LOO and LOYO analyses, the three highest-ranked modeling methods were GBM-CV, GBM-OOB, and the adaptive lasso. The fourth-ranked method was not consistent across the different analyses. By the LOO CV analysis, the adaptive lasso was ranked better than the

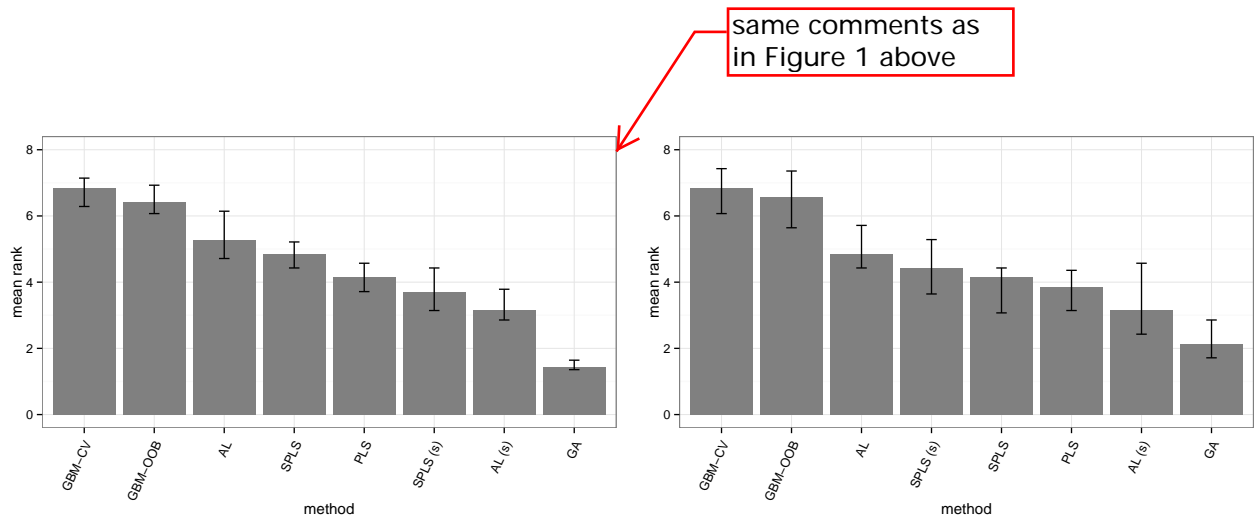not sure why abbreviating GBM but not AL??

Figure 2: Mean ranking of the methods by predictive error sum of squares (PRESS) across all sites (higher is better). The error bars are 90% confidence intervals computed by the bootstrap. At left are the PRESS rankings from the leave-one-year-out cross validation, at right are the PRESS rankings from the leave-one-out cross validation.

	GBM-OOB	AL	SPLS (s)	SPLS	PLS	AL (s)	GA
GBM-CV	0.55	1.00	1.00	1.00	1.00	1.00	1.00
GBM-OOB		1.00	0.91	1.00	1.00	1.00	1.00
AL			0.82	0.91	1.00	1.00	1.00

Table 5: Under leave-one-out cross validation, how often the mean PRESS rank of GBM-OOB, GBM-CV, or the adaptive lasso (in the rows) exceeded that of the other methods (in the columns).

fourth-ranked method by AUROC, adalasso.weighted, on 72.7% of bootstrap samples and better than the fourth-ranked method by PRESS, spls.select, on 81.8% of bootstrap samples. And by the LOYO CV analysis, the adaptive lasso was ranked better than the fourth-ranked method by AUROC, adalasso.weighted, on 72.7% of bootstrap samples and better than the fourth-ranked method by PRESS, spls, on 72.7% of bootstrap samples.

Therefore, we conclude that only the GBM methods and the adaptive lasso are worth considering further because they consistently outperform the other methods. We further narrow our study to GBM-OOB and the adaptive lasso because the GBM-OOB and GBM-CV methods are very similar and fitting a GBM-CV takes many times longer than a GBM-OOB model. While we focus on the AL and GBM-OOB

Classification of responses

The real-world performance of any model for predicting exceedances will be measured by how well it distinguishes between exceedances and nonexceedances. This is quantified similar to AUROC except that AUROC is an average over all possible thresholds instead of an assessment of performance for a specific choice of threshold. Combined with a realistic carefully selected decision threshold, the LOYO CV results were used to simulate real-world use of the models.

Intuitively, the decision threshold should adapt to the conditions that are common in the training data. If, for instance, exceedances are rare in the training data, then we expect them to be rare in the future, and should set the threshold high to reflect this prior expectation. On the other hand, if the bacterial concentration often exceeds the regulatory standard, then the decision threshold should be set lower to properly flag more of those exceedances. This intuition was encoded into how the decision threshold was set for the LOYO models. Specifically, the decision threshold was set to the qth quantile of the predicted values of non-exceedances in the training set, where $1 - q$ is the proportion of exceedances in the training set.

In Figure [fig:counts-barcharts], we look at the counts on a per-beach basis of

positives (correctly posting an advisory), true negatives (correctly not posting an advisory), false positives (wrongly posting an advisory) and false negatives (wrongly not posting an advisory). In most cases, the counts are similar between the two techniques, with GBM-OOB tending to make a few more correct decisions. There are exceptions where adaptive lasso makes more correct decisions (e.g., Hika and Red Arrow).

Variable selection

It was noted in Section Narrowing the focus that GBM-OOB and the adaptive lasso are the two of the three best-ranked methods. One difference between the two is that the adaptive lasso does variable selection while GBM-OOB uses all of the available covariates. We look here at how many covariates were used in the adaptive lasso models compared to the GBM-OOB models.

The covariate counts are displayed in Figure [fig:varselect-barchart]. At most of the sites, the adaptive lasso uses only a small fraction of the available covariates, but at Point the adaptive lasso uses almost all of the available covariates. That's because the variable selection criterion we used (AICc) is intended to minimize prediction error. As the amount of data increases, we accumulate enough information to begin to discern the effect even of covariates that are only slightly correlated with the response. As our dataset grows, then, we should expect more covariates to be selected for an adaptive lasso model, and Point has far more observations than the other sites.

Discussion

yes! I like that wording!

The GBM-CV, GBM-OOB, and AL methods showed the best results by both PRESS and AUROC, under LOO and LOYO cross validation. Though GBM-CV was a bit more accurate than GBM-OOB in all the settings, the small improvement in accuracy was perhaps not worth the large additional cost in time to fit the model. However, the additional computation cost is incurred once when the model is fitted - given a new observation of beach data, both the GBM-CV and GBM-OOB models produce predictions nigh-instantaneously. Where predictive accuracy is the most important consideration and no difficulty is anticipated in acquiring the data, it is hard to argue against using a GBM-type model.

The predictive performance of the AL models was somewhat worse than that of the GBM models, but by including a variable selection step the AL models reduce the number of covariates that must be measured in order to make daily predictions. A model that requires fewer covariates is less expensive and more robust (as the probability of encountering some missing data increases with the number of required covariates). This is particularly important for manually-collected covariates because collecting data by hand takes more time and costs more money than accessing publically available data from a web service. Across all the sites, the ratio of manually-collected to automatically collected covariates in the AL models seems to mirror the ratio among all available covariates, indicating that neither the manually- nor automatically-collected covariates are systematically more important to predicting the bacterial concentration. Some covariates tended to appear at every site in the AL models (and other models that include a variable selection step). The manually-collected covariates that were consistently selected for the models were the (log) turbidity in the beach water, and wave height at the beach.

Where minimizing the number of covariates is important, the selection criterion used here (corrected AIC) may not be appropriate. In that case, the Bayesian information criterion (BIC) is more stingy about including covariates in the model and does not exhibit the property of the AIC (or AICc) where more covariates are included in the model as the number of observations increases. However, the BIC is derived from the standpoint of identifying the most probably model, rather than minimizing the prediction error. It is likely that an AL model using the BIC for variable selection will have slightly worse predictive performance than one using the AICc.

Another advantage of the AL over GBM-type models is interpretability. As a linear regression technique, fitting an AL model means generating a set of coefficients, which can be interpreted as the marginal effect of a change in the corresponding covariate. On the other hand, GBM produces black-box models that make
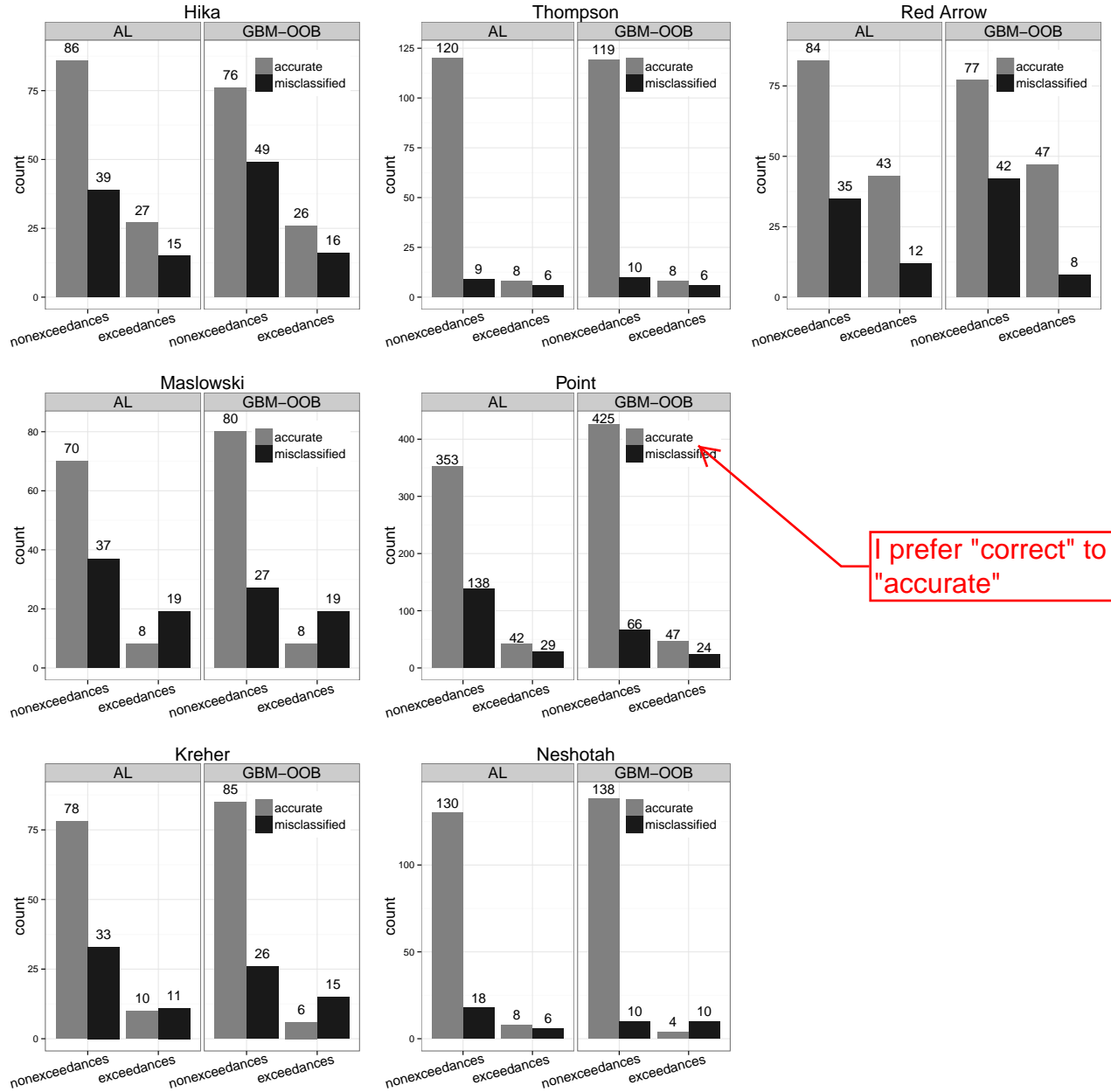
Figure 3: At each site, the number of predictions from the adaptive lasso (AL) and GBM-OOB that fell into four categories, from left: correctly predicted exceedance, incorrectly predicted exceedance, correctly predicted non-exceedance, and incorrectly predicted non-exceedance.
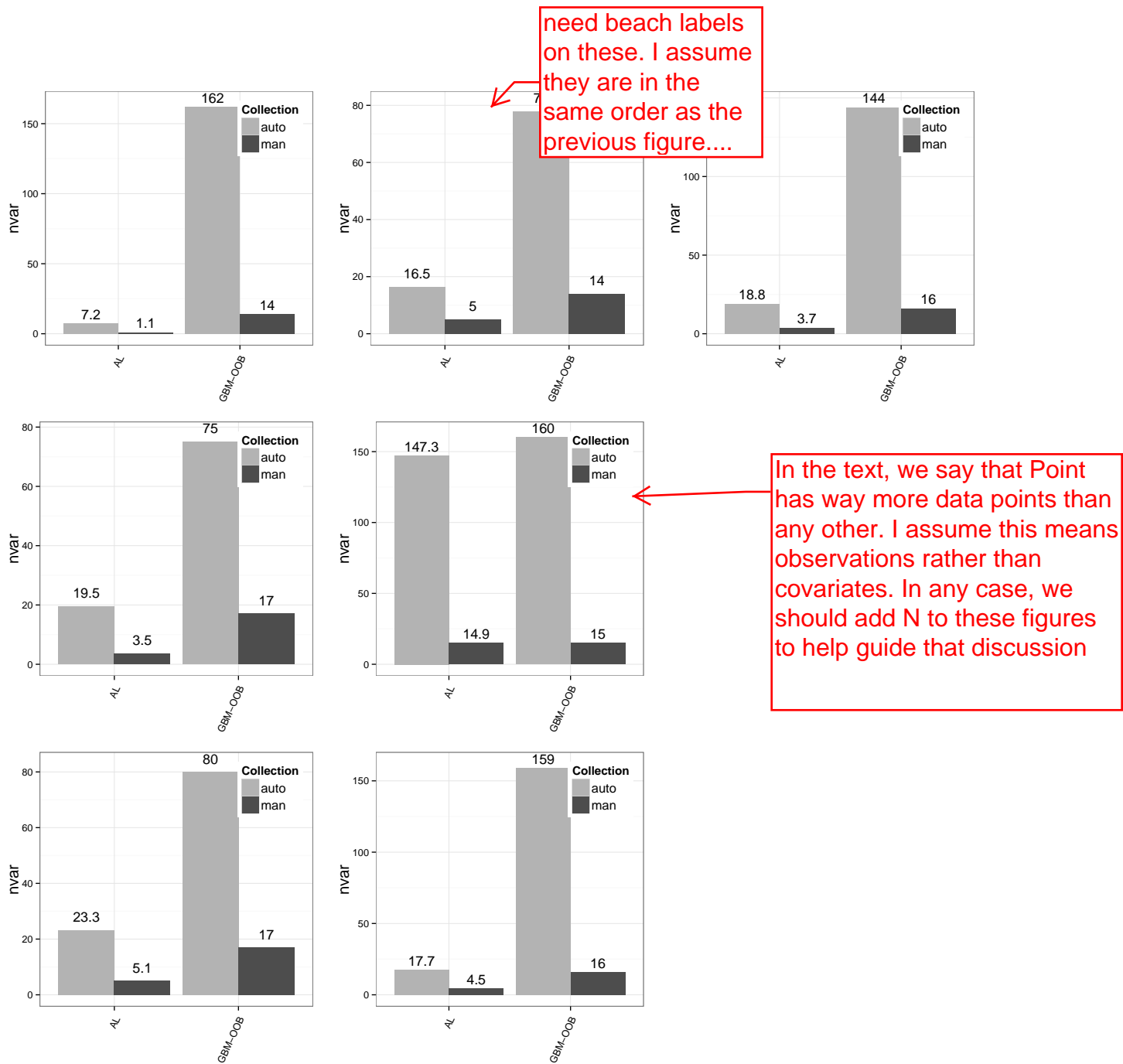
Figure 4: At each site, the mean number of covariates that were selected for the adaptive lasso (AL) model, and the total number of covariates, all of which were used in the gradient boosting machine with an out-of-bag estimate of the optimal tree count (GBM-OOB) models. For both AL and GBM-OOB, the covariate counts are broken down by whether the covariate values were collected automatically from web services or manually at the beach.

very accurate predictions but are difficult to intrepret. One common way to interpret a GBM-type model (which consists of a plethora of regression trees) is to observe the proportion of splits in the underlying trees that involve a particular covariate. The split proportion is a measurement of that covariate's importance to the model but gives no indication of how that covariate affects the bacterial concentraion.

This comparison was carried out in the R statistical software environment. As of version 3.0, the U.S. Environmental Protection Agency's Virtual Beach software can use GBM, GA, or PLS models for prediction of bacterial concentration. A version that can use the AL is anticipated. Scripts and details of the how the modeling methods were implemented are in the online supplement.

Acknowledgments

References

Akaike, Hirotugu. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." In *Second International Symposium on Information Theory*, 267–281. Akademinai Kiado.

Brandt, S., D. Schwab, T. Croley, D. Belestky, and R. Whitman. *Ecosystem Forecasting: Integrating Science to Reduce the Risks to Human Health.* Vol. 87. 36; Suppl. American Geophysical Union.

Brauwere, Anouk de, Nouho Koffi Ouattara, and Pierre Servais. 2014. "Modeling Fecal Indicator Bacteria Concentrations in Natural Surface Waters: a Review." *Critical Reviews in Environmental Science and Technology.* http://www.tandfonline.com/doi/pdf/10.1080/10643389.2013.829978.

Brooks, Wesley R., Michael N. Fienen, and Steven R. Corsi. 2013. "Partial Least Squares for Efficient Models of Fecal Indicator Bacteria on Great Lakes Beaches." *Journal of Environmental Mana* 114: 470–475.

Cabelli, V. J. "Health Effects Criteria for Marine Recreational Waters." Research and Development EPA-600/1-80-031. Vols. EPA-600/1-80-031. U.S. Environmental Protection Agency.

Cabelli, V. J., A. P. Dufour, M. A. Levin, L. J. McCabe, P. W. Haberman, and L. D. Jensen. "Relationship of Microbial Indicators to Health Effects at Marine Bathing Beaches." *Am.J.Public Health, 69(7), 690-696.*

Chun, Hyonho, and Sunduz Keles. 2007. "Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection."

Dufour, A. P. "Health Effects Criteria for Fresh Recreational Waters." Vols. EPA-600/1-84-004.

Fleisher, J. M., L. E. Fleming, H. M. Solo-Gabriele, J. K. Kish, C. D. Sinigalliano, L. Plano, S. M. Elmir, et al. "The BEACHES Study: health Effects and Exposures from Non-Point Source Microbial Contaminants in Subtropical Recreational Marine Waters." *International Journal of Epidemiology*: 8–8.

Fogel, David B. 1998. *Evolutionary Computation: the Fossil Record.* Wiley-IEEE Press.

Francy, Donna S., and Robert A. Darner. "Nowcasting Beach Advisories at Ohio Lake Erie Beaches." *Open-File Report.* Vols. Open-File Report 2007-1427. U.S. Geological Survey.

Friedman, Jerome. 2001. "Greedy Function Approximation: a Gradient Boosting Machine." *The Annals of Statistics* 29: 1189–1232.

Ge, Z., and W. E. Frick. "Some Statistical Issues Related to Multiple Linear Regression Modeling of Beach Bacteria Concentrations." *Environ.Res.* 103 (3): 358–364.

He, Li-Ming Lee, and Zhen-Li He. "Water Quality Prediction of Marine Recreational Beaches Receiving Watershed Baseflow and Stormwater Runoff in Southern California, USA." *Water Research* 42 (10-11): 2563–2573.

Hosmer Jr, David W, and Stanley Lemeshow. 2004. *Applied Logistic Regression.* John Wiley & Sons.

Hou, D., S. J. M. Rabinovici, and A. B. Boehm. "Enterococci Predictions from Partial Least Squares Regression Models in Conjunction with a Single-Sample Standard Improve the Efficacy of Beach Management Advisories." *Environ. Sci. Technol.* 40 (6): 1737–1743.

Hurvich, Clifford M., Jeffrey S. Simonoff, and Chih-Ling Tsai. 1998. "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion." *Journal of the Royal Statistical Society: Series B (Methodology)* 60: 271–293.

Jin, G., and A. J. Englande Jr. "Prediction of Swimmability in a Brackish Water Body." *Manage.Environ.Qual.* 17 (2): 197–208.

Kashefipour, S. M., B. Lin, and R. A. Falconer. 2005. "Neural Networks for Predicting Seawater Bacterial Levels." *Proceedings of the Institution of Civil Engineers-Water Management* 158: 111–118. http://www.icevirtuallibrary.com/content/article/10.1680/wama.2005.158.3.111.

Nevers, M. B., and R. L. Whitman. "Nowcast Modeling of Escherichia Coli Concentrations at Multiple Urban Beaches of Southern Lake Michigan." *Water Research* 39 (20): 5250–5260.

Olyphant, G. A., and R. L. Whitman. "Elements of a Predictive Model for Determining Beach Closures on a Real Time Basis: the Case of 63rd Street Beach Chicago." *Environ.Monit.Assess.* 98 (1-3): 175–190.

Parkhurst, D. F., K. P. Brenner, A. P. Dufour, and L. J. Wymer. "Indicator Bacteria at Five Swimming Beaches - Analysis Using Random Forests." *Water Res.* 39 (7): 1354–1360.

Stidson, R. T, C. A. Gray, and C. D. McPhail. 2012. "Development and Use of Modelling Techniques for Real-Time Bathing Water Quality Predictions." *Water and Environment Journal* 26: 1747–6593. http://onlinelibrary.wiley.com/doi/10.1111/j.1747-6593.2011.00258.x/abstract;jsessionid=BD291BB8B49C5EBBA2F8CABD79A23F2E.f03t02.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society, Series B (Methodological)*: 267–288.

USEPA. 1986. "Ambient Water Quality Criteria for Bacteria." EPA440/5-84-00.

———. 2012. "Recreational Water Quality Criteria." EPA-820-F-12-058. USEPA. http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/RWQC2012.pdf.

———. "Critical Path Science Plan for the Development of New or Revised Recreational Water Quality Criteria."

Wade, T. J., R. L. Calderon, E. Sams, M. Beach, K. P. Brenner, A. H. Williams, and A. P. Dufour. "Rapidly Measured Indicators of Recreational Water Quality Are Predictive of Swimming-Associated Gastrointestinal Illness." *Environmental Health Perspectives* 114 (1): 24–28.

Wade, Timothy J., Rebecca L. Calderon, Kristen P. Brenner, Elizabeth Sams, Michael Beach, Richard Haugland, Larry Wymer, and Alfred P. Dufour. "High Sensitivity of Children to Swimming-Associated Gastrointestinal Illness: results Using a Rapid Assay of Recreational Water Quality." *Epidemiology (Cambridge, Mass.)* 19 (3): 375–383.

Waschbusch, R., S. Corsi, K. Sorsa, J. Walker, J. Standridge, and T. Schnieder. "Data Collection and Modeling of Enteric Pathogens, Fecal Indicators and Real-Time Environmental Data at Madison, Wisconsin Recreational Beaches for Timely Public Access to Water Quality Information." *Stormwater: The Journal for Surface Water Quality Professionals* Final report for the EMPACT Project R-82933901-0.

WDNR. 2012. WDNR. http://dnr.wi.gov/topic/Beaches/documents/BeachReport2012.pdf.

Whitman, R. L., and M. B. Nevers. "Summer E. Coli Patterns and Responses Along 23 Chicago Beaches." *ENVIRONMENTAL SCIENCE & TECHNOLOGY* 42 (24): 9217–9224.

Whitman, R. L., M. B. Nevers, G. C. Korinek, and M. N. Byappanahalli. "Solar and Temporal Effects on Escherichia Coli Concentration at a Lake Michigan Swimming Beach." *Appl.Environ.Microbiol.* 70 (7): 4276–4285.

Wold, Svante, Michael Sjostrom, and Lennart Eriksson. 2001. "PLS-Regression: a Basic Tool of Chemometrics." *Chemometrics and Intelligent Laboratory Systems* 58: 109–130.

Zou, Hui. 2006. "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101: 1418–1429.