

Comparing methods for predicting health advisories for beach water: supplemental material

Wesley Brooks, Rebecca Carvin, Steve Corsi, Mike Fienen

	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
AL (w,b)				0.64	0.91	0.82	0.91	0.91	1.00	1.00	1.00
AL (s)					0.73	0.82	0.73	0.91	0.91	1.00	1.00
AL (l)						0.36	0.64	0.73	0.82	0.82	1.00
PLS							0.64	0.55	0.73	0.82	1.00
SPLS								0.82	0.64	0.64	0.91
SPLS (s)									0.55	0.64	0.91
AL (b,s)										0.55	0.91
AL (w,b,s)											0.82
GA											
GA (w,b)											

Table 1: 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
AL (w,b)				0.82	0.64	0.91	1.00	1.00	1.00	1.00	1.00
AL (s)					0.45	0.64	0.55	0.64	1.00	1.00	0.82
GA						0.36	0.64	0.73	1.00	0.91	0.91
AL (w,b,s)							0.45	0.64	0.73	0.91	1.00
SPLS								0.36	0.64	0.82	0.73
SPLS (s)									0.73	0.82	0.82
PLS										0.55	0.64
AL (l)											0.55
AL (b,s)											
GA (w,b)											

Table 2: 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).



Figure 1: 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.

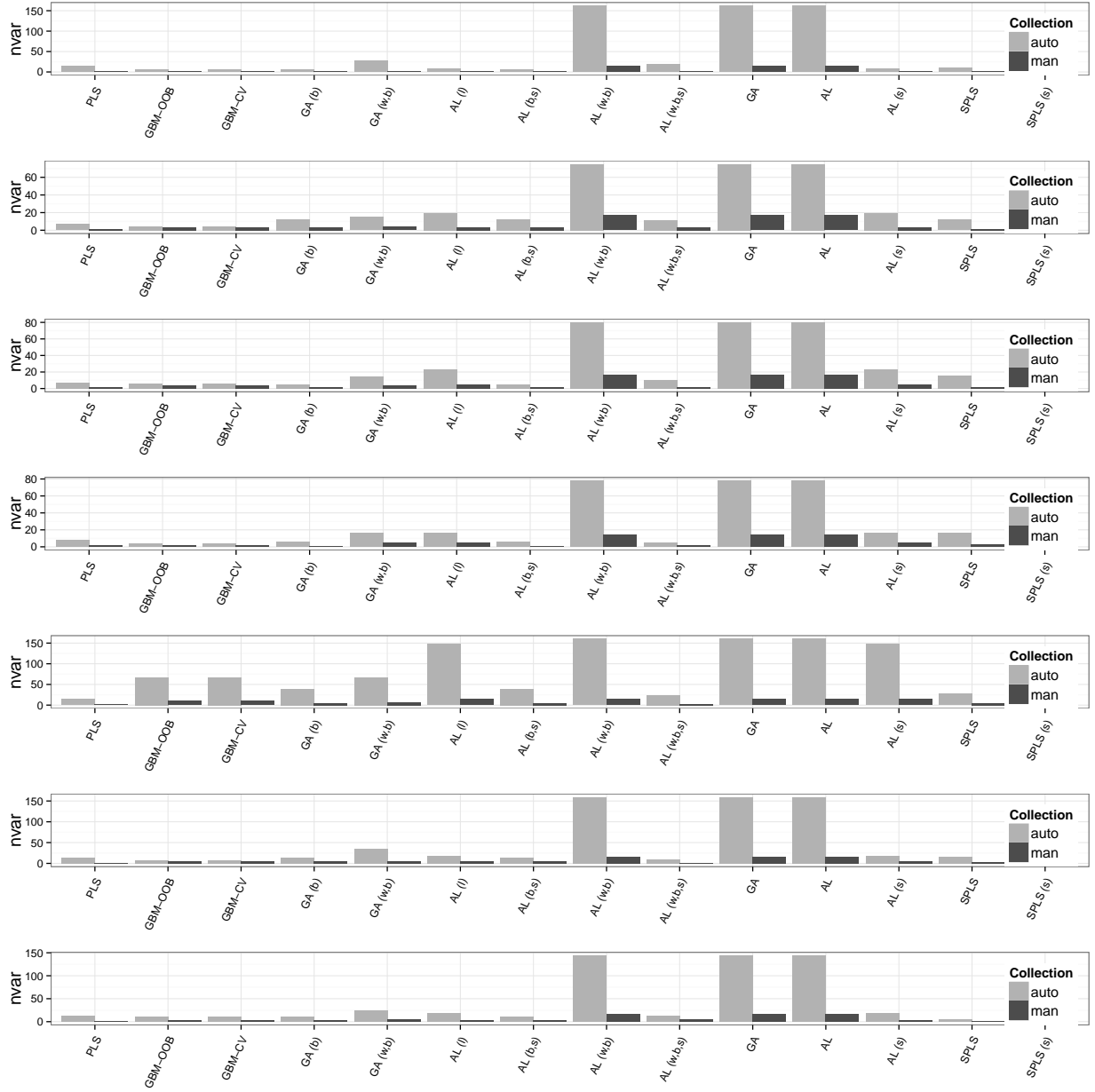


Figure 2: At each site, the mean number of covariates that were selected for the adaptive lasso (AL) model, and the total number of covariates, all of which were used in the gradient boosting machine with an out-of-bag estimate of the optimal tree count (GBM-OOB) models. For both AL and GBM-OOB, the covariate counts are broken down by whether the covariate values were collected automatically from web services or manually at the beach.

	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
SPLS				0.91	1.00	1.00	1.00
PLS					0.55	1.00	1.00
SPLS (s)						0.73	1.00
AL (s)							1.00

Table 3: 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).

	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
SPLS (s)				0.64	0.73	0.73	1.00
SPLS					0.64	0.73	1.00
PLS						0.82	1.00
AL (s)							0.91

Table 4: 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).