e-Xtra*

Predicting Fusarium Head Blight Epidemics with Boosted Regression Trees

D. A. Shah, E. D. De Wolf, P. A. Paul, and L. V. Madden

First and second authors: Department of Plant Pathology, Kansas State University, Manhattan 66506; and third and fourth authors: Department of Plant Pathology, The Ohio State University, Wooster 44691.

Accepted for publication 16 January 2014.

ABSTRACT

Shah, D. A., De Wolf, E. D., Paul, P. A., and Madden, L. V. 2014. Predicting Fusarium head blight epidemics with boosted regression trees. Phytopathology 104:702-714.

Predicting major Fusarium head blight (FHB) epidemics allows for the judicious use of fungicides in suppressing disease development. Our objectives were to investigate the utility of boosted regression trees (BRTs) for predictive modeling of FHB epidemics in the United States, and to compare the predictive performances of the BRT models with those of logistic regression models we had developed previously. The data included 527 FHB observations from 15 states over 26 years. BRTs were fit to a training data set of 369 FHB observations, in which FHB epidemics were classified as either major (severity ≥ 10%) or non-major (severity < 10%), linked to a predictor matrix consisting of 350 weather-based variables and categorical variables for wheat type (spring or winter), presence or absence of corn residue, and cultivar resistance. Predictive performance was estimated on a test (holdout) data set consisting of the remaining 158 observations. BRTs had a misclassification rate of 0.23 on the test data, which was 31% lower than the average misclassification rate over 15 logistic regression models we had presented earlier. The strongest predictors were generally one of mean daily relative humidity, mean daily temperature, and the number of hours in which the temperature was between 9 and 30°C and relative humidity ≥ 90% simultaneously. Moreover, the predicted risk of major epidemics increased substantially when mean daily relative humidity rose above 70%, which is a lower threshold than previously modeled for most plant pathosystems. BRTs led to novel insights into the weather–epidemic relationship.

Additional keywords: disease modeling, disease forecasting, machine learning, plant disease epidemiology, wheat scab.

Major epidemics of Fusarium head blight (FHB), caused primarily by *Fusarium graminearum* sensu stricto (59) of the *F. graminearum* species complex (43,53), are a recurring obstacle to successful wheat (*Triticum aestivum* L. em. Thell) production worldwide. Epidemics are responsible for large direct (35,36) and indirect (42) economic losses. Foliar fungicide applications timed to coincide with anthesis, if the environment is conducive for FHB, are one component of an effective disease management strategy (56). Accurate forecasts help growers recognize when their wheat crops are at a high risk of a major FHB epidemic, and most likely to benefit from a fungicide. When the risk of a major epidemic is low, growers could forgo inessential fungicide applications. Management decisions (and, hence, forecasts) need to be made by anthesis so that producers have sufficient time to spray should they choose that option.

The Fusarium Head Blight Risk Assessment Tool (http://www.wheatscab.psu.edu) is a publicly funded service that provides local-level, empirical FHB predictions across several of the wheat-growing regions of the United States which have historically experienced FHB epidemics (35). The forecasts are based on logistic regression models developed by De Wolf et al. (13), with subsequent revisions (37,38). We recently reexamined those models with more up-to-date data and analytical tools, in the process developing 15 new logistic regression models with improved predictive performance on a test data set (51). That latter effort was not without some statistical challenges, however. There were high correlations among candidate predictors, which were mitigated by restricting modeling to within fixed-length windows. We used bootstrapping with a leaps-and-bounds subset selection algorithm to further deal with the predictor correlations and the tendency of the algorithms to select noise variables in some cases. Models were restricted to no more than four weather-based predictors. With logistic regression applied to a relatively large number of predictors, it was difficult to identify or consider interactions. For some predictors, there was a nonlinear relationship with the response on the logit scale, which we accounted for with a generalized additive model basis. Although our models (51) offered a better sensitivity-specificity balance than currently deployed models (13,37,38) on the test data set, we did question whether the improvements offered by the newer models were sufficient to justify replacing the current models used in the Risk Assessment Tool.

Boosted regression trees (BRTs) is a classification algorithm which originated in the machine-learning community and was later shown to have a statistical interpretation (17). The boosting algorithm works by fitting individual regression trees in a forward, additive manner, with each added tree focusing on the (reweighted) observations that are still misclassified (21). In this way, it combines many (up to thousands) individually weak classifiers to create an overall classifier with improved predictive performance. Some salient features of the algorithm (6,15,18,33) include the ability to handle any type of predictor (i.e., continuous, categorical, or binary), immunity to the effect of outliers, identification of important predictors, automatic modeling of complexity in the data (interactions and nonlinear functions), and the ability to handle highly correlated predictors. For these reasons, BRTs have grown rapidly in popularity in the ecological sciences, providing not only predictive utility but also insights into

Corresponding author: D. A. Shah; E-mail address: quinnconsulting@verizon.net

* The *e*-Xtra logo stands for "electronic extra" and indicates that the online version contains one supplementary file.

http://dx.doi.org/10.1094/PHYTO-10-13-0273-R

ecological relationships (1,24,40). However, despite the apparent exuberant embrace of BRTs in the ecological sciences, not all researchers are convinced by claims of their superior predictive performance over more traditional methods (4,16). Algorithmic performance is linked to the inherent properties of the data set being modeled (e.g., number of and correlations among continuous, ordinal and categorical response and predictor variables) and the skill of the modeler in "tuning" the algorithm (19), and one algorithm that performs "better" on a given data set may not do as well on another. In addition, Hand (19) cautioned that the more complex algorithms such as BRTs may be increasing predictive performance by only a very small amount over what can be achieved by more traditional methods. Nevertheless, even with these caveats in mind, the growing number of studies in ecology which have used BRTs does indicate that the method is proving itself useful. Within the field of plant pathology, BRTs are still a novelty because examples are sparse. Landschoot et al. (32) used a BRT algorithm as one modeling technique in their investigation of cross-validation strategies in the predictive modeling of FHB in Belgian winter wheat. Shah (50) used BRTs on a snap bean survey database to explore factors associated with white mold.

In this article, we explored BRTs as a tool for predictive modeling of FHB in the United States. We investigated, for parsimony reasons, whether fitted BRTs could be simplified without loss of predictive performance, and whether model-averaging individual BRTs improved the predictive performance given by any one BRT model. We then compared their predictive performances with those of the logistic regression models developed earlier on the same data set (51).

MATERIALS AND METHODS

The data matrix. Details on the construction of the data set were given in a previous article (51). Briefly, there were 527 FHB severity observations, collected in 15 states over 26 years, linked to a matrix of 350 weather-based predictors summarizing conditions in windows of 5, 7, 10, 14, or 15 days in length, either before anthesis (i.e., the pre-anthesis period) or after anthesis (i.e., the post-anthesis period); and to binary indicators for corn residue presence or absence (CORN), spring or winter wheat (TYPE), and an ordinal variable representing four cultivar resistance levels (RESIST). FHB severity (*S*), also called the FHB index (52), is the mean percentage of a wheat spike's surface area with FHB symptoms, where the mean is estimated over all sampled spikes. The response variable of interest was a binary categorization of *S* (on a percentage scale):

$$\text{fhb}_i = \begin{cases} 0 \text{ if } S_i < 10 \\ 1 \text{ if } S_i \geq 10 \end{cases} \quad (1)$$

where $\text{fhb}_i = 0$ represented a non-major epidemic and $\text{fhb}_i = 1$ a major epidemic of FHB, for all observations 1,…,*N*. Variable names are identical to those used by Shah et al. (51), with additional abbreviated acronyms (Table 1) used for simpler labeling in graphics presented later. There were six different weather-based predictor types summarizing (in some defined way) dew-point depression (DD), rainfall (R), relative humidity (RH), temperature (T), vapor pressure deficit (VPD), or simultaneous temperature and relative humidity conditions (TRH) within windows and pre- or post-anthesis periods. We define a predictor group as a set of predictors that varied only in the window and period for the conditions being summarized. For example, at the top of Table 1, group 1 consisted of any predictor which summarized the mean overnight VPD, where that summary could be made over a 5-, 7-, 10-, 14-, or 15-day window in either one of the pre- or post-anthesis periods.

We were interested in correctly classifying the two classes of fhb from the available weather-based predictors plus the categorical predictors RESIST, CORN, and TYPE. Five imputed versions of the data matrix were available because of missing values for some predictors (51). Each imputed data set was divided into training (70%) and test (30%) sets by the same split used previously (51). Model building and validation were done with the training set. The test (holdout) set was not used in model development but was used in estimating the expected model error (rate at which a model misclassifies observations) (21).

BRTs. Features of BRTs have been discussed sufficiently elsewhere and are only summarized here. For the applied practitioner, Buston and Elith (6) and Leathwick et al. (33) provide good nontechnical explanations of the principles behind BRTs, while Elith et al. (15) give further details and practical examples of fitting BRTs to ecological data. The more mathematically inclined should consult Bühlmann and Hothorn (5), Friedman et al. (17), and Hastie et al. (21). A discussion of BRTs within the broader class of machine-learning algorithms is in Crisci et al. (11). The Appendix contains more information on the boosted regression model framework used in the current article.

BRT models were fit to the set of predictors within each window; these models were termed individual-window BRT models (brt_i). There were 10 brt_i models, one for each of the five pre- and five post-anthesis windows. Each window contained 35 weather-based predictors, plus the categorical predictors CORN, RESIST, and TYPE. Another BRT model was fit to the unrestricted set of all 350 weather-based predictors plus the three categorical predictors (i.e., no subsetting of predictors by window). We called this model the unrestricted BRT model (brt_u).

Fitting BRTs requires tuning parameters for the total number of trees (*nt*), the tree complexity (*tc*), the learning rate (*lr*), and the bag fraction (η). At each iteration of the algorithm, a fraction η of the training data is sampled at random without replacement and used to grow the next tree in the sequence. However, the BRT algorithm can continue adding trees until the data are completely overfitted, which reduces a model's ability to predict to new data (20,21). Therefore, some form of regularization is desirable, and is accomplished with BRTs by (i) limiting *nt* and (ii) downweighting (shrinking) the contribution of each successive tree added to the model, thereby slowing the rate at which the algorithms "learns" the data. The amount of shrinkage is controlled by the *lr* parameter. The *tc* parameter controls the level of interaction allowed (*tc* = 1 no interactions [i.e., a strictly additive model], *tc* = 2 allows up to pairwise interactions, and so on). Suitable values for *tc*, *lr*, and η together determine *nt*.

We used *tc* = 3 and η = 0.75 for all models; *lr* = 0.005 was used for the pre-anthesis brt_i models plus the brt_u model and *lr* = 0.01 was used for the post-anthesis brt_i models. These parameter settings consistently gave a minimum of 1,000 *nt* across all fitted models and were initially guided by practical suggestions for a binary response with a relatively small data set (15), followed by exploratory model fittings (data not shown). Model fitting was done with the dismo package (ver. 0.7-17) in R (64-bit version 2.15.0; R Foundation for Statistical Computing, Vienna) on Windows 7 Professional ((Microsoft Corp., Redmond, WA). The model-fitting process was repeated with each of the five imputed training data sets.

Model simplification. Our previous experience with this data set (51) was that not all predictors were likely to contribute to predictive performance. Also, with small data sets redundant, noninformative predictors can degrade model performance by increasing the variance (15,19). In the interest of parsimony, we investigated simplified brt_i models by dropping up to 33 predictors (one at a time) per model from the original 35 while, at the same time, always retaining RESIST as an epidemiologically meaningful predictor (14,56). The simplifications were handled with the gbm.simplify function in the dismo package. Ten-fold cross-validation was used to determine the sequence in which predictors could be removed from the model (starting with the lowest contributing predictor), based on the change in model deviance relative to that of the full model. Note that, because of the stochasticity built into the algorithm (via η and cross-validation), the set of predictors retained after each simplification is not necessarily the same across repeated runs of the algorithm or across the different imputed versions of the training data. We then refit BRTs to the sets of retained predictors (one predictor having been removed at a time) and recorded the cross-validated area under the receiver operating characteristic (ROC) curve (cv AUC score) and associated standard error for each refit model. The parameters *tc*, *lr*, and η were the same as above, and *nt* was still a minimum of 1,000 trees.

A plot of the cv AUC score suggested that brt$_\text{i}$ models could be simplified to six predictors without any loss in predictive performance (Fig. 1A). The visual impression was confirmed statistically by comparing the cv AUC scores of the full and simplified brt$_\text{i}$ models by a *Z* test calculated under the conservative assumption of perfect correlation between the AUC estimates for ROC curves built from the same data (28). We also simplified the (unrestricted) brt$_\text{u}$ model down to 10 predictors (again with RESIST included in every model). In this latter case, the cv AUC scores indicated performance degradation with higher numbers of retained predictors and an "optimal" number of predictors of 35 to 50 (Fig. 1B).

There is the problem of additional stochasticity introduced by having five different imputed versions of the data set. Let m_j represent a BRT model *m* fitted to the *j*th imputed version of the training data (j = 1,…,5). For a brt$_\text{i}$ model fitted to predictors within a given window and period, we tabulated the frequencies of predictor occurrence when the m_j were simplified to five weather predictors plus RESIST. For the brt$_\text{u}$ model, we tabulated the frequency of predictor occurrence when the m_j were simplified to 49 weather-based predictors plus RESIST. Not all predictors were found at a frequency of five across imputations within simplified models, perhaps because of the built-in stochasticity in the BRT algorithms introduced via η, the variability among imputations, and noise within the data.

The brt$_\text{i}$ models were then refitted to the subset of predictors which appeared more than once in the frequency tabulations described above, all the time retaining RESIST as a predictor. The brt$_\text{u}$ model was refit to the subset of predictors (including RESIST) with a frequency of five. The deliberate focus on "stronger" predictors should lead to more stable models (19). This last subset of predictors is shown in Table 1 and consisted of weather-based predictors only; TYPE and CORN were not among the predictors

TABLE 1. Weather-based predictors in boosted regression tree models

Predictor type, full acronym[a]	Description	Window (w)	Abbreviated acronym[b]	Group[c]	Model[d]
Vapor pressure deficit (VPD; kPa)					
VPD.A.w.12H	Mean VPD per overnight period	pre.7	vpd.4	1	brt$_\text{i}$
		pre.15	vpd.13	1	brt$_\text{i}$
		post.7	vpd.19	1	brt$_\text{i}$
		post.14	vpd.25	1	brt$_\text{i}$
VPD.L45.w.12H	Number of h VPD ≤ 0.45 kPa overnight	pre.5	vpd.3	3	brt$_\text{i}$
		pre.7	vpd.6	3	brt$_\text{u}$
		post.15	vpd.30	3	brt$_\text{i}$
Dew-point depression (DD; °C)					
DD.A.w.12H	Mean DD per overnight period	pre.5	dd.1	4	brt$_\text{i}$
		pre.7	dd.3	4	brt$_\text{i}$
		pre.10	dd.5	4	brt$_\text{i}$ and brt$_\text{u}$
		pre.14	dd.7	4	brt$_\text{i}$
		pre.15	dd.9	4	brt$_\text{i}$
		post.10	dd.15	4	brt$_\text{i}$
DD.L1.w.12H	Number of h DD < 1°C overnight	pre.14	dd.8	5	brt$_\text{i}$
Relative humidity (RH; %)					
RH.A.w.12H	Mean RH per overnight period	pre.10	rh.21	6	brt$_\text{i}$
		pre.15	rh.41	6	brt$_\text{i}$
		post.5	rh.51	6	brt$_\text{i}$
		post10	rh.71	6	brt$_\text{i}$
RH.A.w.24H	Mean RH per day	pre.7	rh.12	7	brt$_\text{i}$ and brt$_\text{u}$
		pre.10	rh.22	7	brt$_\text{i}$ and brt$_\text{u}$
		pre.14	rh.32	7	brt$_\text{i}$ and brt$_\text{u}$
		post.5	rh.52	7	brt$_\text{i}$ and brt$_\text{u}$
		post.7	rh.62	7	brt$_\text{i}$
RH.G80.w.12H	Number of h RH ≥ 80% overnight	pre.14	rh.33	8	brt$_\text{i}$ and brt$_\text{u}$
		pre.15	rh.43	8	brt$_\text{i}$
RH.MXRLG80.w.24H	Maximum (run length [number of h RH ≥ 80%])[e]	post.7	rh.67	12	brt$_\text{u}$
		post.10	rh.77	12	brt$_\text{u}$
		post.15	rh.97	12	brt$_\text{i}$

(*continued on next page*)

[a] Acronyms follow the naming convention a.b.w.c, where a indicates whether the variable is summarizing vapor pressure deficit (VPD; kPa), dew-point depression (DD; °C), relative humidity (RH; %), temperature (T; °C), rainfall (R; mm), or T and RH conditions being met simultaneously (TRH); b indicates the type of summary measure (see the Description column), such as S for sum, A for average (mean), L for less than, and so on; w is a placeholder for one of 10 vectors of hourly time series weather data ($\boldsymbol{w}_{\text{pre.5}}$, $\boldsymbol{w}_{\text{pre.7}}$, $\boldsymbol{w}_{\text{pre.10}}$, $\boldsymbol{w}_{\text{pre.14}}$, $\boldsymbol{w}_{\text{pre.15}}$, $\boldsymbol{w}_{\text{post.5}}$, $\boldsymbol{w}_{\text{post.7}}$, $\boldsymbol{w}_{\text{post.10}}$, $\boldsymbol{w}_{\text{post.14}}$, and $\boldsymbol{w}_{\text{post.15}}$), where the subscript in the preceding list indicates if the weather data span 5, 7, 10, 14, or 15 days pre- or post-anthesis; c indicates a 24-h day (24H = 0800 h to 0800 h), or a 12-h overnight period (12H = 2000 h to 0800 h). These are the same acronyms used by Shah et al. (51).

[b] An abbreviated predictor name used for clarity in labeling Figures.

[c] Within groups, predictors vary only over the defining window. For example, group 1 contains predictors measuring mean VPD per day, with five in the pre-anthesis period (5-, 7-, 10-, 14-, and 15-day windows) and five in the post-anthesis period (5-, 7-, 10-, 14-, and 15-day windows). The input predictor matrix had 35 groups and 10 predictors per group (51). Note that not all 35 groups or predictors per group appeared in the final models; only those that do so are shown in the Table 1. See Shah et al. (51) for a full listing of predictor groups.

[d] Abbreviations: brt$_\text{i}$ = individual-window boosted regression tree model and brt$_\text{u}$ = boosted regression tree model built on the unrestricted set of predictors.

[e] The maximum run length of the number of consecutive hours in which RH ≥ 80%.

retained during the simplification stages. There were five to nine weather-based predictors per brt$_i$ model, and 18 such predictors for the brt$_u$ model.

Model summaries. The simplified brt$_i$ and brt$_u$ models were summarized by their cv AUC scores on the training data. The Youden index (YI), defined as the maximum difference between the true-positive and false-positive predictive rates (28), was used as an optimal cut-point in converting predicted probabilities to classifications as major or non-major FHB epidemics. We used bubble plots to succinctly summarize several model features, including predictor type, group, the number of predictors per group, and the mean relative influence per group, where these measures were plotted separately for the pre- and post-anthesis periods. In BRTs, a predictor's relative influence is a scaled measure of its contribution based on the number of times the predictor is selected as a criterion for splitting the data within a tree and the improvement in performance resulting from the split (18). Predictors with stronger influence on the response variable have higher relative influences. All measures described above were means over the five m_j for a given brt$_i$ model or the brt$_u$ model. Mean predicted probabilities were calculated before estimating YI. For each model, we also plotted the partial-dependence plots and identified the main pairwise interactions (21). Partial-dependence plots are a graphical representation of the effect of a given predictor on the response function (logit in our case) after accounting for the average effects of all other predictors in the model. For pairwise interactions, the magnitude of the metric measuring the interaction increases (unbounded) with the size of the interaction; a value of zero indicates no interaction (15). The interaction metrics are relative, which allows one to rank the pairwise interactions within any given model. They should not be compared across different models, however.

Model averaging. For each observation, the predicted probabilities returned by the $m = 1, 2,\ldots,M$ brt$_i$ models ($M = 10$ for the five window lengths in each of the pre- and post-anthesis periods) were averaged by estimating the weighted mean predicted probability, where the weights (w) were estimated from the model deviance (d):

$$w_m = \frac{e^{-0.5d_m}}{\sum_{i=1}^{M} e^{-0.5d_i}} \qquad (2)$$

YI was then estimated from the model-averaged predicted probabilities, and the observations classified as major or non-major epidemics accordingly.

Test performance. The predictive performances of the BRT models were evaluated on a holdout (test) data set of 158 observations not used in model development or validation, and

TABLE 1. (*continued from preceding page*)

Predictor type, full acronym[a]	Description	Window (w)	Abbreviated acronym[b]	Group[c]	Model[d]
Temperature (t; °C)					
T.A.w.24H	Mean T per day	pre.5	t.1	16	brt$_i$
		pre.7	t.7	16	brt$_i$
		pre.10	t.13	16	brt$_i$
		pre.14	t.19	16	brt$_i$
		pre.15	t.25	16	brt$_i$
		post5	t.31	16	brt$_i$ and brt$_u$
		post.7	t.37	16	brt$_i$ and brt$_u$
		post.10	t.43	16	brt$_i$
		post.14	t.49	16	brt$_i$
T.15T30.w.24H	Number of h (15°C ≤ T ≤ 30°C)	post.7	t.39	18	brt$_i$ and brt$_u$
		post.10	t.45	18	brt$_i$
		post.14	t.51	18	brt$_i$
T.L9.w.24H	Number of h (T < 9°C)	post.10	t.46	19	brt$_i$
		post.14	t.52	19	brt$_i$
T.L15.w.24H	Number of h (T < 15°C)	pre.10	t.17	20	brt$_i$
T.G30.w.24H	Number of h (T > 30°C)	post.10	t.48	21	brt$_u$
Rainfall(R; mm)					
R.S.w.24H	Total rainfall	pre.5	r.2	23	brt$_i$
		pre.14	r.23	23	brt$_i$
		post.5	r.37	23	brt$_i$ and brt$_u$
		post.7	r.44	23	brt$_i$
R.AD.w.24H	Mean rainfall per day	pre.14	r.26	26	brt$_i$
		pre.15	r.33	26	brt$_i$
		post.14	r.61	26	brt$_i$
		post.15	r.68	26	brt$_i$ and brt$_u$
Simultaneous T and RH conditions					
TRH.9T30nRHG80.w.12H	Number of h (9°C ≤ T ≤ 30°C & RH ≥ 80%) overnight	pre.10	trh.9	29	brt$_i$
		pre.14	trh.13	29	brt$_i$
		post.15	trh.37	29	brt$_i$
TRH.15T30nRHG80.w.12H	Number of h (15°C ≤ T ≤ 30°C & RH ≥ 80%) overnight	pre.7	trh.7	31	brt$_i$
		post.5	trh.23	31	brt$_i$
		post.10	trh.31	31	brt$_i$
		post.14	trh.35	31	brt$_i$
TRH.15T30nRHG90.w.12H	Number of h (15°C ≤ T ≤ 30°C & RH ≥ 90%) overnight	post.7	trh.28	32	brt$_i$
		post.14	trh.36	32	brt$_i$
TRH.9T30nRHG90.w.24H	Number of h (9°C ≤ T ≤ 30°C & RH ≥ 90%)	pre.14	trh.74	36	brt$_i$
		pre.15	trh.78	36	brt$_i$ and brt$_u$
TRH.15T30nRHG80.w.24H	Number of h (15°C ≤ T ≤ 30°C & RH ≥ 80%)	pre.5	trh.63	37	brt$_i$
		pre.10	trh.71	37	brt$_i$
		pre.15	trh.79	37	brt$_u$
		post.15	trh.99	37	brt$_i$ and brt$_u$
TRH.15T30nRHG90.w.24H	Number of h (15°C ≤ T ≤ 30°C & RH ≥ 90%)	post.7	trh.88	38	brt$_i$

which were randomly selected from the original full data set. For each model, the following test performance measures were calculated: (i) AUC, (ii) the classification matrix resulting from using YI as the cut-point, (iii) sensitivity (the proportion of epidemics correctly classified), (iv) specificity (the proportion of non-epidemics correctly classified), (v) Kappa (a measure of the proportion of correctly classified observations after accounting for the probability of chance correct classification), and (vi) the overall misclassification rate.

Comparison with logistic regression models. We had previously developed additive logistic regression and classical logistic regression models on the same training and test data (51). Because the same test data set was used in the current and previous article, the predictive performances of the logistic and BRT model classes could be compared by a McNemar test (2).

RESULTS

Full BRT models. The cross-validated AUCs for the full (i.e., fit to all available predictors in a given window) brt$_i$ models built on the training data were 0.82, 0.84, 0.84, 0.87, and 0.86 for the 5-, 7-, 10-, 14-, and 15-day pre-anthesis windows, respectively. For the 5-, 7-, 10-, 14-, and 15-day post-anthesis windows, the cross-validated AUCs were 0.87, 0.88, 0.91, 0.87, and 0.87, respectively. *Z* tests comparing the AUCs of full and simplified brt$_i$ models (28) indicated that the number of weather-based predictors could be dropped to five with no loss in predictive performance. For the brt$_u$ model, predictive performance degraded when >50 predictors were retained in the model (Fig. 1B).

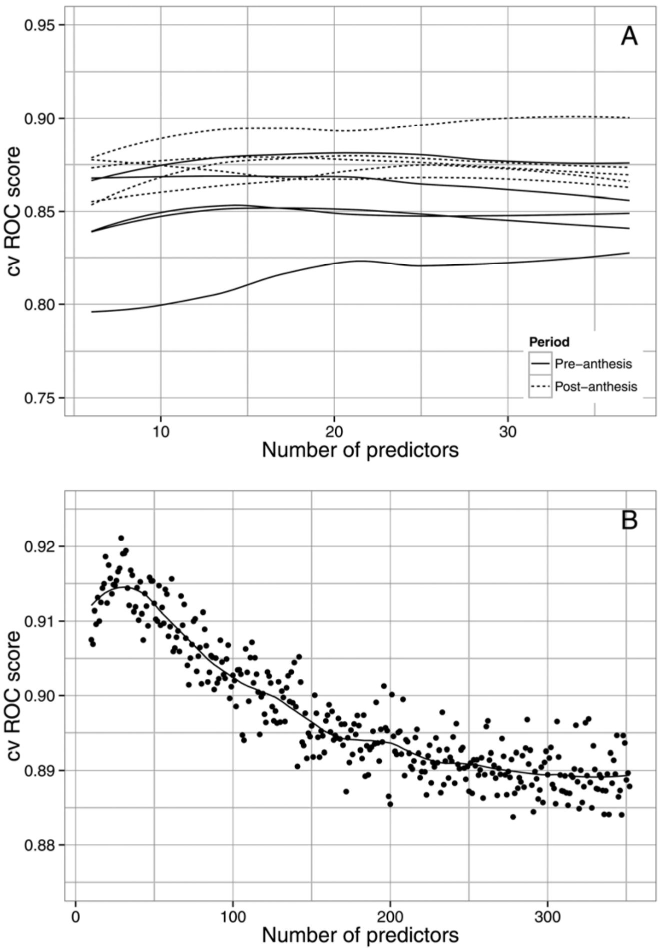

Fig. 1. A, Generalization performance (estimated by the cross-validated receiver operating characteristic [ROC] score) when the number of predictors in individual-window (of length 5, 7, 10, 14, or 15 days) boosted regression tree (BRT) models is reduced from 38 per window to 6, with cultivar resistance always being retained as a predictor. Each loess smooth represents mean model-fitted values over five imputed versions of the training data. There are five curves each for the pre- and post-anthesis periods, corresponding to the five windows. **B,** Generalization performance when the number of predictors in a BRT model (given access to all available predictors) is reduced from the full complement of 353 to a minimum of 10, with cultivar resistance always being retained as a predictor. Points are the mean values over five imputed versions of the training data. A loess smooth (solid black line) is shown.

Simplified BRT models. There were five to nine weather-based predictors per simplified brt$_i$ model (Table 1). Across all the pre-anthesis brt$_i$ models, there were 33 weather-based predictors belonging to 15 predictor groups. Across the post-anthesis brt$_i$ models, there were 30 weather-based predictors belonging to 16 predictor groups. In all, 11 predictor groups (Table 1, groups 1, 3, 4, 6, 7, 16, 23, 26, 29, 31, and 37) were common to both the pre- and post-anthesis periods. All six predictor types (DD, R, RH, T, TRH, and VPD) were represented in both the pre- and post-anthesis brt$_i$ models.

The simplified brt$_u$ model (across all window lengths pre- and post-anthesis) had 18 weather-based predictors (Table 1), of which 8 were from the pre-anthesis period and the remaining 10 from the post-anthesis period. These predictors belonged to 12 groups and all six predictor types were represented. In all, 13 of the 18 weather-based predictors in the brt$_u$ model were in common with the brt$_i$ models; 5 (vpd.6, rh.67, rh.77, t.48, and int.79) were present in the brt$_u$ model only (Fig. 2).

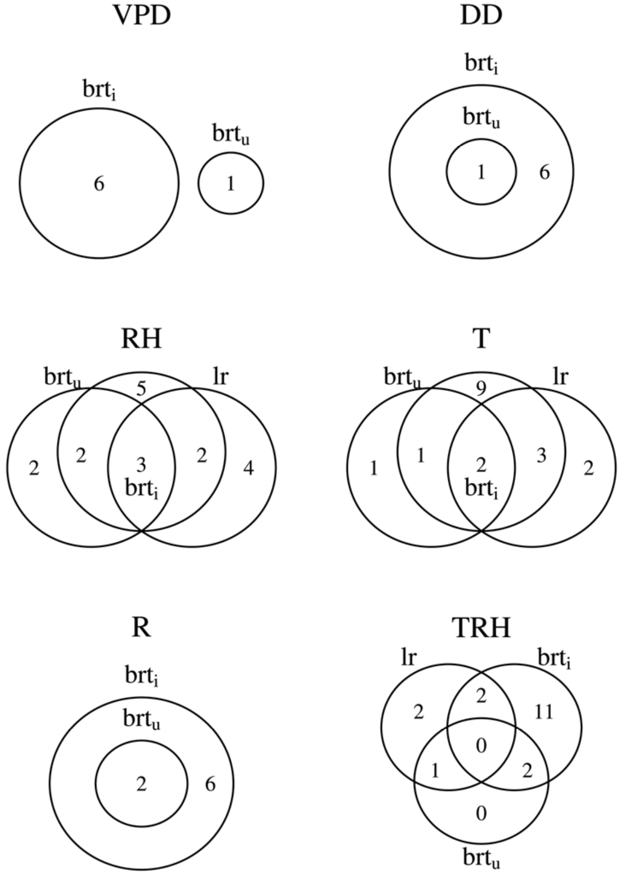

Fig. 2. Venn diagram depictions of the sets of weather-based predictors in (i) 15 logistic regression models (lr) (51), (ii) 10 boosted regression trees built on individual pre- and post-anthesis windows (brt$_i$), and (iii) a boosted regression tree model built by predictor selection from the unrestricted candidate set (brt$_u$). Predictors are graphed by type: VPD = vapor pressure deficit; DD = dew-point depression; RH = relative humidity; T = temperature; R = rainfall; TRH = simultaneous T and RH conditions. Number of predictors per type is shown within the circles. Overlaps among circles indicate predictors common to lr, brt$_i$, or brt$_u$.

The cross-validated AUCs (and standard errors) on the training data for the 5-, 7-, 10-, 14-, and 15-day simplified pre-anthesis brt$_i$ models were 0.802 (0.026), 0.832 (0.026), 0.843 (0.023), 0.875 (0.015), and 0.872 (0.020), respectively. The corresponding numbers for the 5-, 7-, 10-, 14-, and 15-day simplified post-anthesis brt$_i$ models were 0.867 (0.020), 0.852 (0.022), 0.881 (0.019), 0.851 (0.023), and 0.879 (0.018), respectively. For the simplified brt$_u$ model, the cross-validated AUC (plus standard error) on the training data was 0.908 (0.016).

Model performance statistics on the test data are shown in Figure 3. The AUCs trended higher with increasing window length in both the pre- and post-anthesis periods. YI (optimized on the training data) was 0.32 to 0.45 for the brt$_i$ models and, therefore, was higher than the naïve (unoptimized) probability estimate of 0.31 representing the observed proportion of major epidemics in the data set (51). With the YI (estimated from the fitted probabilities on the training data) serving as the classification cut-points on the test data, a trade-off between sensitivity and specificity was apparent; in that, for a given model, a higher sensitivity (compared with another model) was associated with a lower specificity. The brt$_i$ models were overall similar in their misclassification rates.

Predictor relative influences. We first described the full unrestricted BRT model fit to the complete predictor set (350 weather-based plus TYPE, CORN, and RESIST). The three categorical predictors had a total combined relative influence of 3.67%; CORN and TYPE together had a combined relative influence of 0.17%. CORN and TYPE were dropped from further consideration when simplifying the full BRT model but RESIST was retained because of its epidemiological relevance (56). Three predictor groups stood out from the rest in the full unrestricted BRT model: group 7 (mean RH per day), group 16 (mean T per day), and group 36 (number of hours 9°C ≤ T ≤ 30°C and RH ≥ 90%). Predictors belonging to these three groups had total relative influences of 11.1, 9.7, and 8.6%, respectively, together accounting for 29.4% of the total relative influence

The simplified BRT models. The influence of predictor groups 7, 16, and 36 was also evident in the simplified brt$_u$ and brt$_i$ models (Figs. 4 and 5). Figure 4 summarizes key features (predictor type and group, total number of predictors per group, and the mean relative influence per group), in which summaries were

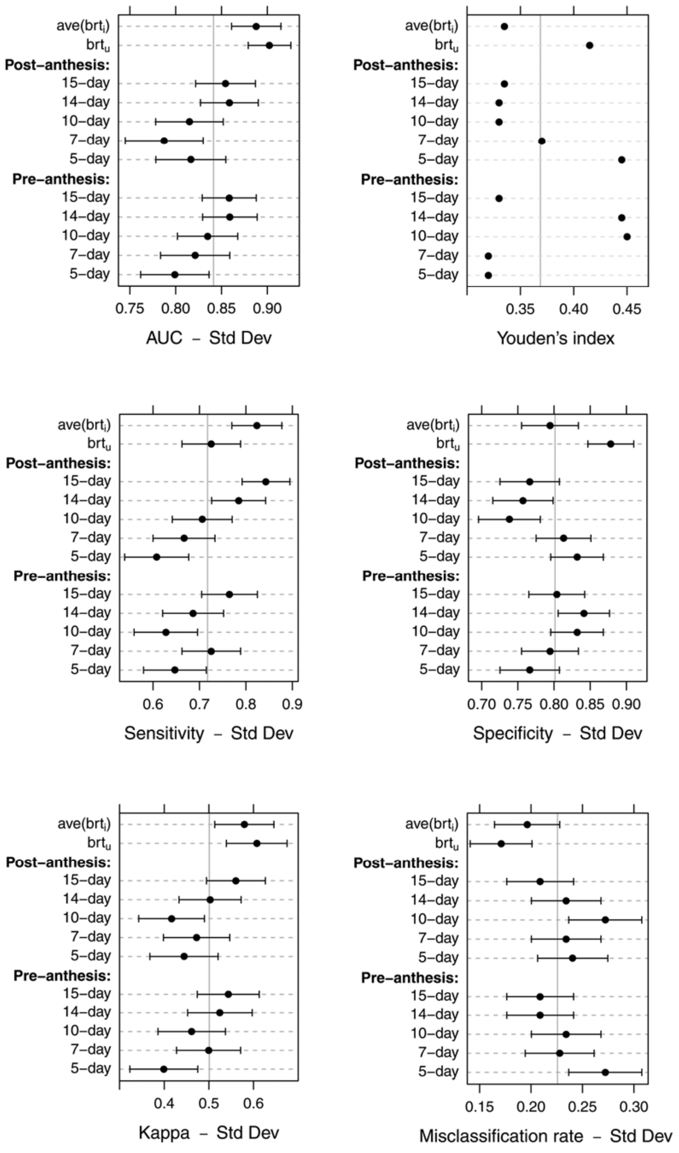

Fig. 3. Predictive performance statistics of fitted boosted regression tree models on the test data set. Models were fitted to predictors within fixed-length (5-, 7-, 10-, 14-, or 15-day) windows within the pre- and post-anthesis periods (i.e., 10 brt$_i$ models); brt$_u$ = model fit to the unrestricted set of predictors (i.e., no subsetting by window or period); ave(brt$_i$) = model-averaged performance over the 10 brt$_i$ models. Means over the brt$_i$, brt$_u$, and ave(brt$_i$) models are indicated by the vertical gray lines.

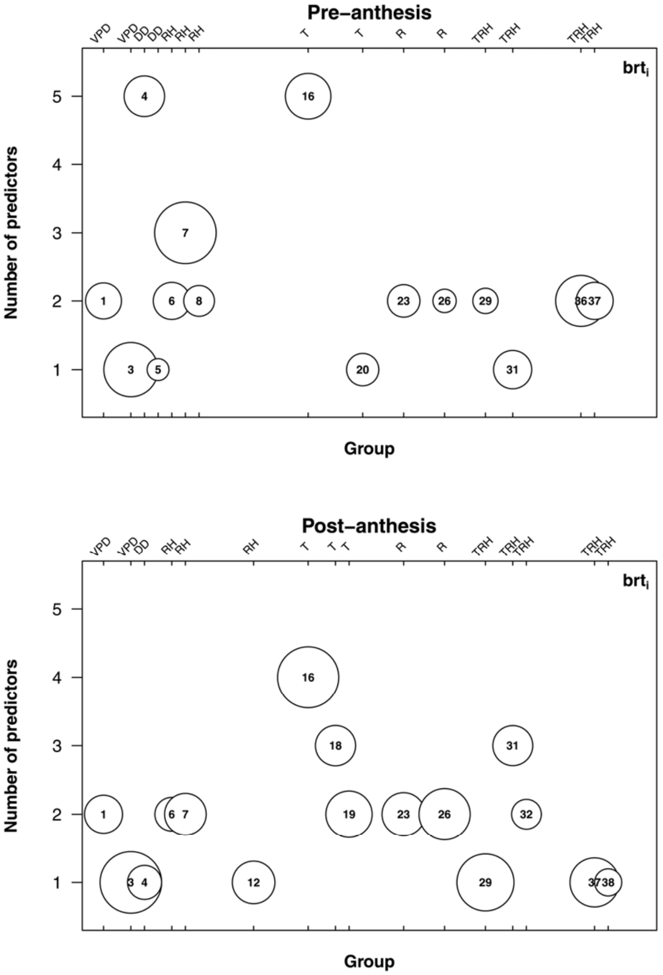

Fig. 4. Summary of weather-based predictors over 10 fitted brt$_i$ models (five each in the pre- and post-anthesis periods). Bubble size is proportional to the mean relative influence of predictors within a group (Table 1). Groups are indicated by the numeric labels within the bubbles. Labels along the top border indicate predictor type (DD = dew-point depression, R = rainfall, RH = relative humidity, T = temperature, VPD = vapor pressure deficit, TRH = simultaneous T and RH conditions). The *y*-axis shows the total number of predictors per group summed over the five individual models per period.

made over the five windows (i.e., five $\mathrm{brt_i}$ models) per period; data are presented separately for the pre- and post-anthesis periods. Figure 5 does the same for the simplified $\mathrm{brt_u}$ model. Shifts in the mean relative influences of predictor types were apparent when transitioning from the pre- to post-anthesis period. Among the more noticeable shifts in mean relative influence were (i) a decrease for the moisture-summarizing predictor types DD, RH (particularly group 7), and VPD; and (ii) an increase for R (groups 23 and 26), T (particularly group 16), and TRH-type predictors (groups 29 and 31), the latter two groups representing relatively warm temperatures combined with high relative humidity. For the $\mathrm{brt_u}$ model, the total relative influence was 49.7% for the pre-anthesis weather-based predictors and 45.2% for the post-anthesis weather-based predictors. The relative influence of RESIST was 5.1%.

Partial-dependence plots. The interpretation of partial-dependence plots is illustrated with the 7-day pre-anthesis $\mathrm{brt_i}$ model (Fig. 6). This particular model had six predictors. Smoothed representations of the model-predicted values on the logit scale were nonlinear, also true of many predictors in the other fitted BRT models. The vertical lines in the panels for rh.12, t.7, and vpd.4 indicate potential cut-points delineating predictor ranges associated with (substantially) reduced or increased risk of major FHB epidemics based on changes in the predicted values. The two vertical lines for t.7 (mean T per day), for example, draw attention to the range 14°C ≤ t.7 ≤ 22°C associated with an increased risk of major FHB epidemics compared with other values of t.7. There also appeared to be a quadratic effect of t.7 on the fitted logit. The risk of a major FHB epidemic increased sharply when the group 7 predictor rh.12 (mean daily RH) was >70%. Partial-dependence plots for the remaining $\mathrm{brt_i}$ models and the $\mathrm{brt_u}$ model also supported a 70% threshold for other group 7 pre-anthesis predictors (i.e., rh.22 and rh.32) (Table 1) (data not shown). However, there was no evident support for a 70% threshold among group 7 post-anthesis predictors. The relative influence of mean daily RH (group 7) in the pre-anthesis period was higher than during the post-anthesis period (Figs. 4 and 5). The partial-dependence plots for the $\mathrm{brt_u}$ model (data not shown) were consistent with those for the $\mathrm{brt_i}$ models.

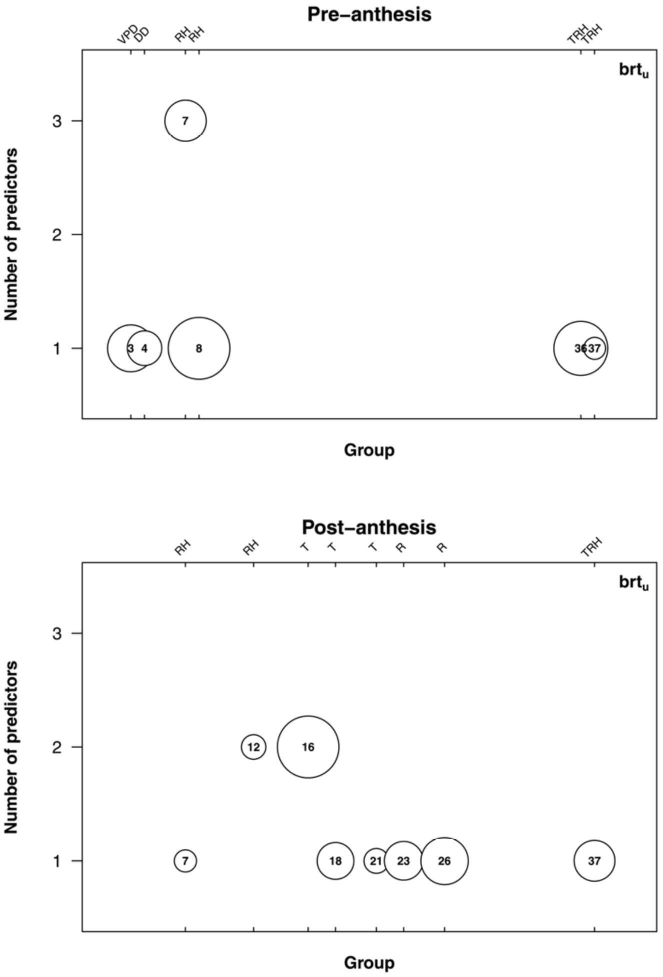

Fig. 5. Summary of the simplified $\mathrm{brt_u}$ model, where weather-based predictor selection was not restricted by window or period, in contrast to the $\mathrm{brt_i}$ models, which were built from predictor sets restricted to individual pre- and post-anthesis windows. Data are presented by predictors in the pre- and post-anthesis periods. Groups are indicated by the numeric labels within the bubbles. Labels along the top border indicate predictor type (DD = dew-point depression, R = rainfall, RH = relative humidity, T = temperature, VPD = vapor pressure deficit, TRH = simultaneous T and RH conditions). The *y*-axis shows the total number of predictors per group.

Pairwise interactions. Pairwise interactions among the predictors for each of the $\mathrm{brt_i}$ models are shown in Figure 7. Among the more prominent interactions in the pre-anthesis period are those between RESIST and moisture-based predictors (dd.1, rh.12, rh.22, and rh.32) (e.g., see the light-colored box for the combination of rh.12 and RESIST in the 7-day pre-anthesis case). An example of interaction effects on the fitted logit is depicted in Figure 8, which shows the mitigating effect of cultivar resistance on the risk of major FHB epidemics even when high RH would usually favor such epidemics. Group 16 predictors (t.1, t.7, t.13, t.25, t.31, t.37, t.43, and t.49; mean T per day) were involved in several of the strongest interactions in both pre- and post-anthesis periods (Fig. 7). The largest pairwise interactions for the $\mathrm{brt_u}$ model were rh.22–RESIST, int.99–t.39, rh.32–rh.33, and rh.32–RESIST (data not shown).

Model averaging. The sensitivity of the model-averaged BRT on the test data was better than that of several of the $\mathrm{brt_i}$ models, with little change in specificity (ave[$\mathrm{brt_i}$]) (Fig. 3) and with an overall lower misclassification rate. The *P* values of McNemar tests for all pairwise model comparisons are graphically depicted in Figure 9. Low *P* values (darker colors) indicate a bigger difference between two models in terms of the classification errors they make. In terms of discordant classification errors, the model-averaged BRT was similar to and significantly better than two (5-day pre-anthesis and 10-day post-anthesis) of the $\mathrm{brt_i}$ models (Fig. 9). The performance of the model-averaged BRT on the test data was close to the performance of the $\mathrm{brt_u}$ model in all metrics except specificity (Figs. 3 and 9).

Comparison with logistic regression models. TYPE, CORN, and RESIST were core categorical predictors in the logistic regression models developed previously (51) from these same data but had a combined relative influence of 3.67% in the full unrestricted BRT model. DD, R, and VPD predictors, absent from the logistic regression models, were included in the simplified $\mathrm{brt_i}$ and $\mathrm{brt_u}$ models (Fig. 2). Among the RH- and T-type predictors, there was partial overlap among the $\mathrm{brt_i}$, $\mathrm{brt_u}$, and logistic regression models. The T-based predictors were more prominent in the $\mathrm{brt_i}$ models than in the logistic regression models. The TRH-type predictors (combinations of T and RH conditions) were more frequent in the $\mathrm{brt_i}$ models than in either the $\mathrm{brt_u}$ or logistic regression models.

There were 21 weather-based predictors over the 15 logistic regression models we had developed earlier (51); these same 21 predictors had a total relative influence of 20.2% in the full unrestricted BRT model. By comparison, the top 21 weather-based predictors in the full unrestricted BRT model had a total relative influence of 46.5%.

The BRT models in general made fewer classification errors than the logistic regression models (Fig. 9). The two exceptions were the 5-day pre-anthesis $\mathrm{brt_i}$ model, which made, statistically, fewer classification errors than only three logistic regression models (ID 3, 11, and 14); and the 10-day post-anthesis $\mathrm{brt_i}$

models, which made statistically fewer classification errors than logistic regression models 3, 10, 11, and 14 (Fig. 9).

DISCUSSION

An analysis based on BRTs deepened our general understanding of the relationship between weather-based predictors and major FHB epidemics in the United States. The BRT models suggested new weather-based predictor formulations which may be more strongly associated with FHB epidemics or which may have a higher signal-to-noise ratio. Averaging the brt_i-predicted probabilities resulted in a predictive performance that was superior to that of any of the brt_i models individually, and similar to that of the brt_u BRT model. On average, the brt_i models had a 31% lower misclassification rate than the 15 logistic regression models we had developed earlier (51). Logistic regression is a popular and useful analytical tool for fitting binary response models to data, and remains a cornerstone in predictive modeling. The model structure is relatively simple and linear, and (via odds ratios) the fitted model is often directly interpretable (21). Yet, logistic regression models are often too simplistic for the predictive modeling of complex data involving numerous intercorrelated variables (15). BRTs can work with complex data and are capable of generating highly predictive yet interpretable models (15,21), although interpretation requires somewhat different analytical and graphical tools compared with the approaches used for logistic or linear models.

Results indicated that the BRT algorithm was more capable than subset selection in identifying which predictors in the candidate set were most associated with FHB epidemics (51). The 21 weather-based predictors identified through the subset selection-logistic regression approach (51) had a combined relative influence that was some 56% lower than the combined relative influence of the top 21 weather-based predictors identified by a BRT model fit to the same training data. In comparison with some of the previously developed logistic regression models, the BRT approach did not require separate models or parameters for spring and winter wheat types or the presence or absence of corn residue (37,38,51). In the logistic regression models, we kept TYPE and CORN a priori as predictors, but it was not straightforward to gauge their influence on prediction. The BRT models showed that both TYPE and CORN had very low relative influence and, consequently, were dropped during model simplification, indicating the overarching influence of weather on FHB epidemics regardless of spring or winter wheat or local inoculum pressure from infested crop debris. Although infested host crop residue in a field suggests a higher local risk of FHB epidemics (55), in general, our results and those of others (49) suggest that the residue effect is far less influential than weather effects on FHB risk. The regional atmospheric transport of *Gibberella zeae* ascospores (34) may be one reason that we were unable to detect a local effect of corn residue in our results.

Fitted BRT models gave insights into the FHB–weather relationship not discernible with the linear regression methods typically used to model FHB epidemics (45). These insights were gained from the ability of BRTs to (i) automatically model nonlinear predictor functions, displayed via partial-dependence plots; (ii) automatically model predictor interactions; and (iii) estimate the contribution of predictors to the fitted model via relative influences. For example, when building logistic regression models (51), we had noticed that the pre-anthesis group 7 RH-type predictors (rh.12, rh.22, and rh.32; mean RH per day) were nonlinear-in-the-logit with respect to the response, and attempted to model the nonlinearity with a penalized spline approach (51). What became clearer with the BRT partial-dependence plots (and previously unnoticed) was the large jump in the risk of major FHB epidemics with mean RH ≥ 70% for these three group 7 pre-anthesis predictors. In fact, most FHB predictive models to date have used RH-type predictors based on RH ≥ 80 to 85% (7,31,39,41,47,54) or RH ≥ 90% (13,23). Notable exceptions are the use of an RH > 75% criterion by Schaafsma and Hooker (48) in toxin prediction in winter wheat and an RH ≥ 70% criterion in toxin prediction in oat (57) and wheat (58) in Europe. In addition to this RH-specific example, the BRT models automatically handled several other nonlinear predictor functions. Several interactions were identified which, as Landschoot et al. (31) point out, is cumbersome to do with traditional linear regression applied to high-dimensional data. The interactions between culti-

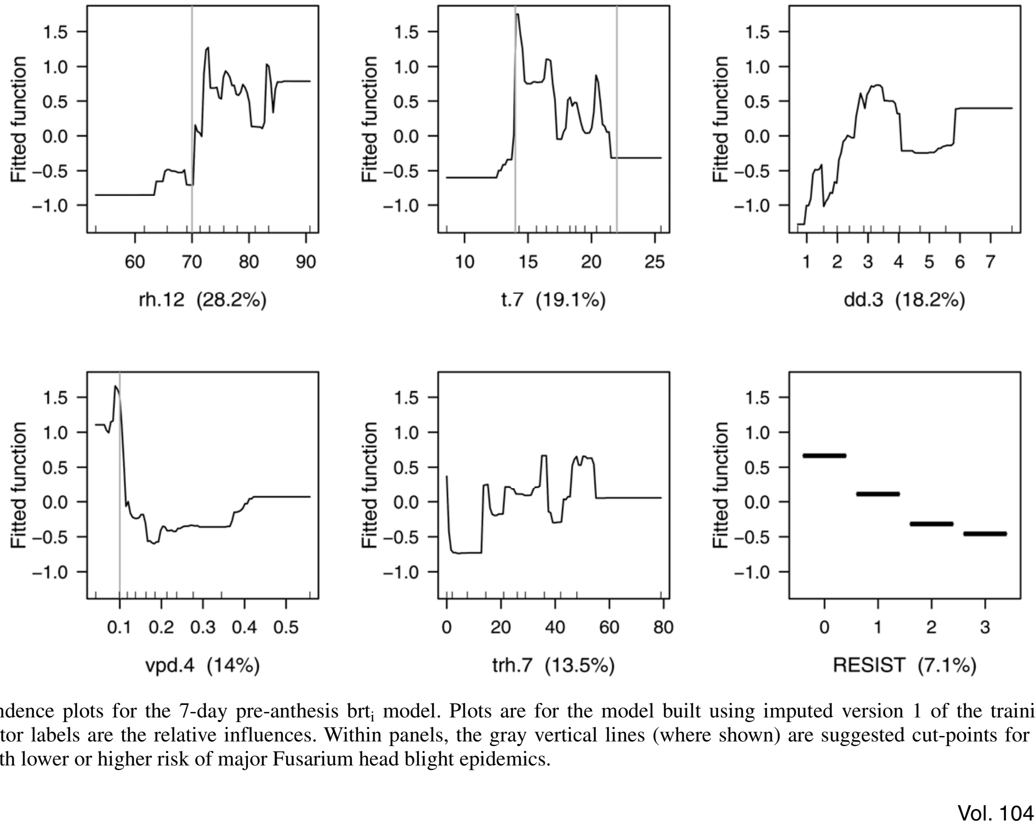

Fig. 6. Partial dependence plots for the 7-day pre-anthesis brt_i model. Plots are for the model built using imputed version 1 of the training data. Percentages following the predictor labels are the relative influences. Within panels, the gray vertical lines (where shown) are suggested cut-points for delineating predictor ranges associated with lower or higher risk of major Fusarium head blight epidemics.

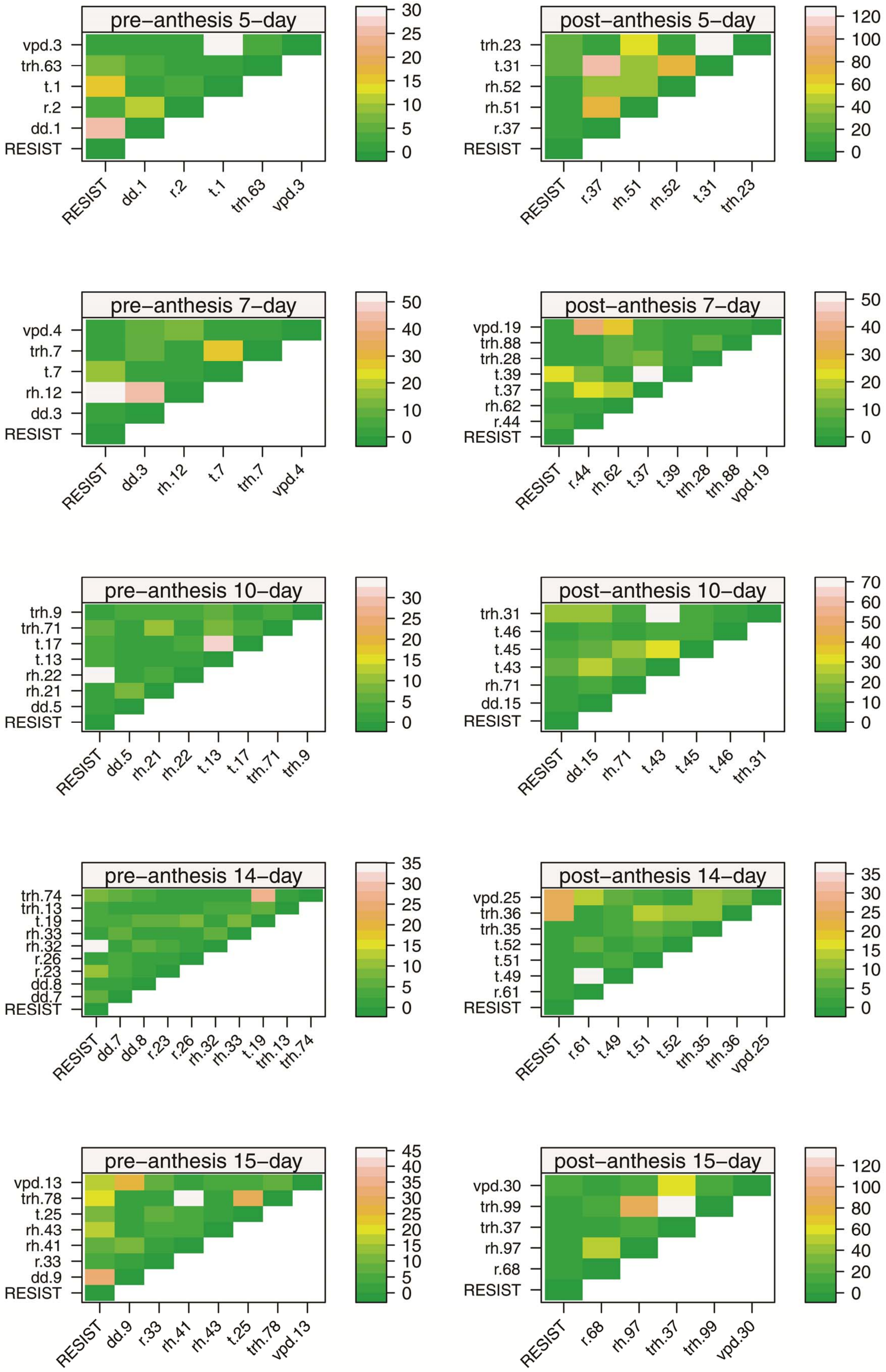

Fig. 7. Pairwise interactions for the brt$_i$ models fitted to the 5-, 7-, 10-, 14-, and 15-day windows in the pre- and post-anthesis periods. Within each model, the value of the metric quantifying the interaction increases linearly and unbounded with the size of the interaction, and zeroes indicate no interaction. Because the interaction metrics are relative, they allow the ranking of pairwise interactions within any given model but should not be compared across models. Plotted data are the mean values for a specific model fitted to each of the five imputed versions of the training data.

var resistance (an ordinal variable) and moisture-based predictors (in particular, rh.12, rh.22, and rh.32) were noteworthy. Cultivar resistance mitigated the effect of RH on the risk of major FHB epidemics but only to a point; the risk of epidemics still dramatically increased once mean RH per day was >70%, though not as much for moderately susceptible and moderately resistant cultivars.

A given weather-based predictor type did not necessarily have the same relative predictive importance across the pre- and post-anthesis periods. Moisture-based variables (especially RH type), acting as surrogates for actual wetness were better predictors of the risk of FHB in the pre-anthesis period than in the post-anthesis period. In the post-anthesis period, predictors summarizing favorable temperature rose to prominence, as well as rainfall as a direct moisture indicator. One possible hypothesis is that free moisture as rain is more crucial in the post-anthesis period than in the pre-anthesis period, and is supported by the empirical observations from experimental trials (9,10). The overall inclusion of R-based predictors by the BRT modeling algorithm contrasts with the lack of this predictor type in logistic regression models we presented earlier (51), where R-type predictors were dropped due to low selection frequency by the algorithms used. The majority of FHB models do, in fact, include rainfall-based predictors (7,12, 13,23,25,29,31,39,41,47,54). Nevertheless, the strongest predictors of FHB epidemics, based on the full brt_u model, were, in general, from one of three groups: mean RH per 24-h day (group 7), mean T per day (group 16), and the number of hours (24-h day) in which T was 9 to 30°C and RH ≥ 90% simultaneously (group 36).

Model averaging is an area of active research from both the frequentist (8) and Bayesian (22) perspectives, and can result in better predictive performance than any single model. Model averaging did not improve accuracy with our logistic regression models because of dominance by two models (51). With the brt_i models in the current article, model averaging using a simple weighting based on model deviance resulted in a predictive performance that was better than each of the 10 brt_i models individually and comparable with that of the brt_u model. In one sense, averaging the brt_i models represented an intermediary step on the way to the brt_u model, which was developed using predictors across all windows. Although BRTs and model averaging were promising, we can surmise from the wider cumulative modeling efforts to date (12,13,25,30,39,47,48,54,58) that no one modeling algorithm is likely to suffice in describing or predicting the highly sporadic FHB epidemics (26). As more algorithms are considered, FHB researchers may do well to consider ensemble prediction methods which combine several algorithms into a single predictive algorithm (46).

The BRT partial-dependence plots suggested new predictor formulations and, in the Supplementary file, we give one example where following those suggestions led to a simple logistic regression model that had a test data misclassification rate not much higher than the mean misclassification rate of the 15 (more complex) logistic regression models built previously (51). Therefore, the BRT modeling effort can be used to inform novel predictor formulations for logistic regression models. In this sense, we do not regard BRTs and logistic regression as mutually exclusive efforts in FHB modeling. Many RH-based predictors are formulated on a 80 to 90% threshold (7,12,13,23,27,31,39,41,47,54), after deliberating the empirical links between weather and FHB epidemics (44). However, our BRT results challenged that threshold assumption for the pre-anthesis period by showing that the risk of FHB epidemics rises most dramatically when mean RH is >70%, a lower threshold than previously considered (48,57,

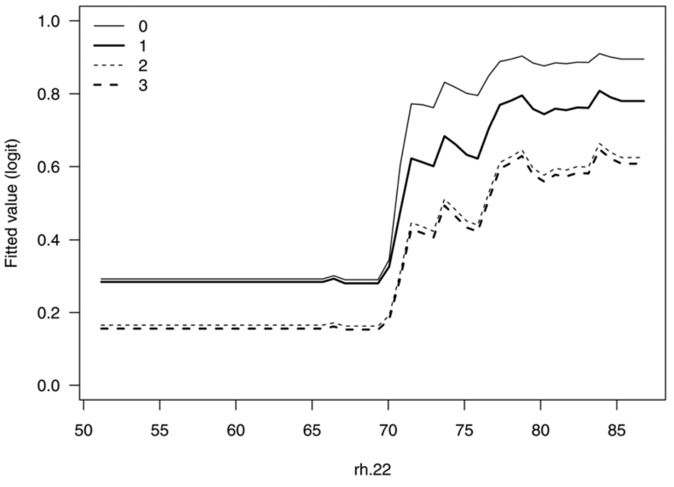

Fig. 8. Effect of the interaction between rh.22 and cultivar resistance level (0 = very susceptible, 1 = susceptible, 2 = moderately susceptible, and 3 = moderately resistant) on the fitted logit for the 10-day pre-anthesis brt_i model fit to imputed version 1 of the training data.

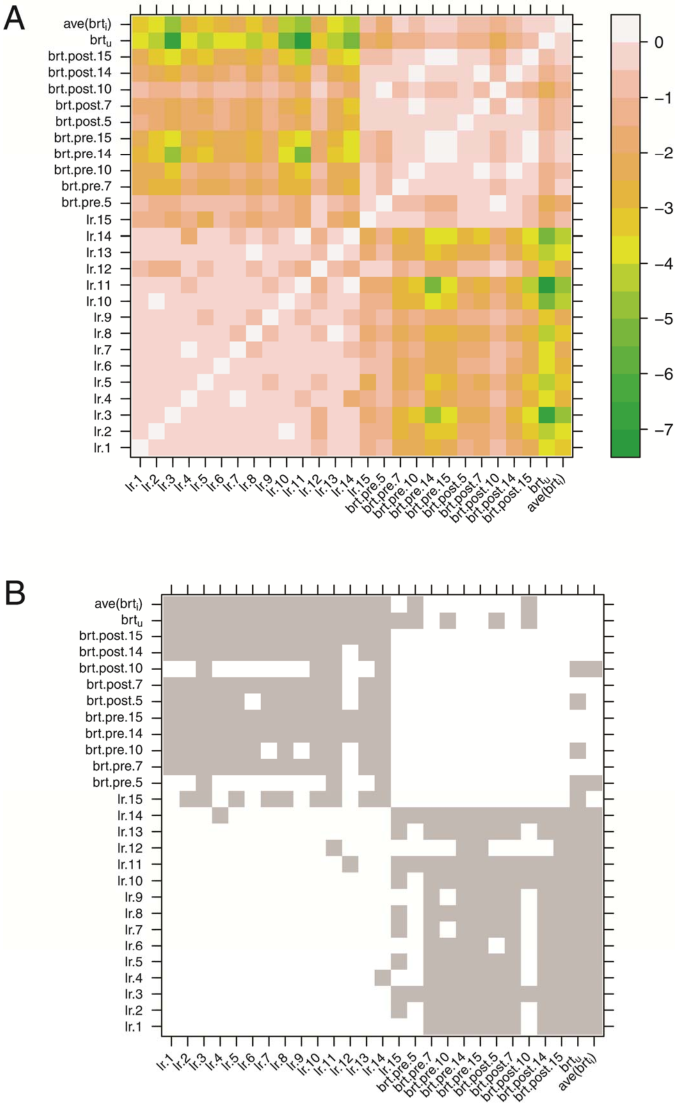

Fig. 9. Matrix plots of McNemar test *P* values in pairwise comparisons of 15 logistic regression models (lr prefix; suffix follows the ID enumeration in Table 2 of Shah et al. [51]) and boosted regression tree (brt prefix; see Figure 3 caption) model test error rates. **A,** Actual two-sided $\log_{10}(P)$ value of the McNemar test. Darker squares represent lower *P* values. On the base 10 logarithmic scale depicted, $\log_{10}(P = 1) = 0$, $\log_{10}(P = 0.01) = –2$, $\log_{10}(P = 0.001) = –3$, and $\log_{10}(P = 0.0001) = –4$. **B,** Matrix plot in which the McNemar test *P* values have been dichotomized, where white squares correspond to *P* values ≥ 0.05 and gray squares to *P* values < 0.05.

58). Weather-based predictors based on a 70% RH threshold should be considered in future logistic regression modeling efforts. Turning to temperature, the BRT pre-anthesis partial-dependence plots indicated higher risks of major FHB epidemics when mean daily T was 14 to 22°C. The optimum temperature for FHB development is ≈25°C (3). Of course, mean daily T does not reflect the diurnal fluctuations in T and, therefore, is not inconsistent with an optimum of ≈25°C. For the post-anthesis period, there was an approximately linear decrease in the logit with mean daily T of 14 to 25°C, suggesting that mean temperatures >25°C during the post-anthesis period are suboptimal (48). As with RH-based predictors, our results suggest that T-based logistic regression predictors should likewise be revisited and recast.

New classification methods are in continual stages of development in both the statistical and machine-learning communities in response to data-driven needs. Not all of these methods will necessarily be suitable for epidemiological applications. BRTs have been embraced by the ecological sciences because of their ability to model complex data sets and provide new insights (6,15,24). Our experience in using BRTs to model FHB epidemics has likewise been the same. The approach may prove useful to other pathosystems as well.

The predictive performances of the brt_i models, seen as lower misclassification rates on the test data, were better than those of the logistic regression models (51). The brt_i models also offered a better sensitivity–specificity balance (optimized using the YI as the cut-point for classification). We are not claiming that the BRT methodology is superior to logistic regression, in general, but that, for this particular data set, the BRT approach was more appropriate for modeling given the data's dimensionality. Additionally, the ability to handle those complexities led to new insights. The trade-off is computational time. Even though we simplified the input predictor set (without loss of predictive power), fitting BRT models is computationally much more expensive than fitting logistic regression models. There is the question of whether or not current computational resources can accommodate BRT models in a rapid-update online system such as the Risk Assessment Tool for FHB; although this is an important question, it is beyond the scope of this article.

APPENDIX

In this section, we provide additional details on BRTs within the context of this article, following the general notations by Elith et al. (15) and Hastie et al. (21). We are interested in modeling the probability of a major FHB epidemic given a set of predictor covariates, $\boldsymbol{X}$:

$$\mu(\boldsymbol{X}) = \text{P}(\text{fhb} = 1|\boldsymbol{X}) \tag{A1}$$

where fhb = 1 represents a major epidemic and fhb = 0 a non-major epidemic. The actual modeling is done on the logit scale, as with standard logistic regression:

$$\text{logit}\ (\mu) = \log\ [\mu/(1 - \mu)] = f(\boldsymbol{X}) \tag{A2}$$

We now look at the form of $f(\boldsymbol{X})$. Consider a single binary classification tree, $T(x;\theta)$, which partitions the joint predictor values x into R_j (j = 1, 2,…,J) disjoint regions. Here, θ is a set of parameters determining the number of partition regions R_j and the rules for their creation. For example, we may have two partitions with rules R_1:RH ≥ 80% $\Rightarrow f(x) = y_1$; R_2:RH < 80% $\Rightarrow f(x) = y_2$ (i.e., a single decision stump with two terminal nodes, where y_i is a constant assigned to region R_j). The number of "levels" in a tree (tree depth) is controlled by the *tc* parameter, which the modeler is free to specify. With *tc* > 1, interactions are considered in the regression tree model. For example, the effect of a T-type predictor in the second tree level would depend on whether the higher level assignment was based on RH ≥ 80% or RH < 80%. A BRT model is a sum of the individual trees:

$$f_A(x) = \Sigma^A_{a=1} T(x;\theta_a) \tag{A3}$$

Thus, logit (μ) = $f_A(x)$.

Trees are added to the model in a forward stagewise manner. What this means is that if we have, for example, $f_2(x) = T(x;\theta_1) + T(x;\theta_2)$ then, in the next iteration of the algorithm, we will obtain $f_3(x) = T(x;\theta_1) + T(x;\theta_2) + T(x;\theta_3)$, where $T(x;\theta_1)$ and $T(x;\theta_2)$ remain unchanged from the previous iteration. Within the boosted logistic regression tree framework we used in this article, any new tree $T(x;\theta_{i+1})$ added to the model is one that best reduces the binomial deviance. When considering a new tree $T(x;\theta_{i+1})$ after the first, the algorithm does not work with the actual observations but focuses on the unexplained variation given by the residuals $y - p$ for model $f_i(x)$, where p is the predicted probability given by the already fitted function $f_i(x)$. The tree to be added to the model next, $T(x;\theta_{i+1})$, is the tree that best fits the residuals left after fitting $f_i(x) = \Sigma^i_{a=1} T(x;\theta_a)$, with the fitted values then being added to the current logit(p). Therefore, the final fitted BRT model is a linear combination of many trees, analogous to the individual terms in a linear regression model except that, in the BRT case, each term is a tree.

One can continue adding trees to the model until some stopping rule is reached. For example, the algorithm may be stopped after a fixed number of trees have been added. To determine the optimal *nt*, we used a stopping rule based on the estimated cross-validated residual deviance as a function of the number of trees in the model. The final model is more robust (better able to predict observations not used in the model development) if the contribution of each new tree added to the model is shrunk by a certain amount, so that the learning proceeds at a slower pace. That is, θ_{i+1} for each new tree is constrained in the estimation so that improvements in prediction accuracy are slow, which helps avoid overfitting of the model to the training data. The amount each newly added tree is allowed to contribute to the model is controlled by the learning rate (*lr*), which the modeler is free to specify, but should be low enough so that *nt* is large (at least 1,000 in our case).

ACKNOWLEDGMENTS

This is a cooperative project with the U.S. Wheat & Barley Scab Initiative (Agreements 59-0206-2-087 [Kansas State University] and 59-0206-9-071 [The Ohio State University]). Salaries and research support for L. V. Madden and P. A. Paul were provided by state and federal funds to the Ohio Agricultural Research and Development Center. Salary and research support for E. D. De Wolf were provided by federal funds to Kansas State University. We thank the many collaborators who have contributed data over the years, especially E. A. Adee and C. A. Bradley (University of Illinois); S. Ali, M. McMullen, and J. K. Ransom (North Dakota State University); S. Beanzinger and S. N. Wegulo (University of Nebraska-Lincoln); G. C. Bergstrom (Cornell University); W. W. Bockus and M. Nita (Kansas State University); D. Brown-Rytlewski (Michigan State University); P. D. Esker (University of Wisconsin-Madison); A. Grybauskas (University of Maryland); D. E. Hershman (The University of Kentucky); C. Hollingsworth and R. Dill-Mackey (University of Minnesota); P. Lipps (The Ohio State University); L. Osborne, K. Ruden, and J. Stein (South Dakota State University); L. E. Sweets (University of Missouri); G. Shaner and K. Wise (Purdue University); and B. G. Young (Southern Illinois University). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the United States Department of Agriculture.

LITERATURE CITED

1. Aertsen, W., Kint, V., De Vos, B., Deckers, J., Van Orshoven, J., and Muys, B. 2012. Predicting forest site productivity in temperate lowland

from forest floor, soil and litterfall characteristics using boosted regression trees. Plant Soil 354:157-172.

2. Agresti, A. 2002. Categorical Data Analysis, 2nd ed. John Wiley & Sons, Inc., Hoboken, NJ.
3. Andersen, A. L. 1948. The development of *Gibberella zeae* headblight of wheat. Phytopathology 38:599-611.
4. Austin, P. C., Lee, D. S., Steyerberg, E. W., and Tu, J. V. 2012. Regression trees for predicting mortality in patients with cardiovascular disease: what improvement is achieved by using ensemble-based methods? Biomed. J. 54:657-673.
5. Bühlmann, P., and Hothorn, T. 2007. Boosting algorithms: regularization, prediction and model fitting. Stat. Sci. 22:477-505.
6. Buston, P. M., and Elith, J. 2011. Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. J. Anim. Ecol. 80:528-538.
7. Chandelier, A., Nimal, C., Andre, F., Planchon, V., and Oger, R. 2011. Fusarium species and DON contamination associated with head blight in winter wheat over a 7-year period (2003-2009) in Belgium. Eur. J. Plant Pathol. 130:403-414.
8. Claeskens, G., and Hjort, N. L. 2008. Model Selection and Model Averaging. Cambridge University Press, Cambridge.
9. Cowger, C., and Arrellano, C. 2013. *Fusarium graminearum* infection and deoxynivalenol concentrations during development of wheat spikes. Phytopathology 103:460-471.
10. Cowger, C., Patton-Ozkurt, J., Brown-Guedira, G., and Perugini, L. 2009. Post-anthesis moisture increased Fusarium head blight and deoxynivalenol levels in North Carolina winter wheat. Phytopathology 99:320-327.
11. Crisci, C., Ghattas, B., and Perera, G. 2012. A review of supervised machine learning algorithms and their applications to ecological data. Ecol. Model. 240:113-122.
12. Del Ponte, E. M., Fernandes, J. M. C., and Pavan, W. 2005. A risk infection simulation model for Fusarium head blight of wheat. Fitopatol. Brasil. 30:634-642.
13. De Wolf, E. D., Madden, L. V., and Lipps, P. E. 2003. Risk assessment models for wheat Fusarium head blight epidemics based on within-season weather data. Phytopathology 93:428-435.
14. Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27-46.
15. Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77:802-813.
16. Ennis, M., Hinton, G., Naylor, D., Revow, M., and Tibshirani, R. 1998. A comparison of statistical learning methods on the GUSTO database. Stat. Med. 17:2501-2508.
17. Friedman, J., Hastie, T., and Tibshirani, R. 2000. Additive logistic regression: a statistical view of boosting. Ann. Stat. 28:337-374.
18. Friedman, J. H., and Meulman, J. J. 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. 22:1365-1381.
19. Hand, D. J. 2006. Classifier technology and the illusion of progress. Stat. Sci. 21:1-14.
20. Harrell, F. E., Jr. 2001. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer-Verlag, New York.
21. Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning, 2nd ed. Springer, New York.
22. Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14:382-401.
23. Hooker, D. C., Schaafsma, A. W., and Tamburic-Ilincic, L. 2002. Using weather variables pre- and post-heading to predict deoxynivalenol content in winter wheat. Plant Dis. 86:611-619.
24. Kint, V., Vansteenkiste, D., Aertsen, W., De Vos, B., Bequet, R., Van Acker, J., and Muys, B. 2012. Forest structure and soil fertility determine internal stem morphology of Pedunculate oak: a modelling approach using boosted regression trees. Eur. J. For. Res. 131:609-622.
25. Klem, K., Vanova, M., Hajslova, J., Lancova, K., and Sehnalova, M. 2007. A neural network model for prediction of deoxynivalenol content in wheat grain based on weather data and preceding crop. Plant Soil Environ. 53:421-429.
26. Kriss, A. B., Madden, L. V., Paul, P. A., and Xu, X. 2012. Heterogeneity of Fusarium head blight of wheat: multi-scale distributions and temporal variation in relation to environment. Plant Health Progress. Online publication. doi:10.1094/PHP-2012-0723-01-RS
27. Kriss, A. B., Paul, P. A., and Madden, L. V. 2010. Relationship between yearly fluctuations in Fusarium head blight intensity and environmental variables: a window-pane analysis. Phytopathology 100:784-797.
28. Krzanowski, W. J., and Hand, D. J. 2009. ROC Curves for Continuous Data. CRC Press, Boca Raton, FL.
29. Landschoot, S., Waegeman, W., Audenaert, K., Haesaert, G., and De Baets, B. 2013. Ordinal regression models for predicting deoxynivalenol in winter wheat. Plant Pathol. 62:1319-1329.
30. Landschoot, S., Waegeman, W., Audenaert, K., Van Damme, P., Vandepitte, J., De Baets, B., and Haesaert, G. 2013. A field-specific web tool for the prediction of Fusarium head blight and deoxynivalenol content in Belgium. Comput. Electron. Agric. 93:140-148.
31. Landschoot, S., Waegeman, W., Audenaert, K., Vandepitte, J., Baetens, J. M., De Baets, B., and Haesaert, G. 2012. An empirical analysis of explanatory variables affecting Fusarium head blight infection and deoxynivalenol content in wheat. J. Plant Pathol. 94:135-147.
32. Landschoot, S., Waegeman, W., Audenaert, K., Vandepitte, J., Haesaert, G., and De Baets, B. 2012. Toward a reliable evaluation of forecasting systems for plant diseases: a case study using Fusarium head blight of wheat. Plant Dis. 96:889-896.
33. Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., and Taylor, P. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Mar. Ecol. Prog. Ser. 321:267-281.
34. Maldonado-Ramirez, S. L., Schmale, D. G., Shields, E. J., and Bergstrom, G. C. 2005. The relative abundance of viable spores of *Gibberella zeae* in the planetary boundary layer suggests the role of long-distance transport in regional epidemics of Fusarium head blight. Agric. For. Meteorol. 132:20-27.
35. McMullen, M., Bergstrom, G., De Wolf, E., Dill-Macky, R., Hershman, D., Shaner, G., and Van Sanford, D. 2012. A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Dis. 96:1712-1728.
36. McMullen, M., Jones, R., and Gallenberg, D. 1997. Scab of wheat and barley: a re-emerging disease of devastating impact. Plant Dis. 81:1340-1348.
37. Molineros, J., De Wolf, E., Madden, L., Paul, P., and Lipps, P. 2005. Incorporation of host reaction and crop residue level into prediction models for Fusarium head blight. Pages 119-122. in: National Fusarium Head Blight Forum, Milwaukee, WI. S. M. Canty, T. Boring, J. Wardwell, L. Siler, and R. W. Ward, eds. Michigan State University, East Lansing.
38. Molineros, J. E. 2007. Understanding the challenges of Fusarium head blight forecasting. Dissertation, The Pennsylvania State University, University Park.
39. Moschini, R. C., and Fortugno, C. 1996. Predicting wheat head blight incidence using models based on meteorological factors in Pergamino, Argentina. Eur. J. Plant Pathol. 102:211-218.
40. Müller, D., Leitão, P. J., and Sikor, T. 2013. Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees. Agric. Syst. 117:66-77.
41. Musa, T., Hecker, A., Vogelgsang, S., and Forrer, H. R. 2007. Forecasting of Fusarium head blight and deoxynivalenol content in winter wheat with FusaProg. Bull. OEPP 37:283-289.
42. Nganje, W. E., Bangsund, D. A., Leistritz, F. L., Wilson, W. W., and Tiapo, N. M. 2004. Regional economic impacts of Fusarium head blight in wheat and barley. Rev. Agric. Econ. 26:332-347.
43. O'Donnell, K., Ward, T. J., Geiser, D. M., Kistler, H. C., and Aoki, T. 2004. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. Fungal Genet. Biol. 41:600-623.
44. Osborne, L. E., and Stein, J. M. 2007. Epidemiology of Fusarium head blight on small-grain cereals. Int. J. Food Microbiol. 119:103-108.
45. Prandini, A., Sigolo, S., Filippi, L., Battilani, P., and Piva, G. 2009. Review of predictive models for Fusarium head blight and related mycotoxin contamination in wheat. Food Chem. Toxicol. 47:927-931.
46. Rose, S. 2013. Mortality risk score prediction in an elderly population using machine learning. Am. J. Epidemiol. 177:443-452.
47. Rossi, V., Giosuè, S., Pattori, E., Spanna, F., and Del Vecchio, A. 2003. A model estimating the risk of *Fusarium* head blight on wheat. Bull. OEPP 33:421-425.
48. Schaafsma, A. W., and Hooker, D. C. 2007. Climatic models to predict occurrence of Fusarium toxins in wheat and maize. Int. J. Food Microbiol. 119:116-125.
49. Schaafsma, A. W., Tamburic-Ilinic, L., Miller, J. D., and Hooker, D. C. 2001. Agronomic considerations for reducing deoxynivalenol in wheat grain. Can. J. Plant Pathol. 23:279-285.
50. Shah, D. A. 2010. Forecasting white mold. Pages 63-65 in: Proc. 2010 Mid-Atlantic Fruit Veg. Conv. Hershey, PA. Pennsylvania Vegetable Growers Association, Richfield, PA.
51. Shah, D. A., Molineros, J. E., Paul, P. A., Willyerd, K. T., Madden, L. V., and De Wolf, E. D. 2013. Predicting Fusarium head blight epidemics with weather-driven pre- and post-anthesis logistic regression models. Phytopathology 103:906-919.

52. Stack, R. W., and McMullen, M. P. 1998. A visual scale to estimate severity of Fusarium head blight in wheat. NDSU Extension Service: Small Grains Publications. Online publication PP-1095. http://www.ag.ndsu.edu/pubs/plantsci/smgrains/pp1095.pdf
53. Starkey, D. E., Ward, T. J., Aoki, T., Gale, L. R., Kistler, H. C., Geiser, D. M., Suga, H., Toth, B., Varga, J., and O'Donnell, K. 2007. Global molecular surveillance reveals novel Fusarium head blight species and trichothecene toxin diversity. Fungal Genet. Biol. 44:1191-1204.
54. Van Der Fels-Klerx, H. J., Burgers, S. L. G. E., and Booij, C. J. H. 2010. Descriptive modelling to predict deoxynivalenol in winter wheat in the Netherlands. Food Addit. Contam. Part A Chem. 27:636-643.
55. Wegulo, S. N. 2012. Factors influencing deoxynivalenol accumulation in small grain cereals. Toxins 4:1157-1180.
56. Willyerd, K. T., Li, C., Madden, L. V., Bradley, C. A., Bergstrom, G. C., Sweets, L. E., McMullen, M., Ransom, J. K., Grybauskas, A., Osborne, L., Wegulo, S. N., Hershman, D. E., Wise, K., Bockus, W. W., Groth, D., Dill-Macky, R., Milus, E., Esker, P. D., Waxman, K. D., Adee, E. A., Ebelhar, S. E., Young, B. G., and Paul, P. A. 2012. Efficacy and stability of integrating fungicide and cultivar resistance to manage Fusarium head blight and deoxynivalenol in wheat. Plant Dis. 96:957-967.
57. Xu, X., Madden, L. V., and Edwards, S. G. 2014. Modelling the effects of environmental conditions on HT2 andT2 toxin accumulation in field oat grains. Phytopathology 104:57-66.
58. Xu, X., Madden, L. V., Edwards, S. G., Doohan, F. M., Moretti, A., Hornok, L., Nicholson, P., and Ritieni, A. 2013. Developing logistic models to relate the accumulation of DON associated with Fusarium head blight to climatic conditions in Europe. Eur. J. Plant Pathol. 137:689-706.
59. Zeller, K. A., Bowden, R. L., and Leslie, J. F. 2004. Population differentiation and recombination in wheat scab populations of *Gibberella zeae* from the United States. Mol. Ecol. 13:563-571.