Predicting Fusarium Head Blight Epidemics With Weather-Driven Pre- and Post-Anthesis Logistic Regression Models

D. A. Shah, J. E. Molineros, P. A. Paul, K. T. Willyerd, L. V. Madden, and E. D. De Wolf

First and sixth authors: Department of Plant Pathology, Kansas State University, Manhattan 66506; second author: Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater 74078; and third, fourth, and fifth authors: Department of Plant Pathology, The Ohio State University, Wooster 44691.
Current address of D. A. Shah: 4662 Vrooman Drive, Lewiston, NY 14092.

ABSTRACT

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.

Our objective was to identify weather-based variables in pre- and post-anthesis time windows for predicting major Fusarium head blight (FHB) epidemics (defined as FHB severity $\geq 10\%$) in the United States. A binary indicator of major epidemics for 527 unique observations (31% of which were major epidemics) was linked to 380 predictor variables summarizing temperature, relative humidity, and rainfall in 5-, 7-, 10-, 14-, or 15-day-long windows either pre- or post-anthesis. Logistic regression models were built with a training data set (70% of the 527 observations) using the leaps-and-bounds algorithm, coupled with bootstrap variable and model selection methods. Misclassification rates were estimated on the training

and remaining (test) data. The predictive performance of models with indicator variables for cultivar resistance, wheat type (spring or winter), and corn residue presence was improved by adding up to four weather-based predictors. Because weather variables were intercorrelated, no single model or subset of predictor variables was best based on accuracy, model fit, and complexity. Weather-based predictors in the 15 final empirical models selected were all derivatives of relative humidity or temperature, except for one rainfall-based predictor, suggesting that relative humidity was better at characterizing moisture effects on FHB than other variables. The average test misclassification rate of the final models was 19% lower than that of models currently used in a national FHB prediction system.

Additional keywords: additive logistic regression, data mining, multiple imputation.

In the United States, Fusarium head blight (FHB) of wheat (*Triticum aestivum* L. em. Thell) is caused primarily by *Fusarium graminearum* sensu stricto of the *F. graminearum* species complex (44). Major FHB epidemics have occurred somewhere in the United States in every decade since the disease was formally described by W. G. Smith in 1884 (60) although, in any given location, epidemics tend to occur sporadically. During the last two decades, U.S. wheat experienced large direct production losses because of FHB (35,36) and even larger indirect losses in other sectors of the economy (43), contributing to the characterization of FHB as a reemerging disease of importance (36,53). Increased corn (*Zea mays*) production in wheat-growing regions, concurrent with wider adoption of reduced tillage for soil conservation, were likely contributory factors to severe epidemics beginning in the latter part of the 19th century (36,60), as pathogen survival in corn residue is an acknowledged FHB risk factor (13,27).

FHB epidemiological research includes (i) basic documentation of epidemics and observed weather conditions at the time, a mainly descriptive effort, followed by quantification of optimal (usually controlled) conditions for various epidemiologically relevant processes (7,14,45,59,62,63); (ii) synthesis of basic epidemiological results into generalized, qualitative risk algorithms predicting FHB epidemics (47,59,63); and (iii) translation of the generalized risk algorithms into quantitative risk models (10,12,18,25,32,34,39,41,58), several of which were reviewed elsewhere

(55). Within the United States, models originally developed by De Wolf and colleagues (12) and subsequently refined (39,40) are the central risk algorithms behind the Fusarium Head Blight Risk Assessment Tool (http://www.wheatscab.psu.edu), a publicly funded service providing local-level, empirical FHB predictions across 31 states (35). The models are simple, in that they contain a few easily calculated predictors, an asset when deploying models at the regional scale. The Risk Assessment Tool's purpose is to provide growers enough advanced notification, if the forecasted risk of an FHB epidemic is high, to optionally apply a fungicide at anthesis (51,52), an effective mitigation strategy, especially when combined with cultivar resistance (2,67,69).

The models driving the Risk Assessment Tool were created with a relatively small set of observations collected prior to 2006 (12,40), which possibly limits inference to new data from a wider set of environmental permutations. In this article, we describe a larger data set of FHB observations, available because of active collaborator participation in U.S. Wheat & Barley Scab Initiative (USWBSI) projects (35). The objective of the present investigation was to identify, within a parsimonious logistic regression framework, weather-based predictors in pre- and post-anthesis windows of fixed length that were associated with FHB epidemics in this observational data set, and which could be used to improve the Risk Assessment Tool. Hence, this work expanded on previous modeling efforts (12,39,40) by considering a larger data set, an expanded set of potential predictors summarizing weather conditions in ways not considered in the previous work, and some improved analytical techniques that have been shown to be of value in the development and testing of risk models in other disciplines (23,24,30).

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

MATERIALS AND METHODS

The observational data matrix. FHB severity and incidence were rated in fungicide-untreated plots maintained by collaborators participating in USWBSI-funded projects. FHB severity (S) is defined as the mean percentage of a wheat spike's surface area with FHB symptoms, where the mean is estimated over *all* sampled spikes (n); if s_h is the severity of FHB on an individual spike h, then

$$S = \frac{1}{n}\sum_{h=1}^{n} s_h$$

S is equivalently called the FHB index (61). Furthermore, let

$$y_h = \begin{cases} 0 \text{ if } s_h = 0 \\ 1 \text{ if } s_h > 0 \end{cases} \tag{1}$$

Then, I is the mean incidence of spikes with FHB symptoms, and is given by

$$I = \frac{1}{n}\sum_{h=1}^{n} y_h$$

Note that I is on a 0-to-1 scale, and can be converted to a percent scale by multiplying I by 100. The related quantity X (diseased-spike severity) (38) is defined as the conditional severity (mean FHB severity over *only* those spikes with FHB symptoms; asymptomatic spikes are excluded), and is related to S and I by

$$S = I \cdot X \tag{2}$$

The FHB observations were collected over several years and subsets of the data were used for different purposes (40). Therefore, data verification and integrity checks were crucial. Some typical situations encountered are described. At the elementary level, some observations of S or I had been entered on a 0-to-1 scale, whereas others were recorded as percentages. S was derived from equation 2 if collaborators recorded I and X only. More difficult cases required verifying whether values entered under the generic label "index" were truly S, or X instead; graphical representations of the functional relationships between S and I (49) were exploited together with equation 2. The most difficult cases were the few in which X alone was reported. Briefly, one knows that S is an empirical function of I (49):

$$S = f(I) \tag{3}$$

From equation 2, we have $X = S/I$. Plugging equation 3 into equation 2 gives

$$X = [f(I)/I] = g(I) \tag{4}$$

X is now a function of I only and, because the relationship in equation 4 is monotonic, I can be estimated using the inverse function:

$$I = g^{-1}(X) \tag{5}$$

Hence, given X alone, an estimate $\hat{I}$ is obtained from equation 5 and, hence, $\hat{S}$ by equation 2. The above-described checks and verifications were applied to all FHB observations, but equations 3 to 5 were needed in ≈5% of cases only.

With S represented as a percentage, we considered observations with $S \geq 10\%$ as major epidemics and $<10\%$ as non-major epidemics. The choice of a 10% discriminatory point between major and non-major epidemics was consistent with past logistic regression models (12,40), and is also used in the current models of the Risk Assessment Tool. The response variable of interest was a binary categorization of S:

$$\text{fhb}_i = \begin{cases} 0 \text{ if } S_i < 10 \\ 1 \text{ if } S_i \geq 10 \end{cases} \tag{6}$$

Thus, $\text{fhb}_i = 0$ represented a non-major-epidemic observation, and $\text{fhb}_i = 1$ a major epidemic of FHB, for all observations $1,\ldots,m$.

The fhb_i were then linked to states in which plots were located and harvest year; and to categorical variables for wheat type (TYPE; binary variable, where 0 = spring wheat and 1 = winter wheat), cultivar resistance level (RESIST; ordinal variable, where 0 = very susceptible, 1 = susceptible, 2 = moderately susceptible, and 3 = moderately resistant), and presence or absence of corn residue in plots (CORN; binary variable, where 0 = no residue and 1 = some level of residue present). A date variable represented flowering date (fdate; the calendar date on which at least 50% of anthers were extruded). The completed observational data matrix consisted of 527 rows ($m = 527$). Observations were collected in 1982 to 2009 in 15 U.S. states (Fig. 1). Most of the data were collected in 2003 to 2005 (37%) and 2009 (30%). In all, 30% of the observations were made on spring wheat, with the remaining 70% coming from winter wheat. Spring wheat observations were mainly from Minnesota, North Dakota, and South Dakota (Fig. 1). Corn residue was present in 35% of the plots. The observations were collected over 150 site-years; multiple observations within a site-year were because of plot differences in cultivar resistance level, flowering date, or corn residue presence.

Hourly meteorological time series data. Hourly temperature (t; °C), relative humidity (rh; %) and rainfall (r; mm) were recorded for each site-year either at the plot site or within 10 km of sites. Most weather data files were supplied by collaborators. The remainder were downloaded from weather data archives maintained by Wolfram Research via Mathematica scripts we wrote (version 7.0; Wolfram Research, Champaign, IL). Some weather stations collected other variables, but $w = \{t,rh,r\}$ were common to all weather data sets; moreover, predictors associated with FHB epidemics in the United States (12,29,40,50) are related in one form or another to the elements of the vector w. All w_j, $j = 1,\ldots,150$, were checked for out-of-range (e.g., rh > 100) and missing data (5). Time series plots identified values that were within expected ranges but, nevertheless, were erroneous due to either instrument failure or transcriptional error (e.g., consecutive 12 h for which t = 15°C). Out-of-range or erroneous values were set to missing. There were 80 w vectors in which at least one value of t, rh, or r was missing.

Hourly data imputations. Multiple imputations are superior to case-wise deletion and missing-indicator analysis (64). Missing values in a w_j were multiply imputed using the R package Amelia II (version 1.2-14), with the following specified options: logistic transformation of rh (on a 0-to-1 scale); square root transformatioin of r; a third-order polynomial over time; and a 1% ridge prior (assists with numerical stability by shrinking the covariances toward zero without affecting means or variances). Imputations were restricted to within the date-times defining the beginning (x) and end (y) of a time series; that is, no attempt was made to impute values outside of (x,y); for example, for day $x - 1$ or $y + 1$. Missing values were successfully imputed in all cases, within plausible ranges, and adhered to the distributions of the actual observed data. Each missing value was imputed 10 times, and the final imputed value was taken as the mean of the 10 individual imputations.

Dewpoint depression and vapor pressure deficit. Dewpoint depression (dd; °C), and vapor pressure deficit (vpd; kPa) were computed from t and rh post-imputation of hourly data, thus expanding w to $\{t,rh,r,dd,vpd\}$. The dd was calculated by Bosen (3):

$$dd = t - [(112 + 0.9t)rh^{0.125} - 112 + 0.1t] \tag{7}$$

The vpd was estimated by Prenger and Ling (56):

$$vpd = \exp(-A/B) \times C \qquad (8)$$

where $A = 2.60131 \times 10^{-8}[2.0365 \times 10^{6} + t(t - 2151.62)] \times [170886 + t(354.342 + t)]$, $B = 491.67 + 1.8t$, and $C = (1.46476 \times 10^{-10} - 1.46476 \times 10^{-12}rh) \times (491.67 + 1.8t)^{6.456}$.

The weather-based predictor matrix. Each complete w (with the imputed values substituted for any missing values) was subset into fixed-length pre- and post-anthesis "windows" using 0800 h on fdate as the reference point. The pre-anthesis windows were periods prior to 50% of anthers extruded, and the post-anthesis windows were the periods beginning after 50% of the anthers have been extruded (and thus covered the flowering and early grain development stages). Let the symbols pre.x and post.x represent series of x days pre- and post-anthesis, respectively, anchored at fdate. Then, for example, $w_{\text{pre.7}}$ represented the hourly time series counting chronologically back 7 days from fdate, and $w_{\text{post.7}}$ represented the hourly time series counting chronologically forward starting from fdate. It follows that $w_{\text{pre.5}}$ is a subset of $w_{\text{pre.7}}$ and, likewise, $w_{\text{pre.7}}$ is a subset of $w_{\text{pre.10}}$. Fixed-length windows of 5, 7, 10, 14, and 15 days pre- and post-anthesis were created, the decision being influenced by previous research (6,7, 12,28,29,40,41,50,71). The windows were discrete, a subset of all possible windows, and were nested on either side of fdate. Pre-anthesis windows were independent of post-anthesis windows.

The hourly time series of t and rh within windows contained typical diurnal patterns. A 24-h day was defined as 0800 h to
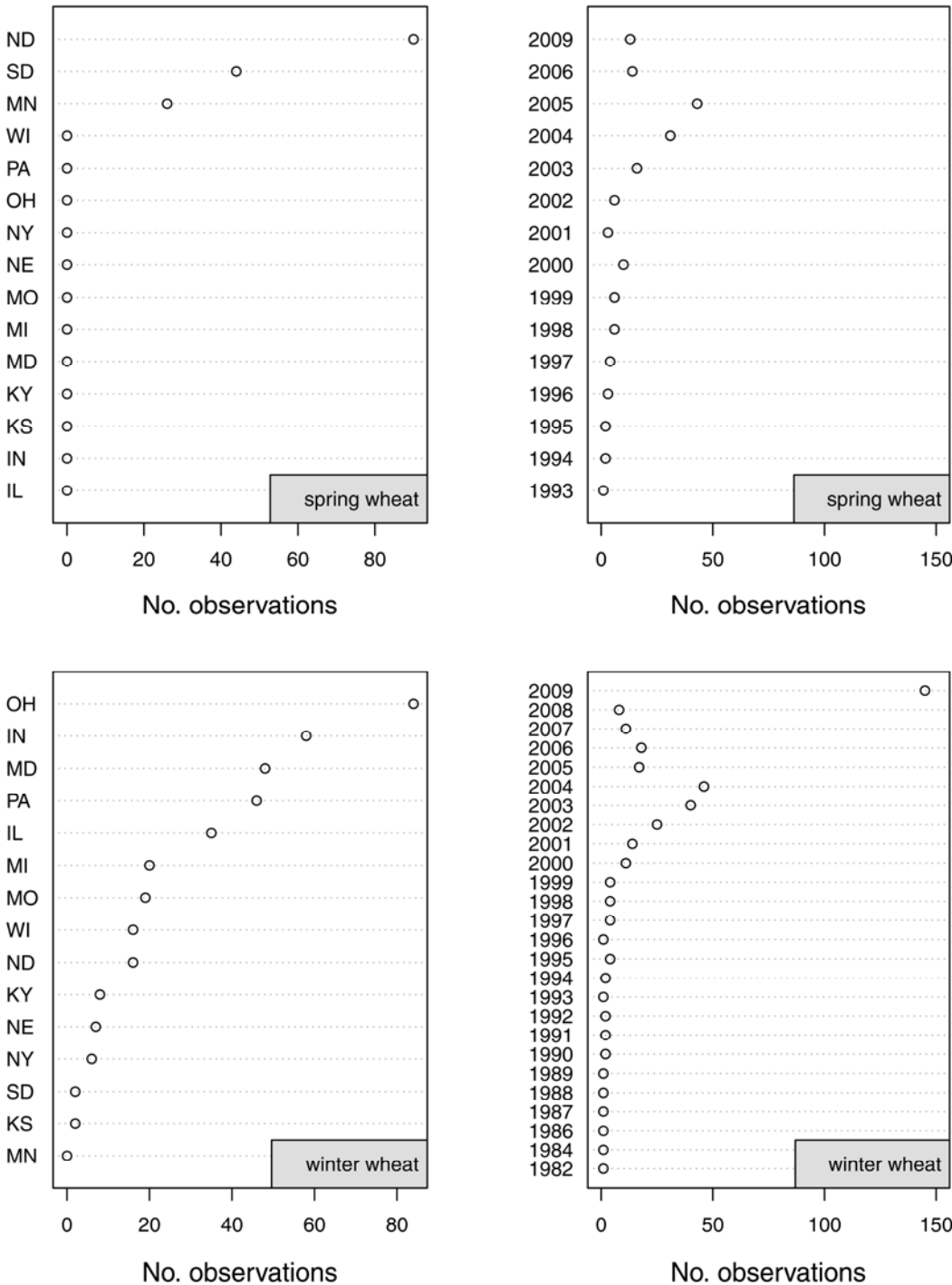

Fig. 1. Number of Fusarium head blight observations in the data matrix, described by wheat type, year, and U.S. state in which the observation was made.

0800 h so that overnight events would not be broken. Some predictors were defined on a 12-h daily period, from 2000 h to 0800 h, which also kept overnight events unbroken.

Previous research (7,12,28,29,40,41,50,71) guided the creation of 380 weather-based predictors from $w_{pre.5}$, $w_{pre.7}$, $w_{pre.10}$, $w_{pre.14}$, $w_{pre.15}$, $w_{post.5}$, $w_{post.7}$, $w_{post.10}$, $w_{post.14}$, and $w_{post.15}$ (Table 1). Some predictors were simple summaries, such as the number of hours per day in which rh $\geq$ 90%; others summarized two target conditions being met simultaneously, such as the number of hours in a window in which both $15 \leq t \leq 30$ and rh $\geq$ 90%.

Weather-based predictor imputations. Ideally, all w would have been a minimum of 30 days in length, having 15 days on either side of fdate. However, 20 of the pre-anthesis hourly weather series were shorter than 15 days; likewise, 15 of the post-anthesis hourly weather series were <15 days long. Shorter-than-required w created missing values in the weather-based predictor matrix (minimum = 7, maximum = 55, and mean = 24 missing values per predictor), with predictors created from $w_{pre.15}$ and $w_{post.15}$ affected most. The missing values warranted a second round of imputations (with Amelia II; this time for weather-based predictors not hourly data values). Imputation algorithms took advantage of the fact that any predictor was part of a correlated series (of five points) across nested subsets of w (Table 1). All imputed values were verified post-imputation to be within a variable's theoretical bounds. Five imputed versions of the weather-based predictor matrix were created.

TABLE 1. Weather-based predictor set considered as input for models of Fusarium head blight epidemics

Predictor class, acronym[a]	Description	Group[b]
Vapor pressure deficit (vpd; kPa)		
VPD.A.w.12H	Mean vpd per overnight period	1
VPD.L20.w.12H	Number of h vpd $\leq$ 0.20 kPa overnight	2
VPD.L45.w.12H	Number of h vpd $\leq$ 0.45 kPa overnight	3
Dewpoint depression (dd; °C)		
DD.A.w.12H	Mean dd per overnight period	4
DD.L1.w.12H	Number of h dd < 1°C overnight	5
Relative humidity (rh; %)		
RH.A.w.12H	Mean rh per overnight period	6
RH.A.w.24H	Mean rh per day	7
RH.G80.w.12H	Number of h rh $\geq$ 80% overnight	8
RH.G90.w.12H	Number of h rh $\geq$ 90% overnight	9
RH.MXRLG80.w.12H	Maximum[rl (number of h rh $\geq$ 80%)] overnight[c]	10
RH.MXRLG90.w.12H	Maximum[rl (number of h rh $\geq$ 90%)] overnight	11
RH.MXRLG80.w.24H	Maximum[rl (number of h rh $\geq$ 80%)]	12
RH.MXRLG90.w.24H	Maximum[rl (number of h rh $\geq$ 90%)]	13
RH.C6RLG80.w.24H	Count[rl (number of h rh $\geq$ 80%) $\geq$ 6]	14
RH.C6RLG90.w.24H	Count[rl (number of h rh $\geq$ 90%) $\geq$ 6]	15
Temperature (t; °C)		
T.A.w.24H	Mean t per day	16
T.9T30.w.24H	Number of h (9°C $\leq$ t $\leq$ 30°C)	17
T.15T30.w.24H	Number of h (15°C $\leq$ t $\leq$ 30°C)	18
T.L9.w.24H	Number of h (t < 9°C)	19
T.L15.w.24H	Number of h (t < 15°C)	20
T.G30.w.24H	Number of h (t > 30°C)	21
Rainfall(r; mm)		
R.PA.w.24H	0 = no rain, 1 = rain	22
R.S.w.24H	Total rainfall	23
R.D.w.24H	Number of days with rain	24
R.H.w.24H	Number of h with rain	25
R.AD.w.24H	Mean rainfall per day	26
R.PD.w.24H	Mean proportion of days with rain	27
R.AH.w.24H	Mean number of h with rain per day	28
Interaction (t and rh)		
TRH.9T30nRHG80.w.12H	Number of h (9°C $\leq$ t $\leq$ 30°C & rh $\geq$ 80%) overnight	29
TRH.9T30nRHG90.w.12H	Number of h (9°C $\leq$ t $\leq$ 30°C & rh $\geq$ 90%) overnight	30
TRH.15T30nRHG80.w.12H	Number of h (15°C $\leq$ t $\leq$ 30°C & rh $\geq$ 80%)overnight	31
TRH.15T30nRHG90.w.12H	Number of h (15°C $\leq$ t $\leq$ 30°C & rh $\geq$ 90%)overnight	32
Interaction (rh and r)		
RHR.RHG80nR.w.24H	Number of days (mean(rh) $\geq$ 80% & rain)	33
RHR.RHG90nR.w.24H	Number of days (mean(rh) $\geq$ 90% & rain)	34
Interaction (t and rh)[d]		
TRH.9T30nRHG80.w.24H	Number of h (9°C $\leq$ t $\leq$ 30°C & rh $\geq$ 80%)	35
TRH.9T30nRHG90.w.24H	Number of h (9°C $\leq$ t $\leq$ 30°C & rh $\geq$ 90%)	36
TRH.15T30nRHG80.w.24H	Number of h (15°C $\leq$ t $\leq$ 30°C & rh $\geq$ 80%)	37
TRH.15T30nRHG90.w.24H	Number of h (15°C $\leq$ t $\leq$ 30°C & rh $\geq$ 90%)	38

[a] Acronyms follow the naming convention a.b.w.c, where a indicates whether the variable summarizing vapor pressure deficit (VPD), dewpoint depression (DD), relative humidity (RH), temperature (T), rainfall (R), or pairwise interactions between temperature and relative humidity (TRH) or between relative humidity and rainfall (RHR); b indicates the type of summary measure (see the Description column), such as A for average (mean), L for less than, and so on; w is a placeholder for 1 of 10 vectors of hourly time series weather data ($w_{pre.5}$, $w_{pre.7}$, $w_{pre.10}$, $w_{pre.14}$, $w_{pre.15}$, $w_{post.5}$, $w_{post.7}$, $w_{post.10}$, $w_{post.14}$, and $w_{post.15}$), where the subscript in the preceding list indicates if the weather data span 5, 7, 10, 14, or 15 days pre- or post-anthesis; and c indicates a 24-h day (24H: 0800 to 0800 h) or a 12-h overnight period (12H: 2000 to 0800 h).

[b] 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). Therefore, there are 10 predictors per group and, hence, 380 total predictors.

[c] Run length (rl) predictors are explained via a worked example in the Appendix.

[d] These additional interactions between t and rh were added after the original predictor set had been conceptually formulated and parallel groups 29 to 32, except that groups 35 to 38 are estimated over a 24 h day period rather than a 12 h overnight period.

The FHB data matrix. Merging the FHB observation matrix with the weather-based predictor matrix created the FHB data matrix. The ordinal variable RESIST was expressed as binary indicator variables: res0 (= 1 if RESIST = 0, 0 otherwise), res1 (= 1 if RESIST = 1, 0 otherwise), res2 (= 1 if RESIST = 2, 0 otherwise), and res3 (= 1 if RESIST = 3, 0 otherwise). Additional binary predictors represented combinations of wheat type and corn residue presence: wc1 (= 1 if spring wheat, regardless of corn residue presence or absence, 0 otherwise), wc2 (= 1 if winter wheat and corn residue is absent, 0 otherwise), and wc3 (= 1 if winter wheat and corn residue is present, 0 otherwise). There were five versions of the FHB data matrix, each corresponding to a different imputed version of the weather-based predictor matrix.

In all, 30 weather-based predictors were removed (leaving 350) because they were identified as having only a few unique values, often at or near zero, with some of those occurring at low frequencies. These identified predictors can cause problems with analytical methods that use cross-validation or bootstrap sub-sampling (23), which were part of the analysis described later. Removed predictors were all in groups 10, 11, and 22 (Table 1).

Data exploration. We did two-way tabulations, mosaic plots of fhb against categorical predictors, and tests of the equality of the proportion of epidemics (42) in categorical groups (RESIST, TYPE, and CORN). Because the time series in each window were nested (i.e., $w_{pre.5} \subset w_{pre.7} \subset w_{pre.10} \subset w_{pre.14} \subset w_{pre.15}$, and, similarly, $w_{post.5} \subset w_{post.7} \subset w_{post.10} \subset w_{post.14} \subset w_{post.15}$, where $\subset$ indicates "is a subset of"), it was likely that some weather-based predictors were highly correlated; 15.7% of the 61,075 pairwise correlations between the 350 weather-based predictors were $\left| > 0.5 \right|$.

Training and test data sets. Each version of the FHB data matrix was partitioned into training and test data sets, with 70% of the observations going into the training data and the remaining 30% into the test data. The training data were used for model building. Test data were used to estimate the prediction error of finalized models.

Base models. We wanted to identify predictors of FHB epidemics for both wheat types in a holistic model framework, in contrast to earlier efforts which developed independent spring and winter wheat models (12,40). There was no a priori model structure for the categorical predictors RESIST, TYPE, and CORN, and five different logistic regression models incorporating these three predictors were explored. Let $\mu(X) = \Pr(\text{fhb} = 1 \mid X)$ be the probability of a major FHB epidemic given a linear combination of predictors X (dependent on the model), and logistic link function of the probability. Then the base models are given by the linear logistic equations:

base.0: $\text{logit}(\mu) = \beta_0 + \beta_1\text{res1} + \beta_2\text{res2} + \beta_3\text{res3}$ (9a)

base.1: $\text{logit}(\mu) = \beta_0 + \beta_1\text{res1} + \beta_2\text{res2} + \beta_3\text{res3} + \beta_4\text{wc2} + \beta_5\text{wc3}$ (9b)

base.2: $\text{logit}(\mu) = \beta_0 + \beta_1\text{res1} + \beta_2\text{res2} + \beta_3\text{res3} + \beta_4\text{TYPE}$ (9c)

base.3: $\text{logit}(\mu) = \beta_0 + \beta_1\text{res1} + \beta_2\text{res2} + \beta_3\text{res3} + \beta_4\text{CORN}$ (9d)

base.4: $\text{logit}(\mu) = \beta_0 + \beta_1\text{res1} + \beta_2\text{res2} + \beta_3\text{res3} + \beta_4\text{CORN} + \beta_5\text{TYPE} + \beta_6\text{CORN} \times \text{TYPE}$ (9e)

In equations 9a to 9e, the reference levels are res0 = 1, wc1 = 1, TYPE = 0 (i.e., spring wheat), and CORN = 0 (i.e., corn residue absent). Equations 9a to 9e were called "base" models because the goal was to add weather-based predictors to the logistic linear equation, given that those categorical predictors were already present, with the goal of further improving the predictive performance of the base models. Three models (equations 9a to 9c) were selected, after checking model coefficients, residual plots, model fit (Akaike's information criterion [AIC] and receiver operating characteristic [ROC] curves), and classification performance on the training data. These types of model checks are described in more detail in later sections.

Dimension reduction in modeling. Some form of variable selection was desirable to adhere to the objective of creating parsimonious models. Forward and backward stepwise regression methods violate several statistical principles (22,46,68) and are now highly discouraged for model development (23). Subset selection is a more statistically principled option. If all possible subsets of a full logistic model were fitted, information-theoretic methods could be used to select the "best" models. This approach becomes computationally impossible as the number of predictors grows. For example, a full logistic model with 350 predictors (no interactions) means fitting ($2^{350} - 1$) candidate subset models.

To address the collinearity and large predictor set issues, the FHB data matrix was partitioned so that logistic regression models were fitted separately in each pre- and post-anthesis window. The partitioning reduced the input weather-based predictor set from the full 350 to 35 within each time window, mitigating the predictor correlation problem. Fitting logistic regression models (base model plus 35 weather-based predictors) within each window led to 50 more predictors being dropped (5 per window, all in groups 20, 21, 26, 27, and 28) (Table 1), due to problems with singularity (perfect correlation between predictors in the data matrix). With 30 weather-based predictors per window, the number of possible model subsets is >1 billion but is within the operable range of the leaps-and-bounds subset-selection algorithm (17).

Some variable selection algorithms are susceptible to including predictors not correlated with the outcome (i.e., noise variables) (11). The tendency of the leaps-and-bounds algorithm to include noise variables was confirmed by applying the algorithm to bootstrapped samples of a simulated data set, in which three predictors were truly correlated with a binary outcome and seven other predictors were noise (data not shown). With these limitations in mind, the leaps-and-bounds algorithm was used in two rounds: to select a subset of weather variables and then a subset of possible models. For each base model and window combination, the algorithm was run on bootstrapped samples of the training data (200 bootstrapped samples from each of the five imputed training sets), requesting the best model with k weather-based predictors out of the 30 available in each window, added to the base model. Building on exploratory analyses (not shown), $k = 8$ for all pre-anthesis models and $k = 9$ for post-anthesis models. The number of times a weather-based predictor was selected, over all 1,000 runs of the algorithm, was tabulated and summarized graphically. The process with base.0 and the 10-day pre-anthesis window is illustrated in Figure 2, showing that all 30 available weather-based predictors appeared at least once in a selected model; however, there were six predictors (RH.C6RLG90.PRE10.24H, RH.A.PRE10.24H, TRH.9T30nRHG80.PRE10.24H, RH.G90. PRE10.12H, T.L9.PRE10.24H, and T.A.PRE10.24H; Table 1 provides explanations of the acronyms) which were selected in >50% of the algorithm runs. Repeating the process on the other combinations of base models and windows generated selection frequency profiles like Figure 2; by culling those predictors with a selection frequency <50%, the candidate weather-based predictor set in each base model–window combination was reduced from 30 to 2 to 8.

A second round of subset selection used the reduced set of weather-based predictors as the input candidate predictor set, and requested $k = 1,...,4$ weather-based predictors added to a base model. The leaps-and-bounds algorithm was run 1,000 times on bootstrapped samples of the training data as input (200 bootstrapped samples for each of the five imputed training data sets). Model selection frequencies were tabulated (as opposed to individual predictor selection frequencies in the previous stage). Figure 3 continues with the base.0, 10-day pre-anthesis window

combination as the illustrative example. In this particular case, models with $k = 3$ weather-based predictors were selected from all such models which could be built from a set of six weather-based predictors (Fig. 2). Twelve different models were returned; however, only one model (with weather-based predictors RH.A.PRE10.24H, RH.G90.PRE10.12H, and RH.C6RLG90.PRE10.24H added to base.0) would be considered further; the other models were selected too infrequently, indicating relative unimportance. The process was repeated for all 30 base model-window combinations, giving 37, 38, and 38 candidate models for base.0, base.1, and base.2, respectively, or 113 candidate models total.

Combining across multiple imputations. Coefficients for each model were adjusted to their respective means over the five versions returned from fitting to each of the imputed data sets (57). Coefficient-adjusted models were used to get the predicted probability of a FHB epidemic for an input observation row in each imputed training and test data set, and these individual predictions were averaged to obtain a final predicted probability adjusted for the multiply imputed data. This process was done before evaluating the candidate models for accuracy (measured in several different ways) and goodness of fit.

Evaluating candidate models. Each of the 113 candidate models (with anywhere from one to four weather-based predictors) was compared with the corresponding base model (no weather-based predictors), where such comparisons were made qualitatively and statistically. The goal was to cull any model that was no better at classifying FHB epidemics than the underlying base model. At the qualitative level, the ROC curve (21) of each

base model was compared visually with the respective curve of any candidate model built on that base (not shown). Descriptive measures included (i) area under the ROC curve (AUC) (21); (ii) the H (an alternative to AUC) measure (20); (iii) the Youden Index (YI), defined as the maximum difference between the true positive and false positive predictive rates (30), which is a global measure of classification accuracy and is used widely as an optimal cut-point (16,26); (iv) the Kolmogorov-Smirnov (K-S) test statistic, which is equivalent to the estimate of the maximum vertical distance between the ROC curve and the chance diagonal line (30); (v) net reclassification index (NRI) (54); and (vi) integrated discrimination improvement (IDI) (54). Each of these global measures individually captures certain features of the model (AUC versus H, for example) (20) but, together, give a more complete comparison of models, albeit at the expense of some overlap.

Statistical tests further culled models which either gave no improved classification performance over the underlying base model or gave no improved classification performance when compared with a simpler version with fewer weather-based predictors. Pairwise tests accounted for correlated ROC curves (as they were fit to the same data) and included comparisons of (i) AUCs (8), (ii) entire ROC curves (65), (iii) a paired-sample test of K-S statistics (31), and (iv) P values for the NRI and IDI comparing model classification performance. Model selection relied on a combined assessment of the P value-based tests for entire ROC curves, K-S statistics, NRI, and IDI. We found the DeLong test (8) to be overly sensitive to minor differences between ROC curves.
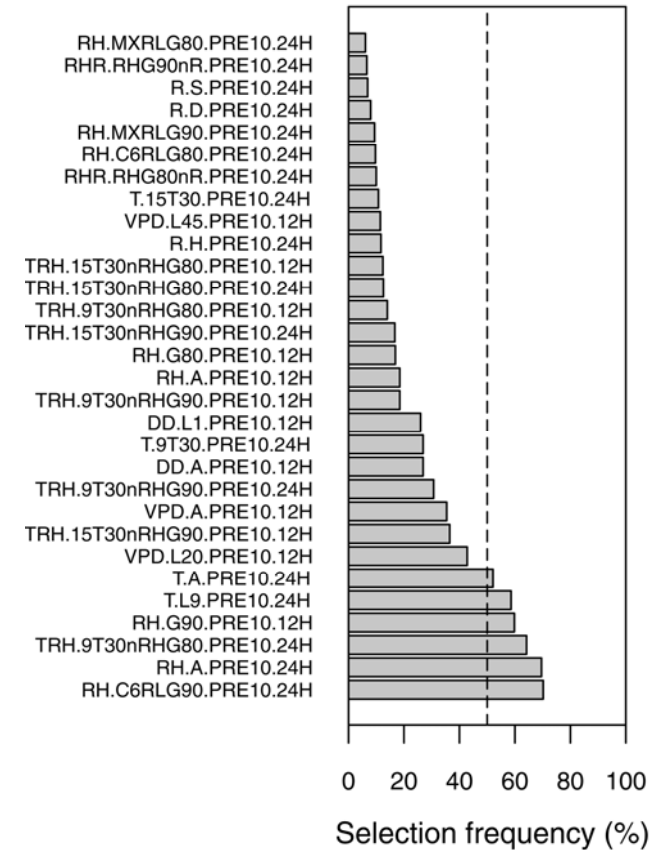

Fig. 2. Weather-based predictor selection frequencies in the 10-day pre-anthesis window after subset selection on 1,000 bootstrapped samples of the training data. The subset selection algorithm was asked to return models with eight weather-based predictors added to base.0 (a model with binary indicators for resistance level). All 30 available weather-based predictors were selected in at least one model returned by the algorithm. The vertical dashed line indicates a selection frequency of 50%, below which a predictor was culled from further consideration. Table 1 provides full descriptions of predictor acronyms.
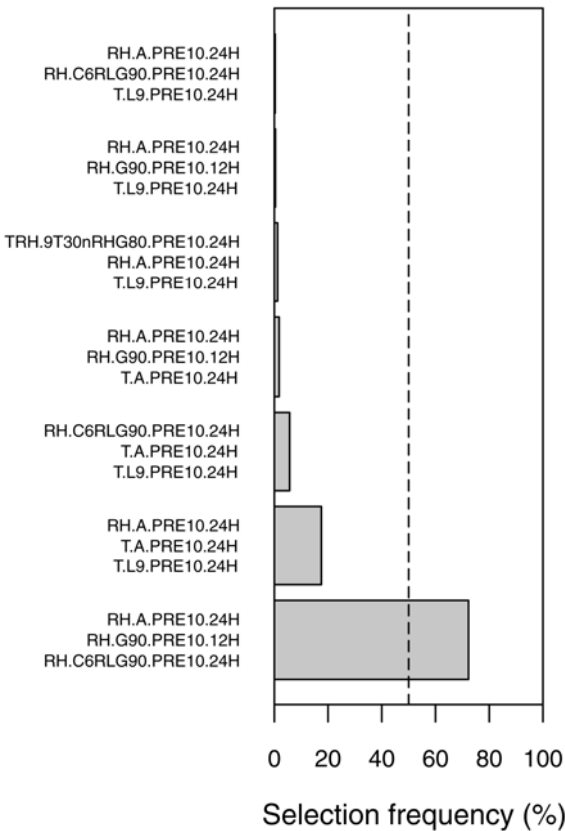

Fig. 3. Model selection frequencies in the 10-day pre-anthesis window after subset selection on 1,000 bootstrapped samples of the training data. The algorithm requested models with three weather-based predictors added to base.0, out of six such predictors to choose from (Fig. 2). Twelve different models were selected over all algorithm runs; the top seven are shown. The vertical dashed line indicates a selection frequency of 50%. Each horizontal bar refers to a different model, and the variables listed on the vertical axis are the weather-based predictors that comprise the models. Table 1 provides variable naming conventions.

Logistic regression diagnostics. The evaluation procedures described above narrowed the field from 113 to 15 candidate models. We then turned our attention to diagnostics of whether the 15 remaining models adhered to the assumptions of logistic regression, and to identify potential problems (such as outliers). Three types of diagnostics were done: (i) a check of model coefficients, (ii) residual plots, and (iii) tests of the linearity-of-the-logit assumption. Model coefficients were examined for sign and magnitude, and simple odds ratios calculated as a check that model parameters were consistent with established biological understanding of FHB epidemics (e.g., reduced odds of FHB epidemics with increasing levels of genetic resistance). Residual checking in binary-response logistic regression models is not as straightforward as with Gaussian-distributed data, and residual plots must be interpreted carefully (70). For binary data residual checking, one common recommendation is to simulate data from a given model, then use the simulated data to examine how residuals should behave when the distributional assumption is correct and the data are really independent (70). Following this suggestion, each of the 15 models was assumed to be a "correct" model (with a binomial distribution for fhb), data were simulated from the model (using the estimated parameters as the true ones), and the logistic regression model was refitted to the simulated data. The resulting residual plots then showed what patterns could be expected with binary data in a data set of this size and in which model assumptions were met. These simulation results (not shown) helped in interpreting the actual residuals for each of the 15 fitted models.

Residual plots examined were (i) residuals versus fitted probabilities, (ii) Q-Q plots, (iii) scale-location plots, (iv) Cook's distance, (v) residuals versus leverage, (vi) Cook's distance versus leverage, (vii) Anscombe residuals, (viii) Studentized residuals versus hat values, and (ix) DFFITS versus predicted probabilities (24). Residual plots i to vi are the default in base R. Anscombe residuals were estimated with the surveillance package (ver. 1.2-1). R code was written to produce residual plots viii and ix. Residual plots were graphed (lattice package ver. 0.19-33) with separate labels for each wheat type as a check for potential pattern discrepancies between spring and winter wheat.

The final check was on the linearity-of-the-logit assumption. Three different linearity-of-the-logit tests (one P value-based and two graphically based) were used for the continuous, weather-based predictors, because each test has its own strengths and nuances (no omnibus procedure is foolproof) (24). The first was the Tukey-Pregnibon test. If the P value of the test is significant, the linearity-of-the-logit assumption for that predictor has been violated. The first graphically based test was with partial residual plots (24). In this test, the continuous predictor is evaluated in terms of its linearity in the logit when other predictors are present in the model (the Tukey-Pregnibon test is a univariate test). A smoothed line that is (approximately) linear with slope of 1 or –1 indicates adherence to the linearity-of-the-logit assumption. The second graphically based test was built on a generalized additive model (GAM) (see Appendix for more details) version of the logistic model and smoothing spline fit of the GAM residuals (24,70). Again, a straight smoothed line of slope 1 or –1 indicated adherence to the linearity-of-the-logit assumption. For both graphically based tests, data points were distinguished by the factors TYPE, CORN, and RESIST as checks into potential violations of the linearity-of-the-logit assumption within subsets of the training data. None of the residual plots indicated any potential problems or outliers of concern with either wheat type (not shown). Linearity-of-the-logit diagnostics did indicate potential curvature with several (mainly rh–type) weather-based predictors. The nonlinearity-in-the-logit issues were solved by using additive logistic regression models (Appendix), with smooth functions for those weather-based predictors so indicated.

Model performance. The AIC, sensitivity (proportion of major epidemics classified correctly), specificity (proportion of non-major epidemics classified correctly), and overall misclassification rate (proportion of fhb observations classified incorrectly) of the logistic and additive logistic (Appendix) versions of each model were compared (not shown), resulting in 7 of the 15 logistic regression models being replaced by their additive logistic regression counterparts (Table 2). Note that, in models with more than one weather-based predictor, not all such terms were necessarily represented by smooth functions (Table 2).

For a given model, any observation with predicted probability >YI for that model was classified as a major epidemic, whereas any predicted probability ≤YI was classified as a non-major epidemic. That is, YI was used as the optimal cut-point for classifying epidemics from model-predicted probabilities. Each of the final 15 models (Table 2) was evaluated for sensitivity, specificity, and the overall misclassification rate on the training set. Those same evaluations were done on the test data not used in model development, thereby enabling estimation of the expected error when applied to new data (23).

We made two-way tables summarizing the number of test data observations correctly or incorrectly classified for all pairwise combinations of the 15 models. The tables were used to compare the discordant errors (model A makes an error while model B is correct, and vice versa) by the McNemar test (1).

Comparison with currently deployed models. Two separate models (one for spring wheat, the other for winter wheat), built with data collected before 2006, are used in forecasting the risk of FHB epidemics in the United States (http://www.wheatscab.psu.edu). The models use weather-based predictors derived from hourly t and rh within the 7-day pre-anthesis window, where each day is defined on a 24-h midnight-to-midnight cycle (note the difference from the 0800-to-0800 h 24-h cycle on which the weather-based predictors in the current article were defined). The model equations (not previously reported in peer-reviewed literature) are

$$\text{logit}\,(\mu) = -16.9369 - 0.71704\,\text{RESIST}_C + 0.23839\,\text{H1} \qquad (10a)$$

for spring wheat, and

$$\text{logit}\,(\mu) = -6.3906 + 0.0746\,\text{TH2} \qquad (10b)$$

for winter wheat, where RESIST_C is a continuous version of RESIST (the ordinal levels 0, 1, 2, and 3 are treated as real numbers within the 0-to-3 range, not as categories as in our current models), H1 is the mean hourly rh, and TH2 is the number of hours during which the following two conditions are met simultaneously within a given hour: t is 9 to 30°C and rh ≥ 90%.

It was unrealistic to expect a low test error with the current models, which were developed on a far more limited subset of the now-available data. Therefore, the coefficients of equations 10a and 10b were updated by fitting to the appropriate (spring or winter wheat) training data subset accordingly. The updated spring wheat model was

$$\text{logit}\,(\mu) = -11.008 - 0.9578\,\text{RESIST}_C + 0.1516\,\text{H1} \qquad (10c)$$

with standard errors of 3.4717, 0.3008, and 0.0469 for the intercept, RESIST_C and H1, respectively.

The updated winter wheat model was

$$\text{logit}\,(\mu) = -1.7954 + 0.0245\,\text{TH2} \qquad (10d)$$

with standard errors of 0.2798 and 0.0055 for the intercept and TH2, respectively. The YI (used as optimal cut-points) were 0.37 for the spring wheat model (equation 10c) and 0.23 for the winter wheat model (equation 10d). Performances of both updated models were evaluated on the training and test sets.

Software. SAS 9.2 (SAS Institute, Cary, NC) was used for importing weather time series data from Excel 2002 spreadsheets (Microsoft Corporation, Redmond, WA), checking for missing and out-of-range values (5), and visualizing time series plots. The weather data were then exported from SAS to R (version 2.13.1; R Foundation for Statistical Computing, Vienna, Austria), where several add-on packages were used in the analysis. These included Hmisc (ver. 3.8-3) for reading SAS xpt data files into R; chron (ver. 2.3-42) for handling date-time data; reshape (ver. 0.8.4) for flexibly manipulating data between long and wide formats, and for some data processing; Amelia II (ver. 1.2-14) for multiple imputations of missing values in weather hourly time series, and for multiple imputations of missing values in the weather-based predictors; epicalc (ver. 2.11.1.0) for testing the equality of proportions of epidemics in categorical variable groups; the nearZeroVar function in the caret package (ver. 5.13-20) for identifying zero and near-zero predictors; vcd (ver. 1.2-12) for visualizing mosaic plots; subselect (ver. 0.10-1) for implementing the leaps-and-bounds algorithm in subset selection; mitools (ver. 2.0.1) for combining parameter estimates from models fit on multiply imputed data sets; pROC (ver. 1.4.4) for plotting ROC curves and for implementing the DeLong and Venkatraman tests; PredictABEL (ver. 1.1) for calculating the NRI and IDI statistics and the *P* values for these tests; PresenceAbsence (ver. 1.1.5) for calculating the YI and classification matrices; and mgcv (1.7-6) for fitting GAMs.

TABLE 2. Final 15 weather-based logistic and additive logistic regression models of Fusarium head blight epidemics

ID[a]	Weather-based predictors[b]	Training data measures[c]							Test data measures[d]			
		AIC	AUC	YI	Kappa	fp	fn	mc	Kappa	fp	fn	mc
Base.0[e]		445	0.63	0.25	0.13	0.67	0.16	0.51	0.03	0.68	0.27	0.55
1	**RH.A.PRE7.24H**	394	0.77	0.38	0.39	0.26	0.32	0.28	0.24	0.28	0.47	0.34
2	RH.A.PRE10.24H	…	…	…	…	…	…	…	…	…	…	…
	RH.G90.PRE10.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.C6RLG90.PRE10.24H	389	0.78	0.29	0.37	0.34	0.23	0.31	0.28	0.40	0.27	0.36
3	**RH.A.PRE14.24H**	383	0.80	0.31	0.40	0.33	0.21	0.30	0.26	0.42	0.27	0.37
4	TRH.15T30nRHG80.POST7.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.A.POST7.12H	…	…	…	…	…	…	…	…	…	…	…
	T.A.POST7.24H	369	0.82	0.34	0.48	0.24	0.24	0.24	0.29	0.30	0.39	0.33
5	TRH.15T30nRHG80.POST10.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.A.POST10.12H	…	…	…	…	…	…	…	…	…	…	…
	T.A.POST10.24H	365	0.82	0.26	0.46	0.32	0.13	0.27	0.29	0.36	0.31	0.35
Base.1		425	0.70	0.25	0.24	0.46	0.26	0.40	0.13	0.56	0.27	0.47
6	RHR.RHG90nR.PRE7.24H	…	…	…	…	…	…	…	…	…	…	…
	RH.A.PRE7.24H	…	…	…	…	…	…	…	…	…	…	…
	T.A.PRE7.24H	…	…	…	…	…	…	…	…	…	…	…
	T.L9.PRE7.24H	366	0.83	0.27	0.47	0.33	0.12	0.26	0.34	0.39	0.22	0.34
7	TRH.15T30nRHG80.POST5.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.A.POST5.12H	…	…	…	…	…	…	…	…	…	…	…
	T.A.POST5.24H	369	0.82	0.40	0.50	0.17	0.33	0.22	0.27	0.27	0.45	0.33
8	TRH.15T30nRHG80.POST7.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.A.POST7.12H	…	…	…	…	…	…	…	…	…	…	…
	T.A.POST7.24H	351	0.84	0.33	0.49	0.24	0.23	0.24	0.24	0.33	0.41	0.35
9	TRH.15T30nRHG80.POST10.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.A.POST10.12H	…	…	…	…	…	…	…	…	…	…	…
	T.A.POST10.24H	350	0.84	0.31	0.51	0.25	0.19	0.23	0.31	0.31	0.35	0.32
Base.2		447	0.62	0.25	0.13	0.67	0.16	0.51	0.03	0.68	0.27	0.55
10	**RH.A.PRE10.24H**	385	0.78	0.33	0.36	0.38	0.20	0.32	0.30	0.41	0.24	0.35
11	**RH.A.PRE14.24H**	385	0.80	0.32	0.41	0.32	0.21	0.29	0.27	0.41	0.27	0.37
12	TRH.15T30nRHG80.PRE15.24H	…	…	…	…	…	…	…	…	…	…	…
	T.A.PRE15.24H	…	…	…	…	…	…	…	…	…	…	…
	T.L9.PRE15.24H	402	0.77	0.34	0.41	0.25	0.31	0.27	0.36	0.28	0.33	0.30
13	TRH.15T30nRHG80.POST5.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.A.POST5.12H	…	…	…	…	…	…	…	…	…	…	…
	T.A.POST5.24H	370	0.85	0.25	0.49	0.32	0.10	0.25	0.29	0.39	0.27	0.35
14	TRH.15T30nRHG80.POST7.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.A.POST7.12H	…	…	…	…	…	…	…	…	…	…	…
	T.A.POST7.24H	367	0.82	0.29	0.48	0.27	0.19	0.24	0.23	0.37	0.37	0.37
15	**TRH.15T30nRHG80.POST10.12H**	…	…	…	…	…	…	…	…	…	…	…
	RH.A.POST10.12H	…	…	…	…	…	…	…	…	…	…	…
	RH.C6RLG90.POST10.24H	…	…	…	…	…	…	…	…	…	…	…
	T.A.POST10.24H	333	0.88	0.28	0.50	0.28	0.15	0.24	0.45	0.26	0.25	0.26

[a] A simple way of identifying a model.

[b] Predictor coefficients represented by a smooth function are shown in bold. Any model with at least one smooth function is an additive logistic regression model (i.e., IDs 1, 3, 6, 10, 11, 13, and 15). The others are logistic regression models.

[c] AIC = Akaike's information criterion; AUC = area under the receiver operating characteristic curve; YI = the Youden index, defined as the maximum difference between the true positive and false positive rates (YI is used here as an optimal cut-point for classifying observations as major or non-major Fusarium head blight epidemics); Kappa = Cohen's kappa statistic, a measure of agreement between model-predicted and actual classification of major and non-major epidemics, taking into account chance agreement; fp = false positive rate, which is the proportion of non-major epidemics wrongly classified as major epidemics (= 1 – specificity); fn = false negative rate, which is the proportion of major epidemics wrongly classified as non-major epidemics (= 1 – sensitivity); mc = the overall misclassification rate, which is 0 if all observations are perfectly classified and 1 if all observations were wrongly classified.

[d] Test metrics were based on YI as the optimal cut-point.

[e] Base models to which weather-based predictors were added. Categorical predictors included in base models were: base 0 (res1 res2 res3); base 1 (res1 res2 res3 wc2 wc3); base 2 (res1 res2 res3 TYPE). res1 = 1 if resistance level = susceptible, 0 otherwise; res2 = 1 if resistance level = moderately susceptible, 0 otherwise; res3 = 1 if resistance level = moderately resistant, 0 otherwise; wc2 = 1 if winter wheat *and* corn residue is absent, 0 otherwise; wc3 = 1 if winter wheat *and* corn residue is present, 0 otherwise; TYPE = 1 if winter wheat, 0 if spring wheat.

RESULTS

FHB severity (S) was 0 to 86% (Fig. 4), with a mean of 10.4%. Of the 527 observations, 15.4% were S = 0. Some 50% of S values were <3.2% and 80% were <19.3% (Fig. 4, inset). After the binary categorization of S, major epidemics were represented by 31% of the observations, with that proportion being approximately equal in spring wheat (30%) and winter wheat (32%) (test for equality of proportions P value = 0.792). Cultivar resistance was associated with decreased risk of major FHB epidemics, as expected (Fig. 5A and B). As resistance level increased from category 0 to 3, the proportion of observations classified as epidemics fell from 0.45 to 0.30 to 0.23 to 0.19, and the differences in proportions were statistically significant ($P = 0.0003$). Across all observations, the presence of corn residue was associated with an increased risk of epidemics; 26% of observations with corn = 0 were classified as epidemics, whereas 40% of observations with corn = 1 were epidemics ($P = 0.001$). However, when the data were subset by wheat type, presence of corn residue was associated with an increased epidemic risk in winter wheat only (Fig. 5C). By comparison, there were not many observations in the category of spring wheat with corn residue (32 such observations); therefore, there was not enough supporting evidence indicating an increased risk of major epidemics in spring wheat due to localized (within-plot) corn residue.

Base model performance. Each of the three base models considered was capable of predicting FHB epidemics to varying degrees of accuracy (Table 2). Resistance alone was, in fact, a significant predictor as expected, because the resistance-level variable is based on previous observations of disease severity. To see this, consider the base.0 model, which contained three binary indicator variables (res1, res2, and res3) representing susceptible, moderately susceptible, and moderately resistant cultivars, respectively. The respective odds ratios (with RESIST = 0 [very susceptible] as the baseline) were 0.53, 0.27, and 0.22. Interpreting the odds ratios meant that the odds of a major FHB epidemic ($\geq 10\%$ severity) on susceptible, moderately susceptible, and moderately resistant cultivars were 47, 73, and 78% lower, respectively, than the odds of a major FHB epidemic on a very susceptible cultivar (RESIST = 0).

Final model set. The two-stage predictor and model selection protocols applied to each base model–window combination, followed by culling based on performance diagnostics, resulted in the selection of 15 models. This model set contained both logistic and additive logistic regression models (Table 2). There were five models built on base.0, four on base.1, and six on base.2. Seven were pre-anthesis and the remaining eight were post-anthesis models. The pre-anthesis models were from the 7-, 10-, 14-, and 15-day windows. The post-anthesis models were, by contrast, concentrated in the shorter windows (5-, 7-, and 10-day). All 15 models were relatively close in measures such as AUC and YI. The AUCs did not exceed 0.9 (Table 2).

Model performance. Overall misclassification rates averaged 0.26 on the training data and 0.34 on the test data. On the training data, post-anthesis models seemed, on average, to be the better classifiers than pre-anthesis models (mean Cohen's kappa of 0.49 versus 0.40 for pre-anthesis). However, that advantage was not apparent in the test data metrics where, for example, the average misclassification rate for post-anthesis models (0.33) was close to that for the pre-anthesis models (0.35). A trade-off between sensitivity and specificity was apparent, in that higher sensitivities were achieved at the expense of specificity, and vice versa. This phenomenon is indicated by the sloped lines in Figure 6, which are (unweighted) linear regressions of specificity on sensitivity over all 15 models. The selected models all offered a better sensitivity-specificity balance than the updated versions of the currently deployed models (Fig. 6), as well as incorporating both spring and winter wheat types into the same models. According to the McNemar tests (not shown), model 12 was the best pre-anthesis model in making fewer misclassification errors than other pre-anthesis models, and the same could be said for model 15 among the post-anthesis models. There was no indication of a trend (either increasing or decreasing) in misclassification rate with window length. However, lower misclassification rates were associated with higher numbers of weather-based predictors in a model.
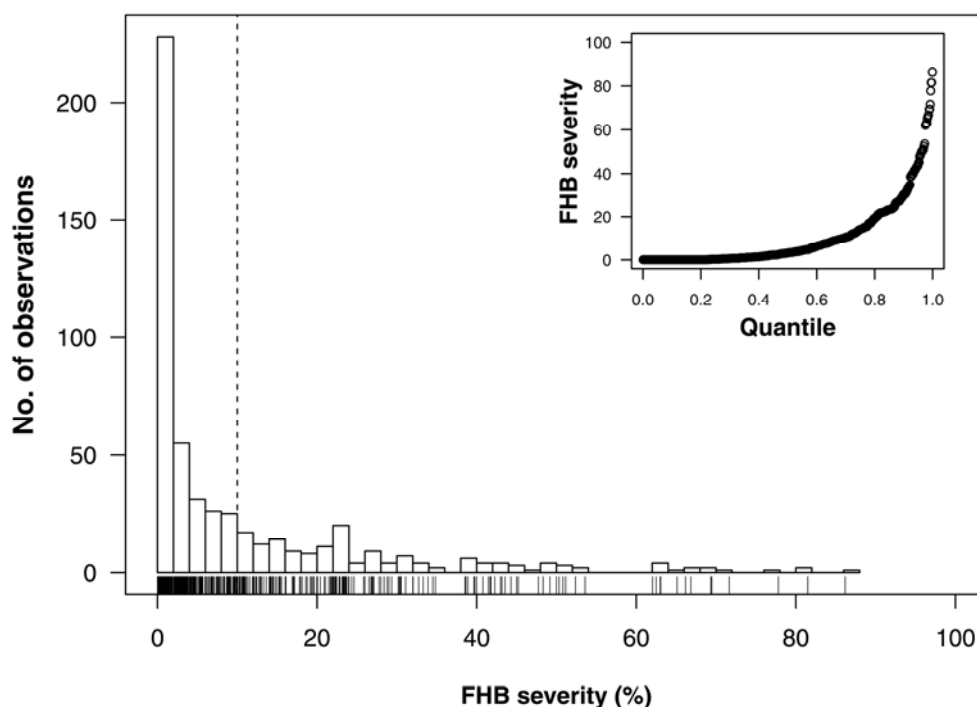

Fig. 4. Histogram of the 527 Fusarium head blight severity (S) values in the data matrix. The vertical dashed line represents the severity cutoff of 10% for dichotomization of observations into non-major epidemics (S < 10) or major epidemics (S ≥ 10). Inset: a sample quantile plot showing the proportion of severity values (x-axis) that are less than an observed severity value (y-axis).

Selected weather-based predictors. Although we started with 380 weather-based predictors (Table 1), only 21 were retained in the 15 final selected models, after eliminating predictors because of collinearity or because they did not pass the first or second stage of predictor or model selection (Table 2). Moreover, these 21 fell into 9 predictor groups, where a group consisted of a set of predictors varying only in the length of the defining window (Appendix, Table 3); the original candidate predictor set contained 38 predictor groups. All 21 selected weather-based predictors were some variant of t, rh, or a combination of the two. The one exception was RHR.RHG90nR.PRE7.24H, which was a combination of rh and r (Appendix, Table 3). None of the dd or vpd predictors appeared in the final 15 models.

DISCUSSION

Different research groups have translated the generalized FHB risk algorithms into quantitative models in various ways (55), thereby producing numerous representations of three basic variables (t, rh, and r). Some differences in predictor formulations reflect the limitations of data available to modelers. Once a specific objective had been defined (e.g., predicting mycotoxin levels versus disease intensity), the properties of the available disease and weather data sets defined the set of possible candidate weather-based predictors, while also being consistent with the generalized risk algorithm and underlying basic epidemiology. For example, Moschini and Fortugno (41) were limited to heading date as the available phenological reference point, and basic weather variables measured were available on a daily basis (as opposed to hourly). Kriss et al. (29) had to rely on a generalized calendar date for crop maturity as the phenological reference point. The weather data available for their empirical efforts were a mixture of daily and hourly measurements. Del Ponte et al. (10) built a mechanistic simulation model, with inputs being restricted to available daily weather summaries. Therefore, when each model is examined within the context of its development, it is not surprising to see so many different representations of weather-based predictors. Basically, all FHB predictive models attempt to capture how two sets of conditions (moisture and t) influence processes making up the FHB disease cycle. They just differ in how these conditions are represented numerically, for the reasons discussed above. The analyses described here benefited from (i) a larger than typical data set of 527 observations; (ii) flowering date estimates serving as the phenological reference point for each observation; (iii) hourly t, rh, and r data; and (iv) analytical insights from contemporary statistical sciences for the modeling of binary data that were either not available or widely appreciated at the time that previous empirical models (12) were being developed. We started with more ways of representing moisture and t than considered before (12), within short windows that either started or ended at anthesis. Subset selection algorithms then identified weather-based predictors, enhancing the predictive accuracy provided by cultivar resistance and corn residue indicators. Several possible models that included up to four weather variables were selected, and no single model could be considered best based on multiple diagnostic criteria.

All 15 models contained at least one predictor that was based on rh. Of the 15 models, 5 had only rh-type predictors. The other 10 models each had predictors based on t, rh, and interactions between t and rh. By contrast, none of the weather-based predictors in the final 15 models were dd- or vpd-based. An r-based predictor (RHR.RHG90nR.PRE7.24H) appeared in only one model; furthermore, this predictor was an interaction-type variable with rh. Our results suggest that (i) rh-type predictors are better than other moisture-based summaries, considered here as representing the association between major FHB epidemics and underlying wetness requirements, and (ii) moisture is more of a limiting factor than t for major FHB epidemics in the United
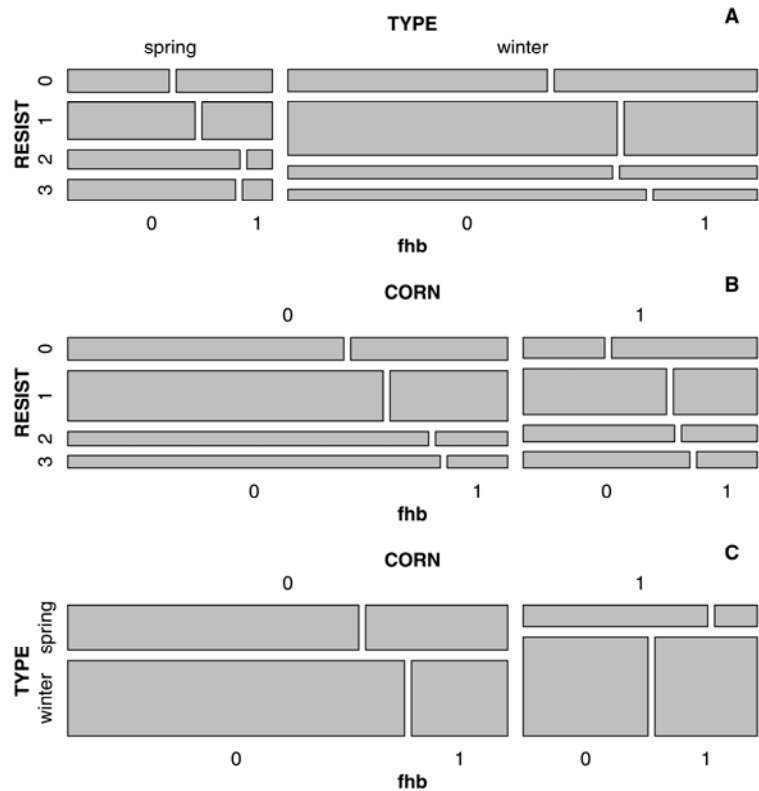

Fig. 5. Mosaic plots of the number of major Fusarium head blight epidemics (fhb = 1) and non-major epidemics (fhb = 0), where observations are further grouped as in a two-way contingency table **A,** by cultivar resistance level (RESIST) and wheat TYPE (spring or winter); **B,** by RESIST and presence (CORN = 1) or absence (CORN = 0) of corn residue; **C,** by TYPE and CORN. Within each mosaic plot, the lengths of the bars are proportional to the number of observations in the classes fhb = 0 and fhb = 1. Bar sizes are proportional to the number of observations in each two-way classification.

States. The greater relative importance of moisture-related predictors over t-related ones was also a qualitative observation made in previous empirical models (12,29,41,50,71). One may be tempted to relate weather-based predictors to specific processes within the FHB cycle but these empirical predictors only show correlation or association with the disease, and no proof of causation or mechanism should be made. However, the modeling results are consistent with known aspects of the disease cycle. Given that wheat is most susceptible to infection at anthesis (47,63), one may presume that the pre-anthesis weather-based predictors capture conditions associated with inoculum production and dispersal up to anthesis (50). The post-anthesis predictors would likewise represent additional favorable conditions for infection and colonization of the spikes occurring within 15 days post-anthesis (9).

Most FHB predictive models to date do contain some representation of r as a moisture-based predictor. Perhaps this is not unexpected, given the emphasis on r in the generalized risk algorithms (47,59,63), which would have influenced modelers to create r-based predictors. However, in our investigation, r predictors were selected much less frequently than rh predictors with the bootstrap procedure. Thus, one is led to question why other empirical model sets contain r-type predictors at a higher frequency than found for our model set. Previous empirical models (12,41) were developed under an apparent parsimony objective, which meant feature selection (such as stepwise regression) was done. One shortcoming with feature selection is the candidate set itself: selection algorithms will generally select *something* from

the candidate set and, if the latter happened to be relatively small and possibly weighted with r-derived predictors (e.g., more r-based predictors than rh-based predictors in the data matrix), one would expect r showing up in at least some of the models returned by the selection algorithm. Rainfall is certainly important for conidial dispersal (48,50) but other effects of r may be more related to the creation of wet surfaces or to increases in atmospheric moisture. With the (sometimes large) correlations between weather variables, we believe that r is probably a surrogate for other moisture-related variables that are more directly related to components of the FHB cycle. In fact, this is what Moschini and Fortugno (41) recognized and tried to capture in their NPPRH predictor, which represents 2-day periods in which moisture was high, because they did not have data at a resolution that would allow a more direct estimation of the duration of spike wetness.

It would seem logical to use both pre- and post-anthesis weather-based predictors in the same model, as previously attempted (12), but the reason for not doing so in the current investigation goes back to the purpose of the online Fusarium Head Blight Risk Assessment Tool. For one major use of the Tool, predictions need to be made sufficiently early so that growers have time to optionally apply a fungicide. Fungicide application at anthesis provides the highest percent control (52). Applications during the milk stage may also be beneficial (72) but must adhere to pre-harvest interval regulatory requirements. Post-anthesis predictors (and models which incorporate them) could only be used at or near anthesis if driven with extended forecasted weather, which adds additional uncertainty to predicted risks, more so the longer out the forecast. Additional research should clarify how uncertainty in forecasted input predictors can be mathematically represented in predicted risk. Nevertheless, post-anthesis models can be useful for predicting the level of disease (and ultimately deoxynivalenol and yield loss) at crop maturity, of benefit in marketing and post-harvest grain-handling decision making. Overall, we found greater accuracy (using different measures) for the post-anthesis models compared with the pre-anthesis ones, at least based on the training data set.

Multiple predictive models and multiple versions of predictor variables are the norm for FHB (12,41,50). Kriss et al. (29) point out that this is expected because weather-based predictors are correlated; even the underlying time series in t, rh, and r are autocorrelated (which was exploited in the multiple imputations described in the Materials and Methods). Multiple, equally accurate empirical models do present a simple but not easily answered question: which of the models does one use in practice? Earlier researchers simply picked a few that were better according to some objective criterion (12,41) on the training set; however, model selection is far more complicated (4). McNemar tests on the discordant errors provided some guidance by suggesting models 12 (pre-anthesis) and 15 (post-anthesis) were the better ones in terms of overall misclassification on the test data. We also investigated model averaging (data not shown), which sometimes leads to better predictive accuracy than any single model in a group (23). However in this case, there was no advantage, because model-averaged predicted probabilities were dominated by models 6 and 15 (D. A. Shah, *unpublished data*).

Relatively short windows risk not capturing conditions associated with FHB processes requiring longer times; however, longer windows can include too much noise in the predictors. A 5-day pre-anthesis duration was apparently too short for capturing associations with FHB, because none of the pre-anthesis 5-day window models made it to the final set of models. Conversely, on the post-anthesis side, windows >10 days appeared to have diluted predictors with too much noise, because none of the 14- or 15-day post-anthesis models were in the final model set.

The YI measure of accuracy (which balances sensitivity and specificity) is just one of several choices (16,26) for determining the cut-off threshold for translating predicted probabilities on a
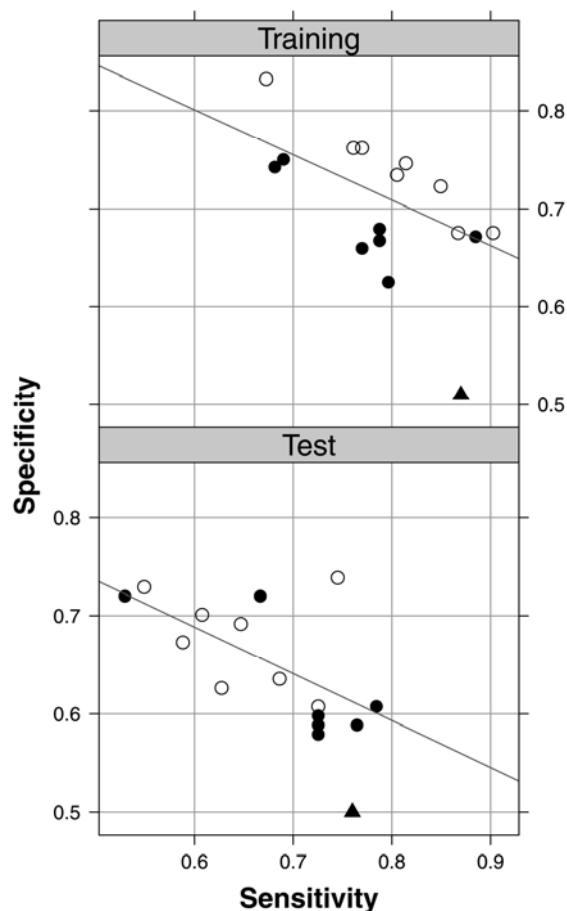

Fig. 6. Sensitivity and specificity of the final pre-anthesis (filled circles) and post-anthesis (open circles) models on the training and test data sets. Test data were not used in model development. Sloped lines are (unweighted) linear regressions of specificity on sensitivity over all 15 models. The performance of currently deployed models, after updating model coefficients on the training data (see equations 10c and 10d), is represented by the triangle symbol, and is the combined performance over both spring and winter wheat.

continuous scale to classifications of site-years as major or non-major FHB epidemics on a binary scale. For any of the selected models, the sensitivity-specificity balance can be controlled by simply moving the classification cut-point along the 0 to 1 probability scale (33); the choice depends on one's objectives (e.g., equal sensitivity and specificity, maximize specificity, minimize costs of false positives, and so on). Interestingly, in this study, there was a relation between model sensitivity and specificity across all 15 final models, where each model was optimized individually by its respective YI. One may postulate that, qualitatively, the better-performing models (in balancing sensitivity and specificity) were those above the linear regression lines of specificity on sensitivity. However, predictive accuracies alone are not sufficient to recommend a model; there needs to be a link to practical decision consequences, which depends on the relative costs associated with false positives and false negatives (37,66). Model optimization for minimizing expected costs is currently being investigated L. V. Madden and N. McRoberts, *unpublished*).

The models in this article were restricted to time windows of fairly short duration (≤15 days either before or after anthesis) and four weather-based predictors, and these are likely reasons why models did not exceed an AUC of 0.9 or a YI of 0.4. One could not force better predictive accuracy on these models without relaxing the number of weather-based predictors allowed in a single model, given the realized variation in the disease and weather data and the limited time span of the weather data set. However, correlation between the multiple predictors used in this study would have caused problems in model fitting if more predictor variables were included in a single model. There also was some curvature between logit of the probability of a major epidemic and some of the predictor variables. This curvature was accounted for by using a generalized additive logistic model, which allowed us to continue using linear models (on the logit scale) for analysis. Although fit and accuracy were improved with smoothing spline functions in the additive logistic model, those models lost some interpretability because there is no longer a single parameter

multiplying a predictor variable (for the predictors represented as splines). There are other data-mining methods better able to handle nonlinear or highly correlated predictors, such as gradient boosting (15) or support vector machines (19), both of which we are currently researching.

We showed that several weather-based, logistic regression models with equivalent predictive performance can be created for FHB epidemics, depending on how one chooses to summarize three basic variables—t, rh, and r—within relatively narrow periods (±15 days) around anthesis. All models we developed contained four or fewer weather-based predictors. They also made fewer misclassification errors than updated versions of the models currently deployed in the United States National Fusarium Head Blight Risk Assessment Tool. Because of their relative simplicity, the new models, or the weather-based predictors identified through the model-selection process, could lead to improvements in the current Risk Assessment Tool, after evaluating their predictive outcome costs.

APPENDIX

On run length predictors. Suppose we have a sequence vector (**seq**) of hourly rh values: **seq** = {77, 62, 61, 73 ,89, 92, 93, 93, 90, 80, 74, 70, 68, 63, 56, 65, 74, 80, 83, 85, 90, 86, 83, 83, 83, 82, 84, 84, 74, 73}. Convert the elements of **seq** to $x \in \{0, 1\}$, where $x = 0$ if rh < 80 and $x = 1$ if rh ≥ 80%. Then, **seq** becomes **seq**$_b$, a binary vector: **seq**$_b$ = {0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0}. A run length of length r is defined as r consecutive 1s in **seq**$_b$. So, for example, $r = 6$ would represent six consecutive hours in which rh ≥ 80%. For **seq**$_b$, we have the set of corresponding run lengths {6, 11}. Therefore, the value of the predictor "max[rl (No. h ≥ 80%)]", which can be stated descriptively as "the maximum run length of the number of consecutive hours in which rh was greater or equal to 80%", is 11. The value of the predictor "count(rl (No. h ≥ 80%) ≥ 6)", which can be stated descriptively as "the number of times

TABLE 3. Weather-based predictors included in the final 15 models

Model IDs[a]	Predictor acronym[b]	Description	Group[c]
7 & 13	RH.A.POST5.12H	Mean rh during overnight period, 5-day post-anthesis	6
4 & 8 & 14	RH.A.POST7.12H	Mean rh during overnight period, 7-day post-anthesis	6
5 & 9 & 15	RH.A.POST10.12H	Mean rh during overnight period, 10-day post-anthesis	6
1 & 6	RH.A.PRE7.24H	Mean rh per day, 7-day pre-anthesis	7
2 & 10	RH.A.PRE10.24H	Mean rh per day, 10-day pre-anthesis	7
3 & 11	RH.A.PRE14.24H	Mean rh per day, 14-day pre-anthesis	7
2	RH.G90.PRE10.12H	Number of h rh ≥ 90% during overnight period, 10-day pre-anthesis	9
2	RH.C6RLG90.PRE10.24H	Number of non-overlapping intervals with ≥6 continuous h in which rh ≥ 90%,10-day pre-anthesis	15
15	RH.C6RLG90.POST10.24H	Number of non-overlapping intervals with ≥6 continuous h in which rh ≥ 90%, 10-day post-anthesis	15
6	T.A.PRE7.24H	Mean daily temperature, 7-day pre-anthesis	16
12	T.A.PRE15.24H	Mean daily temperature, 15-day pre-anthesis	16
7 & 13	T.A.POST5.24H	Mean daily temperature, 5-day post-anthesis	16
4 & 8 & 14	T.A.POST7.24H	Mean daily temperature, 7-day post-anthesis	16
5 & 9 & 15	T.A.POST10.24H	Mean daily temperature, 10-day post-anthesis	16
6	T.L9.PRE7.24H	Number of h (t < 9°C), 7-day pre-anthesis	19
12	T.L9.PRE15.24H	Number of h (t < 9°C), 15-day pre-anthesis	19
7 & 13	TRH.15T30nRHG80.POST5.12H	Number of h (15°C ≤ t≤ 30°C & rh ≥ 80%) during overnight period, 5-day post-anthesis	31
4 & 8 & 14	TRH.15T30nRHG80.POST7.12H	Number of h (15°C ≤ t≤ 30°C & rh ≥ 80%) during overnight period, 7-day post-anthesis	31
5 & 9 & 15	TRH.15T30nRHG80.POST10.12H	Number of h (15°C ≤ t ≤ 30°C & rh ≥ 80%) during overnight period, 10-day post-anthesis	31
6	RHR.RHG90nR.PRE7.24H	Number of days (mean(rh) ≥ 90% & rainfall occurred), 7-day pre-anthesis	34
12	TRH.15T30nRHG80.PRE15.24H	Number of h (15°C ≤ t≤ 30°C & rh ≥ 80%), 15-day pre-anthesis	37

[a] A simple way of identifying a model.

[b] Acronyms follow the naming convention a.b.w.c, where a indicates whether the variable summarizing vapor pressure deficit (VPD), dewpoint depression (DD), relative humidity (RH), temperature (T), rainfall (R), or pairwise interactions between temperature and relative humidity (TRH) or between relative humidity and rainfall (RHR); b indicates the type of summary measure (description column), such as A for average (mean), L for less than, and so on; w is a placeholder for 1 of 10 vectors of hourly time series weather data ($w_{pre.5}$, $w_{pre.7}$, $w_{pre.10}$, $w_{pre.14}$, $w_{pre.15}$, $w_{post.5}$, $w_{post.7}$, $w_{post.10}$, $w_{post.14}$, and $w_{post.15}$), where the subscript in the preceding list indicates if the weather data span 5, 7, 10, 14, or 15 days pre- or post-anthesis; and c indicates a 24-h day (24H: 0800 to 0800 h) or a 12-h overnight period (12H: 2000 to 0800 h).

[c] Within groups, predictors vary only over the defining window. For example, group 7 contains predictors measuring mean rh 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). Therefore, there are 10 predictors per group and, hence, 380 total predictors.

there were six or more consecutive hours in which rh was at least 80%", is 2.

On additive logistic regression. We characterized nonlinearity-in-the-logit by generalizing the linear logistic regression model. Following the notation of Hastie et al. (23), the linear logistic regression model relates the logit link of μ to p predictors with the equation:

$$\text{logit}(\mu) = \alpha + \beta_1 X_1 + \ldots + \beta_p X_p \tag{A1}$$

The β_j are fixed (constant) coefficients. The additive logistic regression model, a special case of a GAM, generalizes equation A1 by relaxing the assumption of fixed coefficients:

$$\text{logit}(\mu) = \alpha + f_1(X_1) + \ldots + f_p(X_p) \tag{A2}$$

where the $f_j(\bullet)$ are smooth functions. Equation A2 is still linear, but each term is now more flexible than a constant β_j multiplied by a predictor.

Additive logistic regression versions of the 15 logistic regression models were fit using a thin plate spline smoother for $f_j(x_j)$ with a slight over-smoothing in the generalized cross validation score ($\gamma = 1.4$), useful to prevent overfitting in binomial models (70).

ACKNOWLEDGMENTS

LITERATURE CITED

1. Agresti, A. 2002. Categorical Data Analysis, 2nd ed. John Wiley & Sons, Inc., Hoboken, NJ.
2. Blandino, M., Haidukowski, M., Pascale, M., Plizzari, L., Scudellari, D., and Reyneri, A. 2012. Integrated strategies for the control of Fusarium head blight and deoxynivalenol contamination in winter wheat. Field Crop Res. 133:139-149.
3. Bosen, J. F. 1958. An approximation formula to compute relative humidity from dry bulb and dewpoint temperature. Mon. Weather Rev. 86:486.
4. Burnham, K. P., and Anderson, D. R. 2010. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer-Verlag, New York.
5. Cody, R. 2008. Cody's Data Cleaning Techniques Using SAS. SAS Publishing, Cary, NC.
6. Cowger, C., and Arrellano, C. 2010. Plump kernels with high deoxynivalenol linked to late *Gibberella zeae* infection and marginal disease conditions in winter wheat. Phytopathology 100:719-728.
7. 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.
8. DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44:837-845.
9. Del Ponte, E. M., Fernandes, J. M. C., and Bergstrom, G. C. 2007. Influence of growth stage on Fusarium head blight and deoxynivalenol production in wheat. J. Phytopathol. 155:577-581.
10. 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.
11. Derksen, S., and Keselman, H. J. 1992. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. Br. J. Math. Stat. Psychol. 45:265-282.
12. 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.
13. Dill-Macky, R., and Jones, R. K. 2000. The effect of previous crop residues and tillage on Fusarium head blight of wheat. Plant Dis. 84:71-76.
14. Dufault, N. S., De Wolf, E. D., Lipps, P. E., and Madden, L. V. 2006. Role of temperature and moisture in the production and maturation of *Gibberella zeae* perithecia. Plant Dis. 90:637-644.
15. Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77:802-813.
16. Freeman, E. A., and Moisen, G. G. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol. Model. 217:48-58.
17. Furnival, G., and Wilson, R. 1974. Regression by leaps and bounds. Technometrics 16:499-511.
18. Gourdain, E., Piraux, F., and Barrier-Guillot, B. 2011. A model combining agronomic and weather factors to predict occurrence of deoxynivalenol in durum wheat kernels. World Mycotoxin J. 4:129-139.
19. Hamel, L. 2009. Knowledge Discovery with Support Vector Machines. John Wiley & Sons, Inc., Hoboken, NJ.
20. Hand, D. J. 2010. Evaluating diagnostic tests: The area under the ROC curve and the balance of errors. Stat. Med. 29:1502-1510.
21. Hanley, J. A., and McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29-36.
22. Harrell, F. E., Jr. 2001. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer-Verlag, New York.
23. Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning, 2nd ed. Springer, New York.
24. Hilbe, J. 2009. Logistic Regression Models. CRC Press, Boca Raton, FL.
25. 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.
26. Jiménez-Valverde, A., and Lobo, J. M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecol. Int. J. Ecol. 31:361-369.
27. Khonga, E. B., and Sutton, J. C. 1988. Inoculum production and survival of *Gibberella zeae* in maize and wheat residues. Can. J. Plant Pathol. 10:232-239.
28. Kriss, A. B., Paul, P. A., Xu, X., Nicholson, P., Doohan, F. M., Hornok, L., Rietini, A., Edwards, S. G., and Madden, L. V. 2012. Quantification of the relationship between the environment and Fusarium head blight, Fusarium pathogen density, and mycotoxins in winter wheat in Europe. Eur. J. Plant Pathol. 133:975-993.
29. 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.
30. Krzanowski, W. J., and Hand, D. J. 2009. ROC Curves for Continuous Data. CRC Press, Boca Raton, FL.
31. Krzanowski, W. J., and Hand, D. J. 2011. Testing the difference between two Kolmogorov-Smirnov values in the context of receiver operating characteristic curves. J. Appl. Stat. 38:437-450.
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. Madden, L. V. 2006. Botanical epidemiology: Some key advances and its continuing role in disease management. Eur. J. Plant Pathol. 115:3-23.
34. Martínez, M., Moschini, R., Barreto, D., and Comerio, R. 2012. Effect of environment on Fusarium head blight intensity and deoxynivalenol content in wheat grains: development of a forecasting system. Cereal Res. Commun. 40:74-84.
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. McRoberts, N., Hall, C., Madden, L. V., and Hughes, G. 2011. Perceptions of disease risk: From social construction of subjective judgments to rational decision making. Phytopathology 101:654-665.
38. McRoberts, N., Hughes, G., and Madden, L. V. 2003. The theoretical basis and practical application of relationships between different disease intensity measures in plants. Ann. Appl. Biol. 142:191-211.
39. 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.
40. Molineros, J. E. 2007. Understanding the challenges of Fusarium head blight forecasting. Ph.D. Dissert. The Pennsylvania State University, University Park.
41. 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.
42. Newcombe, R. G. 1998. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat. Med. 17:873-890.
43. 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.
44. 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.
45. Osborne, L. E., and Stein, J. M. 2007. Epidemiology of Fusarium head blight on small-grain cereals. Int. J. Food Microbiol. 119:103-108.
46. Pace, N. L. 2008. Independent predictors from stepwise logistic regression may be nothing more than publishable *P* values. Anesth. Analg. 107:1775-1778.
47. Parry, D. W., Jenkinson, P., and McLeod, L. 1995. *Fusarium* ear blight (scab) in small grain cereals–A review. Plant Pathol. 44:207-238.
48. Paul, P. A., El-Allaf, S. M., Lipps, P. E., and Madden, L. V. 2004. Rain splash dispersal of *Gibberella zeae* within wheat canopies in Ohio. Phytopathology 94:1342-1349.
49. Paul, P. A., El-Allaf, S. M., Lipps, P. E., and Madden, L. V. 2005. Relationships between incidence and severity of Fusarium head blight on winter wheat in Ohio. Phytopathology 95:1049-1060.
50. Paul, P. A., Lipps, P. E., De Wolf, E., Shaner, G., Buechley, G., Adhikari, T., Ali, S., Stein, J., Osborne, L., and Madden, L. V. 2007. A distributed lag analysis of the relationship between *Gibberella zeae* inoculum density on wheat spikes and weather variables. Phytopathology 97:1608-1624.
51. Paul, P. A., Lipps, P. E., Hershman, D. E., McMullen, M. P., Draper, M. A., and Madden, L. V. 2007. A quantitative review of tebuconazole effect on Fusarium head blight and deoxynivalenol content in wheat. Phytopathology 97:211-220.
52. Paul, P. A., Lipps, P. E., Hershman, D. E., McMullen, M. P., Draper, M. A., and Madden, L. V. 2008. Efficacy of triazole-based fungicides for Fusarium head blight and deoxynivalenol control in wheat: A multivariate meta-analysis. Phytopathology 98:999-1011.
53. Paulitz, T. C. 1999. Fusarium head blight: A re-emerging disease. Phytoprotection 80:127-133.
54. Pencina, M. J., D'Agostino, R. B., Sr., D'Agostino, R. B., Jr., and Vasan, R. S. 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat. Med. 27:157-172.
55. 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.
56. Prenger, J. J., and Ling, P. P. 2001. Greenhouse condensation control: Understanding and using vapor pressure deficit (VPD). Ohio State Univ. Ext. Fact Sheet AEX-804.
57. Rubin, D. B. 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc., New York.
58. 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.
59. Shaner, G. 2003. Epidemiology of Fusarium head blight of small grain cereals in North America. Pages 84-119 in: Fusarium Head Blight of Wheat and Barley. K. J. Leonard and W. R. Bushnell, eds. American Phytopathological Society Press, St. Paul, MN.
60. Stack, R. W. 2003. History of Fusarium head blight with emphasis on North America. Pages 1-34 in: Fusarium Head Blight of Wheat and Barley K. J. Leonard and W. R. Bushnell, eds. American Phytopathological Society, St. Paul, MN.
61. Stack, R. W., and McMullen, M. P. 1998. A visual scale to estimate severity of Fusarium head blight in wheat. N. D. State Univ. Ext. Serv. Small Grains Publ. PP-1095. Online publication.
62. Stein, J. M., Osborne, L. E., Bondalapati, K. D., Glover, K. D., and Nelson, C. A. 2009. Fusarium head blight severity and deoxynivalenol concentration in wheat in response to *Gibberella zeae* inoculum concentration. Phytopathology 99:759-764.
63. Sutton, J. C. 1982. Epidemiology of wheat head blight and maize ear rot caused by *Fusarium graminearum*. Can. J. Plant Pathol. 4:195-209.
64. Van der Heiden, G. J. M. G., Donders, A. R. T., Stijnen, T., and Moons, K. G. M. 2006. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. J. Clin. Epidemiol. 59:1102-1109.
65. Venkatraman, E. S., and Begg, C. B. 1996. A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. Biometrika 83:835-848.
66. Vickers, A. J. 2008. Decision analysis for the evaluation of diagnostic tests, prediction models, and molecular markers. Am. Stat. 62:314-320.
67. Wegulo, S. N., Bockus, W. W., Nopsa, J. H., De Wolf, E. D., Eskridge, K. M., Peiris, K. H. S., and Dowell, F. E. 2011. Effects of integrating cultivar resistance and fungicide application on Fusarium head blight and deoxynivalenol in winter wheat. Plant Dis. 95:554-560.
68. Whittingham, M. J., Stephens, P. A., Bradbury, R. B., and Freckleton, R. P. 2006. Why do we still use stepwise modelling in ecology and behaviour? J. Anim. Ecol. 75:1182-1189.
69. 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.
70. Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. CRC Press, Boca Raton, FL.
71. Xu, X. M., Nicholson, P., Thomsett, M. A., Simpson, D., Cooke, B. M., Doohan, F. M., Brennan, J., Monaghan, S., Moretti, A., Mule, G., Homok, L., Beki, E., Tatnell, J., Ritieni, A., and Edwards, S. G. 2008. Relationship between the fungal complex causing Fusarium head blight of wheat and environmental conditions. Phytopathology 98:69-78.
72. Yoshida, M., Nakajima, T., Tomimura, K., Suzuki, F., Arai, M., and Miyasaka, A. 2012. Effect of the timing of fungicide application on Fusarium head blight and mycotoxin contamination in wheat. Plant Dis. 96:845-851.