Predicting Infection Risk of Hop by *Pseudoperonospora humuli*

David H. Gent and Cynthia M. Ocamb

First author: United States Department of Agriculture–Agricultural Research Service, Forage Seed and Cereal Research Unit, and Oregon State University, Department of Botany and Plant Pathology, Corvallis, OR 97331; and second author: Oregon State University, Department of Botany and Plant Pathology, Corvallis 97330. Accepted for publication 10 June 2009.

**ABSTRACT**


Downy mildew, caused by *Pseudoperonospora humuli*, is one of the most destructive diseases of hop. Weather factors associated with infection risk by *P. humuli* in the maritime region of western Oregon were examined for 24- and 48-h periods and quadratic discriminant function models were developed to classify periods as favorable for disease development on leaves. For the 24-h data sets, the model with superior predictive ability included variables for hours of relative humidity >80%, degree-hours of wetness, and mean night temperature. The same variables were selected for the 48-h data sets, with the addition of a product variable for mean night temperature and hours of relative humidity >80%. Cut-points (Pr) on receiver operating characteristic curves that minimized the overall error rate were identified by selecting the cut-point with the highest value of Youden's index. For the 24- and 48-h models these were \( Pr = 0.49 \) and 0.39, respectively. With these thresholds, the sensitivity and specificity of the models in cross validation by jackknife exclusion were 83.3 and 88.8% for the 24-h model and 87.5 and 84.4% for the 48-h model, respectively. Cut-points that minimized the average costs associated with disease control and crop loss due to classification errors were determined using estimates of economic damage during vegetative development and on cones near harvest. Use of the 24- and 48-h models was estimated to reduce average management costs during vegetative development when disease prevalence was <0.31 and 0.16, respectively. Using economic assumptions near harvest, management decisions informed by the models reduced average costs when disease prevalence was <0.21 and 0.1 for the 24- and 48-h models, respectively. The value of the models in management decisions was greatest when disease prevalence was relatively low during vegetative development, which generally corresponds to the normally drier period from late spring to midsummer in the Pacific Northwest of the United States.

Downy mildew of hop, caused by *Pseudoperonospora humuli*, is among the most important diseases of hop (*Humulus lupulus*) (1,29). The disease continues to threaten the economic viability of the U.S. hop industry because of widespread fungicide resistance in the pathogen population (5,20,28) and lack of host resistance to the disease in commercially acceptable cultivars (29,33).

Economic damage from the disease may occur in numerous ways because *P. humuli* can infect cones, leaves, and shoots, and also may cause systemic infections of the perennial crown and root system (33,38). Systemic infection of shoots leads to the formation of the so-called 'basal spikes,' which are stunted and have brittle, downward-curved leaves from which masses of purple to black sporangiophores and sporangia are produced. Infection of shoots can result in a lack of healthy shoots that can be trained for production (39). Infection of shoots after they have been trained onto strings to enable production of lateral branches can cause bines to fall away from strings, necessitating retaining and reducing yield (29,33). Similarly, infection of lateral branches (termed 'lateral spikes') prevents the normal development of the cone-bearing branches. The most serious yield losses typically result from infection of developing cones, which can cause complete crop loss due to cone abortion, reductions in α-acid content, and poor cone quality (33). In certain cultivars, the disease also may cause a crown rot and subsequent plant death (15,38).

Disease development and severity is favored by extended periods of wetness, high humidity, and temperatures of 5 to 29°C (24,32). Various combinations of weather-related variables have been used in empirical forecasting systems in Europe (21,33) and Washington State (13,16,17), with varying success.

Quantitative analysis of weather factors associated with daily infection risk have not been conducted in the western United States, although models for sporulation potential and long-range forecasts of epidemic occurrence have been developed for semiarid production environments in Washington State (12,13,16,17). Because downy mildew epidemics do not occur annually in Washington, these models were promoted as tools to enhance sampling efficiency and provide long-range predictions of epidemics. However, such models are of limited use in hop production regions where downy mildew is endemic, such as western Oregon and northern Idaho.

Consequently, disease control measures typically are applied prophylactically in regions where the disease occurs on an annual basis (11,39), which likely results in inefficient fungicide use. Therefore, the objectives of this research were to identify weather factors associated with infection risk by *P. humuli* in the maritime region of western Oregon, and explicitly model conditions favorable for infection. Optimal management action thresholds for discriminant functions were derived by receiver operating-characteristic curve analysis to identify action thresholds for the models that minimize average disease control costs and crop loss due to classification errors.

**MATERIALS AND METHODS**

**Plant material.** Plants of the downy mildew susceptible cv. Nugget (8) were propagated clonally from soft wood cuttings (9) and maintained in a greenhouse devoid of downy mildew. The greenhouse was maintained at 20 to 25°C with a 14-h photo-
period. Plants were grown in Sunshine Mix no. 1 (SunGro Horticulture, Bellevue, WA) in 440-cm³ pots for approximately 14 days and then repotted in 1,000-cm³ pots for an additional 14 days of growth. Plants were watered regularly and supplied with Champion 17-17-17 (N-P₂O₅-K₂O) fertilizer with micronutrients (McConkey’s, Portland, OR) at each irrigation to promote vigorous growth. Plants were treated with imidaclopid (Marathon 1% G; OHP, Inc., Mainland, PA) at the manufacturer’s recommended rate to control aphids after plants were repotted in 1,000-cm³ pots.

**Infection bioassay.** A bioassay similar to that described by Royle (32) was used to identify days associated with *P. humuli* infection. Approximately every week (May through September 2005, March through October 2006, and March through June 2007), 10 to 20 potted plants were placed in an experimental hop yard near Corvallis, OR (44.5696, –123.2380). Hills (i.e., a group of one or more rhizome pieces planted together) in the hop yard were arranged on a 2.1-m grid (narrow spacing) under a 5-m trellis. The youngest, fully unfurled leaves on the bioassay plants were marked for later identification as plants were placed in the field at 4:00 p.m. in 2005 and 8:00 p.m. in 2006 and 2007. In all years, bioassay plants were arranged in a circular pattern around several plants with numerous downy mildew spikes in an area of the hop yard that did not receive fungicide applications. The plants used in the bioassay were bordered by at least one diseased hop plant on all sides to minimize variability in the assay related to (potentially) increased deposition of inoculum on some bioassay plants due to the effect of the prevailing wind direction. Preliminary analyses also investigated the effect of plant position in the field on disease incidence and found no evidence of a systematic bias for disease incidence or severity due to position in the field. Sets of bioassay plants were exposed to environmental conditions and natural inoculum for a 24-h period or one 48-h period in order to account for the potential for deposition of sporangia occurring on one day and infection on the second day. In total, 28, 59, and 38 sets of plants (data sets) were collected for a 24-h period in 2005, 2006, and 2007, respectively, and 29 and 19 data sets were collected for a 48-h period in 2006 and 2007, respectively.

After 24 or 48 h of exposure, plants were placed in a growth chamber or greenhouse facility maintained at 18 to 20°C and approximately 70% relative humidity. Plants were irrigated from the bottom to avoid wetting the leaves, thus reducing the chance of infection during the incubation period in the growth chamber or greenhouse. After the 7- to 10-day incubation period, plants were misted to just before runoff with sterile 18 ohm water (nanopure water; Barnstead, Dubuque, IA) and enclosed in plastic bags overnight to induce sporulation. The bags were removed the following day and the procedure was repeated the following night. After the second night of incubation, the incidence of leaves with downy mildew was assessed. Infection severity (lesions per square centimeter) was measured by counting the number of lesions on leaves marked previously. Leaves were then detached and their area was determined with a Li-Cor LI-3000 leaf area meter (Li-Cor Inc., Lincoln, NE). Data from all of the plants were averaged to produce an estimate of disease severity for each 24- or 48-h period. At least one control plant that was not deployed to the field was included with each set of bioassay plants to examine the potential for infection during the growth chamber or greenhouse incubation. Other control plants were inoculated regularly with multiple monosporangial isolates of *P. humuli* (5) to test plant susceptibility to infection.

Inoculum density in the hop yard was measured with a Burkhard volumetric spor trap (Burkhard Manufacturing Co., Ricksmanworth, United Kingdom) operated continuously near the bioassay plants. The sampler orifice was located 1 m from the soil surface at approximately the top of the first node of the bioassay plants. Air flow was maintained at 10 liters/min and was verified at least weekly. Sporangia of *P. humuli* were quantified by staining the Melinex tape with several drops of aniline blue and examining the tape under ×160 magnification.

**Weather data.** Weather variables at the field site were monitored and recorded every 15 min with a Campbell Scientific CR10X datalogger (Campbell Scientific, Inc., Logan, UT). Air temperature and relative humidity was measured at 1.5 m from the ground with a model HMP45C temperature and relative humidity probe. Leaf wetness duration was measured with a model 237 artificial leaf-type sensor coated with latex paint and mounted at 1 m from the ground at a 45° angle facing west (41). Solar flux was measured with a Li-Cor LI-200X pyranometer (Li-Cor) positioned at 6 m from the ground at the top of the hop trellis. Rainfall was measured with a model TE525 tipping bucket rain gauge mounted 6 m from the ground at the top of the hop trellis. All sensors and equipment were purchased from Campbell Scientific.

Temperature, relative humidity, and solar flux for a 15-min period were obtained by averaging three measurements taken at 5-min intervals. Leaf wetness was recorded as a single-point measurement every 15 min. Leaf wetness measurements were calibrated by observing leaves on several mornings and noting when wetness was present and the time of leaf drying. Leaves were considered wet during a 15-min period if the sensor resistance was <4,000, precipitation was recorded, or relative humidity was >95%.

**Infection model development.** Preliminary analysis of potential predictor variables for infection was conducted by creating scatter plots of the incidence of leaves with downy mildew and the weather variables. The independent variables selected after inspection of scatter plots were cumulative rain, hours of relative humidity >80 or 90%, cumulative hours of leaf wetness, leaf wetness associated with rain, morning leaf wetness or daylight leaf wetness, cumulative degree-hours of leaf wetness, cumulative degree-hours of leaf wetness during daylight, mean daily temperature, and mean night temperature. For the 48-h data sets, cumulative degree-hours during nighttime humidity >80 or 90% also were calculated. In calculating the weather variables, a 15-min time period was defined as “daylight” if the average solar flux was at least 1 W m⁻² and “night” otherwise. Cumulative degree-hours were calculated similar to Pfender (31). A low temperature threshold of 8°C was used in the calculations based on earlier reports that 8°C is the minimum temperature for shoot infection (32) and near the threshold for sporulation by *P. humuli* (16). For calculating degree-hours of nighttime humidity >80 or 90%, observations from 12:00 midnight to 6:00 a.m. were used based on previous studies on the diurnal cycle of sporangia appearance and development (34,45).

Box plots and scatter plots of weather and inoculum variables were constructed to compare the distribution of variables on days without and with infection. Median values of the predictor variables for data sets without and with infection were compared using a nonparametric sign test (35). Associations between disease incidence and predictor variables were measured by Spearman’s rank correlation coefficient (5) (42). Correlations also were calculated after disease incidence was normalized for inoculum density by dividing by the logarithm of sporangia M⁻³ h⁻¹ as determined from the number of sporangia collected by the Burkhard sporre trap. The distribution of predictor variables on days without or with infection were compared with the nonparametric Kolmogorov-Smirnov (K-S) test using the NPAR1WAY procedure in SAS (version 9.1; SAS Institute, Cary, NC). Monte Carlo estimation was used to derive exact P values for the K-S test.

Discriminant analysis was used to develop classification functions to predict whether downy mildew would or would not occur on the bioassay plants. Discriminant analysis is a multivariate method used to classify observations into two or more categorical
groups ("group membership"). The method is similar to linear regression because both techniques use one or more predictor variables and a single response variable. However, in discriminant analysis, the response variable is categorical. Categorical predictor variables were not included in the discriminant function because parametric discriminant analysis assumes that the data has a multivariate normal distribution (18). Numerous preliminary models were constructed using forward, backward, and stepwise selection. Many of the weather variables considered were related (e.g., hours of leaf wetness and cumulative degree-hours of wetness), and weather variables with redundant information were considered separately during model development to minimize collinearity. For stepwise selection, the significance level for entry into the model was set to $\alpha = 1.0$ and the significance level for remaining in the model was set to $\alpha = 0.1$ to provide information about a large number of possible models. Model accuracy, sensitivity ($Se$), and specificity ($Sp$) in resubstitution and cross-validation by jackknife exclusion (18) were calculated and used to evaluate preliminary models. To perform cross-validation by jackknife exclusion, data sets were classified using a discriminant function model determined when each observation was omitted individually from the analysis and only the remaining data sets were used to fit the model. An advantage of this validation procedure compared with analysis by resubstitution is that jackknife exclusion reduces bias caused by classifying the same data sets used during model development.

Accuracy was defined as the proportion of data sets classified correctly, $Se$ was the proportion of data sets where disease occurred that were classified correctly, and $Sp$ was the percentage of data sets where disease did not occur that were classified correctly. Final models were selected based on examination of scatter plots and box plots of individual variables, model accuracy, $Se$, and $Sp$, the number of variables in the model, and biological considerations. Models with the least number of independent predictor variables were preferred to avoid overspecification. Discriminant analysis was conducted using the DISCRIM procedure in SAS.

Bartlett’s modified likelihood ratio test (27) indicated that the assumption of homogeneous variance-covariance matrices was violated, which is common with large data sets because of the high power of this test (18). Johnson (18) suggested that differences between variance-covariance matrices, although significant by Bartlett’s modified likelihood ratio test, may not be large enough to be of practical importance for discriminant analysis. Therefore, both pooled (linear discriminant rule) and within-group covariance matrices (quadratic discriminant rule) were calculated and used to evaluate preliminary models. To perform cross-validation by jackknife exclusion, data sets were classified using a discriminant function model determined when each observation was omitted individually from the analysis and only the remaining data sets were used to fit the model. An advantage of this validation procedure compared with analysis by resubstitution is that jackknife exclusion reduces bias caused by classifying the same data sets used during model development.

Receiver operating characteristic curve analysis. Classification of data sets by discriminant analysis requires an arbitrary threshold probability, $p_T$ (often nominally set to 0.5) for prediction of group membership. However, the threshold chosen clearly will influence classification accuracy, $Se$, and $Sp$. Receiver operating characteristic (ROC) curve analysis was used to evaluate the discriminant models and identify decision thresholds, specifically the coordinates of $Se$ and $Sp$ associated with model-estimated probability of group membership. ROC curve analysis provides a graphical method to evaluate a diagnostic test or predictive system, and is a routine analysis in clinical medicine (46). An advantage of ROC curve analysis over other methods of evaluating predictive systems is that it can provide a measure of accuracy for a predictor for both positive and negative predictions which is independent of the prevalence of disease in the data sets (26).

Data sets for the 24- and 48-h infection models were classified as “cases” where disease occurred ($D+$) or “controls” where no disease occurred ($D$). In ROC curve analysis, the term control is used to define the disease status of a data set and should not be confused with an experimental control. The $Se$ and $Sp$ of each model was determined by calculating the proportion of data sets that were classified correctly in cross-validation, assuming a probability of group membership above a decision threshold, or “cut-point.” A cut-point is a point on an ROC curve used to dichotomize predictions of a continuous variable into categories, with each cut-point corresponding to a predictive threshold for a model. To construct empirical ROC curves, model $Se$ (referred to as true positive proportion, TPP) was plotted against 1 – $Sp$ (referred to as false positive proportion, FPP) by allowing the cut-point to vary over the range of its possible values (10,26). To compare models, the area under the ROC curve (AUROC) and its standard error were calculated (7) for each curve using the software MedCalc (MedCalc Software, Mariakerke, Belgium). AUROC is a summary measure of the predictive ability of a predictive system and may vary from 0 to 1 (perfect prediction). A standard normal z-statistic and associated probability was calculated to test the null hypothesis that the AUROC was >0.5, which is the area under the line representing a noninformative predictor (i.e., the diagonal of the unit square).

For each model, the cut-point where the overall error rate was smallest was determined by Youden’s index, $J$, which identifies the point on the ROC curve at the greatest geometric distance from the line representing a noninformative predictor (26). Youden’s index is commonly used as a measure of overall diagnostic effectiveness and is calculated by $J = Se + Sp - 1$ (26).

Likelihood ratios ($LR$s) for positive and negative predictions were depicted graphically for the cut-points identified by Youden’s index. The likelihood ratio of a positive prediction is represented as $LR(\text{+})$ and is calculated as $TPP/FPP$. The likelihood ratio of a negative prediction is represented as $LR(\text{-})$ and is calculated as $FNP/TNP$. $LR(\text{+})$ and $LR(\text{-})$ are useful for comparing cut-points or predictors and inferring “predictive value”, which can be derived by expressing the likelihood ratio as a posterior probability for a given prevalence of disease ($Prev$). In this context, the posterior probability is the probability of infection by $P$. humuli subsequent to a positive model prediction of infection. The $Prev$ represents the probability of downy mildew exceeding a threshold, in this case the occurrence of infection and disease at any level of disease incidence $>0$. This definition of $Prev$ is distinct from the typical use of the term “disease prevalence” in plant pathology to refer to the proportion of fields with diseased plants (23). For a given cut-point, the slope of the line extending from the origin to the cut-point is $LR(+)$. The line with steepest slope has the highest $LR(\text{+})$ and, thus, a higher probability of infection occurring given a positive prediction of infection by the discriminant function (referred to as “positive predictive value”). The line extending from the cut-point to the point $(1,1)$ is $LR(\text{-})$. The line with the shallower slope has the higher $LR(\text{-})$ and, thus, greater probability of infection not occurring given a negative prediction (referred to as “negative predictive value”).

Derivation of optimal decision thresholds. A potential limitation of Youden’s index is that model $Se$ and $Sp$ are weighted equally and, for a discriminant function, Youden’s index identifies a cut-point that minimizes overall misclassification error rate. In most instances, $Se$ and $Sp$ are not of equal importance economically for diagnostic tests (26) or plant disease management (43) because growers of high-value crops tend to be risk adverse (23,30).

Fabre et al. (4) described how a cost function can be utilized to derive an optimal threshold probability ($p_T$) where average management costs are minimized relative to errors, taking into account the cost of false-positive and false-negative predictions. If $p$ is the probability that a fungicide application for downy mildew will provide a positive net return, $p_T$ is defined as a threshold value such that, if $p > p_T$, a fungicide treatment is recommended and, if $p < p_T$, a fungicide treatment is not recommended. Pre-
serving the notation of Fabre et al. (4), $T^+$ and $T^-$ represent prediction that a treatment is recommended or not recommended, respectively, by the discriminant function. $D^+$ and $D^-$, as defined previously, are assumed to be infection events where a treatment is required or not required, respectively, independent of a model prediction. The \( \text{Prev} \), as described above, represents the probability of downy mildew exceeding a threshold, in this case the occurrence of infection and disease at any level. Costs associated with $T$ and $D$ are $C^{++}$ (true positive), $C^{+-}$ (false positive), $C^{-+}$ (false negative), and $C^{--}$ (true negative). The expected cost associated with $p_T$, denoted by $C(p_T)$, for a given level of disease prevalence is

$$C(p_T) = \text{Prev} \times [S \times C^{++} + (1 - S) \times C^{+-}] + (1 - \text{Prev}) \times [S \times C^{-+} + (1 - S) \times C^{--}]$$

(1)

Cost associated with $C^{++}$, $C^{+-}$, $C^{-+}$, and $C^{--}$ were estimated for crop damage during vegetative development and cone development based on information obtained from previous studies. Infection of shoots during early vegetative development can lead to systemic infection of the crown and root system (2,3,38), which has been associated with a 28% reduction in yield compared with noninfected plants. Conversely, infection of cones by \( P. \text{humuli} \) can reduce cone quality (based on subjective assessment by the contracting brewery or dealer), which may result in crop devaluation (33) of ≈50% (D. H. Gent, unpublished). The cost of a false positive and true positive was estimated to be $150 per hectare based on an unnecessary fungicide application of Aliette WDG (Bayer CropScience, Research Triangle Park, NC), given $100 per hectare for the fungicide and $50 per hectare in application costs. The cost of a true negative was assumed to be zero. It was also assumed that a crop of aroma hops was valued at $20,000 per hectare based on 2007 market conditions (6) and 2008 contract prices. These estimates assume costs for $C^{++}$, $C^{+-}$, $C^{-+}$, and $C^{--}$ associated with a single fungicide application, which is likely a worse-case scenario because a false-negative prediction of disease may not lead to crop loss if additional control measures can be implemented. Clearly, actual values associated with the cost of $T^+$ and $D^+$ vary with crop value and a particular sequence of future weather and infection events; therefore, the cost presented are not appropriate for all situations. However, they provide a reasonable and conservative approximation for estimating $p_T$ because multiple values of $D^+$ and $\text{Prev}$ were considered.

To derive optimal decision thresholds, $C(p_T)$ was minimized for a range of potential values of $\text{Prev}$ by setting the first derivative of $C$ to zero with respect to $p_T$, as described by Fabre et al. (4). Algebraic approximations for the first derivatives of $\text{Sp}$ ($\Delta \text{Sp}$) and $\text{Se}$ ($\Delta \text{Se}$) were estimated by fitting (independently) the following models to the data for $\text{Sp}(p_T)$ and $\text{Se}(p_T)$:

$$\text{Sp}(p_T) = rp \times \ln(kp + p_T + 1)$$

(2)

$$\text{Se}(p_T) = 1 - ke \times \exp(re \times p_T - 1)$$

(3)

Equations 2 and 3 yield curves for $\text{Sp}$ and $\text{Se}$ expressed in a functional (parametric) form. Estimates of the shape parameters $rp$, $kp$, $re$, and $ke$ for equations 3 and 4 were obtained by nonlinear regression using PROC NLIN in SAS (version 9.1; SAS Institute). The optimum $p_T$ for a range of values of $\text{Prev}$ was then identified by minimizing:

$$[\Delta \text{Sp} \times (1 - \text{Prev}) \times (C^{--} - C^{+-}) + \Delta \text{Se} \times \text{Prev} \times (C^{++} - C^{--})]^2$$

(4)

Calculations involving equation 4 were solved using the Solver add-in in Microsoft Excel (Microsoft Corp., Redmond, WA).

**RESULTS**

**Infection bioassay.** For the 24-h data sets, infection occurred on 18 of 125 days over the 3 years of study. The mean incidence of leaves with downy mildew was 0 to 0.7 among the data sets. The distribution of disease incidence on leaves was heavily right-skewed, with mean 0.023 (standard deviation 0.088) and median 0. For the 48-h data sets, infection occurred on 16 of the 48 sets of plants. Mean incidence of leaves with downy mildew was 0 to 0.89. Similarly to the 24-h data sets, the distribution of disease incidence was heavily right-skewed, with mean 0.083 (standard deviation 0.22) and median 0.

Among the weather variables for the 24-h data sets, most of the variables, with the exception of mean daily temperature and mean night temperature, were significantly associated with the disease incidence. Expectedly, variables including a measurement of wetness tended to be highly correlated with disease incidence, with cumulative rain ($S = 0.566; P < 0.0001$) and hours of leaf wetness associated with rain ($S = 0.571; P < 0.0001$) being the most strongly correlated. For the 48-h data sets, the strongest correlations were observed with degree-hours of humidity >80% ($S = 0.554; P < 0.0001$), degree-hours of humidity >90% ($S = 0.525; P < 0.0001$), and degree-hours of leaf wetness ($S = 0.535; P < 0.0001$).

Sporangia were detected in the Burkhard trap during 66 of the 125 periods for the 24-h data sets and 30 of the 48 periods for the 48-h data sets. Sporangia density was correlated with disease incidence for the 24-h data sets ($S = 0.491; P < 0.0001$) and 48-h data sets ($S = 0.427; P < 0.003$). In the 24-h data sets, the significance of the correlation between a weather variable and disease incidence did not change when the analysis was conducted with disease incidence standardized by spore density, although the magnitude of $S$ often changed. The exception to this trend was mean night temperature, which was significantly associated with standardized disease incidence ($S = 0.346; P < 0.004$). The variables most highly associated with standardized disease incidence were degree-hours leaf wetness ($S = 0.620; P < 0.0001$), rain leaf wetness ($S = 0.597; P < 0.0001$), cumulative rain ($S = 0.589; P < 0.0001$), and degree-hours of daylight wetness ($S = 0.584; P < 0.0001$). For the 48-h data sets, all variables that were significantly associated with disease incidence also were significant using standardized disease incidence. The variables most highly correlated with the standardized disease incidence were degree-hours of humidity >80% ($S = 0.766 P < 0.0001$), degree-hours of humidity >90% ($S = 0.721; P < 0.0001$), degree-hours of leaf wetness ($S = 0.726; P < 0.0001$), and mean night temperature ($S = 0.702; P < 0.0001$).

Weather variables that were significantly associated with disease incidence as measured by Spearman's rank correlation generally were significant by the K-S test (Table 1). For the 24-h data sets, exceptions to this trend were for mean night temperature and relative humidity >90%. Mean night temperature was not significantly correlated with disease incidence ($P = 0.191$), although the distribution of this variable was significantly different on days without and with infection ($P = 0.013$), as determined by the K-S test. Hours of relative humidity >90% was correlated with disease incidence ($P = 0.033$), whereas the distributions were similar on days without and with infection ($P = 0.100$). For the 48-h data sets, cumulative rain and hours of rain leaf wetness were correlated with disease incidence ($P = 0.002$ and 0.010, respectively); however, the distributions of these variables among data sets were not significantly different by the K-S test ($P = 0.054$ and 0.176, respectively).

**Infection model development.** After inspection of numerous preliminary models, a quadratic discriminant function was developed for the 24-h data sets that included the predictor variables hours of relative humidity >80%, degree-hours of wetness, and mean night temperature (Fig. 1). The distribution of these vari-
ables varied significantly between days without and with infection by P. humuli (Table 1) (P ≤ 0.0131). Median values of hours relative humidity >80%, degree-hours of wetness, and mean night temperature were 12.5, 44.6, and 11.6 on days with infection, respectively, and 9.5, 6.2, and 10.6 on days without infection. Median values of the predictor variables were significantly different for data sets without and with infection based on the non-parametric sign test, with the exception of mean night temperature (Fig. 1).

In cross-validation, a quadratic discriminant function with these variables classified 88.8% of days into the correct category, with 83.3% Se and 89.0% Sp, assuming a nominal \( p_T = 0.5 \) for designating a day as an infection day. A similar model assuming homogeneity of covariance matrices (i.e., a linear discriminant function) had 85.6% accuracy, 72.2% Se, and 87.9% Sp. Due to the superior Se of the quadratic discriminant function compared with the linear discriminant function, the quadratic model was selected for further evaluation (described below).

For the 48-h data sets, the final model included predictor variables for hours of relative humidity >80%, degree-hours of wetness, mean night temperature, and an interaction variable for the product of mean night temperature and hours of relative humidity >80% (Fig. 2). The distribution of degree-hours of wetness, mean night temperature, and an interaction variable for the product of mean night temperature and hours of relative humidity varied significantly between data sets without and with infection (Table 1) \((P ≤ 0.002)\). The distribution of hours of relative humidity >80% were similar \((P = 0.9445)\) between infection and noninfection data sets (Fig. 2). However, this variable was retained in the model because the interaction variable with mean night temperature and relative humidity was significant, and also because of biological considerations. Median values for hours of relative humidity >80%, degree-hours of wetness, mean night temperature, and an interaction variable for the product of mean night temperature and hours of relative humidity >80% for data sets without infection were 19.6, 9.8, 5.9, and 117.7, respectively. For data sets with infection, the median values were 20.5, 64.0, 10.3, and 202.0, respectively. Median values of the variables, with the exception of hours of relative humidity >80%, were significantly different for data sets without and with infection based on the sign test (Fig. 2).

In cross-validation, a quadratic discriminant function for the 48-h data sets had 77.1% accuracy, 68.8% Se, and 81.3% Sp, assuming a nominal \( p_T = 0.5 \). A similar model for the 48-h sets based on pooled covariance matrices had 75% accuracy, 56.3% Se, and 78.1% Sp. The quadratic discriminant function was selected for further evaluation based on the superior Se, Sp, and accuracy of this model.

### DISCUSSION

The primary purpose of this study was to identify and model weather variables associated with daily risk of infection of hop by P. humuli in the unique environment of western Oregon, and identify when use of the models could reduce average management costs for downy mildew. Madden et al. (23) suggest that the first step in formulating a disease management strategy is to “…identify the most important risk factors among those on the long list of possible candidates.” For the 48-h data sets, the model with superior predictive ability included the variables hours of relative humidity >80%, degree-hours of wetness, and mean night temperature. Similar variables were selected for the 48-h data

<table>
<thead>
<tr>
<th>Weather variable</th>
<th>24-h bioassay</th>
<th>48-h bioassay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>P-value</td>
</tr>
<tr>
<td>Daylight leaf wetness (h)</td>
<td>0.396</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Degree-hours daylight wetness</td>
<td>0.467</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Degree-hours humidity &gt;80%</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>Degree-hours humidity &gt;90%</td>
<td>0.466</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Leaf wetness (h)</td>
<td>0.289</td>
<td>0.001</td>
</tr>
<tr>
<td>Mean night temperature (°C)</td>
<td>0.118</td>
<td>0.191</td>
</tr>
<tr>
<td>Mean daily temperature (°C)</td>
<td>-0.039</td>
<td>0.663</td>
</tr>
<tr>
<td>Morning leaf wetness (h)</td>
<td>0.277</td>
<td>0.002</td>
</tr>
<tr>
<td>Rain (mm)</td>
<td>0.566</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Rain leaf wetness (h)</td>
<td>0.571</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Relative humidity &gt;80% (h)</td>
<td>0.285</td>
<td>0.001</td>
</tr>
<tr>
<td>Relative humidity &gt;90% (h)</td>
<td>0.191</td>
<td>0.033</td>
</tr>
</tbody>
</table>

| Spore density (sporangia M⁻³ h⁻¹) | 0.491 | <0.0001 | <0.0001 | 0.427 | 0.003 | 0.025 |

* K-S = Kolmogorov-Smirnov test that the weather variables on days without and with infection belong to the same distribution. Monte Carlo estimation was used to derive exact \( P \) values.
sets, with the addition of a product variable for mean night temperature and hours of relative humidity >80%. The weather variables identified as associated with infection by *P. humuli* generally are in agreement with previous studies conducted in the diverse climates of central Washington (13,16,17), England (32), and Germany (21), but with important distinctions. However, we made no attempt in this study to quantify the effect of weather factors on individual components of disease development, specifically sporulation, spore dispersal, germination, or infection. The models developed in the current study likely include variables that may affect all of these processes but the models only consider these factors implicitly and estimate an overall probability that conditions were suitable for some level of disease development.

Leaf wetness was identified as an important predictor of infection and infection severity, which was expected because moisture is required for infection by nearly all pathogenic oomycetes and is accounted for directly or indirectly in numerous models (14,19, 22,36). Weather variables that incorporated an explicit or surrogate measure of wetness (e.g., rain) were significantly associated

**Fig. 1.** Box plots of 24-h weather variables selected for discriminant function construction for classifying days as favorable (1) or unfavorable (0) for infection of hop plants by *Pseudoperonospora humuli*. Box plots show the median (line), middle 50% of the data (open box), confidence interval for the median based on the nonparametric sign-test (solid bar inside box) (33), extremes of the data points (whiskers), and outliers (solid circles). The probability level of the Kolmogorov-Smirnov test is presented numerically on the graphs. RH = relative humidity.

**Fig. 2.** Box plots of 48-h weather variables selected for discriminant function construction for classifying days as favorable or unfavorable for infection of hop plants by *Pseudoperonospora humuli*. Box plots show the median (line), middle 50% of the data (open box), confidence interval for the median based on the nonparametric sign-test (solid bar inside box) (35), extremes of the data points (whiskers), and outliers (solid circles). The probability level of the Kolmogorov-Smirnov test is presented numerically on the graphs. RH = relative humidity.
TABLE 2. Characteristics and performance of discriminant functions for classifying days as favorable or unfavorable for infection of hop by *Pseudoperonospora humuli*

<table>
<thead>
<tr>
<th>Model</th>
<th>Cases</th>
<th>Controls</th>
<th>AUROC</th>
<th>SE</th>
<th>$T^f$</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>24-h</td>
<td>18</td>
<td>107</td>
<td>0.88</td>
<td>0.053</td>
<td>0.49</td>
<td>0.86</td>
<td>0.83</td>
<td>0.89</td>
</tr>
<tr>
<td>48-h</td>
<td>16</td>
<td>32</td>
<td>0.86</td>
<td>0.063</td>
<td>0.39</td>
<td>0.86</td>
<td>0.88</td>
<td>0.84</td>
</tr>
</tbody>
</table>

- Independent quadratic discriminant functions were derived from predictors variables for 24 or 48-h periods.
- Number of data sets where infection did occur.
- Number of data sets where infection did not occur.
- Area under the receiver operating characteristic curve (AUROC).
- Standard error of the AUROC.
- Optimal cut-point identified by Youden’s index.
- Proportion of data sets classified correctly at the cut-point identified by Youden’s index.
- Proportion of cases classified correctly at the cut-point identified by Youden’s index.
- Proportion of controls classified correctly at the cut-point identified by Youden’s index.

The role of dew wetness in epidemic development is unresolved. The importance of dew in infection has been discounted in Europe and only infections following rain wetness were reported as being economically important (21,32), although length of dew duration after sunrise was not considered in these studies. Extended dew wetness, particularly during the morning, was implicated in infection in California (40) and for other downy mildew diseases (36). Royle (32) also stated that “relatively light” and “moderate levels of infection” of leaves in England can occur in association with dew wetness. In this study, we found that leaf wetness expressed in units of thermal time of total wetness duration, and not only wetness associated with rain events, was a better predictor of disease incidence and occurrence than rain wetness for the 48-h data sets. However, rain wetness was the most highly correlated variable with disease incidence in the 24-h data sets. Scherm and van Bruggen (37) demonstrated that lettuce may be infected by sporangia of *Bremia lactucae* produced that day during a morning with extended dew. Given the results of the current study, this may suggest that, for hop downy mildew, dew wetness may be more important for infection by sporangia that survive from the previous day than by sporangia that are produced and released during the dew wetness. In other studies, temperature has not been considered a limiting factor for the *P. humuli* infection process and downy mildew development (21,32,40). Night temperature and temperature during wetness clearly were important predictors in this study, similar to previous studies in Washington (16).

The importance of humidity (at night or otherwise) was an important factor associated with sporulation and infection of *P. humuli*, as reported in England (32), continental Europe (33), and the semiarid environment of central Washington (16). In the current study, model Se and accuracy were greater when using a humidity threshold of 80 instead of 90%, which has been suggested as the minimum threshold for sporulation by *P. humuli* (33). The weather measurements used in developing the models are proxies for the actual microclimate conditions on the leaf surface where the pathogen interacts with the plant. An ambient
relative humidity of 80 instead of 90% may be a better surrogate for conditions at the leaf surface favorable for sporulation and the infection process. Alternatively, the improved model Se and accuracy with the lower relative humidity threshold may be related to minimum temperature. Calculation of relative humidity is dependent on temperature, and humidity may be greater on days that temperature drops below the minimum temperature required for infection or sporulation. Nonetheless, the empirical humidity threshold selected appears to be a suitable approximation for prediction of conditions suitable for infection and can be recorded simply by an automated weather station.

The discriminant function models developed in this research purposefully were developed to be relatively simple and conservative. Leaf infection is known to occur at temperature thresholds and wetness durations lower than that of shoots (33), although leaf infections are of minimal economic consequence. Potentially, positive predictions of leaf infection could occur when conditions are unsuitable for shoot or cone infection. However, leaf and shoot infection are correlated and conditions favorable only for leaf infection likely would be limited because the degree-hour calculation was based on a lower temperature threshold of 8°C. Similarly, the models were developed for a single, highly susceptible cultivar and other cultivars may behave differently. Because the discriminant functions do not estimate the level of infection, we expect that only cultivars with exceptionally high or low susceptibility to downy mildew would respond markedly differently than cv. Nugget.

Additionally, the models do not provide a quantitative estimate of the level of infection but, rather, simply estimate the probability that infection has occurred at any level. Although this level of precision likely is not adequate to optimize fungicide use efficiency in all circumstances, we recognize that growers of high-value crops such as hops are risk averse (23,30) and, thus, may make management decisions that do not minimize long-term average costs. Optimum economic decision thresholds indicated that conservative management of the disease is cost effective, on average, with several economic assumptions, except for when disease prevalence is relatively low (i.e., 0 to 0.31, depending on the time of the season and model). These conditions would tend to occur during the relatively dry and warm conditions typical of western Oregon in late spring and summer. Indeed, no infections on the 24- or 48-h bioassay plants were observed between 14 July and 7 September. Given the conservativeness of the models, a negative disease prediction during this time could be helpful for informing a grower’s decision if continued fungicide applications for downy mildew are necessary.

Based on the economic assumptions used in this study, the value of the models in management decision making is expected to be greatest when disease prevalence is relatively low (e.g., during relatively dry weather). The economic value of using a predictive model for other high-value crops may also be restricted to periods when disease prevalence is relatively low because the economic cost of crop loss would tend to be much greater than the costs associated with an unnecessary fungicide application. As stated by Maloy (25) in reference to disease forecasting, “Many growers view routine (i.e., calendar) spraying as insurance—better to have it and not need it than the other way around.” The research presented here indicates that predictive systems for hop downy mildew need to be extremely accurate to reduce average management costs compared with routine fungicide applications because the cost of a false negative is much greater than a false positive.

---

**Fig. 4.** Observed and modeled sensitivity and specificity of discriminant function models for classifying the A, 48-h or B, 24-h period as favorable for infection by *Pseudoperonospora humuli* as a function of the optimum decision threshold, $p^*_s$. $p^*_s$ is defined as a threshold value such that if $p > p^*_s$, a fungicide treatment is recommended and if $p < p^*_s$, a fungicide treatment is not recommended, where $p$ is the probability that a fungicide application for downy mildew will provide a positive net return. Solid circles represent sensitivity and open circles represent specificity. **C.** Optimum decision threshold, $p^*_t$, in relation to disease prevalence and crop phenology. ‘Vegetative’ refers to infection by *P. humuli* during early- to midseason vegetative development, and ‘harvest’ refers to infection of cones near harvest. Economic assumptions of the models are described in the text.

---

**TABLE 3.** Parameter estimates for nonlinear regression models describing sensitivity and specificity for discriminant functions classifying a 24- or 48-h period as favorable for infection by *Pseudoperonospora humuli*.

<table>
<thead>
<tr>
<th>Model</th>
<th>$r_p$</th>
<th>$k_p$</th>
<th>$r_e$</th>
<th>$k_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>24-h</td>
<td>0.222</td>
<td>103.9</td>
<td>2.521</td>
<td>0.149</td>
</tr>
<tr>
<td>48-h</td>
<td>0.099</td>
<td>21,614.1</td>
<td>3.406</td>
<td>0.077</td>
</tr>
</tbody>
</table>

* Equations 2 and 3 in the text provide explicit formulas for sensitivity and specificity, respectively. Parameters estimates are provided so that fitted curves for sensitivity and specificity can be reconstructed by the reader.
The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable. Financial support was provided by United States Department of Agriculture–Agricultural Research Service (USDA-ARS) CRIS 5358-21000-035-00, Oregon State University, and USDA-CSREES Western IPM Center. We thank B. Turechek for assistance in developing the macro for equation 4; W. Mahaffee, B. Pfender, and B. Turechek for providing valuable comments during their critical reviews of an earlier draft of the manuscript; and N. Adair, K. King, K. Roseboro, and J. Woods for excellent technical support.

LITERATURE CITED