Forecasting Late Blight in Potato Crops of Southern Idaho Using Logistic Regression Analysis

Donna Henderson, University of Idaho, Aberdeen Research and Extension Center, Aberdeen 83210; **Christopher J. Williams,** University of Idaho, Department of Statistics, Moscow 83844; and **Jeffrey S. Miller,** University of Idaho, Aberdeen Research and Extension Center

ABSTRACT

Henderson, D., Williams, C. J., and Miller, J. S. 2007. Forecasting late blight in potato crops of southern Idaho using logistic regression analysis. Plant Dis. 91:951-956.

Previously published late blight forecasts which predict the threat of disease based on the presence or absence of favorable weather have not been effective in semi-arid potato-producing areas of the Pacific Northwest (Idaho, Oregon, and Washington). Research was conducted to identify weather variables useful for forecasting late blight in southern Idaho. The objectives of this research were to (i) determine if regional weather variables could be related to the occurrence of late blight in southern Idaho, (ii) determine if disease severity (scale of 0 to 4) could be predicted using variables found to be correlated with the annual occurrence of late blight, and (iii) validate the efficacy of this model in predicting disease incidence in regions of the Columbia Basin. Weather data were collected from five locations over a 9-year period (1995 to 2003). A binary logistic regression model (0 = no disease and 1 = disease) indicated that the number of hours with favorable conditions (10ºC ≤ temperature ≤ 27ºC, relative humidity ≥ 80%) in April and May (HF80m) was a significant disease predictor. Logistic regression analysis using an ordinal disease scale (0 = no disease and 4 = severe disease) indicated amount of precipitation (APj) and favorable hours (HF80j) with extended periods from April to June as significant disease predictors. The binary model predicted disease occurrence more accurately, with 67.5% accuracy (27/40 years correctly predicted), 75% sensitivity (12/16 late-blight years predicted), and 62.5% specificity (15/24 non-late-blight years predicted) using a leave-1-year-out error estimate. The binary model was validated with data (1995 to 2003) from the semi-arid Columbia Basin regions, predicting disease with 80.8% accuracy (21/26 years predicted), 84% sensitivity (21/25 outbreak years predicted), and 0% specificity (0/1 non-outbreak years predicted).

Additional keywords: *Phytophthora infestans,* semi-arid climate

Phytophthora infestans (Mont.) De Bary is an oomycete that infects potato crops during cool, wet weather, causing potato late blight. Late blight on potato crops is a concern to potato growers worldwide, and considered to be the most economically important disease to potato crops in North America (7). U.S. growers can experience annual losses of revenue of $210.7 million, $77.1 million being the cost of late blight fungicides alone (8). The pathogen spreads rapidly, is difficult to manage, and can cause massive losses in crop yields, both in the field and in storage.

Forecasting models that predict the likelihood of late blight outbreaks may provide important information for potato producers in southern Idaho, enabling farmers to implement a timely disease management plan. Practical benefits include an advanced warning system for southern Idaho potato farmers. In late-blight-favorable years, fungicide use may increase due to the added control gained by early forecasts. However, when weather is not favorable for late blight, growers could use the forecasts to reduce unnecessary fungicide applications, thereby reducing the monetary and environmental costs associated with traditional calendar-based spray programs.

Several models have been developed to predict the initiation of a late blight epidemic. van Everdingen (25) first proposed using a set of "Dutch rules" based upon presence of dew at night, night-time temperature, mean cloudiness, and rainfall. Late blight was predicted within a fortnight (14 days) of favorable conditions. This model has been modified since then in various ways. The Beaumont method implemented a temperature-humidity rule in lieu of the Dutch rules, where late blight was predicted after 2 days when the minimum temperature was not less than 10ºC and relative humidity (RH) did not drop below 75% (1). Cook (3) developed a simple graph based on daily average temperature and rainfall beginning at the start of the growing season. Growers were advised to spray after the accumulation of rainfall

had passed the "critical rainfall line," a threshold based on the accumulated daily average rainfall and temperatures during the growing season. Efficacy of this model was impaired when above-average or below-average amounts of rainfall occurred early in the growing season (19). Hyre (11) modified Cook's system, classifying days as being "favorable" or "unfavorable." Days were considered favorable when the mean temperature of the previous 5 days was less than 25.6ºC and the cumulative precipitation for the 10 previous days was 3 cm. Late blight was predicted 2 weeks after the accumulation of 10 favorable days. Wallin (26) developed a model based upon the Beaumont rules, where severity values were assigned based on temperature ranges and high RH values. Late blight occurrence first was predicted 7 to 14 days after 18 to 20 severity values had accumulated. BLITECAST was developed as a computerized synthesis of both the Hyre and Wallin models and has been used successfully in many potato-producing regions of North America (17). Regions able to successfully use BLITECAST are more humid, experience more rainfall, and have frequent yearly late blight occurrence compared with southern Idaho. Consequently, using a disease forecast developed for humid regions with regular yearly disease occurrence, such as BLITECAST and other models based upon high humidity (1,2,4–6,9,15,17,19,21–23,25), has not been effective in southern Idaho (*unpublished data*). In southern Idaho, disease inoculum may not always be present and the low humidity and rainfall are unfavorable for disease development.

Given the sporadic late blight occurrence in southern Idaho, development of a qualitative model that predicts yearly presence or absence of late blight may be more useful. Qualitative prediction models have been developed in the Columbia Basin of Washington in the United States (13,14) and in the Netherlands (27) using discriminant and logistic regression to identify relationships between regional weather data variables and disease occurrence on a binary or ordinal response scale.

The Columbia Basin model was constructed for the south-central area of Washington using linear discriminant and logistic regression analyses that described the relationship between weather data collected near Prosser, WA, with prior late

Corresponding author: J. S. Miller
E-mail: jeff@millerresearch.com

blight outbreaks spanning a 25-year period (14). The linear discriminant model used the number of rainy days in April and May (Ram), the occurrence of late blight the preceding year (Yp), and the number of rainy days in July and August (Rja) as well as the total precipitation in May when the daily minimum temperature was ≥5ºC (Pm) in Washington. Additional models were constructed subsequently to include three specific geographic regions of the Columbia Basin (Hermiston, OR; Othello, WA; and Hanford, WA) using logistic regression analyses (13).

The Netherlands model evaluated meteorological data that correspond to categories of disease development (27). Meteorological variables were divided into four categories: A, variables affecting the formation of inoculum in the preceding year; B, variables affecting the over-wintering of inoculum and development of disease on refuse piles; and C, variables limiting or D, enhancing disease development during the growing season (Table 1). Linear discriminant analysis of data was used to identify the variables that were useful in classifying years as late-blight or non-late-blight years. The variables identified as significant in classifying 40 of 47 blight years included: blight status the previous year, number of days with precipitation (DPg), the number of hours with temperatures between 10 and 27ºC and RH above 90% during the growing season (HFg), amount of global radiation (GRg), and hours with temperatures above 27ºC (27).

A model that can be used specifically in the semi-arid climate of southern Idaho has not been developed. Because weather during the potato-growing season of southern Idaho and yearly disease occurrence is different from that of the Columbia Basin or the Netherlands, weather variables that were significant in the Columbia Basin (13,14) and Netherlands (27) models may not be useful for annual forecasting in southern Idaho. The Columbia Basin model (13,14) and the Netherlands model (27) used qualitative analysis of late blight occurrence and weather in specific regions in construction of their models. This approach also may be useful in predicting late blight in regions of southern Idaho that experience a semi-arid climate and sporadic late blight occurrence.

The objectives of this research were to (i) determine if regional weather variables could be related to the occurrence of late blight in southern Idaho, (ii) determine if disease severity (scale of 0 to 4) could be predicted using variables found to be correlated with the annual occurrence of late blight, and (iii) validate the efficacy of this model in predicting disease incidence in regions of the Columbia Basin.

MATERIALS AND METHODS
Weather data were collected from Agrimet Bureau of Reclamation Pacific Northwest Region weather archives (24). Data were collected from the archives of five Agrimet weather station locations across southern Idaho: Aberdeen, Parma, Rupert, Rexburg, and Twin Falls. These five locations are located in the major commercial potato-producing regions of southern Idaho (Fig. 1). Late blight severity for each location was derived from disease scouting notes and interviews with University of Idaho Extension specialists and private consultants who have responsibilities for late blight identification and disease management recommendations. An estimate of the number of fields affected by late blight and the general severity of late blight in those fields was used to assign a disease severity number of 0 to 4 using the scale of Zwankhuizen and Zadoks (27), where 0 = no late blight observed or reported; 1 = late blight reported in a very limited number of fields in only part of the region with, on average, very low disease levels per field; 2 = late blight reported in most of the region in a limited number of fields with, on average, low or intermediate disease levels per field; 3 = late blight reported in many fields in all of the region with, on average, intermediate disease levels per field; and 4 = late blight reported in most fields in all of the region with, on average, high disease levels per field (27).

Weather data from 1995 to 2003 were used to construct 12 weather variables (Table 1). Weather variables were divided into four categories: A, factors affecting the formation of inoculum during the previous harvest; B, factors affecting over-wintering of inoculum; C, factors favoring disease development during the growing season; and D, factors limiting disease development during the growing season (27). Information about the weather sensors and methods of use by Agrimet Bureau of Reclamation Pacific Northwest Region can be found online (24).

The variables constructed using categories C and D (current growing season) were for the period from April to June, in order to create a model that would give predictions of disease occurrence in a timely manner. Late blight has never been found in southern Idaho before 5 July (P. Nolte, *personal communication*). However, in the event that late blight may occur prior to the beginning of July, an additional set of variables were created using only data collected from April to May. All 12 variables for every location–year combination were organized with the late blight severity ratings in Table 2 and analyzed using binary and ordinal logistic regression analysis using SAS PROC LOGISTIC, (SAS 8.2; SAS Institute, Cary, NC).

Binary logistic regression was used to create a dichotomous model that would predict either the presence (=1) or absence (=0) of late blight. Additionally, ordinal logistic regression was used to develop a model to predict the severity of late blight

Table 1. Weather variables quantified from 1995 to 2003 for five locations in southern Idaho for construction of logistic regression models to predict the occurrence and severity of late blight epidemics[a]

Category, variable (unit)	Description
A	
DPh (days)	Number of days with precipitation (>0.25 mm) during September and October
APh (mm)	Amount of precipitation during September and October
DPY (no unit)	Disease presence the previous year, (0 or 1)
B	
APw (mm)	Amount of precipitation in winter (November to April)
HL (no unit)	Hellmann value, sum of the average daily temperatures <0ºC from November to April
FDH (hours)	Hours with temperature ≤–2ºC at 5 cm below soil level, November to April
TS (days)	Temperature sum of average daily temperature >0ºC at 5 cm below soil level, November to April
C	
DPm, DPj (days)	Number of days with precipitation, April to May or April to June
APm, APj (mm)	Amount of precipitation, April to May or April to June
FH80m, FH80j (hours)	Number of favorable hours (10ºC ≤ temperature ≤ 27ºC, relative humidity ≥80%), April to May or April to June
D	
GRm, GRj (J cm^{-2})	Total global radiation, April to May or April to June, over total period
UHm, UHj (hours)	Number of hours with unfavorable conditions (temperature ≥27ºC), April to May or April to June

[a] Variables adapted from Johnson et al. (13,14), and Zwankhuizen and Zadoks (27). Meteorological variables were divided into four categories: A, variables affecting the formation of inoculum in the preceding year; B, variables affecting the over-wintering of inoculum and development of disease on refuse piles; and C, variables limiting or D, enhancing disease development during the growing season.

epidemics on a 0-to-4 scale, where 0 represented no disease and 4 represented severe disease. Logistic regression was chosen over discriminant analysis due to the fact that logistic regression does not assume that multivariate observations are distributed normally (12). The correlation matrix of the explanatory variables was examined and a principal component analysis was performed to rule out collinearity between the variables. Backward selection was used in the logistic regression analyses to select a subset of variables for both the binary and ordinal models. The significance level for dropping variables in the backward selection procedure was 0.05.

The binary logistic regression model had one intercept in the model equation. The ordinal model, which used a disease measure with five levels of disease, had four intercepts. Predictions from either the binary or ordinal model are performed by first calculating estimated logit scores (lf); then, the logit scores are converted into estimated probabilities, using equation 1 (below). For either model, predictions were calculated for a location–year combination of weather data (45 observations):

Probability(x) = exp(lf)/[1 + exp (lf)]   (eq. 1)

$lf = \beta'x$, where β = the estimated coefficients and x = the set of covariate values.

For the binary model, equation 1 gives the probability of disease occurring given a set of covariate values. For the ordinal model, cumulative logit values were calculated for each level of disease severity, yielding four logit scores (lf_1 to lf_4). Data from significant variables were input into the four cumulative logit equations, resulting in four lf values for every observation representing disease levels 1 to 4. The four model logit function equations represented the following: lf_1 = the logit for the sum of probabilities for disease levels 1 to 4; lf_2 = the logit for the sum of the probabilities for disease levels 2 to 4; lf_3 = the logit for the sum of the probabilities for disease levels 3 and 4; and lf_4 = the logit for the probability of disease level 4. The calculation of probability for disease level 0 in the ordinal model was simply one minus the probability of disease levels 1 to 4.

Each of the four cumulative logit ($lf_{1–4}$) values for each of the 45 observations then was converted into a cumulative probability value (equation 1), with the results denoted by P(1), P(2), P(3), and P(4). The individual probability values for the predicted disease levels then were calculated with the following formulas:

P′(0) = 1 – [P(1) + P(2) + P(3) + P(4)]   (eq. 2)
P′(1) = P(1) – [P(2) + P(3) + P(4)]   (eq. 3)
P′(2) = P(2) – [P(3) + P(4)]   (eq. 4)
P′(3) = P(3) – [P(4)]   (eq. 5)
P′(4) = P(4)   (eq. 6)

The appropriateness for using the binary model was assessed by checking the influence of each data point on the model coef-

ficients, as well as by performing a Hosmer-Lemeshow goodness-of-fit test. Individual data point influence was assessed by DFBETA values, which measure the amount that model coefficients change

when a data point is deleted, and by hat matrix values, which measure the degree to which a data points' covariate values are in an unusual part of the covariate space which may allow them to distort the pa-

Fig. 1. Location of weather stations (Aberdeen, Parma, Rexburg, Rupert, and Twin Falls) across southern Idaho used for collection of weather data for logistic regression modeling of late blight occurrence from 1995 to 2003.

Table 2. Severity of late blight epidemics in five locations in southern Idaho from 1995 to 2003[a]

Year	Parma	Twin Falls	Rupert	Aberdeen	Rexburg
1995	3	2	2	0	0
1996	2	1	1	0	0
1997	1	4	4	4	0
1998	1	4	4	4	3
1999	0	0	0	0	2
2000	0	0	0	0	0
2001	0	1	1	0	0
2002	0	0	0	0	0
2003	1	0	0	1	0

[a] Severity was estimated for each location using a 0-to-4 scale adapted from Zwankhuizen and Zadoks (27) where 0 = no late blight observed or reported; 1 = late blight reported in a very limited number of fields in only part of the region with, on average, very low disease levels per field; 2 = late blight reported in most of the region in a limited number of fields with, on average, low or intermediate disease levels per field; 3 = Late blight reported in many fields in all of the region with, on average, intermediate disease levels per field; and 4 = late blight reported in most fields in all of the region with, on average, high disease levels per field.

rameter estimates of the model (10). The Hosmer-Lemeshow goodness-of-fit test is conducted by dividing the observations into groups of predicted probabilities, and comparing observed with expected counts of disease occurrence in these categories (10).

For the binary model, disease was predicted to occur when the predicted probability exceeded a prespecified cutoff value of 0.3 (discussed below). For the ordinal model, the probability function that gave the highest probability value was considered to be the predicted disease intensity level for that year. Sensitivity was favored over specificity because it is considered more beneficial to overpredict disease rather than underpredict disease in regions where late blight does not occur regularly (13). Therefore, models were adjusted to overpredict outbreak years by redefining the prediction cutoff value from 0.5 to 0.3. The cutoff value of 0.3 was chosen based upon the ability of the model to predict with higher sensitivity without compromising specificity. Thus, if the predicted probability was ≥0.3, observations were classified as disease events (disease years), and if the predicted probability was <0.3, observations were classified as nonevents (no disease years).

Performance of both models was evaluated for accuracy (percentage of years correctly identified), sensitivity (percentage of outbreak years correctly identified, DI = 1), and specificity (percentage of nonoutbreak years correctly identified, DI = 0). Predictions from the binary model were evaluated via three different types of cross-validation. The first method for estimating cross-validation errors was produced by SAS PROC LOGISTIC, which calculated a predicted probability of disease for the current observation based on a model that is fitted without the current observation. The second method used was a leaving one out-cross validation method, where error rate estimates were calculated with year as the data left out (for each year, a model was fitted without that year,

and then predictions for the left-out year were compared with the actual disease incidence for that year). These error estimates for each year then were pooled to obtain one set of error estimates. The third method for estimating error rates used external data validation, using an independent data set from three regions across the Columbia Basin (Hermiston, OR; Prosser, WA; and Othello, WA) from 1995 to 2003. Data were collected from Agrimet (Hermiston, OR) and the Washington State University Public Agricultural Weather System for Othello and Prosser, WA. Disease incidence data were obtained for each year and region from Dr. Dennis Johnson (Washington State University). In each case, data were input into the binary model and classified as no disease or disease years based upon the 0.3 probability of disease used in the binary model self-validation tests. For the ordinal model, a proportional-odds test was conducted to assess whether evidence existed of unequal slopes for the four model logit equations, and resubstitution predictions were used to estimate prediction errors. These resubstitution error rates are optimistically biased, but were used to compare predictions between the ordinal model and the binary model.

RESULTS

Binary model. Of the 12 weather variables examined from potato-producing regions across southern Idaho, two were significant in predicting disease occurrence in the logistic regression model: the amount of precipitation (mm) during April and May (APm) ($P = 0.0389$) and number of favorable hours (10ºC ≤ temperature [T] ≤ 27°C, RH ≥ 80%) in April and May (HF80m; Table 3) ($P = 0.0254$). The resulting binary model had a logit score of:

$$lf = -2.9783 + 0.0243 \text{ (HF80m)} + 0.2754 \text{ (APm)} \quad \text{(eq. 7)}$$

Binary model assessment. For the model shown in equation 7, a few observa-

tions had moderately high DFBETA and hat values, and the Hosmer-Lemeshow test had a P value of 0.087, indicating some lack of fit. To further investigate the dependence of the model on influential data points, the five observations with the largest hat values were deleted and the logistic regression analysis with backward selection was repeated. In this analysis, only the number of favorable hours remained significant, yielding a revised model:

$$lf = -2.014 + 0.0352 \text{ (HF80m)} \quad \text{(eq. 8)}$$

This model had acceptable DFBETA and hat values, the intercept and coefficient for HF80m were similar to those in the original model, and the Hosmer-Lemeshow test indicated a better fit to the data ($P = 0.37$); therefore, this model was adopted as the final binary model.

Binary model validation. The three types of cross-validation were applied to the model in equation 8 and are reported here based on the probability cutoff of 0.3. The reduced data set without the five influential data points was used for the first two sets of estimates. The SAS cross-validation error estimates for the model with HF80m was 72.5% accurate (29/40 years of disease incidence correctly predicted), with sensitivity of 75% (12/16 disease presence years correctly predicted) and specificity of 70.8% (17/24 disease absence years correctly predicted). The leave-1-year-out method yielded an estimated accuracy of 67.5% (27/40), sensitivity of 75% (12/16), and specificity of 62.5% (15/24). The error estimates based on external validation, using disease occurrence in regions of the Columbia Basin from 1995 to 2003, gave estimates of 80.8% accuracy (21/26 years of disease incidence correctly predicted), 84% sensitivity (21/25 disease presence years correctly predicted), and 0% specificity (0/1 disease absence years correctly predicted).

Ordinal model. The ordinal model was selected based on the results of the backward selection procedure. The ordinal logistic regression model developed to predict disease severity (0 to 4) using the April to June data set had the following variables: amount of precipitation (APj), and the number of favorable hours (10ºC ≤ T ≤ 27°C, RH ≥ 80%; HF80j). The ordinal model includes four equations for predicting disease levels (0 to 4), constructed using coefficient estimates of variables and intercept values found to be significant in the model (Table 4):

$$lf_1 = -2.9716 + \text{(HF80j) } 0.0123 + \text{(APj) } 0.3973 \quad \text{(eq. 9)}$$

$$lf_2 = -4.1004 + \text{(HF80j) } 0.0123 + \text{(APj) } 0.3973 \quad \text{(eq. 10)}$$

$$lf_3 = -4.8329 + \text{(HF80j) } 0.0123 + \text{(APj) } 0.3973 \quad \text{(eq. 11)}$$

$$lf_4 = -5.2688 + \text{(HF80j) } 0.0123 + \text{(APj) } 0.3973 \quad \text{(eq. 12)}$$

Table 3. Coefficient estimates for predicting the occurrence of potato late blight in southern Idaho from the final binary logistic regression model

Parameter	df	Estimate	Standard error	Walden's χ^2	$P > \chi^2$
Intercept	1	–2.0139	0.6812	8.74	0.0031
HF80m[a]	1	0.0352	0.0128	7.60	0.0058

[a] Number of favorable hours (10ºC ≤ temperature ≤ 27°C, relative humidity ≥80%), April to May.

Table 4. Coefficient estimates for variables identified by an ordinal logistic regression analysis for predicting the level of yearly disease occurrence of potato late blight in southern Idaho

Parameter	df	Estimate	Standard error	Walden's χ^2	$P > \chi^2$
Intercept 4	1	–5.27	1.08	23.72	<0.0001
Intercept 3	1	–4.83	1.03	21.63	<0.0001
Intercept 2	1	–4.1	0.96	17.92	<0.0001
Intercept 1	1	–2.97	0.85	12.04	0.0005
HF80j[a]	1	0.012	0.01	7.80	0.0052
APj[b]	1	0.397	0.16	5.76	0.0163

[a] Number of favorable hours (10ºC ≤ temperature ≤ 27°C, relative humidity ≥80%), April to May.
[b] Amount of precipitation (mm) from April to May.

Although the variables HF80j and APj were significant for the ordinal model, the proportional odds test was significant ($P = 0.043$), indicating that the coefficients for HF80j and APj should vary across logit equations, and the predictions did not improve upon those found with the binary model. The optimistically biased resubstitution error estimates showed some promise for ruling out disease (22 of 25 disease-absent years correctly predicted), but not good performance for predicting disease presence, because only 8 of the 20 years that were positive (disease levels 1 to 4) were predicted to have a positive disease level.

DISCUSSION

The first objective to determine if regional weather variables could be related to the occurrence of late blight in southern Idaho was accomplished by using logistic regression analysis to determine the relationship between weather variables and late blight incidence for five locations in southern Idaho over a 9-year period. Of the 12 variables, 1 was found to relate to late blight occurrence in southern Idaho. This binary model identified the hours of combined occurrence of favorable temperature ($10°C \leq T \leq 27°C$) and humidity (RH > 80%) from April to May (HF80m) as significant predictors of disease occurrence. In the work of Zwankhuizen and Zadoks (27), the 90% RH used to define favorable hours was ineffective for this study and was lowered to an 80% RH level to define favorable periods, similar to the work reported by Gudmestad et al. (6). The number of days with precipitation was not as important as the amount of precipitation, which is different from the results of Johnson (13,14) and Zwankhuizen and Zadoks (27). The Netherlands model identified the number of days with precipitation in September and the total global radiation in September as being important in classifying outbreaks (27). We chose to limit our investigations to periods of time that would be useful in making forecasts which did not include data from September, a period of time that occurs late in the growing season of southern Idaho. Other forecasts that have been used to predict the occurrence of late blight have used periods of high relative humidity (1,4,15,23,26,27). Even though the early Dutch rules didn't directly measure periods of high humidity (25), the four parameters used to identify blight-favorable weather corresponded with conditions of high humidity. Rainfall also has been an important factor in late blight prediction models (4,9,13,14,19,23, 25,27).

The binary model (0 = no outbreak, 1 = late blight outbreak) was tailored to liberally predict disease outbreaks by specifying a predicted probability cut-off value of 0.30. This was done to give a more liberal prediction of disease in an attempt to minimize the probability of not predicting late blight when late blight occurs. Sensitivity (75% accurate in correctly classifying outbreak years) of the binary model was higher than specificity (62.5% accurate in correctly classifying nonoutbreak years), using the leave-1-year-out estimates. Poor model specificity was not problematic from a practical point of view in southern Idaho. Due to the hot, dry climate of the area and the sporadic nature of late blight over the last 10 years, many potato producers in southern Idaho enter each growing season with the default assumption that late blight won't occur. Under these circumstances, a model was not needed to help potato producers decide not to apply fungicides for late blight management. Growers have found that few (two to three) fungicide applications are required in non-late-blight years to manage potato early blight. Due to the decreasing prices being paid for potato, growers typically desire minimum fungicide input during a growing season. The real benefit of a model for use in southern Idaho would be to give potato growers a warning when the probability for late blight is high so that additional fungicide applications could be made early in the season prior to initial late blight identification.

The second objective of the study was to determine if disease severity could be estimated using variables found to be correlated with the annual occurrence of late blight. This objective also was realized using logistic linear regression (ordinal model). Similar to the binary model, the ordinal model found amount of rain (APj) and the number of favorable hours ($10°C \leq T \leq 27°C$, RH $\geq$ 80%; HF80j) during the early growing season to be significant variables in predicting late blight severity. However, lower sensitivity than in the binary model was a result of underpredicting outbreak years. This model was unable to effectively estimate the correct severity of disease, and was not useful for predicting disease severity on an ordinal disease index. The ordinal model had more parameters that had to be accounted for and less data were available (outbreak years were spread into four categories). Additionally, a higher proportion of years were nonoutbreak years (25/45) than outbreak years (20/45), making it harder to construct a model tailored to predict outbreak years.

Factors other than weather variables have the potential to contribute to the severity of late blight in a given year, contributing to reduced accuracy of the binary and ordinal models. Potato growers in southern Idaho in 1997 were not prepared for the severe epidemic based on minimal historical experience with late blight. As a result, fungicide applications were applied later than would be deemed acceptable and late blight was very severe. Also, the late blight pathogen could migrate from areas where late blight establishment was favored by early-season weather to areas where early-season weather was unfavorable. Almost all potato plants in southern Idaho are grown under irrigation, and these conditions can favor pathogen infection during the growing season. Under these conditions, the ordinal model would predict an absence of late blight in the unfavorable area, but late blight would have been reported. Conversely, weather conducive to late blight in one area may motivate growers to apply fungicides in a timely manner, increasing the protection of the crop. These reasons may provide some explanation as to the relatively low sensitivity of the ordinal model.

The third objective was to test our model on an independent data set of the Columbia Basin region. The binary model was more successful than the ordinal model in self-validation tests and, thus, was chosen for testing of the independent data set. The binary model was validated successfully using an independent data set from three regions across the Columbia Basin region of Washington (Prosser and Othello) and Oregon (Hermiston) from 1995 to 2003. The successful use of our model in the Columbia Basin may have been due to the similarities between southern Idaho and the Columbia Basin weather and climate. Additionally, the Columbia Basin area has had a long history of late blight occurrences and, thus, was a useful area for testing the efficacy of our binary model because we could compare disease incidence with predicted disease incidence. The binary model was extremely accurate (80.8%) and had a high sensitivity (84%), but was unable to predict 1 year of no late blight occurrence (0% specificity). Although there are similarities between the weather and climate of the Columbia Basin and southern Idaho, the variables significant for disease forecasting were dissimilar between the regions. The binary model was successful at forecasting late blight in southern Idaho when the weather was favorable according to the variable HF80m ($10°C \leq T \leq 27°C$), which may reflect differences in relative humidity between the two production areas. In the Columbia Basin model, the number of days of rainfall may correspond to field conditions favorable for disease development such as cool temperature and high humidity, but this variable was not significant in predicting disease for southern Idaho. Conditions in the field may be similar between the two regions, but the difference in variables may reflect the unique environments of southern Idaho and the Columbia Basin. Thus, from this external data validation analysis, we can surmise that the binary model was successful in predicting the occurrence of a late blight year in semi-arid climates.

Evaluating weather data from April to June with combined forecasts both from the highly sensitive binary model and the

highly specific ordinal model together may provide a useful tool for more accurate late blight prediction. Binary model output would be available at the end of May. If the model predicted late blight to occur, potato producers could feel somewhat confident that late blight will be present due to the high sensitivity of the binary model. If the ordinal model predicted an absence of late blight, growers could assume that the default assumption of no late blight was true based on the high specificity of the ordinal model. Fortunately, in only 3 of the 45 observations (7%) did late blight occur when both forecasts predicted a "0" year.

Qualitative analysis of disease occurrence in the Netherlands (27) and Columbia Basin model (13,14) were successfully constructed using large amounts of past weather data, usually in the 25- to 50-year range for each region studied. In the qualitative analysis of southern Idaho disease occurrence, there was less weather data available to analyze. Subsequently, a larger geographic region was used for sampling data. Specifically, weather data were collected from five locations across southern Idaho (Fig. 1). Late blight first was identified in southern Idaho in 1995 (18), limiting the number of years available for analysis. Additionally, *P. infestans* is a relatively new pathogen in southern Idaho, with sporadic occurrences. Therefore, it will be useful to continue collecting weather data for further analysis of the model so that this model can be refined and improved.

When using weather data to construct a predictive model, it is necessary to consider the window of time in which a forecast will be useful. Subsequently, weather data that might extend past a window of time from which a prediction can be made and action can be taken was not used. Consequently, this model does not represent a complete description of the pathogen life cycle, but does give a reliable prediction in a time frame useful for potato producers. In both models, only two variables from category C (factors affecting formation of inoculum early in the growing season) were significant in predicting disease: rain and favorable temperature and humidity. During this time frame, potato seed pieces are cut, planted, and emerge as young seedlings. During the cutting process, healthy seed tubers can be contaminated with *P. infestans* sporangia produced by blighted tubers (16). Developing sprouts may become infected even when seed tubers have been treated with protectant seed piece fungicides (20). Once above ground, growing plants also could become infected by sporangia produced on infected volunteer potato plants or from

potato plants growing in cull piles. Free moisture in the form of rain may create environmental conditions favorable for sprout infection either in the soil or after emergence. Precipitation and periods of high humidity and moderate temperatures early in the season would favor these events, and were the variables most closely related to the occurrence of late blight in southern Idaho.

BLITECAST assumes that inoculum is present every year and that late blight epidemics will begin at a very low intensity (15). Although these assumptions may hold true in southern Idaho, weather favorable for the establishment of primary inoculum may not always be present. Consequently, semi-arid regions such as southern Idaho have been in need of a qualitative yearly predictive model to give a preliminary prediction of the probability of late blight development before going into the growing season. If a yearly forecast predicts a given year to have a high probability of late blight development, then a second forecast such as BLITE-CAST could be deployed as a predictive model for scheduling of fungicide sprays during the growing season. Further research into this possibility is currently underway.

LITERATURE CITED
1. Beaumont, A. 1947. The dependence on the weather of the dates of outbreak of potato late blight epidemics. Trans. Br. Mycol. Soc. 31:45-53.
2. Bruhn, J. A., and Fry, W. E. 1981. Analysis of potato late blight epidemiology by simulation modelling. Phytopathology 71:597-601.
3. Cook, H. T. 1949. Forecasting late blight epiphytotics of potatoes and tomatoes. J. Agric. Res. 78:545-563.
4. Fry, W. E., Apple, A. E., and Bruhn, J. A. 1983. Evaluation of potato late blight forecasts modified to incorporate host resistance and fungicide weathering. Phytopathology 73:1054-1059.
5. Grünwald, N. J, Rubio-Covarrubias, O., and Fry, W. E. 2000. Potato late blight management in the Toluca Valley: Forecasts and resistant cultivars. Plant Dis. 84(4):410-416.
6. Gudmestad, N. C., J. W. Enz, D. A. Preston, and Secor, G. A. 1995. Late blight forecasting and dissemination system using an automated weather monitoring network. Pages 209-213 in: Phytophthora 150. L. J. Donley, ed. Boole Press, Ltd., Dublin.
7. Guenther, J. F., Michael, K. C., and Nolte, P. 2001. The economic impact of potato late blight on U.S. growers. Potato Res. 44:121-125.
8. Guenther, J. F., Weise, M. V., Pavlista, A. D.,

Sieczka, J. B., and Wyman, J. 1999. Assessment of pesticide use in the U.S. potato industry. Am. J. Potato Res. 76:25-29.
9. Hansen J. G., Andersson, B., and Hermansen, A. 1995. NEGFRY—A system for scheduling chemical control of late blight in potatoes. Pages 201-208 in: Phytophthora 150. L. J. Donley, ed. Boole Press, Ltd. Dublin..
10. Hosmer, D. W., and Lemeshow, S. 1989. Applied Logistic Regression. John Wiley & Sons, New York.
11. Hyre, R. A. 1954. Progress in forecasting late blight of potato and tomato. Plant Dis. Rep. 38:245-253.
12. Johnson, D. E. 1998. Applied Multivariate Methods for Data Analysts. Duxbury Press, Pacific Grove, CA.
13. Johnson, D. A., Alldredge, J. R., and Hamm, P. B. 1998. Expansion of potato late blight forecasting models for the semiarid environment of south-central Washington. Plant Dis. 82:642-645.
14. Johnson, D. A., Alldredge, J. R., and Vakoch, D. L. 1996. Potato late blight forecasting models for the semiarid environment of south-central Washington. Phytopathology 86:480-484.
15. Krause, R. A., Massie, L. B., and Hyre, R. A. 1975. BLITECAST: A computerized forecast of potato late blight (*Phytophthora infestans*). Plant Dis. Rep. 59:95-98.
16. Lambert, D.H., Currier, A.I., and Alanya, M., 1998. Transmission of *Phytophthora infestans* from infected seed potato tubers to developing sprouts. Am. Potato J. 73:370-371.
17. Mackenzie, D. R. 1981. Scheduling fungicide applications for potato late blight with BLITECAST. Plant Dis. 65(5):394-399.
18. Nolte, P., Forster, Mohan, S. K., and Olsen, N. L. 2001. Potato late blight in Idaho from 1995-1999. (Abstr.) Phytopathology 91:S188.
19. Nugent, T. J. 1950. Three years' experience forecasting late blight in Tidewater Virginia. Plant Dis. Rep. Suppl. 190:9-13.
20. Powelson, M. L., and Inglis, D. A., 1999. Foliar fungicides as protective seed piece treatments for management of late blight of potatoes. Plant Dis. 83:265-268.
21. Raposo, R., Wilks, D. S., and Fry, W. E. 1993. Evaluation of potato late blight forecasts modified to include weather forecasts: a simulation analysis. Phytopathology 1:103-108.
22. Stevenson, W. R. 1993. IPM for potatoes: a multifaceted approach to disease management and information delivery. Plant Dis. 77:309-311.
23. Ullrich, J., and Schrodter, H. 1966. Das problem der vorhersage des aufretens der kartoffelkrautfaule *(Phytophthorainfestans)* und die moglichkeit seiner losung durch eine negativprognose. Nachrichtenbl. Dtsch. Pflanzenschutzdienstes (Braunschweig) 18:33-40.
24. U.S. Department of the Interior: Bureau of Reclamation. Pacific Northwest Region. Agrimet—The Pacific Northwest Cooperative Agricultural Weather Network. Online, accessed 2002-2003.
25. van Everdingen, E. 1926. Het verband tusschen de weersgesteldheid en de aardappelziekte (*P. infestans*). Tijdschr. Plantenziekten 32:129-140.
26. Wallin, J. R. 1962. Summary of recent progress in predicting late blight epidemics in United States and Canada. Am. Potato J. 39:306-312.
27. Zwankhuizen, M. J., and Zadoks, J. C. 2002. *Phytophthora infestans*'s 10-year truce with Holland: a long-term analysis of potato late-blight epidemics in the Netherlands. Plant Pathol. 51:413-423.