Expansion of Potato Late Blight Forecasting Models for the Columbia Basin of Washington and Oregon

Dennis A. Johnson, Plant Pathologist, Department of Plant Pathology, Washington State University, P.O. Box 646430, Pullman 99164-6430; **J. Richard Alldredge,** Associate Professor, Program in Statistics, Washington State University; and **Philip B. Hamm,** Extension Plant Pathologist, Department of Botany and Plant Pathology, Hermiston Agricultural Research and Extension Center, Oregon State University, Hermiston 97838

ABSTRACT

Johnson, D. A., Alldredge, J. R., and Hamm, P. B. 1998. Expansion of potato late blight forecasting models for the Columbia Basin of Washington and Oregon. Plant Dis. 82:642-645.

A regional potato late blight forecasting system for irrigated potatoes in the semiarid environment of the Columbia Basin was expanded by developing specific forecasting models for four vicinities throughout the Basin. Relationships between weather and outbreaks of late blight at the locations over a 27-year period were examined using logistic regression analysis. The response variable was a year either with or without a late blight outbreak. An indicator variable representing the occurrence of an outbreak during the preceding year (*Yp*) and number of days of rain during April and May (*Ram*) correctly classified the disease status (presence or absence of late blight) of 89, 82, 78, and 78% of the years at Prosser, Washington, Hermiston, Oregon, and Hanford and Othello, Washington, respectively. The percentage of years with disease outbreaks correctly classified was 93, 85, 79, and 79% at the four respective locations. All years with late blight outbreaks and 96% of the total years were correctly classified using data from at least one of the four locations. These predictors are particularly important early in the growing season and can be used to make area forecasts. A second set of predictors, *Yp* and number of days of rain in July and August (R*ja*), for Hermiston and Hanford, and a third set, *Yp*, R*am*, and R*ja*, for Prosser and Othello were found effective for making additional late blight forecasts later in the growing season.

Outbreaks of late blight on potato, *Solanum tuberosum L.*, caused by *Phytophthora infestans* (Mont.) de Bary, occurred during 1990 through 1997 in the semiarid environment of the Columbia Basin of Washington and Oregon. This region extends approximately 195 km south to north and 80 km west to east and is cropped to approximately 65,000 ha of potatoes annually. Late blight was first identified in this region during the 1947 growing season when weather was unusually cool, cloudy, and wet. It was next reported 27 years later in 1974 and was observed in fields 7 of 16 years between 1974 and 1989. The monetary cost of managing late blight is now high and approached $30 million in the Columbia Basin in 1995 (12).

Several factors have contributed to the recent outbreaks of late blight in the Columbia Basin. These include wetter than normal weather (11), an increased proportion of the *P. infestans* population insensitive to metalaxyl and more aggressive on potato (5,15,16), and an increased production of potato cultivars extremely susceptible to *P. infestans* (9).

Corresponding author: D. A. Johnson
E-mail:DJOHNSN@wsu.edu

Accepted for publication 14 March 1998.

Publication no. D-1998-0414-01R

Management of late blight has been augmented in several regions of North America and Europe by scheduling fungicide applications using predictive models (1,4,8,13,22). Linear discriminant and logistic regression analyses have been used to develop two regional forecasting models for the Columbia Basin of Washington (11). The first model identifies the risk of late blight early in the growing season, and the second gives the probability of disease occurrence in midseason. Information from the models can be used either independently or in concert with weather forecasts to initiate disease management practices. Weather data from Prosser, Washington, was used in making the original regional late blight models. Weather conditions, especially rainfall, can vary within the region. More accurate late blight forecasts could possibly be made if weather data were collected and considered from several locations throughout the Columbia Basin.

The purpose of this study was to develop late blight forecasting models using past weather data from Hermiston, Oregon, and Hanford and Othello, Washington, and to apply them in conjunction with models developed from Prosser weather data in order to predict late blight in the Columbia Basin. Two additional years of weather data, 1995 and 1996, were added to the 1969 to 1994 data set from Prosser for model development and testing. Logistic regression analysis was selected over discriminant regression analysis to analyze data from the four locations, because the two methods have previously given similar results and the assumption of multivariate normality required for parametric discriminant analysis is not necessary for logistic regression to be valid (20).

MATERIALS AND METHODS

Weather data were collected at the Irrigated Research and Extension Center near Prosser, Othello, the Hermiston Agricultural Research and Extension Center near

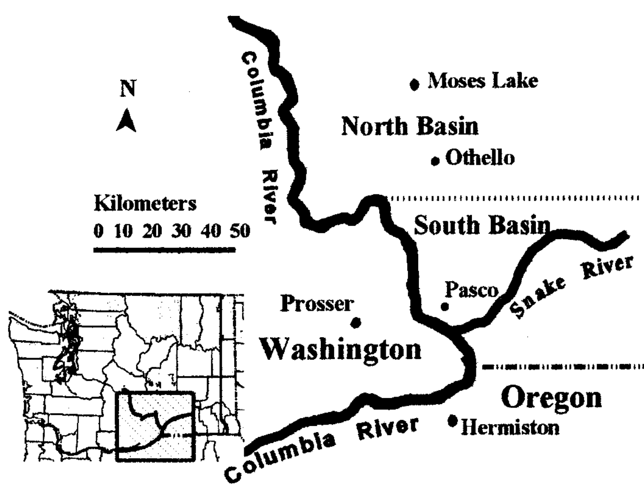

Fig. 1. Map of the potato production area in the Columbia Basin of Washington and Oregon (insert) showing the North and South Basin of Washington and the Basin of Oregon.

Hermiston, and the Hanford Meteorological Station 70 km northwest of Pasco, Washington. These locations are distributed throughout the Columbia Basin (Fig. 1). Variables used in these studies include: daily total rainfall, number of days with rain ≥0.25 mm, number of consecutive days with rain ≥0.25 mm, and daily and mean monthly minimum and maximum temperatures from April through September of each year. Variables included to determine possible effects of winter weather on survival of inoculum were monthly mean temperature, monthly mean minimum temperature, monthly minimum temperature, and monthly precipitation from November through February preceding each growing season. A threshold of 0.25 mm of rain was chosen because this was the least amount of measured rain during the study period. Air temperatures were measured at a height of 1.5 m above the soil surface, and rainfall (irrigation water was not included) was measured from November 1969 through 1987 with a hygrothermograph and a rain gauge, respectively, and from 1988 through 1996 with a probe thermistor and a tipping-bucket rain gauge, respectively, at Prosser. Hygrothermographs and rain gauges were used at the other locations.

Meteorological data from 1969 through 1996 were used to evaluate the effects of weather on late blight development in the Columbia Basin and for model development. This period represents the length of time potatoes in this region have been grown primarily under center-pivot irrigation systems. Potato fields in the study area were monitored regularly by growers and field consultants, and any suspected observations of late blight were brought to the attention of G. E. Easton, P. B. Hamm, or D. A. Johnson. Years in which late blight was observed in any field location within the region and confirmed by either Easton, Hamm, or Johnson were classified as outbreak years. Years in which no late blight was reported were classified as nonoutbreak years.

Logistic regression, which involves relating qualitative variables such as outbreak status (whether or not late blight was confirmed in the Columbia Basin for a given year) to other variables through a logistic cumulative distribution function, was used to develop forecasting models for late blight outbreaks (18,19,20). Variable selection included univariate analysis (7) and univariate assessment of the relationship between a predictor variable and the binary response. A previous study indicated that the prior year's outbreak status was useful in predicting the current year's outbreaks (11). Therefore, the prior year's outbreak status was included in nearly all models examined. To reduce the number of independent variables considered in any one analysis, weather variables containing redundant information were considered in separate analysis runs. However, all meaningful weather variables were considered for their ability to predict outbreaks in at least one analysis.

One method of evaluating prediction models is to reclassify the years used to create the models. Because this method may overestimate model performance, a one-step approximation was used (SAS PROC LOGISTIC; SAS User's Guide: Statistics, SAS Institute, Cary, NC). Final models were selected for their ability to correctly classify years as outbreak or nonoutbreak by resubstitution and cross-classification. Sensitivity and specificity also were considered, where the former term refers to correct classification of outbreak years and the latter refers to correct classification of nonoutbreak years (3). Furthermore, variables were selected on the basis of their predictive ability in combination with other independent variables, not by the strength of their one-at-a-time measure of correlation. To avoid model overspecification, models containing the least number of independent variables were desirable.

RESULTS

Late blight was observed in the Columbia Basin during 15 of 28 years in commercial fields from 1970 through 1997. Years in which outbreaks occurred were 1974 to 1977, 1982 to 1984, and 1990 to 1997. Total areas affected by late blight during years with outbreaks varied from 40 to more than 20,000 ha from 1970 to 1994 (11). The entire area of 65,000 ha was affected in 1995, and about 32,000 ha were affected in 1996.

Three models were selected to describe the occurrence of late blight in four locations in the Columbia Basin. For the first model (model 1), an indicator variable for the presence of a late blight outbreak during the preceding year (Y*p*) and number of days of rain during April and May (R*am*) were selected as predictors. The second model (model 2) used Y*p* and number of days of rain during July and August (R*ja*) as predictors. The third model (model 3) used Y*p*, R*am*, and R*ja* as predictors.

Logistic regression was used for classification by first calculating a likelihood function (l*f*) for each location. For model 1 at Prosser, Hermiston, Hanford, and Othello, respectively, l*f* was computed: $lf = 9.252 - 4.004(Yp$, 1 = yes, 0 = no$) - 0.660(Ram)$, $lf = 3.744 - 2.856(Yp) - 0.188(Ram)$, $lf = -2.987 - 3.062(Yp) - 0.163(Ram)$, and $lf = 4.723 - 3.561(Yp) - 0.293(Ram)$. Accuracy and sensitivity of midseason predictors differed based on location, so models 2 and 3 were used at certain sites. At Hermiston and Hanford, respectively, l*f* for model 2 was computed: $lf = 0.730 - 3.195(Yp) - 0.103(Rja)$ and l*f*

Table 1. Accuracy, sensitivity, and specificity of three sets of variables used to predict potato late blight at four locations when analyzed by logistic regression

Independent variables[a]	Logistic regression			Cross-validation		
Location	Accuracy[b]	Sensitivity[c]	Specificity[d]	Accuracy[b]	Sensitivity[c]	Specificity[d]
Y*p*, R*am* (model 1)						
Prosser	89	93	85	89	93	85
Hermiston	81	79	85	82	85	79
Hanford	78	79	77	78	79	77
Othello	85	86	85	78	79	77
Y*p*, R*ja* (model 2)						
Prosser	81	79	85	82	85	79
Hermiston	81	79	85	82	79	85
Hanford	81	79	85	78	79	77
Othello	81	79	85	78	71	85
Y*p*, R*am*, R*ja* (model 3)						
Prosser	89	93	85	85	86	85
Hermiston	81	79	85	74	79	69
Hanford	81	79	85	74	79	69
Othello	85	86	85	82	86	77

[a] Y*p* = indicator variable for the presence of a late blight outbreak the preceding year; R*am* = number of rainy days (≥0.25mm) during April and May; R*ja* = number of rainy days (≥0.25) during July and August.
[b] Accuracy = percentage of years with or without late blight outbreak classified correctly.
[c] Sensitivity = percentage of years with late blight outbreak classified correctly.
[d] Specificity = percentage of years without outbreaks classified correctly.

= 1.716 – 2.844(Y*p*) – 0.094(R*ja*). At Prosser and Othello, respectively, l*f* for model 3 was computed: l*f* = 11.513 – 3.894(Y*p*) – 0.719(R*am*) – 0.259(R*ja*) and l*f* = 4.960 – 3.084(Y*p*) – 0.273(R*am*) – 0.129(R*ja*). Probability (PROB) was then calculated using the formula: PROB = 1/[1 + exp(l*f*)]. If PROB < 0.5, then the year was classified as nonoutbreak; if PROB ≥ 0.5, then the year was classified as outbreak.

The likelihood ratio chi-square, with 2 degrees of freedom for testing the significance of model 1, ranged from 13.379 to 22.388 and had *P* values less than 0.0012 at the four locations, and for model 2 was 12.157 with a *P* value 0.0023 at Hermiston and 11.979 with a *P* value of 0.0025 at Hanford. The likelihood ratio chi-square, with 3 degrees of freedom for model 3, was 23.683 with a *P* value of 0.0001 at Prosser and 16.375 with a *P* value of 0.0009 at Othello.

Accuracy, sensitivity, and specificity of the logistic regression and cross-validation analyses were good for models 1 to 3 at the four locations (Table 1). Model 1 had the highest accuracy and sensitivity from cross-validation analysis using data collected at Prosser (Table 1). Model 1 also had high accuracy and sensitivity using data from Hermiston. Model 2, using cross-validation analysis, had higher accuracy than model 3 using data from Hermiston and Hanford, but not Prosser and Othello (Table 1).

The nonoutbreak years 1978 and 1988 and the outbreak year 1990 were misclassified by models 1 and 3 using cross-validation with data from Prosser. The year 1978 was misclassified by all selected models using data from the other three locations. However, 1988 was correctly classified by models 1, 2, and 3 using data from the other three locations, and 1990 was correctly classified by models 1 and 3 using data from Othello.

DISCUSSION

Models developed in this study have been used to forecast the likelihood of late blight occurrence in the major potato producing areas of the Columbia Basin in Washington and Oregon. Previous late blight models for the Columbia Basin were derived from weather data over a 25-year period from only the Prosser location. Models were developed in this study with 27 years of weather data from the four locations. The location where weather data is collected is important in developing regional disease forecasting models (10,11). The number of locations for collecting weather data is also important; since weather factors such as rainfall vary within a region, several locations would more accurately represent the region than only one location. The number of rainy days in April and May, for example, varied from 6 at Othello to 13 at Hermiston in 1983, and from 5 at Hanford to 12 at Prosser in 1992 (Table 2). More confidence can be placed on disease forecasts when weather data are used from the four locations, which are distributed throughout the Columbia Basin. Late blight forecasts can also be made for specific areas within the Basin.

Accuracy and sensitivity in this study from the cross-validation analysis for model 1 with Prosser weather data were 89 and 93%, respectively. These were similar to when the early season model with variables Y*p*, R*am*, and total precipitation in May when the daily minimum temperature was greater than or equal to 5°C (P*m*) was cross-validated in the previous study (11). Specificity was high for the selected models. However, a high sensitivity was preferred over specificity because in the Columbia Basin it is better to expect an outbreak and not experience one than not to expect an outbreak and have one occur.

Model 1 can be used to forecast the likelihood of late blight each year, usually before 31 May. That date is generally 4 to 10 weeks after planting and 5 and 14 days before late blight has been observed in any year in the Columbia Basin of Oregon and Washington, respectively. Accurate forecasts were made before 10 May in 1995 and 1996 and on 16 May in 1997. Sufficient time can therefore be given growers to make initial fungicide applications before late blight has been observed in commercial fields. The intensity of fungicide applications, such as timing of initiation and interval between applications, can be based on the probability of late blight occurrence from model 1, by itself or in concert with weather forecasts.

Models 2 and 3 can be used through July

Table 2. Area of potato production in south-central Washington and north-central Oregon affected by *Phytophthora infestans*, number of days with rain during April and May (R*am*), and number of days with rain during July and August (R*ja*), at Hermiston, Oregon, and Prosser, Hanford, and Othello, Washington, from 1970 through 1997

Year	Area affected (ha)	Prosser		Hermiston		Hanford		Othello	
		R*am*	R*ja*	R*am*	R*ja*	R*am*	R*ja*	R*am*	R*ja*
1970	0	8	1	7	5	5	1	5	2
1971	0	9	4	14	5	9	2	8	1
1972	0	9	6	10	9	8	7	13	4
1973	0	6	1	8	2	3	2	3	0
1974	50	16	6	15	6	12	4	11	3
1975	810	10	7	12	9	8	7	8	8
1976	120	12	12	13	13	9	11	9	12
1977	40	10	4	11	5	8	7	7	5
1978	0	11	10	17	14	13	9	8	10
1979	0	8	9	17	13	11	11	6	7
1980	0	13	1	17	3	19	1	16	3
1981	0	8	3	12	4	6	3	12	2
1982	10,100	15	6	15	5	10	4	10	3
1983	14,150	9	12	13	11	7	10	6	7
1984	150	17	1	16	3	12	1	16	1
1985	0	5	4	8	3	3	2	6	3
1986	0	8	3	7	7	6	5	9	4
1987	0	5	5	8	5	4	7	8	5
1988	0	15	3	15	2	14	1	14	2
1989	0	12	8	16	10	13	5	13	5
1990	250	12	10	12	5	13	6	18	8
1991	15,000	11	4	14	6	12	5	13	7
1992	25,000	12	6	11	5	5	7	9	9
1993	31,000	20	9	26	6	18	9	22	10
1994	770	16	2	22	1	16	4	16	4
1995	65,000	20	10	15	9	17	6	19	10
1996	32,000	17	4	17	4	18	5	17	7
1997	48,000	10	7	16	12	9	6	12	20
Mean		11.6	5.6	13.7	6.5	10.3	5.3	11.2	5.8
±SE		±0.8	±0.6	±0.8	±0.7	±0.9	±0.6	±0.9	±0.8

and August if late blight was not observed earlier. Model 2 would best be used with data from Hermiston and Hanford, and model 3 with data from Prosser and Othello. The value of one of the variables of each model, R*ja*, would not be available until 31 August, but the model could be used by solving for the value of R*ja* needed for an outbreak to be predicted and comparing it with the normal for each location and expected occurrences of rainy days during July and August based on weather forecasts. The calculations can be repeatedly made and the forecast updated as weather and crop conditions change during July and August.

A regional forecasting system as used in the Columbia Basin is best utilized with input from an "expert" late blight epidemiologist, careful monitoring of short and up to 30-day weather forecasts, and methods to rapidly communicate with growers. The probability of an outbreak, weather forecasts, and crop canopy development are used to determine initiation and intervals between fungicide applications. A noted benefit of the late blight models used in the Columbia Basin is that growers have become more aware of environmental conditions that favor late blight outbreaks.

Meteorological factors have a major influence on late blight development in the Columbia Basin of Washington and Oregon (10). The number of rainy days during April and May and the number of rainy days during July and August were good indicators of late blight in this and a previous study (10). Early season rain is probably important for inoculum to multiply in fields containing volunteer potatoes, in fields with infected seed, and in locations with cull tubers, as well as for the dissemination of sporangia to additional fields. Sporangia are sensitive to drying (17,23) and are disseminated most effectively from field to field during rainy periods (6). Once potato plants in commercial irrigated fields in the Columbia Basin are infected after row closure, microclimate conditions generally are favorable for continued late blight development (2,21).

The variable P*m*, when included with Y*p* and R*am*, was a good indicator of late blight in this as well as the previous study (11). However, P*m* was not included as a predictor in the models of this study because either accuracy or sensitivity were slightly less when P*m* was included with Y*p* and R*am* for two of the four locations (Hanford and Hermiston). Furthermore, a two-variable model (Y*p* and R*am*) was considered to be better than a three variable model (Y*p*, R*am*, and P*m*), given the same accuracy and sensitivity.

A relationship of initial inoculum with the development of late blight outbreaks has long been recognized. The forecasting system Blitecast assumes that inoculum is constantly present (14). Initial inoculum was identified as an important factor for the development of late blight outbreaks in the Columbia Basin by the inclusion of the indicator variable for the presence of an outbreak during the preceding year. If an outbreak occurred during the preceding year, there was a greater likelihood that inoculum would survive the winter, presumably in infected tubers buried in soil, in cull piles, or held in storage. Infected seed tubers brought into the Columbia Basin from seed-production areas with late blight during the previous year are important potential sources of initial inoculum. The presence of late blight in a seed producing area for the Columbia Basin may potentially be used in the models developed in this study as the indicator variable representing the occurrence of late blight the preceding year. It is important for growers in the Columbia Basin to use certified seed produced from areas that are free of late blight or produced in areas or farms where late blight was successfully managed. Poor sanitation practices would negate the effectiveness of these as well as other late blight forecasting models (14).

Logistic regression has been useful in quantifying the effects of weather on disease development and in developing disease forecasting models (10,11). It is an alternative method to discriminant analysis for classifying objects into categories such as years with late blight outbreak and those without, based on predictor variables. Discriminant analysis and logistic regression may be easily applied because statistical software is readily available for both. However, logistic regression is often preferred because the predictor variables may be discrete, and hence nonnormally distributed, rather than assuming all predictor variables have a multivariate normal distribution, as required by discriminant analysis (20). Another appealing feature of logistic regression is that the critical probability value for classifying observations may be specified. In this study, we used 0.5 as the critical probability value for classification because approximately half of the years had outbreaks of late blight. An alternative strategy would be to choose a critical probability value less than 0.5 so more years would be classified as outbreak years. This strategy would tend to increase sensitivity and decrease specificity, which is consistent with the belief that it is better to treat for an expected outbreak and not experience it rather than not apply a treatment and experience an outbreak.

ACKNOWLEDGMENTS

PPNS No. 0267, Department of Plant Pathology, College of Agriculture and Home Economics Research Center, Project No. 0678, Washington State University, Pullman 99164-6430.

LITERATURE CITED

1. Beaumont, A. 1947. The dependence on the weather of the dates of outbreaks of potato late blight epidemics. Trans. Br. Mycol. Soc. 31:45-53.
2. Easton, G. D. 1982. Late blight of potatoes and prediction of epidemics in arid central Washington State. Plant Dis. 66:452-455.
3. Friedman, G. D. 1974. Primer of Epidemiology. McGraw-Hill, New York.
4. Fry, W. E., and Doster, M. A. 1991. Potato late blight: Forecasts and disease suppression. Pages 326-336 in: *Phytophthora*. J. A. Lucas, R. C. Shattock, D. S. Shaw, and L. R. Cooke, eds. Cambridge University Press, New York.
5. Hamm, P. B., Fry, B. A., and Jaeger, J. 1994. Occurrence and frequency of metalaxyl insensitivity and mating types of *Phytophthora infestans* in the Columbia Basin of Oregon and Washington. (Abstr.) Phytopathology 84:1123.
6. Hirst, J. M., and Stedman, O. J. 1960. The epidemiology of *Phytophthora infestans* II. The source of inoculum. Ann. Appl. Biol. 48:489-517.
7. Hosmer, D. W., and Lemeshow, S. 1989. Applied Logistic Regression. John Wiley & Sons, New York.
8. Hyre, R. A., and Bonde, R. 1956. Forecasting late blight of potatoes in Aroostook County, Maine, in 1956. Plant Dis. Rep. 40:1056-1090.
9. Inglis, D. A., Johnson, D. A., Legard, D. E., Fry, W. E., and Hamm, P. B. 1996. Relative resistances of potato clones in response to new and old populations of *Phytophthora infestans*. Plant Dis. 80:575-578.
10. Johnson, D. A., Alldredge, J. R., and Allen, J. R. 1994. Weather and downy mildew epidemics of hop in Washington State. Phytopathology 84:524-527.
11. 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.
12. Johnson, D. A., Cummings, T. F., Hamm, P. B., Rowe, R. C., Miller, J. S., Thornton, R. E., Pelter, G. Q., and Sorensen, E. J. 1997. Potato late blight in the Columbia Basin: An economic analysis of the 1995 epidemic. Plant Dis. 81:103-106.
13. Krause, R. A., Massie, L. B., and Hyre, R. A. 1975. BLITECAST: A computerized forecast of potato late blight. Plant Dis. Rep. 59:95-98.
14. MacKenzie, D. R. 1981. Scheduling fungicide applications for potato late blight with BLITECAST. Plant Dis. 65:394-399.
15. Miller, J. S., Hamm, P. B., and Johnson, D. A. 1997. Characterization of the *Phytophthora infestans* population in the Columbia Basin of Oregon and Washington from 1992 to 1995. Phytopathology 87:656-660.
16. Miller, J. S., Johnson, D. A., and Hamm, P. B. 1995. Aggressiveness of isolates of *Phytophthora infestans* from the Columbia Basin of Washington and Oregon. Phytopathology 88:190-197.
17. Minogue, K. P., and Fry, W. E. 1981. Effect of temperature, relative humidity, and rehydration rate on germination of dried sporangia of *Phytophthora infestans*. Phytopathology 71:1181-1184.
18. Morrison, D. F. 1976. Multivariate Statistical Methods. McGraw-Hill, New York.
19. Morrison, D. F. 1983. Applied Linear Statistical Methods. Prentice Hall, Englewood Cliffs, NJ.
20. Press, S. J., and Wilson, S. 1978. Choosing between logistic regression and discriminant analysis. J. Am. Stat. Assoc. 73:699-705.
21. Thornton, R. E., and Johnson, D. A. 1997. In-canopy environment as a factor in late blight management. Spudman. January.
22. Wallin, J. R. 1962. Summary of recent progress in predicting late blight epidemics in the United States and Canada. Am. Potato J. 39:306-312.
23. Warren, R. C., and Colhoun, J. 1975. Viability of sporangia of *Phytophthora infestans* in relation to drying. Trans. Br. Mycol. Soc. 64:73-78.