Logistic Regression Modeling of Prevalence of Soybean Sclerotinia Stem Rot in the North-Central Region of the United States

A. L. Mila, A. L. Carriquiry, and X. B. Yang

First and third authors: Department of Plant Pathology; second author: Department of Statistics, Iowa State University, Ames 50011.
Current address of A. L. Mila: Department of Plant Pathology, University of California-Davis, Kearney Agricultural Center, Parlier 93648.
Accepted for publication 15 September 2003.

ABSTRACT

Mila, A. L., Carriquiry, A. L., and Yang, X. B. 2004. Logistic regression modeling of prevalence of soybean *Sclerotinia* stem rot in the north-central region of the United States. *Phytopathology* 94:102-110.

Regional prevalence of soybean *Sclerotinia* stem rot (SSR), caused by *Sclerotinia sclerotiorum*, was modeled using tillage practices, soil texture, and weather variables (monthly air temperature and monthly precipitation from April to August) as inputs. Logistic regression was used to estimate the probability of stem rot prevalence with historical disease data from four states of the north-central region of the United States. Potential differences in disease prevalence between states in the region were addressed using regional indicator variables. Two models were developed: model I used spring (April) weather conditions and model II used summer (July and August) weather conditions as input variables. Both models had high explanatory power (78.5 and 77.8% for models I and II, respectively). To investigate the explanatory power of the models, each of the four states was divided into small geographic

areas, and disease prevalence in each area was estimated using both models. The R^2 value of the regression analysis between observed and estimated SSR prevalence was 0.65 and 0.71 for models I and II, respectively. The same input variables were tested for their significance to explain the within-field SSR incidence by using Poisson regression analysis. Although all input variables were significant, only a small amount of variation of SSR incidence was explained, because R^2 of the regression analysis between observed and estimated SSR incidence was 0.065. Incorporation of available site-specific information (i.e., fungicide seed treatment, weed cultivation, and manure and fertilizer applications in a field) improved slightly the explained amount of SSR incidence ($R^2 = 0.076$). Predicted values of field incidence generally were overestimated in both models compared with the observed incidence. Our results suggest that preseason prediction of regional prevalence would be feasible. However, prediction of field incidence would not, and a different site-specific approach should be followed.

Soybean *Sclerotinia* stem rot (SSR), caused by *Sclerotinia sclerotiorum* (Lib.) de Bary, was first found in the United States in 1946 and reported in 1951 (9). Initially, the disease was a minor problem with localized epidemics occurring in Michigan, Minnesota, and Wisconsin, where soybean of maturity groups 0 and I is grown (14). Since the beginning of the last decade, the disease distribution changed significantly, expanding to areas where maturity groups II and III are grown (32). Before 1991, in the north-central region, SSR was ranked 12th among the diseases causing soybean yield losses (12). In 1994, however, it was second only to the losses caused by soybean cyst nematode (34), whereas in 1998 it was ranked 11th (35).

Epidemics of SSR are initiated by ascospores released from apothecia produced by sclerotia (8). Sclerotia of *S. sclerotiorum* must be preconditioned (i.e., held under cool, moist conditions) before they are capable of carpogenic germination (2). The time of preconditioning varies among different isolates (2). Most infections of soybean plants occur during the flowering stage, because ascospores attached to flower petals use the petals as a nutrient base to germinate when free water is present on the plant surfaces (2,8). Soil moisture and temperature commonly affect carpogenic germination of sclerotia of *S. sclerotiorum* (1,2). Prolonged periods of low temperature and high soil moisture are favorable for apothecial development (14). SSR severity is associated with differences in the microclimate within the plant canopy (7) and

soil texture (25). Light-textured soils favored the germination and apothecia formation of sclerotia of *S. sclerotiorum*, leading to greater disease incidence and severity (25).

Tillage also has been shown to be associated with SSR incidence (13,31). More apothecia of *S. sclerotiorum* were observed in minimum-tillage than in no-tillage fields, although there were no significant differences in disease incidence among tillage treatments (13). In a study in Alberta (31), plowing reduced sclerotia germination and apothecia production. Use of conservation tillage often is associated with narrow-row soybean planting, which is more favorable to SSR than wide row soybean (15,26).

Much less is known about the regional distribution and prevalence of SSR. In a recent study (32) conducted in Illinois, Iowa, Minnesota, Missouri, and Ohio, the disease was more prevalent when summer temperature was below normal (based on the 30-year average) than when it was above normal. Furthermore, the disease prevalence increased exponentially with latitude and was found further south in 1996, a year with a cooler than normal summer (32). In the same survey, the prevalence of SSR was highest in minimum-tillage, intermediate in conventional-tillage, and lowest in no-tillage fields.

Compared with efforts of modeling disease epidemics in individual fields, fewer modeling studies are available on regional disease prevalence, presumably because appropriate data are scarce. In Alberta, Manitoba, and Saskatchewan, Canada (21,24,29), regional disease prediction models of canola stem rot have been developed and provide weekly disease warnings for appropriate fungicide applications. The weekly forecasting approach, however, may not be applicable to soybean SSR because management decisions in soybean are made a few months before the disease

Corresponding author: X. B. Yang; E-mail address: xbyang@iastate.edu

Publication no. P-2003-1030-01R
© 2004 The American Phytopathological Society

occurrence (i.e., tillage usually during the previous fall, cultivar selection in late winter, and width of row planting in spring).

Most of the published studies on disease prediction focus on disease incidence; thus, there are few or no available comparative studies on disease prevalence and incidence to show how changes in factors affecting one scale may affect changes in another scale. Such studies could help in understanding the relationship between two of the fundamental measures of plant disease intensity: regional disease prevalence and incidence within a field.

The primary objective of this study was to develop explanatory models of SSR prevalence in four states of the north-central region by using tillage, soil texture, and climate variables as input variables. A secondary objective was to investigate whether the within-field SSR incidence could be explained by the same variables used to explain SSR prevalence. The variable selection was based on the biology and epidemiology of the pathogen (1,8, 23,25) and on the preliminary results of our previous study (32).

MATERIALS AND METHODS

Data collection. Data collection has been described in detail (32,33). Briefly, soybean samples were collected in 1995 and 1996 from Illinois, Iowa, Minnesota, Missouri, and Ohio to determine the prevalence of SSR in collaboration with the National Agricultural Statistics Service (NASS). In 1997 and 1998, more samples were collected from Iowa, Minnesota, and Missouri. From each selected field, 20 soybean stems were sampled in a zigzag pattern and shipped to Iowa State University for analyses. The zigzag pattern contained 10 corners, and two stems were collected from each corner. Each stem was observed externally for presence or absence of the typical cottony mycelium. The stems also were split longitudinally and checked for presence or absence of sclerotia in the pith. Isolation was made on acidified potato dextrose agar from stems with signs and symptoms that could not be readily identified. The soil texture of each sample was determined by a commercial laboratory (MVT Laboratories, Inc., Bismarck, ND) with methods described previously (33). SSR was not found in samples from Missouri. Because the disease has not been reported in this state, data from Missouri were not included in model development. Overall, during the 4 years of investigation, 1,545 fields from Illinois, Iowa, Minnesota, and Ohio were sampled. SSR was detected in 87 fields. For fields in which SSR was detected, disease incidence varied from 5 to 80%. In this study, the term "prevalence" is used to describe the percentage of fields in which the disease was found, and the term "incidence" is used to describe the percentage of infected soybean plants in a field.

Tillage categories. Tillage information for each field was obtained from producers during interviews conducted by NASS. Tillage systems are classified into three categories according to the amount of surface residue (4). Conservation-tillage systems maintain >30% surface residue after planting. Tillage practices that maintain 15 to 30% surface residues are categorized as minimum tillage, whereas those that maintain <15% surface residue are classified as conventional tillage.

Weather data. Weather data were obtained from the National Oceanic and Atmospheric Administration online through the National Climatic Data Center in Asheville, NC. Data included mean monthly air temperature and precipitation for April, May, June, July, and August from 1995 to 1998 for Iowa and Minnesota and from 1995 and 1996 for Illinois and Ohio.

Mean monthly temperature and precipitation were obtained for each sampled field from the nearest weather station. The corresponding weather data were obtained from 101 weather stations in Illinois, 107 in Iowa, 51 in Minnesota, and 61 in Ohio. Of the sampled fields, 82% were located less than 20 km away from the nearest weather station. In only 18 fields in Illinois, 8 in Iowa, 7 in Minnesota, and 11 in Ohio was the distance between field and weather station greater than 25 km.

Logistic regression analysis. Logistic regression was used to identify the factors significantly associated with regional prevalence of SSR. The dependent variable was absence or presence of the disease in a field; thus, fields were divided into two categories based on the presence or absence of disease on the collected samples. Logistic regression is used widely in epidemiological research, where the binary response usually is the presence or absence of a disease and the predictor variables are putative risk factors and possible confounding variables (5,18). If Y represents disease presence in a field and only takes on values 0 and 1 (absence or presence of the disease), the probability of disease prevalence can be modeled as follows:

$$P(Y = 1) = \exp(\sum b_i X_i) / (1 + \exp(\sum b_i X_i)) \quad (1)$$

where b_i are parameters to be estimated and X_i are the covariates or predictors. Because the response variable $P(Y = 1)$ is a probability, it is constrained to lie between 0 and 1; note that, for any values of b_i and X_i , the logistic function given in equation 1 also is constrained to the interval 0 or 1. The parameters b_i are similar to the intercept and regression coefficients in an ordinary multiple regression model. However, their interpretations are somewhat different from those in logistic regression (18). They are used to quantify the disease prevalence risk factor of an area.

The LOGISTIC procedure in SAS (Statistical Analysis Systems, version 8.0; SAS Institute, Inc., Cary, NC) was used to fit equation 1 to the data. To select the best set of predictors, the method of backward elimination was used. Starting from a saturated (or full) model, we sequentially dropped those covariates or predictors that did not seem to be significantly associated with the response variable. In the final model, all predictors were significantly associated with disease prevalence.

Poisson regression analysis. Poisson regression commonly is used in the study of rare events when the responses take the form of counts (5). For SSR incidence, the dependent variable was the number of diseased soybean plants out of 20 sampled in a field. Only in very few fields (87 fields) was SSR detected and, usually, it was found only in a small portion of the 20 sampled plants (1 to 5) even in 1996, an epidemic year for soybean SSR in the north-central region. Thus, SSR occurrence could be safely considered a rare event. If y represents the number of diseased plants in a field and only takes values between 0 and 20 (20 is the maximum possible response because 20 stems were collected from each field), the probability of disease incidence can be modeled as follows:

$$P(y) = \lambda^y e^{-\lambda} / y! \quad (2)$$

with $\lambda = \exp(\sum b_i X_i)$ to ensure that y is positive because the Poisson distribution is defined for nonnegative integers.

In equation 2, b_i are parameters to be estimated and X_i are the covariates or predictors. Because the response variable $P(y)$ is a probability, it is constrained to be between 0 and 1. The parameters b_i are used to quantify the disease incidence risk factor of a field. The GENMOD procedure in SAS (SAS Institute) was used to fit equation 2 to the data.

Criteria of goodness of fit. *Logistic regression.* For all pairs of observations with different values of the response variable, a pair is "concordant" when the observation with the larger ordered value of the response has a higher predicted event probability than the observation with the smaller ordered value of the response. A pair is "discordant" when the observation with the larger ordered value of the response has a lower predicted event probability than the observation with the smaller ordered value of the response. A pair is "tied" when it is neither concordant nor discordant.

Correlation indices. Somers' D, Gamma, Tau-a, and c are computed from the numbers of concordant and discordant pairs of observations and are used to assess the predictive ability of a model. Specifically, c is equal to the area under a receiver operating characteristic curve that measures the predictive accuracy of a model and takes values between 0 and 1 (6). A model with higher

values for these indices has better predictive ability than a model with lower values for these indices (3,6). The best fitted models for prevalence were selected on the basis of the amount of variation explained by the predictors as measured by the criteria mentioned above.

Poisson regression. The deviance (equal to $-2 \times \log[\text{likelihood}]$) and the log(likelihood) values were used to compare the two explanatory models tested for SSR incidence. Models with good fit have low deviance and high log(likelihood) values. The difference in deviance was used to compare the two fitted

TABLE 1. Correlation coefficients among temperature and precipitation in the north-central region from 1995 to 1998^a

Variable	APRPR	MAYT	MAYPR	JUNT	JUNPR	AVERAGE	JULPR	AUGPR
APRT								
Pearson ^b	0.39 ^c	0.83 ^c	0.50 ^c	0.37 ^c	0.21	0.63 ^c	-0.07	-0.07
P value ^d	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.003	0.003
Number ^e	1,611	1,626	1,611	1,626	1,616	1,609	1,616	1,616
APRPR								
Pearson	...	0.43 ^c	0.22	0.42	-0.01	0.55 ^c	0.01	-0.21
P value	...	0.0001	0.0001	0.0001	0.655	0.0001	0.71	0.0001
Number	...	1,611	1,644	1,611	1,639	1,606	1,639	1,639
MAYT								
Pearson	...	...	0.26	0.14	0.28	0.50 ^c	-0.06	-0.04
P value	...	...	0.0001	0.0001	0.0001	0.0001	0.008	0.116
Number	...	...	1,611	1,626	1,616	1,609	1,616	1,616
MAYPR								
Pearson	...	...	...	0.44 ^c	-0.06	0.44 ^c	0.02	-0.09
P value	...	...	...	0.0001	0.01	0.0001	0.40	0.0002
Number	...	...	...	1,611	1,639	1,606	1,639	1,639
JUNT								
Pearson	...	...	...	...	-0.25	0.57 ^c	-0.01	-0.13
P value	...	...	...	...	0.0001	0.0001	0.73	0.0001
Number	...	...	...	...	1,616	1,609	1,616	1,616
JUNPR								
Pearson	...	...	...	...	...	-0.09	0.02	0.14
P value	...	...	...	...	...	0.0001	0.34	0.0001
Number	...	...	...	...	...	1,611	1,648	1,648
AVERAGE								
Pearson	...	...	...	...	...	...	-0.14	0.006
P value	...	...	...	...	...	...	0.0001	0.805
Number	...	...	...	...	...	...	1,611	1,611
JULPR								
Pearson	...	...	...	...	...	...	...	-0.17
P value	...	...	...	...	...	...	...	0.0001
Number	...	...	...	...	...	...	...	1,648

^a Mean monthly temperature of April (APRT), May (MAYT), and June (JUNT); average temperature of July and August (AVERAGE); and precipitation of April (APRPR), May (MAYPR), June (JUNPR), July (JULPR), and August (AUGPR).

^b Pearson correlation coefficients.

^c High correlation coefficients.

^d P value for a significance test.

^e Number of observations.

Fig. 1. Monthly air temperature of April, May, and June, and average of July and August for the years between 1960 and 1999 in Iowa.

incidence models. The significance of the difference in deviance was evaluated using a χ^2 test with the appropriate degrees of freedom (10).

Input variables. Prevalence. Input variables were monthly average air temperature and total precipitation for March to August; soil texture; and percentage of sand, clay, and silt in the soil. Also, two indicator variables were used. Indicator variables are used to quantify qualitative differences among groups. One was used to capture tillage effects and the other to capture regional effects. For the former, conventional tillage was arbitrarily selected as the reference group, so that the effect of other types of tillage was estimated relative to the effect of conventional tillage. For region, Illinois was arbitrarily used as the reference group, and the SSR prevalence risk in Iowa, Minnesota, and Ohio was estimated relative to the risk in Illinois. The selection of the reference group does not affect the outcomes; thus, it can be arbitrary. The use of indicator variables allows for the data to be pooled and analyzed together, while differences among subgroups are still distinguishable (such as among states and tillage regimes in our study). However, for simplicity, the results are presented separately for the four states and the three tillage regimes.

Variables were tested for possible correlation (using the CORR procedure of the SAS software) because the presence of collinearity among predictors in a model reduces the accuracy with which parameters can be estimated. The high positive correlation between spring and summer weather variables (Table 1) suggests that incorporation of all weather variables into the model would not be proper. The long-term weather pattern of the area from 1960 to 1990 (A. L. Mila, *unpublished data*) does not support the assumption that the correlations found during the 4-year period of survey represent the general weather pattern of the north-central region. The average air temperatures in April, May, and June and the average in July and August from 1961 to 1999 show that low spring temperature can be followed by either high or low summer temperature (Fig. 1). Therefore, spring and summer variables were

not used simultaneously, and two categories of models were developed separately: one using spring weather variables (candidate model I) and one using summer weather variables (candidate model II).

Incidence. Input summer weather variables used for modeling SSR prevalence were tested also as candidate variables for explaining the within-soybean field SSR incidence because the within-growing season weather is important for SSR incidence (7, 15,25). Furthermore, several other management practices known (by interviewing the growers) to have been applied in the sampled fields were incorporated for modeling SSR incidence. These management practices were as follows: fungicide seed treatment, manure and fertilizer applications, and weed cultivation. All practices were used as indicator variables with no application of a management practice as the reference group. Collinearity between management practices and management practices and weather variables has been addressed previously (22). Most of the input variables used for SSR incidence were highly correlated. Thus, parameter estimates may not be efficient.

Presentation of estimates and explanatory power of models.

The four states in the north-central region were divided arbitrarily into smaller geographic areas. For each of these areas, the estimated probability of disease prevalence for each year from 1995 to 1998 was calculated using model I (April weather variables) and model II (July and August weather variables). The PROB procedure of SAS was used to calculate the probability of disease prevalence and incidence for each sampled field in the area. Estimated probabilities of prevalence then were averaged to produce the estimated probability of disease prevalence for the area. Estimated prevalence for each area was plotted against observed prevalence, and regression analysis was performed between the estimated and observed values. Similarly, estimated incidences calculated with PROB procedure of SAS were regressed against observed incidences in sampled fields so that estimation accuracy could be investigated.

TABLE 2. Parameter estimates of the logistic regression used to explain the prevalence of soybean Sclerotinia stem rot (caused by *Sclerotinia sclerotiorum*) by using spring weather variables and tillage practices (model I) in Iowa, Illinois, Minnesota, and Ohio from 1995 to 1998

Tillage, state	Variable ^a	Parameter estimate ^b	Standard error
Conventional or no tillage Illinois and Ohio	Intercept	-1.874	1.37
	AprilT	-0.162	0.006
	AprilPr	-0.012	0.00035
Iowa	Intercept	-1.33	1.37
	AprilT	-0.13	0.006
	AprilPr	-0.012	0.00035
Minnesota	Intercept	-0.93	1.37
	AprilT	-0.11	0.0063
	AprilPr	-0.012	0.00035
Minimum tillage Illinois and Ohio	Intercept	-1.4	1.37
	AprilT	-0.135	0.0059
	AprilPr	-0.012	0.00035
Iowa	Intercept	-0.85	1.37
	AprilT	-0.164	0.0059
	AprilPr	-0.012	0.00035
Minnesota	Intercept	-0.46	1.37
	AprilT	-0.083	0.0063
	AprilPr	-0.012	0.00035

^a AprilT = temperature of April (°C) and AprilPr = precipitation of April (cm).

^b All parameter estimates are significantly different from 0 ($P = 0.05$). Logistic regression for model I: $P(Y = 1) = \exp(b_0 + b_2\text{State} + b_3\text{AprilT} \times \text{Tillage} + b_4\text{AprilPr}) / (1 + \exp[b_0 + b_2\text{State} + b_3\text{AprilT} \times \text{Tillage} + b_4\text{AprilPr}])$, where Y (no, 0; yes, 1) = disease presence in a field; Tillage (conventional, 0; no tillage, 1; minimum tillage, 2) = indicator variable for tillage practices; and State (Illinois, 0; Iowa, 1; Minnesota, 2; Ohio, 3) = indicator variable for regional effects.

TABLE 3. Parameter estimates of the logistic regression used to explain the prevalence of soybean Sclerotinia stem rot (caused by *Sclerotinia sclerotiorum*) by using summer weather variables and tillage practices (model II) in Iowa, Illinois, Minnesota, and Ohio from 1995 to 1998

Tillage, state	Variable ^a	Parameter estimate ^b	Standard error
Conventional or no tillage Illinois and Ohio	Intercept	5.71	3.66
	Average	-0.46	0.0095
	JulyPr	0.029	0.0002
Iowa	Intercept	6.22	3.66
	Average	-0.43	0.0096
	JulyPr	0.029	0.0002
Minnesota	Intercept	6.44	3.66
	Average	-0.42	0.0096
	JulyPr	0.029	0.0002
Minimum tillage Illinois and Ohio	Intercept	6.06	3.66
	Average	-0.44	0.0096
	JulyPr	0.029	0.0002
Iowa	Intercept	6.57	3.66
	Average	-0.41	0.0096
	JulyPr	0.029	0.0002
Minnesota	Intercept	6.8	3.66
	Average	-0.40	0.0096
	JulyPr	0.029	0.0002

^a Average = mean average temperature of July and August (°C) and JulyPr = precipitation of July (cm).

^b All parameter estimates are significant different from 0 ($P = 0.05$). Logistic regression for model II: $P(Y = 1) = \exp(b_0 + b_2\text{State} + b_3\text{Average} \times \text{Tillage} + b_4\text{JulyPr}) / (1 + \exp[b_0 + b_2\text{State} + b_3\text{Average} \times \text{Tillage} + b_4\text{JulyPr}])$, where Y (no, 0; yes, 1) = disease presence in a field; Tillage (conventional, 0; no tillage, 1; minimum tillage, 2) = indicator variable for tillage practices; and State (Illinois, 0; Iowa, 1; Minnesota, 2; Ohio, 3) = indicator variable for regional effects.

RESULTS

Correlation between weather variables. High correlations were found between April temperature and April precipitation, April temperature and May temperature, April temperature and May precipitation, and April temperature and average temperature of July and August (Table 1). Significant correlations also were found between April precipitation and average temperature of July and August, May temperature and average temperature of July and August, May precipitation and average temperature of July and August, and June temperature and average temperature of July and August.

Logistic regression results for SSR prevalence. The two best fitted models were model I, where April precipitation, April air temperature, and its interactions with the indicator variables for tillage and regional effects were the input variables (Table 2), and model II, which included July precipitation, average air temperature of July and August as well as the interaction between average temperature of July and August, and tillage and state indicator variables (Table 3). For both models, goodness of fit was assessed by the proportion of concordant and discordant pairs, and the values of Somers' D, Gamma, Tau-a, and c statistics (Table 4). The number of concordant pairs was very high, indicating that the models calculated high probabilities when SSR was truly ob-

served and low probabilities when SSR had not been observed and only a few probabilities were falsely predicted (as per the number of discordant pairs). Accordingly, the other criteria (Somers' D, Gamma, Tau-a, and c) that are calculated from the number of concordant and discordant pairs were high as well. Especially, c was very high, indicating that the models had good predictive accuracy.

Model I suggests that low temperature and precipitation in April increase the risk of SSR; whereas, in model II, high precipitation in July and low temperature during July and August increase the disease prevalence in the four states examined. Both models reveal the differences in SSR risk among the four states and the tillage regimes. Assuming that the weather conditions are the same within a state, SSR risk is higher under minimum tillage than conventional or no tillage (Tables 2 and 3; higher parameter estimates of April or average of July and August temperature in model I and II, respectively). Also, if everything else is equal, the risk is highest in Minnesota, second in Iowa, and least in Illinois and Ohio, as is suggested by the magnitude of the intercepts of models I and II for the four states (Tables 2 and 3).

Observed and estimated prevalence of SSR with model II for Illinois and Ohio in 1995 and for Iowa and Minnesota in 1998 are given in Figure 2. Both models I and II gave good estimation for all the areas of Illinois, Iowa, Ohio, and the three lower areas of Minnesota. In the three upper areas of Minnesota, the estimated prevalence was higher than the observed prevalence in 1995, 1996, and 1998 and lower in 1997.

The comparison between observed and estimated SSR prevalence for each geographic area made with models I and II is presented in Figures 3 and 4, respectively. In a large number of areas, observed prevalence was very low, and only two areas showed high SSR prevalence (Fig. 3A). Both models I and II redistribute the very low prevalence values in the low classes of estimation (between 0 and 0.1) and underestimate the very high prevalence values (Fig. 3B and C), smoothing the regional impact of SSR risk, as commonly happens when large-scale modeling is attempted. In 89% of the comparisons in which the actual prevalence was zero, both models I and II gave <0.1 prevalence. In the remainder (11%), the estimated prevalence was between 0.1 and 0.14. The R^2

TABLE 4. Criteria of goodness of fit of soybean *Sclerotinia* stem rot (caused by *Sclerotinia sclerotiorum*) prevalence models (model I and model II) for four north-central states

Criteria	Model I	Model II
Concordant (%) ^a	78.1	77.4
Discordant (%) ^a	21.1	21.7
Tied (%) ^a	0.8	1.0
Somers' D ^b	0.571	0.557
Gamma ^b	0.575	0.562
Tau-a ^b	0.061	0.057
c ^b	0.785	0.778

^a Measurements assess the association of estimated probabilities and observed frequencies.

^b Indices computed from the two first measurements. A model with higher values for these indices have better predictive ability than a model with lower values.

Fig. 2. Observed frequency and estimated probability (parentheses) of soybean *Sclerotinia* stem rot (caused by *Sclerotinia sclerotiorum*) prevalence in Illinois (IL) and Ohio (OH) in 1995 and in Iowa (IA) and Minnesota (MN) in 1998 by using tillage and July and August weather as input variables (model II).

value between the observed and estimated values of prevalence was 0.65 for model I and 0.71 for model II (Fig. 4). The intercept of the regression between estimated and observed values was not significantly different from zero ($P < 0.05$), and the slope was not significantly different from 1 for both model I and model II.

Poisson regression results for SSR incidence. Summer weather variables, state, and tillage indicator variables used for modeling SSR prevalence also were significant for the SSR field incidence (Table 5). Interestingly, parameter estimates from Poisson regression for the SSR incidence were similar to the estimated parameters from logistic regression accounting for the SSR prevalence (Table 3). The main difference was that August precipitation was significant for SSR incidence but not significant for SSR prevalence. Actually, the parameter estimate for August precipitation indicates that its effect is equivalent to the July precipitation effect (Table 5). When estimated incidences were regressed against observed values, the R^2 was very low (0.065) (Fig. 5). The intercept of the regression between estimated and observed values was not significantly different from zero, and the slope was significantly different from 1 ($P < 0.05$).

In addition, in the Poisson regression analysis, we incorporated more information on management practices available from the sampled fields. Results of Table 5 show that all the newly in-

corporated management practices were significant for the SSR field incidence. Use of seed treated with fungicides and weed cultivation had a positive effect and manure and fertilizer application a negative effect on SSR field incidence. Furthermore, the difference in deviance between the two models suggests that the newly incorporated information on management practices accounts significantly more for SSR incidence variability than the previous one, where only weather and tillage information were used (Table 5). When estimated incidences were regressed against observed values, the R^2 again was very low (0.076). The intercept of the regression between estimated and observed values was not significantly different from zero, and the slope was significantly different from 1 ($P < 0.05$).

DISCUSSION

Nonlinear probability models, logistic, regression, and Poisson regression were used to quantify two of the principal measures of plant epidemics, disease prevalence and disease incidence. Air temperature, precipitation, and tillage practices account for a significant portion of the variation in the prevalence of SSR in four north-central states and allow for the development of models with high explanatory power. When the same variables were tested with Poisson regression, only a small portion of SSR field incidence could be explained. Predicted values were consistently higher than the observed. Our results suggest that preseason prediction of regional prevalence would be feasible. However, prediction of field incidence would need a site-specific approach.

Fig. 3. Frequency of **A**, observed, **B**, estimated with model I (April temperature and precipitation), and **C**, estimated with model II (July and August temperature and precipitation) prevalence of *Sclerotinia stem rot* (caused by *Sclerotinia sclerotiorum*) during the 4-year survey (1995–98) in the subareas of Figure 2.

Fig. 4. Comparison of the predicted and the observed prevalence of soybean *Sclerotinia stem rot* (caused by *Sclerotinia sclerotiorum*) during the 4-year survey (1995–98) for **A**, model I (April weather variables) and **B**, model II (July and August weather variables). If a predicted value is equal to the observed value this point should be on the line. Model parameter estimates are presented in Tables 2 and 3.

Two models were selected for explaining SSR prevalence. Model I uses April weather conditions and model II uses July and August weather conditions as input variables. The selection of weather variables was based on results of previous studies (1,2,8,25). Workneh and Yang (32) reported the significance of July and August air temperature to SSR spatial distribution. No reports were found on the effect of April weather on prevalence of SSR. However, April weather data were used in our model development to investigate their possible significance on explaining SSR prevalence. It would be useful to have preseason indicators of the expected SSR prevalence during the upcoming growing season.

We used 4 years of data from Iowa, Illinois, Ohio, and Minnesota to increase the inter-year weather variation. However, the high correlation between monthly weather variables suggests that 4 years of data may be too short of a sequence with regard to weather variation (R. Carlson, *personal communication*). Due to the high correlation between some input variables (Table 1), different combinations of covariates were tested in model selection.

Use of air temperature and precipitation in April resulted in a model with high explanatory power ($R^2 = 0.79$). Low temperature and precipitation in April increase SSR regional risk. Given the fact that planting takes place in late April or May and, thus, no crop is present in April, it is surprising the importance of spring weather on SSR prevalence. A hypothesis that might be interesting for further investigation is that spring weather could be related to the pathogen epidemiology. Only preconditioned sclerotia produce apothecia. Apothecia occur 4 to 8 weeks after completion of preconditioning, depending on the temperature and soil moisture (A. L. Mila, *unpublished data*). Thus, low April temperatures may contribute to sclerotia preconditioning that is necessary for the consequent appearance of apothecia. However, we should mention that the importance of April weather may be a result of the high correlation between April precipitation and July to August air temperatures. The correlation may create confounding between April precipitation and air temperature of July and August, a factor that has been found to be significant to the development of SSR epidemics (32).

The significance of July and August average temperature agrees with previous investigations (23,30,32) on the pathogen biology and epidemiology. July and August air temperatures were used as one variable (i.e., average) because the high positive correlation between them does not allow separating these 2 months as two independent predictors. Workneh and Yang (32) reported that precipitation was not a limiting factor for SSR prevalence in the north-central region in 1995 to 1998. The conclusion was made examining 10 locations (5 in Iowa and 5 in Minnesota). In the present study, we considered 1,545 locations (fields) in a large region that stretched from southern Illinois to northern Minnesota. The significance of precipitation on prevalence as a predictor may be due to the large number of locations distributed in the broad geographic area used in the present study. Precipitation should be an important factor because it affects not only sclerotia germination but also ascospore germination and disease development (8). Soybean is not irrigated in the north-central states; thus, precipitation is the only within-season source of moisture for the crop.

In our analysis, minimum tillage increases the risk of SSR prevalence in all states, whereas no significant difference was found between conventional and no tillage. This result is not in complete agreement with a previous analysis (32), where prevalence of SSR was reported to be significantly less in no-tillage fields than in conventional- or minimum-tillage fields and greater in minimum-tillage than in conventional-tillage fields. In that analysis, tillage was examined independently of any other factors; whereas, in the present study, more factors and their interactions were incorporated into the analysis. Differences between previous and current analysis may be due to the different approaches followed in the two studies.

Although both model I and model II incorporate weather variables from different months, they present a similar structure: in both models, air temperature and precipitation are significant, with the significance of air temperature depending upon tillage practices. Both models predict a higher risk for Iowa and Minnesota (i.e., larger intercepts) and lower for Illinois and Ohio and reveal the importance of minimum tillage in the increased regional prevalence of SSR.

TABLE 5. Parameter estimates of the Poisson regression used to explain the incidence of soybean *Sclerotinia* stem rot (caused by *Sclerotinia sclerotiorum*) using summer weather variables and management practices

Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
Intercept	6.88 ^a	1.3	13.28 ^a	2.08
Average temperature of July and August	-0.46 ^a	0.055	-0.68 ^a	0.09
Iowa ^b	0.71 ^a	0.28	0.48	0.35
Minnesota ^b	0.98 ^a	0.29	0.46	0.37
Ohio ^b	-1.67 ^c	0.75	-2.1 ^a	0.77
July precipitation	0.028 ^a	0.01	0.033 ^d	0.02
August precipitation	0.037 ^a	0.014	0.021	0.02
Average temperature of July and August × no tillage ^c	-0.022 ^d	0.013	-0.027 ^d	0.016
Average temperature of July and August × minimum tillage ^c	0.024 ^a	0.006	0.008	0.01
Clay	...	...	-0.03 ^c	0.016
June precipitation	...	...	0.044 ^d	0.024
Seed treatment	...	...	1.63 ^d	0.87
Manure	...	...	-18.63 ^a	5.44
Weed cultivation	...	...	2.42 ^c	1.14
Fertilizer application	...	...	-16.5 ^a	3.63
Seed treatment × August precipitation	...	...	-0.24 ^c	0.11
Manure × average temperature	...	...	0.83 ^a	0.23
Weed cultivation × June precipitation	...	...	-0.37 ^a	0.14
Fertilizer × clay	...	...	0.63 ^a	0.15
Fertilizer × average temperature	...	...	0.082 ^a	0.03
Deviance	1,195	...	628	...
Log(likelihood)	-528.7	...	-286.8	...

^a Level of significance, 1%.

^b Indicator variable for regional effect (Illinois, 0; Iowa, 1; Minnesota, 2; Ohio, 3).

^c Level of significance, 5%.

^d Level of significance, 10%.

^e Indicator variable for tillage practices (conventional, 0; no-tillage, 1; minimum-tillage, 2). Poisson regression model: $P(y) = \lambda^y e^{-\lambda} / y!$, where y is the number of diseased soybean plants (1, 2, ..., 20) and $\lambda = \exp(\sum b_i X_i)$, with X_i as variables and b_i as parameter estimates presented in the table.

Both models tend to overestimate SSR prevalence when the observed prevalence is low. This overestimation was mainly observed in the upper three areas of Minnesota, where the number of sampled fields was small during all the years of the investigation and, thus, may not be sufficiently informative about actual levels of disease prevalence. For instance, in 1998 in Minnesota, there was only one sampled field from the area with predicted prevalence 0.11 (Fig. 2). Thus, the observed prevalence could be only 0 or 100% and, in no case, could be in close accordance with our predictions. Another possible reason for overestimation is the absence of inoculum in fields that have SSR-favorable conditions, which has been suggested in white mold prediction in bean (19). Variables accounting for the presence of the pathogen were not incorporated into our models because inoculum information was not available.

Studies on soybean and other crops have demonstrated that apothecia of *S. sclerotiorum* occur after closure of the crop canopy, after crop flowering, and when prolonged plant surface wetness is present with air temperatures that are low to moderate (8,15,28,32). Especially in soybean, apothecia were observed during middle to late July under dense, cool canopy around the same time that soybean flowering occurred (8). Disease initially occurred after rain had initiated periods of prolonged plant surface wetness, whereas SSR epidemics were observed later, in early to mid-August (8). Given this information on *S. sclerotiorum* epidemiology and the limitations of the available data, we argue that model II, which includes average air temperature of July and

August, precipitation of July, and tillage practices, provides a more reliable and biologically justifiable explanation than model I for the regional prevalence of SSR in the north-central region.

SSR incidence could not be adequately explained by the same variables used to model SSR prevalence. Use of regional weather variables as inputs to predict incidence generally overestimated SSR incidence, indicating that site-specific environment or production conditions are less conducive to the occurrence of SSR compared with regional weather. There are three possibilities. First, planted cultivars in some of the sampled fields might not be susceptible to SSR. The regional models were based on the assumption that planted cultivars were uniformly susceptible across the region, although some soybean cultivars moderately resistant to SSR are available and possibly were planted in some of the sampled fields. Therefore, SSR might not have occurred, although environmental conditions were conducive. Second, it is well established in the literature that SSR incidence is affected by a number of microclimatic factors (1,22,27,28). Environmental factors such as daily air temperature under the soybean canopy, frequency, and total daily rainfall or soil moisture are more appropriate variables to describe SSR incidence than monthly average temperature and total precipitation (8,23,25,27,30). Furthermore, management practices that were not considered in the study, such as row spacing, may account significantly for the SSR field incidence (15,26,36). Last, information on field inoculum level is important because it has been shown to be highly correlated with the incidence of Sclerotinia rot in several studies (11,16,17,20). Nevertheless, we were not able to integrate these aspects into our model because data were not available.

Another regional prediction system of SSR is available for fungicide applications in canola (21). This system predicts the risk of sclerotia germination based on weather data and the top (0 to 10 cm) soil moisture content. The underlying assumption in this system is that, if sclerotia germinate, ascospores will be released and infection will occur. The risk is considerably higher during flowering stage, because flowers are the main court of infection for *S. sclerotiorum* (1). Similarly to our results, this prediction system suggests that local weather and field conditions may result in different risk levels than those of the regional predictions (21).

Generally, producers' decisions on pest management are based on field-level information (i.e., disease incidence). Thus, information on daily temperature and precipitation under the soybean canopy in a field is needed so that the anticipated level of SSR incidence can be defined with a high degree of accuracy. SSR prevalence is currently very low in north-central states. Under these circumstances, it is questionable whether producers will be willing to spend time and effort collecting within-field information to determine the anticipated SSR incidence or to change their management practices to minimize the risk of a rare disease. In the current literature (22,37), it has been suggested that, for rare diseases, producers may depend mainly on their past experience to make decisions; therefore, prevalence could be practically efficient information for decision making, though less precise and informative than incidence.

ACKNOWLEDGMENTS

The study was supported by Hatch Act and State of Iowa funds. Journal Paper No. J-19474 of the Iowa Agriculture and Home Economics Experimental Station. This manuscript is a portion of the first author's Ph.D. thesis submitted to the Graduate College of Iowa State University. We thank S. Wegulo for critically reviewing the manuscript.

LITERATURE CITED

1. Abawi, G. S., and Grogan, R. G. 1975. Source of primary inoculum and effects of temperature and moisture on infection of beans by *Whetzelinia sclerotiorum*. *Phytopathology* 65:300-309.

Fig. 5. Comparison of the predicted and the observed incidence of soybean Sclerotinia stem rot (caused by *Sclerotinia sclerotiorum*) during the 4-year survey (1995–98) when A, summer weather and tillage and B, summer weather and several management practices are used as model input variables. If a predicted value is equal to the observed value, this point should be on the line. Model parameter estimates are in Table 5.

2. Abawi, G. S., and Grogan, R. G. 1979. Epidemiology of diseases caused by *Sclerotinia* species. *Phytopathology* 69:899-904.
3. Agresti, A. 1996. An introduction to categorical data analysis. John Wiley & Sons, New York.
4. Anonymous. 1998. Crop Residue Management: Midwest Region. Conservation Technology and Information Center. West Lafayette, IN.
5. Armitage, P. 1971. Statistical Methods in Medical Research. 2nd ed. John Wiley & Sons, New York.
6. Bamber, D. 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* 12:387-415.
7. Blad, B. J., Steadman, J. R., and Weiss, A. 1978. Canopy structure and irrigation influence white mold disease and microclimate of dry edible beans. *Phytopathology* 68:1431-1437.
8. Boland, G. J., and Hall, R. 1988. Epidemiology of *Sclerotinia* stem rot of soybean in Ontario. *Phytopathology* 78:1241-1245.
9. Chamberlain, D. W. 1951. *Sclerotinia* stem rot of soybean. *Plant Dis. Rep.* 35:490-491.
10. Chapman, D. G., and Nam, J. 1968. Asymptotic power of chi-square tests for linear trends in proportions. *Biometrics* 24:315-327.
11. Crowe, F. J., Hall, D. H., Greathead, A. S., and Baghott, K. G. 1980. Inoculum density of *Sclerotinia cepivorum* and the incidence of white rot of onion and garlic. *Phytopathology* 70:64-69.
12. Doupnik, B., Jr. 1993. Soybean production and disease loss estimates for north central United States from 1989 to 1991. *Plant Dis.* 77:1170-1171.
13. Garcia-Garza, J. A., Boland, G. J., and Vyn, T. J. 1998. Influence of crop rotation and reduced tillage on white mold of soybean caused by *Sclerotinia sclerotiorum*. (Abstr.) *Phytopathology* 88:533.
14. Grau, C. R. 1989. *Sclerotinia* stem rot. Pages 47-48 in: Compendium of Soybean Diseases. 3rd ed. J. B. Sinclair and P. A. Backman eds. The American Phytopathological Society, St. Paul, MN.
15. Grau, C. R., and Radke, V. L. 1984. Effects of cultivars and cultural practices on *Sclerotinia* stem rot of soybean. *Plant Dis.* 68:56-58.
16. Gugel, R. K., and Morrall, R. A. A., 1986. Inoculum-disease relationships in *Sclerotinia* stem rot of rapeseed in Saskatchewan. *Can. J. Plant Pathol.* 8:89-96.
17. Holley, R. C., and Nelson, B. D. 1986. Effect of plant population and inoculum density on incidence of *Sclerotinia* wilt of sunflower. *Phytopathology* 76:71-74.
18. Hosmer, D., and Lemeshow, S. 1989. Applied Logistic Regression. John Wiley & Sons, New York.
19. Hunter, J. E., Abawi, G. S., and Crosier, D. C. 1978. Effects of timing, coverage, and oil additive on control of white mold of snap bean with benomyl. *Plant Dis. Rep.* 62:633-637.
20. Imolehim, E. D., and Grogan, R. G. 1980. Factors affecting survival of sclerotia and effects of inoculum density, relative position and distance of sclerotia from the host on infection of lettuce by *Sclerotinia minor*. *Phytopathology* 70:1162-1167.
21. Lamey, A., and Bradley, C. A. 2003. *Sclerotinia* stem rot of canola: Biology and management. N.D. State Univ. Ext. Serv.
22. Mila, A. L., Carriquiry, A. L., Zhao, J., and Yang, X. B. 2003. Impact of management practices on prevalence of soybean *Sclerotinia* stem rot in the north-central United States and on farmers' decisions under uncertainty. *Plant Dis.* 87:1048-1058.
23. Moore, W. D. 1955. Relation of rainfall and temperatures to the incidence of *Sclerotinia sclerotiorum* in vegetables in South Florida during the years 1944 to 1954. *Plant Dis. Rep.* 39:470-472.
24. Morrall, R. A. A., and Dueck, J. 1982. Epidemiology of *Sclerotinia* stem rot of rapeseed in Saskatchewan. *Can. J. Plant Pathol.* 4:161-168.
25. Singh, U. P., and Singh, R. B. 1983. The effect of soil texture, soil mixture, soil moisture and depth of soil on carpogenic germination of *Sclerotinia sclerotiorum*. *J. Plant Dis. Prot.* 90:662-669.
26. Steadman, J. R., Coyne, D. P., and Cook, G. E. 1973. Reduction of severity of white mold disease on great northern beans by wider row spacing and determinate plant growth habit. *Plant Dis. Rep.* 12:1070-1071.
27. Sun, P., and Yang, X. B. 2000. Light, temperature and moisture effects on apothecium production of *Sclerotinia sclerotiorum*. *Plant Dis.* 84:1287-1293.
28. Tu, J. C. 1989. Management of white mold of white beans in Ontario. *Plant Dis.* 73:281-285.
29. Turkington, T. K. Morrall, R. A. A., and Baker, R. J. 1988. Sample sizes in relation to forecasting *Sclerotinia* stem rot of canola. *Can. J. Plant Pathol.* 10:159-165.
30. Weiss, A., Kerr, E. D., and Steadman, J. R. 1980. Temperature and moisture influence on development of white mold disease (*Sclerotinia sclerotiorum*) on great northern beans. *Plant Dis.* 64:757-759.
31. Williams, J. R., and Stelfox, D. 1980. Influence of farming practices in Alberta on germination and apothecium production of sclerotia of *Sclerotinia sclerotiorum*. *Can. J. Plant Pathol.* 2:169-172.
32. Workneh, F., and Yang, X. B. 2000. Prevalence of *Sclerotinia* stem rot of soybean in north-central United States in relation to tillage, climate, and latitudinal positions. *Phytopathology* 90:1375-1382.
33. Workneh, F., Yang, X. B., and Tylka, G. L. 1999. Soybean brown stem rot, *Phytophthora sojae*, and *Heterodera glycines* affected by soil texture and tillage relations. *Phytopathology* 89:844-850.
34. Wrather, J. A., Anderson, T. R., Arsyad, D. M., Gai, J., Plopper, L. D., Porta-Puglia, A., Ram, H. H., and Yorinori, J. T. 1997. Soybean disease loss estimates for the top 10 soybean production countries in 1994. *Plant Dis.* 81:107-110.
35. Wrather, J. A., Anderson, T. R., Arsyad, D. M., Tan, Y., Plopper, L. D., Porta-Puglia, A., Ram, H. H., and Yorinori, J. T. 2001. Soybean disease loss estimates for the top 10 soybean producing countries in 1998. *Can. J. Plant Pathol.* 23:115-121.
36. Yang, X. B., Lundeen, P., and Uphoff, M. D. 1999. Soybean varietal response and yield losses caused by *Sclerotinia sclerotiorum*. *Plant Dis.* 83:456-461.
37. Yuen, J. E., and Hughes, G. 2002. Bayesian analysis of plant disease prediction. *Plant Pathol.* 51:407-412.