Source: Plant Protection [Zhiwu Baohu, ISSN 0529-1542; CN 11-1982/S]
(1990) v.16 (6) p.19-21.
Translated by Dunxiao Hu, China Agricultural University; Edited by Xiaorong Wu, Kansas State University, 2003

Study of the Prediction of Developmental Stage and Population Size of Soybean Aphid in Northern Liaoning, China

ZHENGREN TIAN[1], SHUYAN ZHAO[2], CHENXIAO HU[2]
1. Plant protection station, Tieling, Liaoning   2. Tieling Center of computing techniques, Liaoning

ABSTRACT A model for predicting the days for aphids to reach 10 thousand/100 plants in Northern Liaoling province was established by stepwise regression. In order to include enough parameters in the stepwise regression model to predict the aphid population, parameters were carefully selected by path analysis at first, and then a multiple regression model was established (F=0). At the same time, a stepwise discriminant model was established too. Integrated prediction was conducted by the combination of these two methods. Prediction fed with historical data had 100% fitness. Forecasting tests in the past two years were all correct.

KEYWORDS soybean aphid; stepwise regression; path analysis; integrated prediction

Soybean aphids are widely distributed in the soybean production area. It is an important insect pest in the soybean-growing area and usually causes severe damage in Liaoning, Jiling, Heilongjiang, and Hebei provinces. In northern Liaoning the economic threshold was usually set at 10 thousand aphids/ 100 plants.

Methods

From 1988 to 1989, we employed the methods of stepwise regression analysis and stepwise discrimination on an IBMPC/XT compatible computer to predict the days needed for aphids to reach 10 thousand/ 100 plants, and the size of aphid populations.

Results and analysis

1. Forecasting of occurrence date of soybean aphids by stepwise regression

1.1 Soybean number data and meteorological data from 1975 to 1984 (missing the data of 1980) were used in the analysis. Thirty-four sets of soybean aphid data and corresponding meteorological data were considered independent variables, and days for aphid to reach 10 thousand/100 plants was considered the dependent variable (converting 31, May into 0). All data were entered into the computer.
1.2 The F-test was used to determine the retention or deletion of a variable with a standard of $F_1=4$, $F_2=4$. The model was established as follows
$\hat{Y}=62.1267-0.0016x_1-0.6139x_2+0.7162x_3-17.527x_4-0.4335x_5\pm0.2928$
Where, $\hat{Y}$ is the expected value (days) for soybean aphids to reach 10 thousand/100 plants;

X_1 is the aphid number/100 plants on June 6; X_2 is the days with precipitation of ≥ 0.1 mm in the first 10 days of June; X_3 is the precipitation (mm) in the last 10 days of June; X_4 is the temperature/precipitation coefficient in the last 10 days of June; X_5 is hours of sunshine during the second 10 days of June.

1.3 Evaluation of forecast quality

Multiple correlation coefficient R= 0.9999, Sample standard deviation from regression Sy = 0.2928. Historical fitness rate = 100% (see Table 1). Permitted errors (day) for predicting control date are: ± one day for short term forecasting; ± two days for mid-term forecasting; ± three days for long-term forecasting. Accuracy rate was 100% during permitted error range. If the predicted date was one day out of the permitted range, accuracy rate was considered 70%; if it was 2 days out of the permitted range, the accuracy was 50%; 3 days out of the permitted range, the accuracy rate was considered zero.

Table 1. Historical fitness rate 1988 Tieling

Year	Practice value June/days	Predicted value June/days	Fitness
1997	11	11	√
1978	17	17	√
1979	20	20	√
1981	30	30	√
1982	27	27-28	√
1983	8	8-9	√
1984	31	31	√

The date for reaching 10 thousand aphids/100 plants in 1988 was as follows:

$$Y=62.1267 - 0.0016\times 2 - 0.6139\times 2 + 0.7162\times 37.3 - 17.527\times 1.62-0.4345\times 76.5 \pm 0.2928$$
$$=25.977 \pm 0.2928 = 25.6842\sim 26.2698 \approx 26$$

Number of aphids /100 plants was 9746 on June 28 in the observation plots (no control), which was two days from the predicted date, but still was correct. The aphid number on June 6 1989 was 9086/100 plants, very close to 10 thousand/100 plants. No further calculation was needed. Test forecasting was correct in both 1988 and 1989. Therefore, this model could be used in mid-term forecasting of soybean aphid infestation for most areas of northern Liaoning.

2. Forecast of soybean aphid population by multivariate linear regression

2.1 The data of soybean aphids in field surveys and corresponding meteorological observation data that came from 1973-1984 (missing data of 1974, 1975, and 1977) was studied. Simple correlation analysis was employed to examine the correlation coefficients of 71 primary parameters. Thirty-two parameters with correlation coefficient of r > 0.4 were selected as independent variables.

2.2 Parameters were further examined and screened by stepwise regression and path analysis. The results of stepwise regression indicated that when F_1= 2.5, F_2= 2.5, seven factors were selected with R= 0.99999 and Sy= 1000.226. When F value increased to 3.49, still seven factors were retained without any change of the R-value. When F_1 and F_2 values were both increased to 3.5, only one factor was retained with R= 0.7898 and Sy= 48221.24. Thus, the F-value should be between 3.49 and 3.5 with 8 places of decimal, and the number of parameters was still seven. However, no satisfactory results were obtained because of the low computing speed and long computing time. A path analysis method was applied to select three variables with major direct effects from these seven factors and to establish a multivariate linear regression model (stepwise regression analysis with F=0). After analyzing the effects of each parameter on the Y, we may understand direct or indirect effects of each variable (see Table 2).

Table 2. Direct effect and indirect effect of variables

Factors / Corr. coe. / Effe mod / Path coe.		X_1	X_2	X_3	X_4	X_5	X_6	X_7
		-0.2614	-0.5891	0.644	0.6286	-0.5824	0.4202	-0.7637
Direct effect		0.2459	-0.2192	0.7219	0.6308	-0.1759	0.44599	0.1671
Indirect effect	By X_1		-0.05211	-0.5862	0.0543	-0.1735	-0.1529	0.0203
	By X_2	0.0585		-0.1612	-0.3087	-0.0391	0.0237	0.1043
	By X_3	-0.1997	0.0489		-0.01249	0.1366	0.06147	-0.1128
	By X_4	0.0212	0.1073	-0.0143		-0.01098	-0.0414	-0.0479
	By X_5	0.2426	-0.0487	-0.5609	0.0394		0.1768	0.09779
	By X_6	-0.0843	0.0165.	0.0995	-0.0585	0.0697		-0.06373
	By X_7	0.1355	-0.1369	-0.4873	-0.1691	-0.1029	0.1701	

$\hat{Y}$= -299439.2+1951.41x_1+24210.98x_2+43989.16x_3 ± 12452.41 with R= 0.991 and Sy= 12452.41

Where X_1= precipitation (mm) from late April to early May; X_2= average temperature (°C), during middle April; and X_3= temperature-precipitation coefficient (E_R) from late April to early May.

2.3 Evaluation of forecast quality

Historical fitness rate was 100% (see Table 3). This model was used for long-term prediction. The theoretical peak aphid number was between 80211 and 105105 per 100 plants, which forecasted a severe, third-level outbreak. The actual peak aphid number was 75800 per 100 plants, and the outbreak was of mild severity and level two.

Table 3 Historical fitness rate

Year	Practice value	Prediction value	Fitness
1973	5059	0-12153	√
1976	85973	75604-100509	√
1978	34494	35945-60850	Basic fitness
1979	30400	24636-49540	√
1980	226	0-3206	√
1981	37935	29866-54771	√
1982	114250	82766-107671	Basic fitness
1983	232204	220945-345851	√
1984	22100	15277-40181	√

3. Forecasting aphid population by stepwise discrimination

3.1 Infestation level was classified into three levels according to the aphid numbers in each year (Table 4 and 5). Variables were screened by a discriminating program. By referring to the F distribution table, we found $F_{\alpha\ 0.1}$= 3.26 and $F_{\alpha\ 0.05}$= 4.74. When $F_k > F_{\alpha\ 0.05}$ was applied, 25 variables were selected. When $F_k > F_{\alpha\ 0.1}$, 32 variables may be included in the model.

Table 4

Intensity of infestation	(1) slight	(2) middle	(3) Severe
Peak no. of aphid/100 plants ($\times 10^4$)	<1.0	<8.0	>8.0

Table 5. Intensity and level of aphid infestation

Year	1973	1976	1978	1979	1980	1981	1982	1983	1984
Intensity of infestation	slight	heavy	middle	middle	slight	middle	heavy	heavy	middle
Level	1	3	2	2	1	2	3	3	2

3.2 Establishment of a discriminating forecasting model

Discriminating models were established after stepwise discriminant analysis of the 32 variables and levels of aphid infestation in Tieling.

$$f(1)= -1432145+3661.65x_{12}+909.29x_{13}+3621.48x_{15}-1778.415x_{17}-3176.26x_{21}-93764.74x_{32}$$

$$f(2)= -15402+1200.475x_{12}+298.2993x_{13}+1186.996x_{15}-582.9617X_{17}-10415.04X_{21}-30727.69x_{32}$$

$$f(3)= -21395+1415.201x_{12}+351.3064X_{13}+1398.798X_{15}-687.1263X_{17}-12278.45X_{21}-36205.74X_{32}$$

Where, X_{12}= precipitation (mm) in the early and middle May; X_{13}= precipitation from late May to early June; X_{15}= precipitation in middle and late June; X_{21}= E_R (temperature-precipitation coefficient in middle May); X_{32}= E_R (temperature-precipitation coefficient in middle and late June); X_{17}= precipitation (mm) from mid-April to early and mid-May.

3.3 Testing forecast

The predicated values were f(1)= -89199, f(2)=2326, and f(3)= -511 in 1988. Because f(2) was the highest value, a level-two outbreak was forecast. The actual aphid number was 33790 /100 plants, which was a level-two category. The forecast was correct. In 1989, the predicated values were f(1)= -80525*, f(2)=5154, and f(3)= 2830. A level-two outbreak was forecast because f(2) was the largest value. The actual peak aphid number was 75800/100 plants, infestation of level two. The forecast was correct again. Historical fitness rate was 100% (see table 6). This model could be used in mid-term forecasting in northern Liaoning Province.

Table 6. Discriminant matrix

Original classification \ Calculated classification	1	2	3	Σ
1	2	0	0	2
2	0	4	0	4
3	0	0	3	3
Σ	2	4	3	9

4. Integrated forecast

Because the predication results from stepwise regression were different from those by the stepwise discriminant method, a weighted integration formula, $y= Y_1 \times r + Y_2 \times PC/r + PC$, was used to forecast, where y= results of integrated forecast, Y_1= multiple correlation coefficient of stepwise regression analysis, Y_2= forecasting result of stepwise discrimination, PC= rate of historical fitness. The results of integrated forecast were very good for the 1988 and 1989 data. The results of discriminating and integrated forecast in 1989 were both level-two. The forecast was correct. The weighted integration of both forecasting results by using multiple correlation coefficient and historical fitness made it possible to use both correlation of variables and discriminant analysis, and thus gave more accurate results. Therefore, it could be used in mid-term forecasting in the northern area (represented by Tieling) of Liaoning Province.

Discussion

Occurrence of soybean aphid was mainly affected by temperature and humidity. The reproduction of aphids increased rapidly in intermediate humidity and relatively higher temperatures. It was clear that variables in the forecasting model influenced the reproduction of soybean aphid. These variables mainly affected the hatching of overwintering eggs into fundatrices and the number of generations aphids can reproduce.

* Translator’s annotation: Value of f(1)=80525 in the original may be mistake . Its value may be -80525