

## Nonrandom Patterns of Bacterial Brown Spot in Snap Bean Row Segments

B. D. Hudelson, M. K. Clayton, K. P. Smith, D. I. Rouse, and C. D. Upper

Department of Plant Pathology (all authors), Department of Statistics (second author), and Agricultural Research Service, U.S. Department of Agriculture (fifth author), University of Wisconsin, Madison 53706.  
Cooperative research of the College of Agriculture and Life Sciences, University of Wisconsin, Madison, and the Agricultural Research Service, U.S. Department of Agriculture.

Research supported in part by USDA Competitive Grant 85-CRCR-1-1662.

We thank J. Diedrich, K. Hoffman, J. Pitzer, and P. Welch for valuable technical assistance and S. Vican for preparation of the figures.

We thank R. Durbin, S. Hirano, and B. Yandell for reviewing this manuscript.

We also thank the snap bean growers of Wisconsin for their cooperation in providing commercial snap bean fields for this work. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

Accepted for publication 6 February 1989 (submitted for electronic processing).

---

### ABSTRACT

Hudelson, B. D., Clayton, M. K., Smith, K. P., Rouse, D. I., and Upper, C. D. 1989. Nonrandom patterns of bacterial brown spot in snap bean row segments. *Phytopathology* 79:674-681.

Each leaflet on every plant in 37 5-m row segments and a single 12-m row segment from commercial snap bean fields was assessed for bacterial brown spot. Graphs of the proportion of diseased leaflets per plant (disease incidence values) versus plant position along the row suggested two types of nonrandom variability in disease: an extreme jaggedness superimposed on a slow, undulating change in disease. Arcsine square root-transformed disease incidence values were analyzed for spatial nonrandomness using three techniques: runs analysis, autocorrelation analysis, and autoregressive integrated moving average (ARIMA) modeling. All three

techniques detected the slow, undulating change in disease incidence values; however, only ARIMA modeling detected the jaggedness and could quantify both patterns. A "generalized ARIMA(1 0 1) model" was found to describe 35 of the 38 data sets. The biological mechanism generating these patterns is unknown. Knowledge of the existence of such patterns is important for developing effective sampling strategies for this disease. Theoretical characteristics of the generalized ARIMA(1 0 1) model indicate that random start systematic sampling will provide a better estimate of total or mean brown spot in a row than simple random sampling.

*Additional keywords:* adaptive sampling, *Pseudomonas syringae*.

---

Previous work has shown that a quantitative relationship exists between the population size of *Pseudomonas syringae* pv. *syringae* van Hall and incidence of bacterial brown spot on snap beans (*Phaseolus vulgaris* L.) (21,26,38). The hazard of this disease can be predicted approximately 1 wk in advance from estimates of the frequency with which large epiphytic populations of the pathogen occur on bean leaflets. A model has been developed that describes the relationship between pathogen population size and disease incidence (38) by combining the frequency distribution

of population sizes of pathogenic bacteria on individual leaflets and the probability of disease given population size. The lognormal distribution is used to describe population size of pathogenic bacteria on individual leaves or leaflets (20), and the probit function is used to describe the probability of disease given bacterial population size (16). Although these studies have increased our knowledge of the temporal variation of brown spot disease incidence, they have not addressed an equally important component of the epidemiology of this disease: its spatial variation.

Several studies suggest that nonrandom patterns of diseases caused by pathovars of *P. syringae* may exist (15,25,34). As early as 1920, Elliott (15) reported that halo blight of oats occurs in foci within a field. More recently, Lindemann et al (25)

---

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 1989.

demonstrated that both population size of *P. s. syringae* and bacterial brown spot incidence differed among bean plots that were separated spatially within a 64-km transect, even though the plots were sown at the same time and with the same lot of seeds. In addition, Poushinsky and Basu (34) have determined that bacterial blight of soybean, caused by *P. s. pv. glycinea* (Coerper) Young, Dye & Wilkie, can occur in a nonrandom pattern.

To date, most studies of spatial pattern have used individual quadrats as a spatial sampling unit (19,30,37,42). Often quadrat data are analyzed by determining the frequency distribution of quadrats with given levels of disease or inoculum. If the observed frequency distribution is not significantly different from the Poisson distribution, the disease or inoculum is said to be distributed randomly (33). Because this test does not take into account the location of each quadrat, it does not provide any information about the spatial pattern of the disease or the inoculum at a scale greater than the quadrat (31,33). If any additional search for spatial pattern is made, it is made at multiples of the initial quadrat size (30,40). This approach is useful but is unlikely to detect patterns that do not occur at a scale that is an even multiple of the initial quadrat size. In addition, if quadrat data are based upon samples taken randomly from within quadrats, as is frequently done, then any nonrandom pattern within quadrats will not be detected. Furthermore, it is assumed that samples provide an adequate estimate of the disease or inoculum within the quadrats. In most plant disease studies, this assumption has neither been tested nor validated and may be suspect if there are nonrandom patterns within the quadrats.

A number of methods of analysis of spatial pattern do not require quadrat data; instead, the location of each sample is used (1,9,43). For distance methods (12,13), the basic sampling unit is a point, for example, a plant, and the data collected are the distances between diseased plants (18,35). Usually, distance methods note only the presence or absence of disease but not the amount of disease on the sampling unit. In the runs test (17), the number of groups (runs) of diseased and healthy plants along a line is recorded. This method has been used for row crops to determine if disease occurs nonrandomly along rows (28,29). Finally, in correlation analysis, the value of each sample is correlated with that of its neighbors (3,9,32,36,37). Thus far this technique has been used only to study quadrat data, but, as with frequency approaches, these applications may suffer by ignoring spatial patterns that occur within quadrats.

The most intimidating problem facing anyone wishing to measure spatial patterns of disease in agricultural settings centers on the immensity of the task of sampling from an entire field in a way that can detect and quantitate patterns that may occur at several scales. Cost and manpower constraints rule out a census approach that monitors disease or pathogen populations on all leaflets, leaves, or even plants in a field. Thus a theoretically sound approach for sampling, from quantitating, and describing unknown spatial patterns of plant diseases at all scales within a field is needed. The strategy we are currently developing, which we call "adaptive sampling," is an iterative procedure that starts with small areas and proceeds to increasingly larger areas, adjusting sampling methods in each iteration to consider the patterns present in the smaller areas. This paper describes the spatial patterns of brown spot that we have detected during the first iteration of this method.

## MATERIALS AND METHODS

**Sample collection.** Thirty-seven 5-m snap bean row segments and a single 12-m snap bean row segment were monitored for natural infections of bacterial brown spot during the 1985, 1986, and 1987 growing seasons. In 1985, six 5-m  $\times$  4-row experimental plots and a single 5-m  $\times$  5-row experimental plot were established in diseased areas of two commercial snap bean fields (cultivar Goldrush) located near Arlington, WI (Columbia County). During July of that year, all row segments within these plots were assessed for bacterial brown spot as described below. In July of 1986,

two diseased areas were identified within a commercial snap bean field (cultivar Galamore) located near Spring Green, WI (Iowa County), and three adjacent 5-m row segments from the first area and a single 5-m row segment from the second area were assessed for bacterial brown spot. In addition, a 12-m row segment, selected from a diseased area in a commercial snap bean field (cultivar Early Bird) located on the University of Wisconsin experimental farms, Arlington, WI, was assessed. In July of 1987, two adjacent Arlington, WI, 5-m row segments, selected from a diseased area of a commercial snap bean field (cultivar Bush Blue Lake 109) located on the University of Wisconsin experimental farms, Arlington, WI, were assessed for bacterial brown spot. In addition, in August of 1987, two separate 5-m row segments were selected from diseased areas in a second snap bean field (cultivar Eagle) located on the University of Wisconsin experimental farms and assessed. Thus, we have sampled row segments for 3 yr and from six commercial-size plantings of snap beans ( $\geq 2$  ha), five snap bean cultivars, and two diverse locations in Wisconsin.

**Disease assessment.** Within a given row segment, each leaflet on every snap bean plant was assessed for the presence or absence of bacterial brown spot. Bacterial brown spot was distinguished from other foliar diseases of snap bean by its characteristic roughly circular, dark brown lesions with narrow yellow halos. For lesions of questionable etiology, bacterial isolations were made onto King's medium B (23) using the techniques outlined by Schaad (39). A predominance of *P. syringae* was considered evidence that the lesion in question was bacterial brown spot.

**Data analysis.** Within a given row segment, each plant was assigned a positive integer representing its position in the row. For example, the first plant in the row segment was assigned the number 1, the second plant number 2, etc. Disease incidence values were expressed as the proportion of diseased leaflets per plant; these proportions will subsequently be referred to as "untransformed disease incidence values." To stabilize the variance within a given data set, disease incidence values were transformed using an arcsine square root transformation (41). These "transformed disease incidence values" were used for subsequent analyses. The transformed disease incidence value of the  $i$ -th plant in a row segment is denoted  $Y_i$ . For row segments in which diseased plants were clustered near the center of the 5 m (data sets 8.1-8.3 and 9.1 in Table 1), only data from the central diseased portion of the rows were used for analysis. Extended areas of zeros were excluded from analysis in this way. For these data sets, the row length actually analyzed ranged between 1.85 and 2.81 m. All other data sets were analyzed in their entirety.

Transformed disease incidence values within a given row segment were initially analyzed using a two-sided runs test (17). A run was defined as a succession of transformed disease incidence values all of which were either above or below the median transformed disease level for the row segment being analyzed. Transformed disease incidence values equal to the median were omitted from the data set, and the runs analysis was performed on the remaining data points (17). For data sets where the median transformed disease level was zero, a runs test could not be performed. For the remaining data sets, the number of runs ( $U$ ) and the total number of transformed disease incidence values that were above ( $n_1$ ) or below ( $n_2$ ) the median transformed disease level were determined. Because  $n_1$  and  $n_2$  were greater than 10 in these data sets, a standard normal ( $Z$ ) test ( $\alpha = 0.05$ ) was used to test the null hypothesis that transformed disease incidence values were randomly arranged within row segments (17).

In addition to the runs test, transformed disease incidence values were described using autocorrelation (or lag correlation) analysis (4,7). For each data set, the sample autocorrelation function (ACF) was calculated using MINITAB, release 5.1.1 (The Pennsylvania State University, University Park). The sample ACF, denoted  $r_Y(s)$ , estimates the theoretical autocorrelation function, denoted  $\rho_Y(s)$ , which is the correlation in transformed disease incidence values for all pairs of plants that are  $s$  plants apart (for integer values of  $s > 0$ ). The integer  $s$  also is known

as the lag value. The sample ACF is calculated as follows (4):

$$r_Y(s) = \frac{\sum_{t=1}^{T-s} (Y_t - \bar{Y})(Y_{t+s} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

for  $s = 1, 2, 3$ , etc. As noted above,  $Y_t$  and  $Y_{t+s}$  represent the transformed disease incidence values for plants that are located at positions  $t$  and  $t + s$  within a given row segment.  $\bar{Y}$  is the mean transformed disease level, and  $T$  is the total number of plants in a given row segment. For the purposes of this analysis, the direction in which the plants are numbered (that is, either up or down a row segment) is immaterial.

An approximate 95% confidence interval (CI) was calculated for each value of  $\rho_Y(s)$ , for  $1 \leq s \leq 20$ , under the assumption that transformed disease incidence values were randomly arranged on plants within a given row segment. Under this assumption, the expected value of  $r_Y(s)$  is zero for all  $s > 0$ . In addition, the variance of  $r_Y(s)$  is approximately  $1/T$  for all  $s > 0$  (4), and therefore an approximate 95% CI for  $\rho_Y(s)$  is  $0 \pm 2/\sqrt{T}$  for all  $s > 0$ . The presence of an excess number of values of  $s$  for which  $r_Y(s)$  was outside these 95% CIs was considered evidence that transformed disease incidence values were not randomly arranged on plants within the row segment being studied. Because we used a 95% CI and evaluated 20 values of  $r_Y(s)$  for each data set, we expected (under the hypothesis stated above) to observe on average that one value of  $r_Y(s)$  would be outside its corresponding 95% CI (6,8). Therefore, in this context, "an excess number of values of  $s$ " was defined as two or more. In addition, single significant values of  $r_Y(s)$  were considered of interest if they occurred at low lags (that is, lags 1 or 2) or were far outside the CIs.

To obtain a clearer description of the spatial patterns of bacterial brown spot within snap bean row segments, data sets were modeled using autoregressive integrated moving average (ARIMA) models (4). These models are special cases of the spatio-temporal ARIMA (STARIMA) models described by Reynolds and Madden (36) and Reynolds, Madden, and Ellis (37). In contrast to the autocorrelation analysis, the direction in which plants are numbered in the row could have an effect on the results obtained by ARIMA modeling. Therefore analyses were conducted twice, using both orientations.

An ARIMA( $p$   $d$   $q$ ) model, with  $d = 0$ , has the following general form:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} + \delta.$$

The  $Y_s$  in this model are transformed disease incidence values. The  $\epsilon_s$  represent random "noise" components associated with each plant and are assumed to be independently, identically, and normally distributed with expected value,  $E(\epsilon_t) = 0$ , and variance,  $V(\epsilon_t) = \sigma^2$ , for all  $t$ . Thus the ARIMA model given above indicates that the transformed disease incidence value of any given plant can be described by the transformed disease incidence values of the  $p$  preceding plants in the row, as well as the random "noises" associated with the  $q$  preceding plants. The  $\phi_s$  and  $\theta_s$  in this model are constants that quantify this description. In most circumstances,  $\phi_s$  and  $\theta_s$  can be calculated directly from the  $Y_s$ . In other situations however (4),  $\phi_s$  and  $\theta_s$  can be estimated more readily by modeling the first difference of the data. Thus the  $Y_s$  in the equation above are replaced by  $W_s$ , where  $W_t = Y_t - Y_{t-1}$ . This is denoted with  $d = 1$  in the ARIMA( $p$   $d$   $q$ ) specification. The final parameter in the ARIMA model is the constant  $\delta$ , which is related to the expected value of an observation,  $Y_t$ , as follows:

$$E(Y_t) = \delta / (1 - \phi_1 - \dots - \phi_p).$$

Potential models for data sets were chosen based on the sample ACFs that were calculated during autocorrelation analysis. In

addition, sample partial autocorrelation functions (PACFs) (4) were calculated for this purpose via MINITAB using a recursive method developed by Durbin (14). Sample PACFs are denoted  $\hat{\phi}_Y(s)$  and should not be confused with parameters ( $\phi_s$ ) in the ARIMA model described above. PACFs, like sample ACFs, estimate correlations between  $Y_t$  and  $Y_{t+s}$  for integer values of  $s > 0$ . However, PACFs take into account and correct for the fact that correlations at a given distance may result from correlations that occur at shorter distances. An approximate 95% CI for  $\hat{\phi}_Y(s)$  was calculated in a manner similar to that described for  $\rho_Y(s)$  above (4).

Sample ACFs and PACFs were compared with theoretical ACFs and PACFs derived from particular ARIMA models (4). If the sample and theoretical functions were similar, the corresponding ARIMA model was fit to the data set. For these selected models, least squares estimates of  $\phi_s$  and  $\theta_s$  and the standard errors of these estimates were obtained using the ARIMA command of MINITAB. Estimates were considered significantly different from zero if their absolute values were at least twice their standard errors. This procedure is approximately equivalent to performing an appropriate  $t$ -test with  $T - p - q - 1$  degree of freedom and  $\alpha = 0.05$  under the null hypothesis that  $\phi = 0$  (or  $\theta = 0$ ). If a  $\phi$  or  $\theta$  did not appear to be significantly different from zero, then an ARIMA model without this term was fit to the data set. Terms were restored to the model if diagnostic procedures (see below) indicated that the reduced model was inadequate. The general process of model building given above is similar to that outlined by Chatfield (8).

The adequacy of models was evaluated using the diagnostic tools of Box and Jenkins (4). These diagnostic tools are based on the assumption that a model that fits a data set adequately should yield residuals ( $\hat{\epsilon}_t$ ) that exhibit a random pattern. Plots of residuals versus predicted  $Y_s$  and residuals versus plant position along the row were evaluated to determine if any systematic patterns (that is, fan-shaped or curved patterns) were present that might indicate that model assumptions had been violated (4,6,8). In addition, sample ACFs and PACFs for residuals  $r_{\hat{\epsilon}}(s)$  and  $\hat{\phi}_{\hat{\epsilon}}(s)$ , respectively, were calculated in a manner similar to that described above for transformed disease incidence values. The expected values of  $r_{\hat{\epsilon}}(s)$  and  $\hat{\phi}_{\hat{\epsilon}}(s)$  under the assumption of randomness are zero. Corresponding 95% CIs were calculated using the method of Box and Pierce (5). A fitted model was considered inappropriate if, for two or more values of  $s$ ,  $r_{\hat{\epsilon}}(s)$  or  $\hat{\phi}_{\hat{\epsilon}}(s)$  was outside the corresponding 95% CIs (6,8). In addition to calculating sample ACFs and PACFs of residuals, the modified Box-Pierce statistic,  $Q$ , defined as:

$$Q = T(T+2) \sum_{j=1}^K [r_{\hat{\epsilon}}(j)]^2 / (T-j)$$

with  $K = 24$ , was calculated (27). The  $Q$  statistic is used to determine if values of  $r_{\hat{\epsilon}}(s)$  taken as a group rather than individually are larger than expected as compared with random data. If the fitted model is adequate, then  $Q$  is distributed as a  $\chi^2$  variable with  $K - (p + q)$  degrees of freedom, and a  $P$ -value for  $Q$  can be calculated based on this distribution. The choice of  $K$  was based on the recommendations of Brockwell and Davis (6).

## RESULTS

Untransformed disease incidence values were variable both between and within data sets. The mean untransformed disease incidence values of individual data sets (based on unweighted averages of untransformed disease incidence values of individual plants) ranged between 0.4 and 58.3% (Table 1); half of the data sets had average untransformed disease incidence values between 9.3 and 33.1% (the first and third quartiles, respectively). Within data sets, diseased plants were either concentrated within row segments and surrounded by healthy plants as illustrated by Figure 1 or present continuously over the entire 5 or 12 m as illustrated by Figures 2 and 3. Untransformed disease incidence values of individual plants ranged as much as 100 percentage points (for example, from 0 to 100%) within some data sets to as little as

14.3 percentage points (for example, from 0 to 14.3%) in others. Among the 38 data sets, the median number of leaflets per plant ranged from 15.5 to 30.

Visually (Figs. 1 and 2), nonrandom variability appeared to be of two types. The most striking component was an extreme jaggedness of the plots. Plants with relatively high untransformed disease incidence values tended to alternate with plants with untransformed disease incidence values of lower magnitude. Underlying this jaggedness, there appeared to be a slowly undulating change in the local mean untransformed disease level. This phenomenon was particularly evident in data sets where diseased plants were concentrated within row segments (Fig. 1).

**Runs analysis.** Five of the 38 data sets could not be analyzed using the runs test because the median transformed disease incidence value for these data sets was zero. A runs analysis of the remaining 33 data sets indicated that, in general, transformed disease incidence values were not randomly located within row segments (Table 1). For 32 of these 33 data sets, a negative Z-value was obtained, indicating that the observed number of runs was less than the number of runs expected given a random arrangement

of disease within the row segment. Of these 32 data sets, 24 had Z-values of  $-1.65$  or less ( $P \leq 0.10$ ); 21 had Z-values of  $-1.96$  or less ( $P \leq 0.05$ ). The single data set that contained a greater than expected number of runs had a Z-value of only  $+0.59$  ( $P = 0.56$ ), indicating that the observed number of runs was not significantly different from the number expected given random data.

**Autocorrelation analysis.** Sample autocorrelation functions also indicated a nonrandom arrangement of transformed disease incidence values within row segments (see Fig. 4 for the ACF derived from data shown in Fig. 2). Thirty-one data sets exhibited one or more values of  $s$  ( $1 \leq s \leq 20$ ) for which  $r_Y(s)$  lay outside a 95% confidence interval (Table 1). Among these 31 data sets, fewer runs than expected were generally observed (Table 1). Twenty-one of the 31 data sets exhibited large correlations in transformed disease incidence values for plant pairs that were one plant apart ( $s = 1$ ). Of these 21, 18 had runs test Z-values of  $-1.65$  or less and 16 had Z-values of  $-1.96$  or less.

**ARIMA modeling.** The direction in which plants were numbered within a row segment had no effect on the final model

TABLE 1. Results from autoregressive integrated moving average (ARIMA) modeling, autocorrelation analysis, and runs analysis of 37 5-m and one 12-m snap bean row segments

Data set	Mean disease (%)	ARIMA model	$\phi^a$ estimates	$\theta^a$ estimates	Significant <sup>b</sup> autocorrelation function (ACF) lags	Runs test Z-value
4.3 <sup>c</sup>	24.6	CND <sup>d</sup>	...	...	2	-0.20
4.4	22.4	CND	...	...	9	-2.22
7.2	27.4	CND	...	...	5,13	-1.25
1.4	4.6	0 0 0	0	0	None	-0.84
2.3	23.6	0 0 0	0	0	None	-2.07
3.2	22.8	0 0 0	0	0	None	-1.04
6.2	0.4	0 0 0	0	0	None	CND
8.3	3.5	0 0 0	0	0	None	CND
11.1	25.4	0 0 0	0	0	None	-2.48
6.1	1.0	0 0 0	0	0	8	CND
2.1	32.8	0 0 0	0	0	10	+0.59
12.1	50.8	0 0 0	0	0	13	-0.78
4.1	36.4	0 0 0	0	0	15,17	-1.03
4.2	22.1	1 0 0	0.199	0	None	-2.35
3.4	14.7	1 0 0	0.286	0	1,3	-2.20
6.3	2.2	1 0 0	0.305	0	1	CND
5.1	17.1	1 0 0	0.340	0	1	-4.54
3.1	36.2	1 0 0	0.401	0	1,2	-1.84
8.1	6.6	1 0 0	0.402	0	1,3,19	-1.85
5.4	9.5	1 0 0	0.507	0	1-3,17	-2.03
5.3	6.2	1 0 0	0.628	0	1-4,11-13,17	CND
5.5	7.5	1 0 1	0.769	0.416	1-3,5,12,13,18,19	-2.58
1.1	24.5	1 0 1	0.816	0.607	1	-2.05
2.2	29.8	1 0 1	0.825	0.653	1,3	-3.74
3.3	38.3	1 0 1	0.844	0.664	2,3,5	-1.38
8.2	40.5	1 0 1	0.859	0.628	3	-3.36
5.2	13.5	1 0 1	0.861	0.406	1-5,20	-3.56
13.1	58.3	1 0 1	0.868	0.659	1,3,18,19	-1.56
1.3	10.6	1 0 1	0.911	0.539	1-6	-3.13
2.4	19.6	1 0 1	0.915	0.759	1,4,6	-3.52
6.4	8.6	1 0 1	0.926	0.525	1-7,15-19	-4.46
10.1	16.2	1 0 1	0.928	0.671	1-11,14	-4.92
7.4	21.1	1 0 1	0.933	0.767	1,5-8	-2.80
11.2	34.2	1 0 1	0.941	0.648	1-7,9,19	-4.22
9.1	43.8	0 1 1	1	0.662	1-6,18,19	-3.72
1.2	20.8	0 1 1	1	0.684	1-16	-6.09
7.1	34.6	0 1 1	1	0.826	1-9,11,12,14,16	-2.87
7.3	23.0	0 1 1	1	0.867	4,8	-1.82

<sup>a</sup>Parameter estimates for the "generalized ARIMA(1 0 1) model." A  $\phi$  value of 1 implies that differenced data were modeled. See text for further discussion.

<sup>b</sup>ACF lags for which  $\rho_Y(s)$  lay outside a 95% confidence interval.

<sup>c</sup>Numbers to the left of the decimal point refer to the plot or area in a field from which the row segments were collected. Numbers to the right of the decimal point indicate the relative position of the rows within the plots or areas. Data sets 1.1-1.4, 2.1-2.4, 3.1-3.4, 4.1-4.4, 5.1-5.5, 6.1-6.4, and 7.1-7.4 were collected in 1985, cultivar Goldrush; data sets 8.1-8.3 and 9.1 were collected in 1986, cultivar Galamore; data set 10.1 was collected in 1986, cultivar Early Bird; data sets 11.1-11.2 were collected in 1987, cultivar Bush Blue Lake 109; data sets 12.1 and 13.1 were collected in 1987, cultivar Eagle.

<sup>d</sup>Could not determine.

selected for a given data set and had little effect on the value of parameter estimates. The largest difference in a parameter estimate was 0.0005. Parameter estimates discussed below represent an average of the parameter estimates from both orientations rounded to three decimal places.

Among the 38 data sets collected, three yielded sample ACFs and sample PACFs that were not suggestive of any standard ARIMA model, and these data sets could not be adequately fit by any of the numerous ARIMA models tested. For an additional 10 data sets (illustrated by Fig. 3), the pattern of disease appeared random. These data sets (Table 1) were best fit by an ARIMA(0 0 0) model, which has the following form:

$$Y_t = \epsilon_t + \delta.$$

In the remaining 25 data sets (illustrated by Figs. 1 and 2), disease was nonrandomly arranged on plants within row segments. Eight of the data sets (Table 1) were best fit by an ARIMA(1 0 0) model. This model has the following form:

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t + \delta.$$

Estimates of  $\phi_1$  for these data sets ranged from 0.199 to 0.628. Thirteen data sets (Table 1) were fit best by an ARIMA(1 0 1) model, which has the following form:

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} + \delta.$$

Estimates of  $\phi_1$  for these data sets ranged from 0.769 to 0.941 and estimates of  $\theta_1$  ranged from 0.406 to 0.767. For all 13 data sets, estimated values of  $\phi_1$  were greater than estimated values of  $\theta_1$ . For the final four data sets (Table 1), an ARIMA(0 1 1) model provided the best fit. This model has the following form:

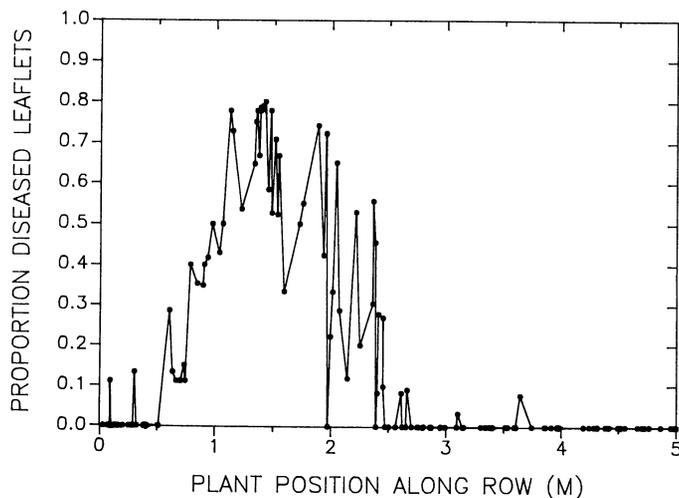
$$W_t = \epsilon_t - \theta_1 \epsilon_{t-1} + \delta.$$

Because  $W_t = Y_t - Y_{t-1}$ , this model also may be written as follows:

$$Y_t = Y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} + \delta.$$

Thus an ARIMA(0 1 1) model is equivalent to an ARIMA(1 0 1) model, with  $\phi_1 = 1$  and nonzero values of  $\theta_1$ . Estimated values of  $\theta_1$  for these four data sets ranged from 0.662 to 0.867.

Results from ARIMA modeling closely paralleled those obtained from the runs test (Table 2). Data sets that could not



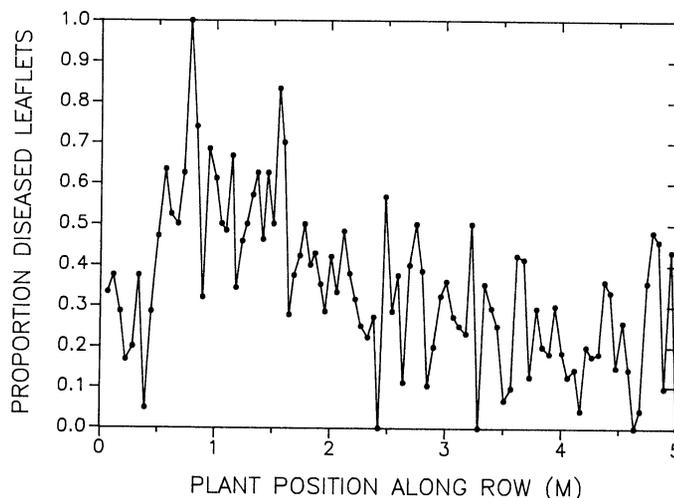
**Fig. 1.** Bacterial brown spot disease incidence values for individual snap bean plants (•) within a 5-m row segment. Disease incidence values are expressed as the proportion of diseased leaflets per plant. Data were collected in July 1986 from a commercial snap bean field (cultivar Galamore) located near Spring Green, WI. See data set 9.1 in Table 1 for additional information.

be distinguished from random data using the runs test tended to be fit best by an ARIMA(0 0 0) model, whereas data sets that exhibited too few runs tended to be fit best by an ARIMA(1 0 0), ARIMA(1 0 1), or ARIMA(0 1 1) model.

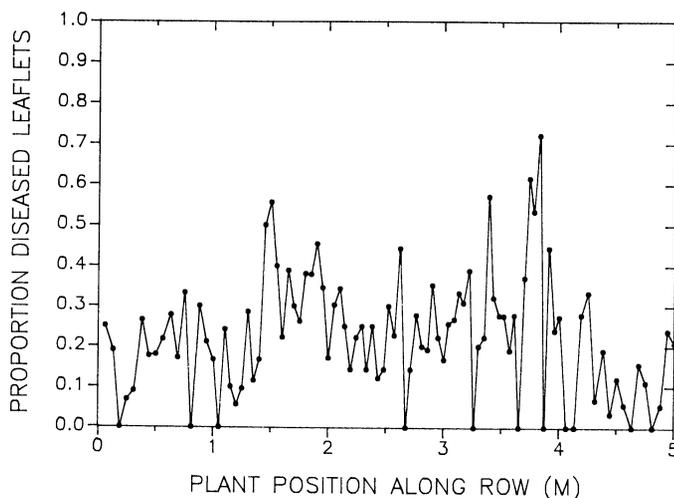
## DISCUSSION

Spatial nonrandomness of bacterial brown spot in the snap bean row segments collected in this study was surprisingly similar across data sets. Only four classes of ARIMA models—ARIMA(0 0 0), ARIMA(1 0 0), ARIMA(1 0 1), and ARIMA(0 1 1)—were required to adequately describe more than 90% of the data sets in this study. In addition, more than 65% of the data sets showed some type of nonrandom variability based on ARIMA modeling.

Nonrandom variability occurred at at least two levels. The dominant component of nonrandom variability was positive correlation between transformed disease incidence values on adjacent plants. This component was visually evident as the slowly



**Fig. 2.** Bacterial brown spot disease incidence values for individual snap bean plants (•) within a 5-m row segment. Disease incidence values are expressed as the proportion of diseased leaflets per plant. Data were collected in July 1985 from a commercial snap bean field (cultivar Goldrush) located near Arlington, WI. See data set 7.1 in Table 1 for additional information.



**Fig. 3.** Bacterial brown spot disease incidence values for individual snap bean plants (•) within a 5-m row segment. Disease incidence values are expressed as the proportion of diseased leaflets per plant. Data were collected in July 1985 from a commercial snap bean field (cultivar Goldrush) located near Arlington, WI. See data set 3.2 in Table 1 for additional information.

undulating changes in disease within a given data set and was detected by all three of the analysis techniques used in this study. In the runs test, it was evident as a smaller than expected number of runs. In the autocorrelation analysis, this component was detected as large, positive values in the sample ACF at lag 1, as well as at additional lags between 2 and 16, for most data sets. Finally, this variability was described and quantified by the  $\phi_1 Y_{t-1}$  term in the ARIMA(1 0 0), ARIMA(1 0 1), and ARIMA(0 1 1) models.

The presence of positive correlation between transformed disease incidence values of adjacent plants is not surprising and is consistent with the idea that disease occurs in localized patches or foci. Although the presence of positive correlation was not entirely unexpected, the four data sets for which an ARIMA(0 1 1) model provided the "best" fit are somewhat disturbing. The theoretical autocorrelation structure of this model is such that, as  $t$  approaches infinity, the variance  $V(Y_t)$  becomes infinitely large. Thus the ARIMA(0 1 1) model is not consistent with the nature of the biological data (proportion of diseased leaflets per plant) in this study. Proportions and their arcsine square root transformations have variances that cannot become infinitely large. Recall however that an ARIMA(0 1 1) model also can be thought of as an ARIMA(1 0 1) model with  $\phi_1 = 1$ . Several other data sets in this study were described by an ARIMA(1 0 1) model with  $\phi_1$ s between 0.9 and 1.0. Such a model is consistent with proportion data because  $V(Y_t)$  for this model is finite for all  $t$ . This suggests that the "true" model for the four anomalous data sets may be an ARIMA(1 0 1) model with  $\phi_1$ s close to but less than 1. It also suggests that the superior fit of an ARIMA(0 1 1) model to these data sets may be due to sampling variability and relatively small sample sizes.

Superimposed on the dominant positive correlation was a

component of negative correlation that gave the graphs of disease versus plant position along the row their jagged appearance. This second type of nonrandom variability was not detected by either the runs test or by autocorrelation analysis. However, it was detected by ARIMA modeling and quantified by the  $-\theta_1 \epsilon_{t-1}$  term in the ARIMA(0 1 1) and ARIMA(1 0 1) models. This negative component of correlation did not appear to be present in those data sets for which an ARIMA(1 0 0) model provided the best fit. However, this may simply reflect an inadequate sample size for detecting a  $\theta_1$  of small magnitude. The situation for those data sets fit by an ARIMA(0 0 0) model is even more complex. Data sets that can be fit by this model can be generated by a process for which  $\phi_1$  and  $\theta_1$  are not zero as long as these parameters are equal. ARIMA modeling cannot effectively estimate  $\phi_1$  and  $\theta_1$  in this situation, and estimates of zero are the result.

At present, the biological origin of both the negative and positive component of nonrandom variability is unknown. However, it is interesting to note that all four classes of ARIMA models that were fit to data sets in this study, including the ARIMA(0 0 0) model, are special cases of a single model that we call the "generalized ARIMA(1 0 1) model." This model has the following form:

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} + \delta$$

where values of  $\phi_1$  and  $\theta_1$  can range between  $-1$  and  $1$  inclusive. An ARIMA(0 0 0) model results from the generalized ARIMA(1 0 1) model when  $\phi_1 = \theta_1$ . When  $-1 < \phi_1 < 1$  (but not 0) and  $\theta_1 = 0$ , an ARIMA(1 0 0) model results. Similarly, when both  $\phi_1$  and  $\theta_1$  are nonzero and  $-1 < \phi_1 < 1$ , an ARIMA(1 0 1) model results. Finally, when  $\phi_1 = 1$  and  $\theta_1$  is nonzero, the generalized ARIMA(1 0 1) model is equivalent to an ARIMA(0 1 1) model. The existence of a general model that describes more than 90% of the data sets collected in this study may indicate that the biological processes that generated these data sets are not extremely diverse but are simply variations of a single underlying process or group of processes.

No matter what its form, any proposed mechanism for the development of brown spot must be consistent with the patterns that we have observed. Mechanisms or models of disease development that rely on mean disease levels or that assume homogeneous disease throughout a field are incomplete in the sense that they ignore these patterns. Any hypothesized mechanism also must be flexible enough to accommodate other patterns as they are discovered. This appears particularly important for brown spot because we have preliminary evidence suggesting that additional patterns of this disease exist both within and across rows (Hudelson et al, *unpublished*).

The work outlined in this paper has stemmed from the application of an iterative sampling strategy, which we call "adaptive sampling," to the study of the spatial patterns of bacterial brown spot. The first step of this strategy, which we now have completed, involves sampling from an area of defined size in a field and describing any spatial patterns within this area. Our initial sampling approach has involved observing every leaflet on every plant in a 5- or 12-m row segment. This method is

TABLE 2. Comparison of results from runs analysis and autoregressive integrated moving average (ARIMA) modeling for 37 5-m and one 12-m snap bean row segments

ARIMA model	Runs Z-value			
	Z > +1.96	-1.96 < Z < -1.65	-1.65 < Z	Unable to determine
	or Z < -1.96	+1.65 < Z < +1.96	< +1.65	
0 0 0	2	0	5	3
1 0 0	4	2	0	2
1 0 1	11	0	2	0
0 1 1	3	1	0	0
Unable to determine	1	0	2	0

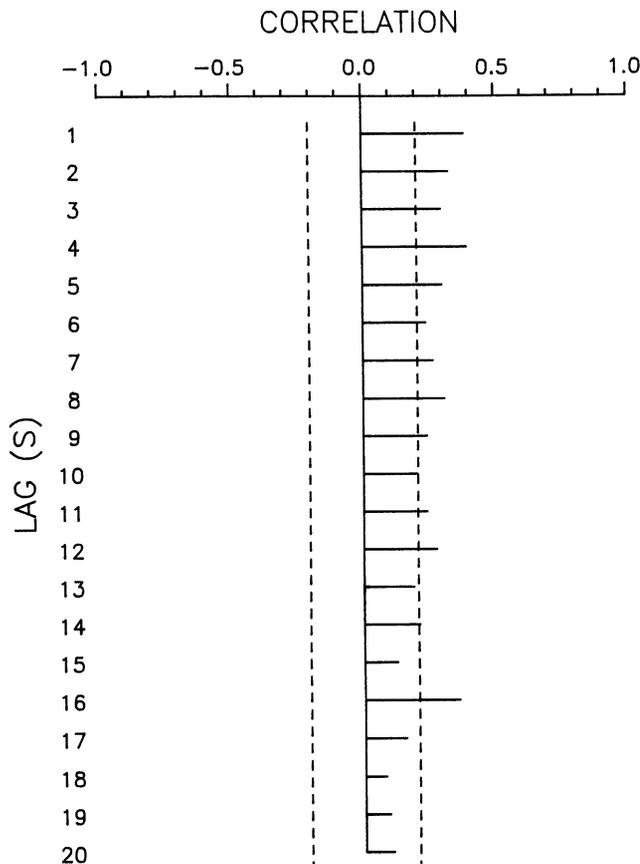


Fig. 4. Autocorrelation function (ACF) for the data shown in Figure 2 after the data were transformed using an arcsine square root transformation. The value of the ACF at lag  $s$  is the correlation of transformed disease incidence values for all pairs of plants that are  $s$  plants apart. Dashed lines delimit approximate 95% confidence intervals.

extremely labor intensive and time consuming; however, it has provided information of extraordinary detail. Such detail typically would not be available using more standard methods that rely on sampling within quadrats.

For descriptive purposes, we have found ARIMA modeling to be particularly useful. As with any modeling approach, ARIMA modeling can provide only an approximation of the complex phenomenon we are observing. In particular this method provides only a unidirectional description of a process that is undoubtedly multidimensional in nature. However, even given this drawback, this technique is superior to runs analysis and autocorrelation analysis because it is able to detect the negative component of nonrandom variability (that is, the jaggedness) in our data sets. In addition, this technique provides a quantitative description of the patterns that we have observed.

The quantitative description provided by ARIMA modeling is valuable in the theoretical development and evaluation of sampling plans. To date, evaluation of sampling plans in the plant pathology literature has relied solely on the use of empirical techniques such as simulation studies (2,11,24,30,34). The patterns observed in this study suggest exploring the use of random start systematic samples for estimating levels of bacterial brown spot in a field. A random start systematic sample involves choosing a starting point (that is, a plant) at random in a given row and then sampling every  $k$ -th plant along the row thereafter (10). Previous work has shown that random start systematic sampling will result in estimates of the mean level of disease that are less variable than estimates based on simple random sampling if the data follow certain forms of an ARIMA(1 0 1) model (10,22). These forms include a generalized ARIMA(1 0 1) model with  $\phi_1 \geq 0$  and  $\phi_1 \geq \theta_1$ .

Finally, ARIMA models also provide information that is useful in evaluating sampling plans for detecting spatial patterns on a variety of scales. In the development of our adaptive sampling strategy, we are exploring the usefulness of systematic and other samples in detecting spatial patterns of bacterial brown spot in snap bean row segments that are greater than 12 m in length (Hudelson et al, *unpublished*).

Our eventual goal is to describe the spatial patterns of bacterial brown spot in commercial-size plantings of snap beans and understand the biological origins of these patterns. The results from this study suggest that spatial patterns of brown spot (and by inference, other plant diseases) may be numerous and complex. We already have identified two patterns of brown spot incidence that occur within relatively short row segments (5 m). These patterns provide an added complexity to the epidemiology of this disease and we can begin now to explore the biological implications of these patterns. In addition, knowledge of these patterns suggests ways of more efficiently estimating brown spot in commercial bean fields which should prove useful to snap bean growers and processors. Finally, through development of our adaptive sampling strategy, we expect to provide a research tool that will be useful for the study of spatial patterns in a variety of biological settings.

#### LITERATURE CITED

- Bartlett, M. S. 1975. *The Statistical Analysis of Spatial Pattern*. Chapman & Hall, London. 90 pp.
- Basu, P. K., Lin, C. S., and Binns, M. R. 1977. A comparison of sampling methods for surveying alfalfa foliage diseases. *Can. J. Plant Sci.* 57:1091-1097.
- Bennett, R. J. 1979. *Spatial Time Series. Analysis-Forecasting-Control*. Pion Limited, London. 674 pp.
- Box, G. E. P., and Jenkins, G. M. 1976. *Time Series Analysis. Forecasting and Control*. Rev. ed. Holden-Day, San Francisco. 575 pp.
- Box, G. E. P., and Pierce, D. A. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.* 65:1509-1526.
- Brockwell, P. J., and Davis, R. A. 1987. *Time Series: Theory and Methods*. Springer-Verlag, New York. 519 pp.
- Campbell, C. L., and Noe, J. P. 1985. The spatial analysis of soilborne pathogens and root diseases. *Annu. Rev. Phytopathol.* 23:129-148.
- Chatfield, C. 1984. *The Analysis of Time Series: An Introduction*. 3rd ed. Chapman and Hall, New York. 286 pp.
- Cliff, A. D., and Ord, J. K. 1981. *Spatial Processes. Models and Applications*. Pion Limited, London. 266 pp.
- Cochran, W. G. 1977. *Sampling Techniques*. 3rd ed. John Wiley & Sons, New York. 428 pp.
- Delp, B. R., Stowell, L. J., and Marois, J. J. 1986. Evaluation of field sampling techniques for estimation of disease incidence. *Phytopathology* 76:1299-1305.
- Diggle, P. J. 1979. On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* 35:87-101.
- Diggle, P. J., and Cox, T. F. 1983. Some distance-based tests of independence for sparsely-sampled multivariate spatial point patterns. *Int. Stat. Rev.* 51:11-23.
- Durbin, J. 1960. The fitting of time-series models. *Rev. Int. Inst. Stat.* 28:233-244.
- Elliott, C. 1920. Halo-blight of oats. *J. Agric. Res. Washington, DC.* 19:139-172.
- Ercolani, G. L. 1973. Two hypotheses on the aetiology of response of plants to phytopathogenic bacteria. *J. Gen. Microbiol.* 75:83-95.
- Freund, J. E. 1981. *Statistics, A First Course*. 3rd ed. Prentice-Hall, Inc., Englewood Cliffs, N. J. 466 pp.
- Gray, S. M., Moyer, J. W., and Bloomfield, P. 1986. Two-dimensional distance class model for quantitative description of virus-infected plant distribution lattices. *Phytopathology* 76:243-248.
- Hau, F. C., Campbell, C. L., and Beute, M. K. 1982. Inoculum distribution and sampling methods for *Cylindrocladium crotalariae* in a peanut field. *Plant Dis.* 66:568-571.
- Hirano, S. S., Nordheim, E. V., Arny, D. C., and Upper, C. D. 1982. Lognormal distribution of epiphytic bacterial populations on leaf surfaces. *Appl. Environ. Microbiol.* 44:695-700.
- Hirano, S. S., Rouse, D. I., and Upper, C. D. 1987. Bacterial ice nucleation as a predictor of bacterial brown spot disease on snap beans. *Phytopathology* 77:1078-1084.
- Iachan, R. 1983. Asymptotic theory of systematic sampling. *Ann. Stat.* 11:959-969.
- King, E. O., Ward, M. K., and Raney, D. E. 1954. Two simple media for the demonstration of pyocyanin fluorescin. *J. Lab. Clin. Med.* 44:301-307.
- Lin, C. S., Poushinsky, G., and Mauer, M. 1979. An examination of five sampling methods under random and clustered disease distribution using simulation. *Can. J. Plant Sci.* 59:121-130.
- Lindemann, J., Arny, D. C., and Upper, C. D. 1984. Epiphytic populations of *Pseudomonas syringae* pv. *syringae* on snap bean and nonhost plants and the incidence of bacterial brown spot disease in relation to cropping pattern. *Phytopathology* 74:1329-1333.
- Lindemann, J., Arny, D. C., and Upper, C. D. 1984. Use of an apparent infection threshold population of *Pseudomonas syringae* to predict incidence and severity of brown spot of bean. *Phytopathology* 74:1334-1339.
- Ljung, G. M., and Box, G. E. P. 1978. On a measure of lack of fit in time series models. *Biometrika* 65:297-303.
- Madden, L. V., Louie, R., Abt, J. J., and Knoke, J. K. 1982. Evaluation of tests for randomness of infected plants. *Phytopathology* 72:195-198.
- Madden, L. V., Louie, R., and Knoke, J. K. 1987. Temporal and spatial analysis of maize dwarf mosaic epidemics. *Phytopathology* 77:148-156.
- Mihail, J. D., and Alcorn, S. M. 1987. *Macrophomina phaseolina*: Spatial patterns in a cultivated soil and sampling strategies. *Phytopathology* 77:1126-1131.
- Nicot, P. C., Rouse, D. I., and Yandell, B. S. 1984. Comparison of statistical methods for studying spatial patterns of soilborne plant pathogens in the field. *Phytopathology* 74:1399-1402.
- Noe, J. P., and Campbell, C. L. 1985. Spatial pattern analysis of plant-parasitic nematodes. *J. Nematol.* 17:86-93.
- Pielou, E. C. 1977. *Mathematical Ecology*. John Wiley & Sons, New York. 385 pp.
- Poushinsky, G., and Basu, P. K. 1984. A study of distribution and sampling of soybean plants naturally infected with *Pseudomonas syringae* pv. *glycinea*. *Phytopathology* 74:319-326.
- Proctor, C. H. 1984. On the detection of clustering and anisotropy using binary data from a lattice patch. *Commun. Stat.-Theory Methods* 13:617-638.
- Reynolds, K. M., and Madden, L. V. 1988. Analysis of epidemics using spatio-temporal autocorrelation. *Phytopathology* 78:240-246.
- Reynolds, K. M., Madden, L. V., and Ellis, M. A. 1988. Spatio-temporal analysis of epidemic development of leather rot of strawberry. *Phytopathology* 78:246-252.
- Rouse, D. I., Nordheim, E. V., Hirano, S. S., and Upper, C. D.

1985. A model relating the probability of foliar disease incidence to the population frequencies of bacterial plant pathogens. *Phytopathology* 75:505-509.
39. Schaad, N. W. 1988. Laboratory Guide for Identification of Plant Pathogenic Bacteria. 2nd ed. American Phytopathological Society, St. Paul, MN. 164 pp.
40. Schuh, W., Frederiksen, R. A., and Jeger, M. J. 1986. Analysis of spatial patterns in sorghum downy mildew with Morisita's index of dispersion. *Phytopathology* 76:446-450.
41. Snedecor, G. W., and Cochran, W. G. 1980. Statistical Methods. 7th ed. The Iowa State University Press, Ames, IA. 507 pp.
42. Thal, W. M., and Campbell, C. L. 1986. Spatial pattern analysis of disease severity data for alfalfa leaf spot caused primarily by *Leptosphaerulina briosiana*. *Phytopathology* 76:190-194.
43. Upton, G. J. G., and Fingleton, B. 1985. Spatial Data Analysis by Example. John Wiley & Sons, New York. 410 pp.