# Spatial Heterogeneity of the Incidence of Grape Downy Mildew

L. V. Madden, G. Hughes, and M. A. Ellis

First and third authors: Department of Plant Pathology, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster 44691; second author: Institute of Ecology and Resource Management, University of Edinburgh, W. Mains Rd., Edinburgh EH9 3JG, Scotland.

Salaries and research support provided by State and Federal funds appropriated to the Ohio Agricultural Research and Development Center, The Ohio State University.

Accepted for publication 18 November 1994.

## ABSTRACT

Madden, L. V., Hughes, G., and Ellis, M. A. 1995. Spatial heterogeneity of the incidence of grape downy mildew. Phytopathology 85:269-275.

Heterogeneity of the incidence of downy mildew of grape, caused by *Plasmopara viticola*, was quantified in an experimental Ohio vineyard. The proportion of diseased leaves on each of 15 shoots (sampling units) per plot was determined for 18 separate plots at two times (generally August and September) during each of 3 yr. The binary data analogue of Taylor's power law, in which the logarithm of the observed variance is regressed on the logarithm of the theoretical variance for a binomial (random) distribution, provided strong and consistent evidence that diseased leaves were aggregated. Year and assessment time did not affect power law parameters. The estimate of the regression slope ($b$), an overall measure of heterogeneity, was 1.30 (SE = 0.04). Heterogeneity in individual plots was measured with variance ratio and $C(\alpha)$ tests and with the aggregation parameter ($\theta$) of the beta-binomial distribution fitted to the data. Except for mean incidence less than 0.05, the majority of the plots had significant heterogeneity, and the data were better described by the beta-binomial than by the binomial distribution. Estimates of $\theta$ were variable but were highest in the middle range of disease incidence. Using the modified power law, sampling curves were generated to precisely estimate disease incidence.

*Additional keywords*: extra-binomial variation, overdispersion, quantitative epidemiology, spatial analysis.

Spatial pattern is a significant epidemiological characteristic of plant disease (4,16). Besides aiding in the understanding of the population dynamics of pathogens and diseases, knowledge of heterogeneity can be used to develop efficient sampling schemes (5,12) and properly analyze the effects of experimental treatments on disease variables (17,38). Depending on whether the location of sampling units (e.g., plants, leaves) is known and the type of disease data observed in the sampling units (e.g., counts of lesions, area of leaves affected), various techniques can be used (16,38) to measure heterogeneity. A common practice in ecology, as well as in plant pathology, is to fit distribution models to the frequency of diseased units and characterize heterogeneity by estimated parameters (4). This approach can be extended to calculate indices of aggregation, usually based on the mean and variance of the variable of interest, such as lesions per sampling unit.

The Taylor empirical power law describes precisely the relationship between the variance and mean on a logarithmic scale for counts of organisms in sampling units (34–37). The slope of the log(variance):log(mean) line is considered an index of aggregation, although debate remains as to how biological processes give rise to resulting slopes (1,7,35). Recently, Hughes and Madden (10) showed that the standard power law is inappropriate for binary data, such as disease incidence, in which a plant (or leaf) is either diseased or healthy. They proposed a new model to be used with incidence data. Their results showed why inconsistent or misleading results occur when fitting the standard Taylor model to incidence data (19,20). Based on the modified power law, it was shown that the beta-binomial distribution is the appropriate model for the frequency of aggregated diseased units, rather than the negative binomial (11). The beta-binomial has been used successfully to represent the frequency distribution of virus diseased plants per quadrat (11,28,31).

In this study, the modified power law, beta-binomial distribution, and related variance-ratio tests were used to characterize the spatial heterogeneity of downy mildew of grape (*Vitis labrusca* L.), caused by *Plasmopara viticola* (Berk. & M.A. Curtis) Berl. & De Toni in Sacc., in an Ohio vineyard. Downy mildew is one of the most important diseases of grapes in Ohio and regions where relative humidity is high and rainfall common (8,25). The pathogen overwinters as oospores in infected leaf residues that germinate to produce sporangia in the spring and summer. These sporangia germinate directly or indirectly to infect leaves. Secondary spread occurs throughout the growing season, although infection and sporangia production are highly dependent on environmental conditions (2,13,14). Because disease intensity is variable among locations and years, presumably because of variation in weather conditions, there is considerable interest in developing predictive systems for timing fungicide applications (25). The proposed systems use either monitored weather variables (2,8,25) or disease level threshold (3) to determine whether to apply a fungicide.

To date, predictive systems for downy mildew are still in the experimental stage. Progress is hindered by lack of information on density and distribution of downy mildew incidence and severity in vineyards. In this study, incidence of downy mildew-infected leaves was assessed in small plots over 3 yr. Incidence was assessed instead of severity because of the relative ease of identifying and counting diseased leaves compared to estimating the area of lesions or proportion of leaves covered by lesions. The objectives of this study were to: quantify the heterogeneity of downy mildew incidence; determine the effects, if any, of assessment time, year, and mean disease incidence on heterogeneity; and use information on heterogeneity to develop an efficient sampling curve. A range of statistical techniques and parameter estimation methods was used to achieve these objectives. This is the first detailed study of the modified power law and beta-binomial distribution for characterizing heterogeneity of a fungal disease.

## MATERIALS AND METHODS

**Data collection.** The experiments were located in vineyards at the experimental farm of the Ohio Agricultural Research and Development Center, The Ohio State University, Wooster. The

grape cultivar Catawba (planted in 1975) was used in 1989 and 1990, and the cultivar Reliance (planted in 1989) was used in 1992. Each vineyard had six rows spaced 3 m apart with 2.1 m between plants within a row. Vineyards were partitioned into 18 plots, with three contiguous plants per plot. There also was one plant between each plot.

The 18 plots were grouped into three blocks with six different treatments assigned randomly to the plots in each block. For the purposes of this study, the treatments were used to obtain different levels of disease incidence, and possibly heterogeneity. The treatments consisted of spraying the plants from zero to eight times throughout the season with fungicides, metalaxyl + mancozeb (Ridomil MZ-58) or captan. Scheduling of fungicide application was based on an experimental predictive system currently being tested (13,14; M. A. Ellis and L. V. Madden, *unpublished data*) that is incorporated into a microprocessor marketed by Neogen Corporation (Lansing, MI).

In each plot, the leaves on five arbitrarily selected shoots per plant were assessed visually for symptoms of downy mildew. Thus, the sampling unit for determination of disease heterogeneity was the shoot, and there were 15 shoots per plot. Total number of leaves ($n$) and diseased leaves ($X$) were recorded for each sampling unit. Because of the nature of grape growth, shoots are intertwined and the exact spatial location of leaves or shoots cannot readily be specified. The vertical layering of shoots also complicates the determination of position. Thus, analyses performed (discussed below) did not rely on spatial location of the observations.

Disease assessment was performed twice per year (Table 1). With the exception of the first assessment in 1989, both times were late enough for substantial secondary spread to have occurred.

**Data analysis—moments.** The mean and variance of disease incidence was determined for each plot, assessment time, and year. Also, the mean number of leaves per shoot (sampling unit) was determined. Mean incidence of diseased leaves ($p$) is given as:

$$p = \Sigma X_i / \Sigma n_i \tag{1}$$

in which $i$ is an index for the $i$-th shoot. The number of shoots per plot is represented by $N$ (=15). The summation, therefore, is for the $i = 1,...,N$ shoots in a plot. The mean number of leaves/shoot in a plot was calculated as:

$$\bar{n} = \Sigma n_i / N \tag{2}$$

The variance of diseased leaves, $v(X)$, was calculated according to the formula of Cochran (5) for variable $n$:

$$v(X) = [\Sigma(X_i - pn_i)^2]/(N - 1). \tag{3a}$$

Note that $pn_i$ is the expected number of diseased leaves on a shoot. The estimated variance of the proportion of diseased leaves ($y_i = X_i/n_i$) was obtained by dividing equation 3a by $\bar{n}^2$. This variance (=$v$) can be written as:

$$v = [\Sigma n_i^2 (y_i - p)^2]/[\bar{n}^2(N - 1)]. \tag{3b}$$

TABLE 1. Summary of mean incidence[a] of downy mildew, caused by *Plasmopara viticola*, in grape cultivars Catawba (1989 and 1990) and Reliance (1992)

| Data set | Assessment date | Mean | Median | Minimum | Maximum |
|---|---|---|---|---|---|
| 1989a | 7/1 | 0.032 | 0.018 | 0.000 | 0.108 |
| 1989b | 9/14 | 0.147 | 0.032 | 0.000 | 0.775 |
| 1990a | 8/28 | 0.120 | 0.039 | 0.004 | 0.618 |
| 1990b | 9/30 | 0.153 | 0.049 | 0.017 | 0.764 |
| 1992a | 8/19 | 0.168 | 0.139 | 0.000 | 0.767 |
| 1992b | 9/22 | 0.280 | 0.178 | 0.040 | 0.873 |

[a] Incidence is the proportion of leaves on a shoot with symptoms of downy mildew. There were 15 sampled shoots per three-plant plot and 18 plots per data set. Statistics shown here are based on mean incidence values per plot not on individual shoot results within plots.

The theoretical variance of $X$ for a random (i.e., binomial) distribution was calculated as:

$$v_r(X) = \bar{n}p(1 - p) \tag{4a}$$

and for $y$ as:

$$v_r = p(1 - p)/\bar{n} \tag{4b}$$

**Data analysis—power law.** The binary form of Taylor's power law (10) was fitted to the variances for the 18 plots at each time and year. The model can be written as:

$$\log(v) = \log(A) + b \log[p(1 - p)/\bar{n}] \tag{5}$$

in which $A$ and $b$ are parameters. The independent variable in equation 5 is the logarithm of the expected variance for a random (binomial) distribution [=$\log(v_r)$]. It should be noted that for unbounded counts, the Poisson distribution represents a random pattern, and $v_r$ equals the mean per quadrat [=$np$]). When $n$ is constant, equation 5 can be written as:

$$\log(v) = \log(a) + b \log[p(1 - p)] \tag{6}$$

with $a = An^{-b}$. Randomness, as represented by a binomial distribution, is indicated when $b = 1$ and $A = 1$. Values of $b > 1$ indicate heterogeneity; therefore, $b$ is used as an index of aggregation for a collection of data sets.

Ordinary least squares regression was used to estimate $\log(A)$ and $b$. Significance in relationships between $\log(v)$ and $\log(v_r)$ was determined with $F$ tests. The coefficient of determination ($R^2$) and the mean square error (MSE) were used to determine the goodness of fit of the model. Residuals (i.e., differences between observed and predicted $\log[v]$) were plotted against predicted values to evaluate the appropriateness of the model (23). Normality of the residuals was assessed using the technique of Looney and Gulledge (15). Equality of $b$ to 1 was tested with a $t$ test, using the parameter estimate and its standard error. The alternative hypothesis was that $b > 1$ (i.e., a one-sided test).

For comparative purposes and because of the importance of obtaining accurate estimates of $A$ and $b$, two other approaches were evaluated for estimating parameters, namely geometric mean and resistant-line regression. Because both $\log(v)$ and $\log[p(1 - p)/\bar{n}]$ are determined with sampling error and neither variable is clearly the dependent variable in a strict sense, geometric mean regression has been advocated for these types of data sets (36). This technique is also called functional regression and central line regression. Although rare in plant pathology, the technique has had some use for analyses based on the traditional Taylor power law. For geometric mean regression, the slope is always greater than the ordinary least squares (=dependent regression) slope, unless $R^2 = 1$.

The slope and intercept of equation 5 also were estimated with resistant-line regression, which is a nonparametric statistical method (39). Parameter estimates are less sensitive to outliers or extreme points than for ordinary least squares regression, and normality of the residuals is not required.

The effect of treatment on the slope and intercept of equation 5 was assessed with covariance analysis (23). In effect, this consisted of determining the reduction in the residual sum of squares when separate intercepts, and then slopes, were calculated for each treatment. $F$ tests were used to test for significant treatment effects.

The consistency of regression results among years and between times was evaluated by a "reduced versus full model" analysis (23). The residual sums of squares for each year/time were totaled (=full model) and compared to the residual sum of squares for a single model fitted to the pooled data set. An $F$ test was used for determining significance. All regression analyses were performed with MINITAB (22).

**Data analysis—aggregation indices.** The index of dispersion ($D$) was calculated as the ratio of observed to theoretical variance, i.e., $D = v(X)/v_r(X)$ (9). (For count data [e.g., number of propa-

gules], the denominator of this expression is simply the mean.) With the modified power law, if $b = 1$ and $A > 1$, then aggregation is not affected by $p$ and $A$ is an overall estimate of $D$. Significant departure from randomness was determined with a chi-square test, in which the test statistic is given by $(N - 1)D$. If there is a random distribution, $(N - 1)D$ has a $\chi^2$ distribution with $N - 1$ degrees of freedom.

A $C(\alpha)$ test, sensu Neyman (24), also was calculated as described in Tarone (33). The test statistic is:

$$Z = \frac{v(X)(N-1)/[p(1-p)] - (\bar{n}N)}{[2\Sigma n_i(n_i - 1)]^{1/2}} \quad (7)$$

in which $Z$ is the standard normal deviate. For both the variance ratio ($[N - 1]D$) and $C(\alpha)$ tests, the null hypothesis is that the pattern is random (i.e., the binomial distribution is appropriate). The alternative hypothesis for the variance ratio test is simply that the pattern is aggregated. For the $C(\alpha)$ test, the alternative hypothesis is more specific, namely, that the data have a beta-binomial distribution (33). Besides testing for significant aggregation, $D$ and $Z$ were used as indices of aggregation.

**Data analysis—discrete distributions.** The beta-binomial and binomial distributions were fitted to each data set that did not have zero mean incidence. The binomial distribution has one parameter ($\pi$) that is the (constant) probability of a leaf being infected. If the binomial distribution was appropriate, $p$ (equation 1) would be the estimate of $\pi$.

The beta-binomial can be generated as a compound or generalized distribution (11,28,31,32) and can be written in several forms. For instance, the probability that a shoot (=sampling unit) has $x$ diseased leaves is:

$$Prob\,(X = x) = \binom{n}{x} \frac{\prod_{i=0}^{x-1}(p + i\theta) \prod_{i=0}^{n-x-1}(1 - p + i\theta)}{\prod_{i=0}^{n-1}(1 + i\theta)} \quad (8)$$

in which $\Pi$ is the product function and $i$ is a counting index. Here, $p$ is the expected (i.e., mean) probability of a leaf being infected, because it is not now assumed that $\pi$ is a constant. The parameter $\theta$ is an index of aggregation, which equals 0 when there is a random pattern (or binomial distribution) and increases as aggregation increases. When equation 8 is appropriate, the data are said to exhibit extra-binomial variation or overdispersion.

Maximum likelihood estimates of $p$ and $\theta$ and their standard errors were determined with the FORTRAN program previously described (18). Because $n$ was not constant, it was not possible to calculate expected frequencies or perform the classic chi-square goodness-of-fit test. However, it was possible to determine if the beta-binomial provided a better fit than the binomial using a likelihood ratio statistic (LRS) (28). The LRS is based on the log-likelihoods for the two distributions. This test is not very powerful when $N < 20$, meaning that the tests may fail to detect true differences (26). The $C(\alpha)$ test, which does not require maximum likelihood estimation, is more powerful than the LRS test at relatively small sample sizes (26).

**Sample size.** Results from the modified power law (equation 5) were used to estimate sample sizes for fixed levels of the coefficient of variation ($C$ = standard error of estimated $p$ divided by $p$). Number of shoots, i.e., sample size, is given by:

$$N = a\,p^{b-2}(1 - p)^b / C^2 \quad (9)$$

in which $a = A\,n^{-b}$ and $A$ and $b$ are from equation 5 (12). For simplicity here, it was assumed that the number of leaves/shoot ($n$) was constant.

## RESULTS

**Disease incidence.** Mean disease incidence of downy mildew in a plot was dependent on the imposed fungicide treatment (data not shown). Across plots, mean incidence (per plot) ranged from lows of 0.0–0.04 and highs of 0.11–0.87, depending on the year and assessment time (Table 1). The first assessment time in 1989 (1989a in Table 1) was much earlier than for the other years, and incidence was very low in all plots. For all other assessments, the range of mean incidence values per plot was at least 0.61. The maxima shown in Table 1 corresponded to plots not sprayed with fungicide; other values in the table reflect the effects of fungicides as well as time and year effects.

**Power law.** There was a significant relationship between $\log(v)$ and $\log(v_r)$ (=$\log(p[1 - p]/\bar{n})$) for each data set (Table 2; Fig. 1). $R^2$ values exceeded 0.86, except for 1992b, and residual plots did not reveal a nonlinear relationship. Tests of the residuals for normality (15) were not significant ($P > 0.10$), indicating that the residuals were normally distributed.

The ordinary least squares estimates of $b$ exceeded 1 for all data sets (Table 2) and were significantly greater than 1 ($P < 0.05$) for all sets except 1992b, in which overall variability was greater than the other data sets. In this case, one could not reject the null hypothesis of $b$ being equal to one in favor of the alternative hypothesis that $b$ was greater than one ($P > 0.05$). The predicted values for equation 5 were greater than the predicted values for a random pattern ($A = 1$; $b = 1$; broken line in Fig. 1) for most of the range of $\log(v_r)$ for each data set. Standard errors of the estimates of $b$ generally were less than 11% of the slope values. Standard errors for the estimates of $\log(A)$ also were low for all data sets, except 1992b (Table 2).

Estimates of the parameters were quite similar among data sets (Table 2). An $F$ test based on the full and reduced models indicated that year and assessment time did not significantly affect regression results ($P > 0.20$). Parameter estimates based on the pooled data (=reduced model), therefore, could be used to describe the $\log(v)$:$\log(v_r)$ relationship. Moreover, analysis of covariance indicated that fungicide treatment did not significantly affect $\log(A)$ or $b$ (Table 2). Therefore, although fungicide treatment affected $p$, it did not directly affect overall spatial heterogeneity

TABLE 2. Results of fitting the binary analogue of the power law equation (equation 5) to the variance and mean incidence of grape downy mildew, caused by *Plasmopara viticola*[a]

| Data set | Ordinary least squares regression | | | | | | | Geometric mean | | Resistant line | | Treatment effects[b] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | log(A) | (SE) | b | (SE) | $R^2$ | MSE | df | log(A) | b | log(A) | b | P(log[A]) | P(b) |
| 1989a | 0.74 | (0.329) | 1.220 | (0.119) | 0.898 | 0.038 | 12 | 0.92 | 1.287 | 0.81 | 1.240 | 0.55 | 0.15 |
| 1989b | 0.87 | (0.269) | 1.258 | (0.103) | 0.920 | 0.040 | 13 | 0.94 | 1.312 | 0.52 | 1.147 | 0.39 | 0.10 |
| 1990a | 0.92 | (0.187) | 1.278 | (0.072) | 0.952 | 0.023 | 16 | 1.00 | 1.310 | 1.25 | 1.395 | 0.48 | 0.54 |
| 1990b | 1.21 | (0.331) | 1.447 | (0.140) | 0.870 | 0.039 | 16 | 1.452 | 1.551 | 1.33 | 1.503 | 0.42 | 0.92 |
| 1992a | 1.46 | (0.342) | 1.503 | (0.156) | 0.861 | 0.045 | 15 | 1.712 | 1.620 | 1.61 | 1.579 | 0.84 | 0.59 |
| 1992b | 1.14 | (0.638) | 1.424 | (0.304) | 0.578 | 0.058 | 16 | 2.076 | 1.837 | 1.11 | 1.410 | 0.29 | 0.30 |
| Pooled | 0.93 | (0.109) | 1.296 | (0.043) | 0.895 | 0.041 | 98 | 1.110 | 1.370 | 0.99 | 1.313 | 0.50 | 0.85 |

[a] log(A) and b are the estimated intercept and slope of the best fitting line, using various estimation procedures (ordinary least squares, geometric-mean, and resistant-line regression); SE is the standard error of the estimated parameter (intercept or slope); $R^2$ is the coefficient of determination; MSE is the mean square error; and df is the error degrees of freedom.

[b] Significance level for the effect of fungicide treatment on the intercept (P[log(A)]) and slope (P(b)) of the power law equation.

as measured by the power law.

Even though there was significant overall spatial heterogeneity, as indicated by a good fit of equation 5 to the data and $b > 1$, the line for equation 5 crossed the line for a random (binomial) pattern at low $v_r$ (Fig. 1). The crossover occurred at $\log(v_r)^* = \log(A)/(1 - b)$. Below this point, observed variance was less than $v_r$ for individual plots, even though $\log(v)$ increased with $\log(v_r)$ at a rate greater than for a random pattern. For the pooled data here, $\log(v_r)^* = -3.10$, corresponding to $v_r^* = 0.0008$. Assuming an $n$ of 15 (on average), this indicates that the lines crossed at $p \approx 0.012$ at the low end of the observed disease incidence values.

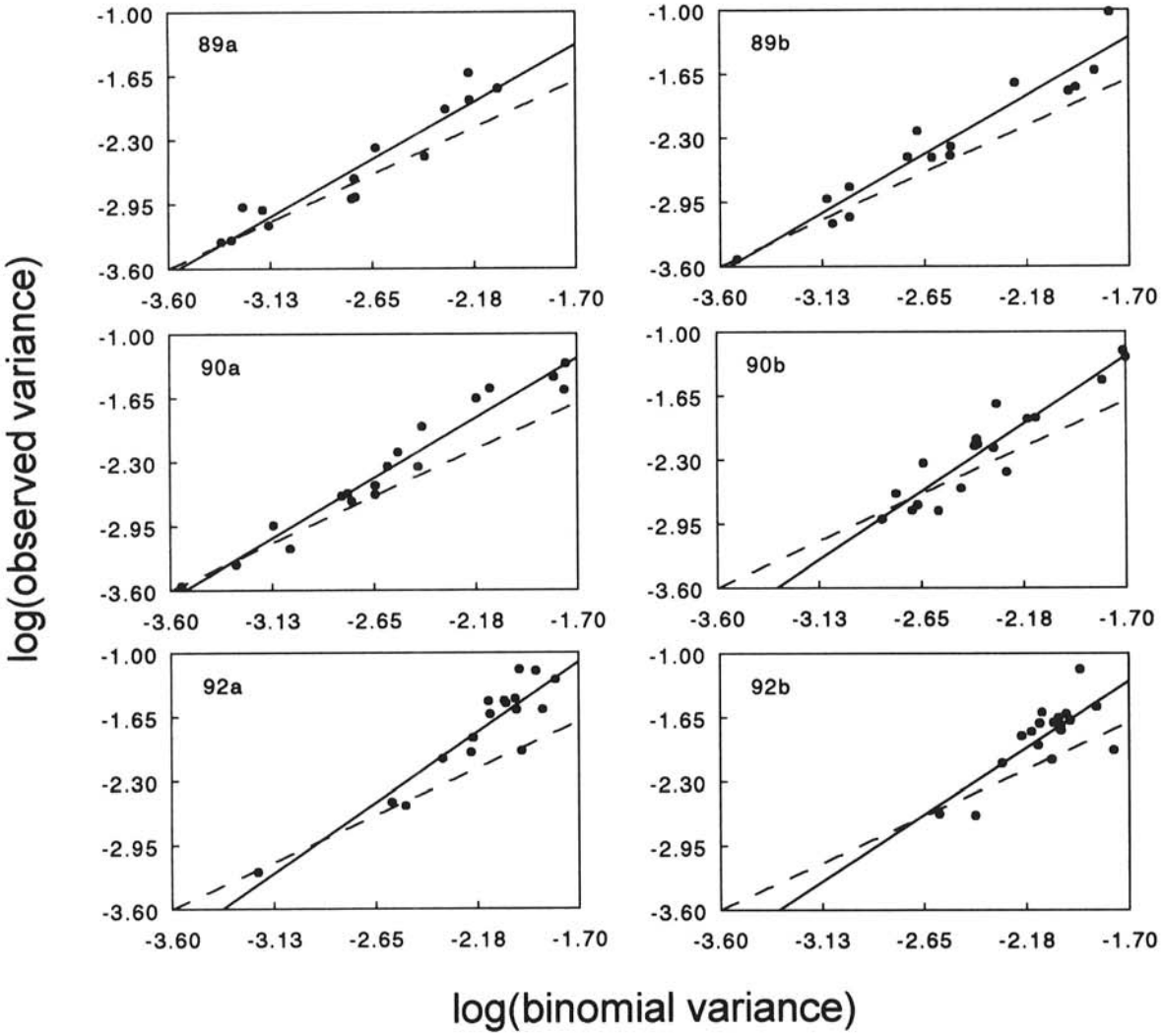


**Fig. 1.** Relationship between the logarithm of observed variance (equation 3b) and the logarithm of theoretical variance for a random pattern (binomial distribution; equation 4b) for incidence of grape downy mildew, caused by *Plasmopara viticola*, at each of two assessment times (lowercase letters) during each of 3 yr. Each point represents a field plot at one sampling time. Solid line represents the ordinary least squares fit to the data (equation 5; Table 2), and broken line represents the binomial line (i.e., observed variance = theoretical variance). Table 1 contains information on each disease assessment.

TABLE 3. Percentage of grape plots with significant variance ratio ($D$) test, $C(\alpha)$ test ($Z$ statistic), convergence of the maximum likelihood estimation (MLE) procedure, significant likelihood ratio statistic (LRS)[a], and the mean value of estimated $\theta$ in relation to categories of incidence of grape downy mildew

| Incidence class[b] | Count | Significant variance ratio ($D$)[c] (%) | Significant $C(\alpha)$ test[c] (%) | MLE convergence (%) | Significant LRS test[c] (%) | Mean $\theta$ |
|---|---|---|---|---|---|---|
| 0 | 8 | ...[d] | ... | ... | ... | ... |
| 0.00–0.05 | 41 | 9.8 | 14.6 | 43.9 | 11.1 | 0.021 |
| 0.05–0.10 | 16 | 56.2 | 50.0 | 81.2 | 46.1 | 0.072 |
| 0.10–0.20 | 21 | 85.7 | 90.5 | 95.2 | 85.0 | 0.141 |
| 0.20–0.40 | 10 | 90.0 | 90.0 | 90.0 | 88.9 | 0.210 |
| 0.40–0.80 | 9 | 55.6 | 55.6 | 88.9 | 66.5 | 0.132[e] |
| 0.80–1.00 | 3 | 100 | 66.7 | 100 | 66.7 | 0.075 |

[a] Test of the beta-binomial versus the binomial.
[b] Classes end with the indicated value (e.g., 0.10). Classes start with the next highest incidence value above the listed value (e.g., any incidence value above 0.10).
[c] $P \leq 0.05$.
[d] No tests were done or distributions fitted when incidence was 0 (no diseased leaves).
[e] Between incidence of 0.40 and 0.60, mean $\theta$ was 0.273 (based on three plots); between 0.60 and 0.80, mean $\theta$ was 0.062 (based on six plots).

Estimates of log($A$) and $b$ using geometric mean regression were only slightly larger than the ordinary least squares estimates for most of the individual data sets and the pooled data (Table 2). The greatest change in parameter estimates was for 1992b, in which $b$ increased from 1.4 to 1.8. A $t$ test of the slope estimate for 1992b revealed that $b$ was greater than 1 ($P < 0.05$), unlike the situation for the ordinary least squares estimate. The resistant-line estimates of log($A$) and $b$ also were very similar to the estimates from ordinary least squares regression (Table 2). The greatest difference in $b$ was for 1990a, in which the resistant-line estimate was 0.12 higher.

**Aggregation indices.** The degree of spatial heterogeneity for individual plots varied considerably. The variance ratio test based on $D$ was significant ($P < 0.05$) in 48% of the plots in which $p > 0$. Likewise, the $C(\alpha)$ test was significant ($P < 0.05$) in 49% of the plots. All plots that had an observed variance below the random or binomial (broken) line in Figure 1 had no indication of aggregation based on the two tests. It is informative to summarize results based on ranges of mean disease incidence (Table 3). At very low $p$ (<0.05), significant aggregation was indicated by $D$ or $C(\alpha)$ in only 10–15% of the plots. At $0.05 < p < 0.10$, about half of the plots indicated significant aggregation. At $p > 0.10$, over 80% of the plots generally had significant aggregation.

A plot of $Z$ from the $C(\alpha)$ test (equation 7) versus $p$ (Fig. 2) reveals the variation in downy mildew heterogeneity across individual grape plots. At low $p$, most of the points were below 1.64, the cut-off for significance. There was a general increase in $Z$ (or $D$; data not shown) as $p$ increased, followed by a decrease at high $p$. The largest $Z$ values were mostly in the range $0.2 < p < 0.6$.

**Beta-binomial distribution.** The maximum likelihood estimation procedure converged for 71% of the data sets with $p > 0$. As with the tests with indices of aggregation, convergence depended on $p$ (Table 3) and never occurred when the observed variance was below the binomial variance (Fig. 1). At $p > 0.05$, convergence occurred for at least 80% of the data sets. Where convergence did occur, the LRS indicated that the beta-binomial provided a better fit to the data than the binomial distribution in the majority of cases (Table 3). Results of the LRS test agreed closely with the $C(\alpha)$ test.

Estimates of $\theta$ for the data sets ranged from 0.0 to 0.56. Standard errors for the estimates were relatively large, reflecting the small sample size for estimating parameters. Estimates of $\theta$ were highly correlated with both $D$ and $Z$ (Fig. 3). At $Z > 0$, there was a linear increase in $\theta$ with an increase in $Z$ (or $D$).

**Sample size.** Sampling curves were generated for four values of $C$ (0.1–0.4) using the pooled estimates of $b$ and log($A$) (Table 2; Fig. 4A and B) and for a random (binomial) distribution of incidence ($A = b = 1$) (Fig. 4B). It was assumed that there were 15 leaves per shoot ($n = 15$), which is close to the overall mean of 14.4. With a mean incidence of 0.01, nearly 200 shoots would need to be sampled for a desired coefficient of variation of 20% (i.e., $C = 0.2$) (Fig. 4A). At $p = 0.20$, however, only about 20 shoots would be required.

Sample size based on the power-law results (for a given level of $C$) was either less than or greater than sample size for the binomial case (Fig. 4B). At $p > 0.012$, more shoots would need to be sampled compared to the binomial. The opposite would

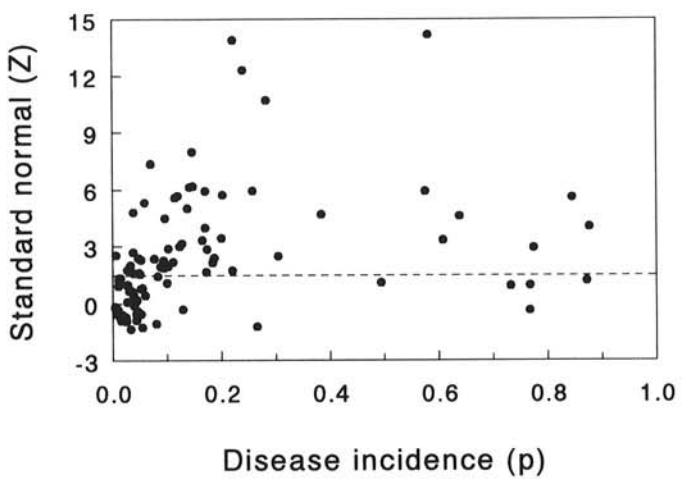


**Fig. 2.** Standard normal statistic ($Z$) of the $C(\alpha)$ test (equation 7) in relation to mean disease incidence ($p$) of grape downy mildew, caused by *Plasmopara viticola*. Each point represents a field plot at one sampling time. Horizontal broken line represents the cutoff (1.64) for significant extra-binomial variation ($P = 0.05$).

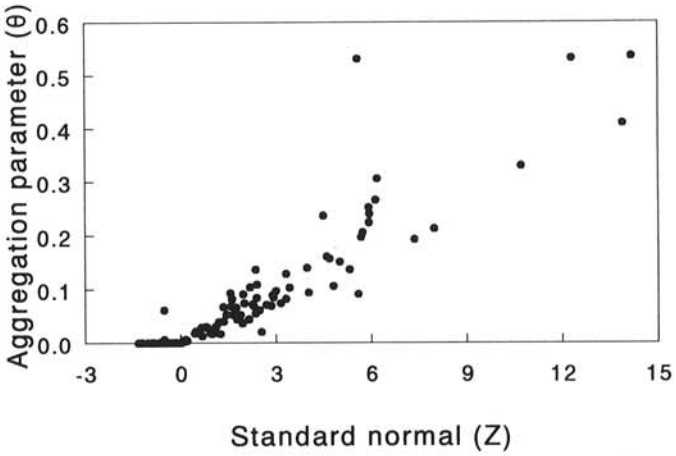


**Fig. 3.** Estimated aggregation parameter ($\theta$) of the beta-binomial distribution (equation 8) in relation to the standard normal statistic ($Z$) of the $C(\alpha)$ test (equation 7) for grape downy mildew incidence, caused by *Plasmopara viticola*. Each point represents a field plot at one sampling time.

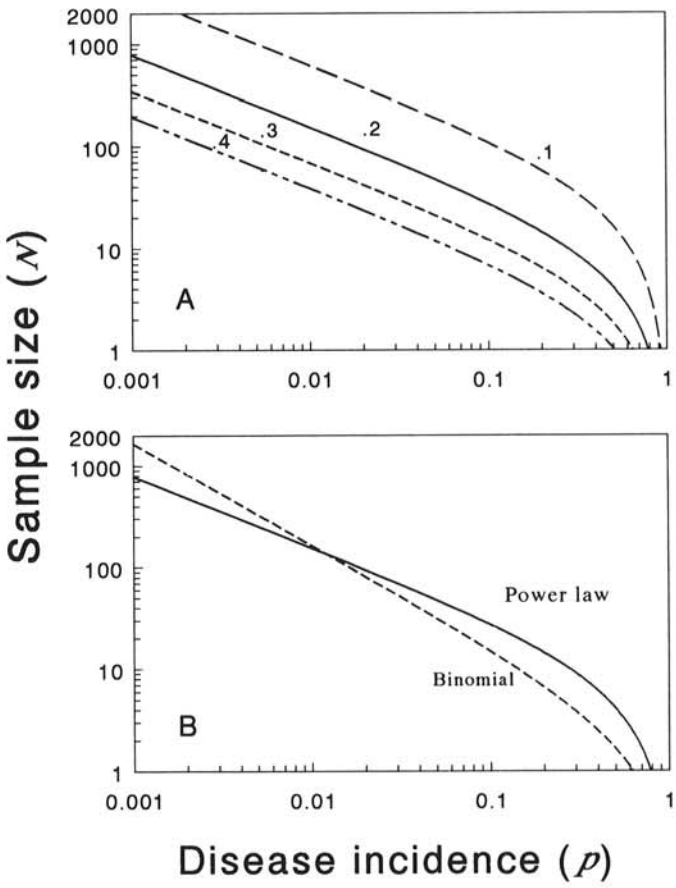


**Fig. 4.** Sampling curves for the number of grape shoots required for estimating mean disease incidence ($p$) with: **A,** $C = 0.1$, 0.2, 0.3, and 0.4, based on the power law (equation 9) with $b = 1.3$ and $a = 0.25 = (8.5)(15^{-1.3})$ (equations 5 and 6); and **B,** binomial ($a = 15^{-1}$; $b = 1$) and power law ($a = 0.25$; $b = 1.3$) cases ($C = 0.2$).

be true at $p < 0.012$, because the variance lines crossed at about this disease incidence (Fig. 1). At $p = 0.2$, for example, assumption of a binomial distribution would imply that six shoots (of $n = 15$ leaves) should be sampled, when, in fact, 20 shoots should be sampled based on the observed heterogeneity.

## DISCUSSION

There are many ways of characterizing the heterogeneity of unmapped or sparsely sampled discrete data such as plant disease incidence (4,16,38). Detection and testing of aggregation for single data sets can be accomplished with statistics based on the index of aggregation ($D$), and standard normal $Z$ from the $C(\alpha)$ test, and the degree of aggregation can be assessed based on the magnitude of $D$ and $Z$. Theoretical frequency distributions, such as the beta-binomial, can be fitted to data, and estimates of certain parameters can then be used to measure aggregation (11). Moreover, when there is a collection of data sets, the modified power law (equation 5) can be used to provide an overall measure of aggregation (10). Using these methods with the incidence data of downy mildew, the incidence of disease on leaves was aggregated, but the degree of aggregation was variable and generally dependent on mean incidence.

It is well established that some indices of aggregation, such as the variance-to-mean ratio (VM), which measures the number of organisms "intimately associated" with a randomly chosen organism (38), are dependent on the mean density of organisms (4). Larger means result in larger VM values, with all other factors held constant. The binary data equivalent of VM, $D$, would also be expected to vary with the mean, although the largest $D$ values would be envisioned in the midrange of disease incidence. In general, this was found for the downy mildew incidence data. A majority of plots had a large and significant variance ratio only when $p > 0.05$ (Table 3) and $D$ had the largest values in the midrange of $p$ values (data not shown). To our knowledge, the $C(\alpha)$ test has not been used previously for assessing aggregation of disease, and the expected relationship between $Z$ of the $C(\alpha)$ test and mean incidence is unknown. However, we found that $Z$ varied with $p$ and gave similar results to the tests based on $D$. This is not surprising given that $Z$ is a function of $v(X)$ and, hence, $D$ (equation 7). High percentages of the plots had significant aggregation, based on $Z$ (and $D$), for $0.10 < p < 1.0$. Overall, the largest values of $Z$ were found for the midrange of incidence. Moreover, maximum likelihood estimates of the beta-binomial $\theta$ parameter were near 0 at $p < 0.05$, somewhat higher at $0.05 < p < 0.10$ and $p > 0.8$, and highest in the midrange of disease incidence (Table 3). Because $Z$ is easy to calculate and gives results similar to $\theta$ (Fig. 3), it can be recommended as a simple and statistically powerful way of estimating overdispersion in disease incidence data.

As predicted by the binary form of the power law (11), the beta-binomial generally provided a good description of the incidence data in each plot. At $p > 0.10$, the parameter estimation procedure was successful in over 90% of the data sets. By definition, estimation was not possible for the plots with $D < 1.0$. At low values of incidence, lack of convergence likely was due to the very few nonzero values in the data set (18). For most plots, and especially for $p > 0.10$, the beta-binomial provided a better representation of the incidence data than the binomial distribution, based on LRS, in agreement with the $C(\alpha)$ test. The data from these plots exhibited extra-binomial variation (overdispersion). That is, there were more shoots with either a very small or a very large proportion of leaves with at least one lesion than would have been expected on the basis of a random pattern, as described by the binomial distribution. These plots also had correspondingly fewer shoots with approximately the mean proportion of diseased leaves. The percentages significant were slightly lower for the LRS than for $C(\alpha)$, reflecting the greater power of the $C(\alpha)$ test compared to the LRS at small sample sizes (26). Although $N = 15$ is a sufficient number of sampling units for estimating variances precisely (21), and, hence, regressing $\log(v)$ on $\log(p[1 - p]/\bar{n})$ (35), this number is relatively low for distribution fitting (38). Estimates of the parameters can have high standard errors in this situation. Nevertheless, it is not uncommon to fit this distribution to as few as five sampling units (6) in other disciplines.

In this study, the binary form of the power law (equation 5) precisely and consistently represented the overall heterogeneity of incidence of grape downy mildew. Parameter estimates were very similar among years and between assessment times within years, and they did not vary much with different estimation procedures. In fact, there was no significant difference in regression results among data sets. Moreover, fungicide treatment of plants in plots did not significantly affect the model parameters. The estimated $b$ for the pooled data based on ordinary least squares regression equaled 1.30, with a 95% confidence interval of 1.21–1.39. Estimates of $b$ were only slightly higher based on geometric-mean and resistant-line regression. With higher variation around the lines, one would expect a greater effect of statistical method on estimate of $b$ (36). Clearly, $b$ was greater than 1, the expected value for a binomial distribution and random dispersion. The estimate of $b$ was typical for a large number of organisms that are known to have aggregated spatial patterns (27,35–37).

The estimate of $A$ for the pooled data equaled $10^{0.93} = 8.5$, with a 95% confidence interval (calculated by back-transforming the confidence interval for $\log[A]$) of 5.2–13.9. When $b = 1$, $A$ is simply the ratio of observed to theoretical variance for a random pattern ($=D$). Because $b > 1$, $A$ would be the expected observed variance ($v$) if $v_r$ equaled 1. Because $v_r$ ($=p[1 - p]/\bar{n}$) cannot equal 1, $\log(A)$ must be interpreted as an overall measure of the height of the $\log(v)$:$\log(v_r)$ line.

Taylor (34) and Taylor et al (35–37) demonstrated that aggregation of a large number of organisms varies consistently with mean density according to the power law and that $b$ is a species-specific character, although the magnitude can change with varying conditions. Our results indicate that $b$ for downy mildew incidence was a stable character, at least for one geographic location. Also, Taylor et al (37) showed that aggregation at individual densities can be predicted with the parameters of the power law. Analogous predictions can be made using the binary version of the power law, namely, that aggregation (measured as $\theta$) is low at $p$ near 0, increases to a maximum, and then declines as $p$ approaches 1.0 (11). Determination of a simple relationship between any of the measures of aggregation and $p$ was complicated here by the magnitude of $N$ and the variable number of leaves (=observations) per sampling unit ($n$), which also influences aggregation. Taylor et al (37) postulated that the observed relationship between aggregation and the mean, such as equation 6 in (11), is a result of the power law. However, it is also likely that the power law is a result of a fundamental dynamic relationship between aggregation and the mean. In either case, the advantage of the power law is that aggregation over all data sets is described by the fewest number of parameters, namely $A$ and $b$. Other approaches require calculation of a parameter (e.g., $D$, $Z$, or $\theta$) for each data set to characterize aggregation.

Despite the economic importance of the disease and the need to make management decisions based on environmental or disease data, there is very little published information on the spatial or temporal distribution of grape downy mildew. In one study, Blaise et al (3) assessed aggregation in four commercial vineyards of *V. vinifera* during 1 yr by counting the number of lesions per vine. A standard power law analysis of these counts gave an estimate of $b = 1.6$. Like our study, data from Blaise et al (3) correspond to the situation in which secondary spread likely occurred. Considering that the sampling units varied between the studies (shoots versus vines), heterogeneity of lesion counts was remarkably similar to heterogeneity of disease incidence, as measured by the appropriate form of the power law. Although either lesion counts or disease incidence could be used to assess aggregation, in many survey situations it may be more efficient and easier to determine incidence, especially when severity is low. If some measure of severity, such as lesion counts, is needed for management decisions, one could then predict severity if there

was a repeatable relationship between the two (29).

Another study, by Seem et al (30), was concerned with the development of a sampling procedure for the detection of downy mildew in vineyards of *V. vinifera*. In that case, it was assumed that when disease incidence was very low, disease was distributed at random. Our finding that the degree of aggregation varies with mean incidence in such a way that it is highest in the midrange of incidence and lowest (the pattern then being essentially random) at low incidence provides support for this assumption. However, our findings show that it would not be valid to assume randomness when estimating disease at higher levels of incidence. Our sampling curve has a different purpose to the scheme of Seem et al (30). Figure 4b allows for determination of the number of shoots to sample for estimation of mean disease incidence with a prespecified degree of precision and takes aggregation into account. Seem et al (30) were concerned with disease detection and were able to ignore the problem of aggregation. The two approaches are complementary, each being appropriate at a different stage in the disease-management process.

## LITERATURE CITED


1. Binns, M. R. 1986. Behavioural dynamics and the negative binomial distribution. Oikos 47:315-318.
2. Blaeser, M., and Weltzien, H. C. 1979. Epidemiologische Studien an *Plasmopara viticola* zur Verbesserung der Spritzterminbestimmung. Z. Pflanzenkr. Pflanzenschutz 86:489-498.
3. Blaise, P., Schumacher, P., and Gessler, C. 1994. Forecast of downy mildew on grapevine: Sampling disease severity in a vineyard. Pages 42-48 in: Proc. 1st Int. Work. Grapevine Downy Mildew. D. M. Gadoury and R. C. Seem, eds. N.Y. Agric. Exp. Stn. Special Rep. 68.
4. Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. Wiley Interscience, New York.
5. Cochran, W. G. 1977. Sampling Techniques. 3rd ed. John Wiley and Sons, New York.
6. Crowder, M. J. 1978. Beta-binomial ANOVA for proportions. Appl. Stat. 27:34-37.
7. Downing, J. A. 1986. Spatial heterogeneity: Evolved behavior or mathematical artefact? Nature (London) 323:255-257.
8. Ellis, M. A., Madden, L. V., and Lalancette, N. 1994. A disease forecasting program for grape downy mildew in Ohio. Pages 92-95 in: Proc. 1st Int. Work. Grapevine Downy Mildew. D. M. Gadoury and R. C. Seem, eds. N.Y. Agric. Exp. Stn. Special Rep. 68.
9. Fisher, R. A. 1970. Statistical Methods for Research Workers. 14th ed. Hafner Publishing Company, New York.
10. Hughes, G., and Madden, L. V. 1992. Aggregation and incidence of disease. Plant Pathol. 41:657-660.
11. Hughes, G., and Madden, L. V. 1993. Using the beta-binomial distribution to describe aggregated patterns of disease incidence. Phytopathology 83:759-763.
12. Hughes, G., and Madden, L. V. 1994. Aggregation and incidence of disease: Some implications for sampling. Aspects Appl. Biol. 37:25-31.
13. Lalancette, N., Ellis, M. A., and Madden, L. V. 1988. Development of an infection efficiency model of *Plasmopara viticola* on American grape based on temperature and duration of leaf wetness. Phytopathology 78:794-800.
14. Lalancette, N., Madden, L. V., and Ellis, M. A. 1988. A quantitative model for describing the sporulation of *Plasmopara viticola* on grape leaves. Phytopathology 78:1316-1321.
15. Looney, S. W., and Gulledge, T. R., Jr. 1985. Use of the correlation coefficient with normal probability plots. Am. Stat. 39:75-79.
16. Madden, L. V. 1989. Dynamic nature of within-field disease and pathogen distributions. Pages 96-126 in: Spatial Components of Plant Disease Epidemics. M. J. Jeger, ed. Prentice Hall, Englewood Cliffs, NJ.
17. Madden, L. V. 1993. Aggregation of *Colletotrichum acutatum* in response to simulated rain episodes. J. Phytopathol. 138:145-156.
18. Madden, L. V., and Hughes G. 1994. BBD—Computer software for fitting the beta-binomial distribution to disease incidence data. Plant Dis. 78:536-540.
19. Madden, L. V., Louie, R., and Knoke, J. K. 1987. Temporal and spatial analysis of maize dwarf mosaic epidemics. Phytopathology 77:148-156.
20. Madden, L. V., Pirone, T. P., and Raccah, B. 1987. Analysis of spatial patterns of virus-diseased tobacco plants. Phytopathology 77:1409-1417.
21. McArdle, B. H., Gaston, K. J., and Lawton, J. H. 1990. Variation in the size of animal populations: Patterns, problems and artefacts. J. Anim. Ecol. 59:439-454.
22. Minitab Incorporated. 1991. Minitab Statistical Software: MINITAB Reference Manual, Release 8. Minitab Inc., State College, PA.
23. Neter, J., and Wasserman, W. 1974. Applied Linear Statistical Models. Richard D. Irwin, Inc., Homewood, IL.
24. Neyman, J. 1959. Optimal asymptotic tests of composite hypotheses. Pages 213-234 in: Probability and Statistics. U. Granander, ed. The Harold Crainer Volume, Wiley, New York.
25. Park, E. W., Seem, R. C., Pearson, R. C., and Gadoury, D. M. 1994. DMCAST: A forecasting model for grape downy mildew development. Pages 96-102 in: Proc. 1st Int. Work. Grapevine Downy Mildew. D. M. Gadoury and R. C. Seem, eds. N.Y. Agric. Exp. Stn. Special Rep. 68.
26. Paul, S. R., Liang, K. Y., and Self, S. G. 1989. On testing departure from binomial and multinomial assumptions. Biometrics 45:231-236.
27. Perry, J. M., and Taylor, L. R. 1986. Stability of real interacting populations in space and time: Implications, alternatives and the negative binomial $k_c$. J. Anim. Ecol. 55:1053-1068.
28. Qu, Y., Beck, G. J., and Williams, G. W. 1990. Polya-Eggenberger distribution: Parameter estimation and hypothesis tests. Biom. J. 2:229-242.
29. Seem, R. C. 1984. Disease incidence and severity relationships. Annu. Rev. Phytopathol. 22:137-150.
30. Seem, R. C., Magarey, P. A., McCloud, P. I., and Wachtel, M. F. 1985. A sampling procedure to detect grapevine downy mildew. Phytopathology 75:1252-1257.
31. Shiyomi, M., and Takai, A. 1979. The spatial pattern of infected or infested plants and negative hypergeometric series. Jpn. J. Appl. Entomol. Zool. 23:224-229. In Japanese.
32. Skellam, J. G. 1948. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between sets of trials. J. R. Stat. Soc. B 10:257-261.
33. Tarone, R. E. 1979. Testing the goodness of fit of the binomial distribution. Biometrika 66:585-590.
34. Taylor, L. R. 1961. Aggregation, variance and the mean. Nature (London) 189:732-735.
35. Taylor, L. R., Perry, J. N., Woiwod, I. P., and Taylor, R. A. J. 1988. Specificity of the spatial power-law exponent in ecology and agriculture. Nature (London) 332:721-722.
36. Taylor, L. R., and Woiwod, I. P. 1982. Comparative synoptic dynamics I. Relationships between inter- and intra-specific spatial and temporal variance/mean population parameters. J. Anim. Ecol. 51:879-906.
37. Taylor, L. R., Woiwod, I. P., and Perry, J. N. 1978. The density-dependence of spatial behaviour and the rarity of randomness. J. Anim. Ecol. 47:383-406.
38. Upton, G., and Fingleton, B. 1985. Spatial Data Analysis by Example. Vol. 1, Point Pattern and Quantitative Data. John Wiley and Sons, Chichester, England.
39. Velleman, P. F., and Hoaglin, D. C. 1981. Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press, Boston.