Ecology and Epidemiology

Spatial and Temporal Stability of the Estimated Parameters of the Binary Power Law

D. H. Gent, W. W. Turechek, and W. F. Mahaffee

First author: U.S. Department of Agriculture-Agricultural Research Service (USDA-ARS), Forage Seed and Cereal Research Unit, Oregon State University, Department of Botany and Plant Pathology, Corvallis 97331; second author: USDA-ARS, U.S. Horticultural Research Laboratory, Fort Pierce, FL 34945; and third author: USDA-ARS, Horticultural Crops Research Laboratory, Oregon State University, Department of Botany and Plant Pathology, Corvallis 97330.

Accepted for publication 9 June 2008.

ABSTRACT

Gent, D. H., Turechek, W. W., and Mahaffee, W. F. 2008. Spatial and temporal stability of the estimated parameters of the binary power law. Phytopathology 98:1107-1117.

The incidence of hop powdery mildew on leaves, caused by *Podosphaera macularis*, collected from 1,606 transects in 77 commercial hop yards in Oregon and Washington over 9 years was used to assess variability in heterogeneity of disease and the estimated binary power law parameters. Spatial analyses of data sets were conducted at the level of individual rows (row level) and multiple rows within a yard (yard level). The binary power law provided a good fit to all data sets, with R^2 values ranging from 0.933 to 0.993. At the row level, the intercept parameter $\ln(A_x)$ was >0 for 8 years, but was not significantly greater than 0 in 2006. The parameter *b* was greater than 1 for all row-level data sets collected from 1999 to 2005, but was <1 in 2006 and not significantly different from 1 in 2007. Covariance analysis indicated the factor ‘region’ affected $\ln(A_x)$ in 3 years, and *b* in 2 years. ‘Cultivar’ had an effect on $\ln(A_x)$ in 3 years and *b* in year. At the yard level, $\ln(A_x)$ was greater than 0 for 6 years, but in 2006 and 2007, $\ln(A_x)$ was not significantly different from 0. The slope parameter *b* was greater than 1 in 6 years, but was not significantly different from 1 in 2006 and 2007. Differences in *b* among years were large enough to have practical implications for sample sizes and precision of fixed and sequential sampling. Although the binary power law parameter tended to be relatively stable, variability of the estimated parameters may have practical consequences for sampling precision and costs.

Additional keywords: quantitative epidemiology, *Sphaerotheca macularis*.

Description and quantification of spatial patterns of plant disease are integral to botanical epidemiology, providing insight into biological processes associated with pathogen dispersal and disease development, and allowing for the development of testable hypotheses to explain observed patterns (18). Quantification of spatial pattern is prerequisite to designing statistically sound sampling methods (3–5,7,14,16,30,31) and controlling the effects of aggregation in analysis of data from designed field experiments (19), and also has been used in development of crop loss models (6).

For disease incidence data, the binary form of Taylor’s power law (23–25) provides a simple model with only two parameters to detect and quantify aggregation at the scale of individual sampling units. Taylor’s power law relates the observed variance (v_{obs}) to the population mean (*m*) for count data with no upper limit through a power function:

$$v_{\text{obs}} = am^b \tag{1}$$

where *a* and *b* are parameters typically estimated by least squares regression after logarithmic transformation. For count data, when $v_{\text{obs}} = m$ (i.e., $a = b = 1$) the data can be described by the Poisson distribution, indicating a random pattern. Parameter *b* is a rate of chance of log variance with log mean, and is considered an index of aggregation (28). However, one can consider Taylor’s power law as describing the empirical relationship between two variances, an observed variance (v_{obs}) and the variance of the population if it were randomly distributed (*m*), since the variance is equal to the mean for the Poisson distribution (21). For binary data, the binomial is the appropriate distribution for a random spatial pattern (13,18), which has a variance of $v_{\text{bin}} = np(1 - p)$, where *n* is the number of observations (individuals) within a sampling unit and *p* is the probability of a plant being diseased. The estimate of *p*, $\hat{p}$, is equivalent to disease incidence, the mean proportion of plants diseased. Thus, the binary power law can be written as

$$v_{\text{obs}} = A_x\, v_{\text{bin}}^b = A_x[n\hat{p}(1 - \hat{p})]^b \tag{2}$$

where A_x and *b* are parameters. Using the notation of Turechek and Madden (32), the binary power law can be written as

$$\ln(v_{\text{obs}}) = \ln(A_x) + b\ln(v_{\text{bin}}) \tag{3}$$

in its linear form after log transformation. When $A_x = 1$ and $b = 1$, equation 3 indicates a random pattern of diseased individuals that can be represented by the binomial distribution. When $A_x > 1$ and $b = 1$, disease incidence has an aggregated pattern that is not dependent on *p*. Values of $b > 1$ indicate that aggregation is systematically related to *p*, which is typical of most plant diseases at multiple scales (2,3,11,13,16,30,32,33,35).

The binary power law provides a convenient method of quantifying and characterizing aggregation of disease for multiple data sets when they are collected at a single or over several time periods (8,13). The power law model in equation 3 can be expanded allowing a covariance analysis to determine the effect of factors or class variables (e.g., cultivar, fungicide application, and year) on the degree of aggregation (3,16,31–33). Gent et al. (3) analyzed the incidence of the hop powdery mildew (caused by

Corresponding author: D. H. Gent; E-mail address: gentd@onid.orst.edu

---

doi:10.1094/PHYTO-98-10-1107

Podosphaera macularis) on hop cones, and found that year of sampling affected parameter estimates, with a tendency for the intercept ($\ln[A_x]$), and slope (b) to increase with time. This raises questions about the temporal stability of the parameters of the binary power law. Explanations for this variability are speculative, but may include natural variability of $\ln(A_x)$ and b caused by seasonal weather patterns, the impacts of different and/or changing management practices, or the variability may represent a temporal process where $\ln(A_x)$ and b are approaching relatively stable, mean values as the pathogen adapts to the new environment. However, in other pathosystems parameter estimates were stable when compared over a period of 2 to 3 years. Madden et al. (16) suggested that for grape downy mildew (caused by *Plasmopara viticola*) b was a "stable character" at the geographic location where data sets were collected.

While instantaneous spatial patterns are transient and rarely identical, the relationship between mean and variance of counts generally are stable and robust across space and time (23-25,27-29). In their study of 156 data sets comprising counts of organisms (with no upper limit) of 102 species, Taylor et al. (28) found that Taylor's power law provided an adequate model to describe the relationship between mean population density and variance over nearly all taxa, scales, and sampling methods investigated in unmanaged systems. Sampling of numerous aphid and moth species at multiple locations over a period of 6 to 14 years also suggested that parameter b was temporally stable (27), and the authors suggested that, "…temporal stability is largely an intrinsic species property." However, the intercept parameter a is reported to differ among environments (26,28,29). With constant b, a change in the parameter a indicates that environment affects variance and, consequently, the magnitude (or degree) of aggregation equally across the range of possible densities (26).

The portability and temporal stability of parameter estimates of the binary form of the power law are unclear, partly due to the lack of available long-term data sets for disease incidence data. Stability of the parameter estimates in managed agroecosystems also is unknown. In this study, incidence of hop powdery mildew on leaves collected from commercial hop yards over 9 years was used to assess variability in heterogeneity of disease and the binary power law parameters. From a practical standpoint we also investigate the impact of parameter variability on fixed and sequential sampling plans for this disease. The importance of proper covariance analysis and sampling over time to estimate parameters of the binary power law are emphasized.

MATERIALS AND METHODS

Field sites and data collection. The incidence of powdery mildew on hop leaves was assessed from 1,606 transects in 77 commercial hop yards in the primary hop growing regions of western Oregon (Willamette Valley) and three separate regions in central Washington, namely the Yakima Indian Reservation, near the town of Moxee City (Moxee), and in the eastern extent of the Yakima Valley (Lower Valley) from 1999 to 2007. Data sets were constructed from the disease incidence data from all sampling units within an individual row and also all sampling units assessed in that yard, and are herein referred to as row- and yard-level data sets, respectively (33). Data sets from 1999 to 2001 were reported previously (33). Climate, cultural practices, cultivars, and disease severity vary among these regions (3,20,33), and typically the greatest disease severity is observed near the Yakima Indian Reservation, followed by Moxee, the Lower Valley, and Oregon (34).

Yards surveyed in Oregon were planted exclusively with the aroma cultivars Glacier (41 yard-level data sets), Liberty (3 data sets), Perle (71 data sets), and Willamette (187 data sets) that are moderately to highly susceptible to hop powdery mildew. Yards surveyed in Washington were planted primarily with the bittering cultivars Chelan (12 yard-level data sets), Columbus, Tomahawk, or Zeus (genetically indistinguishable and collectively referred to as CTZ) (426 yard-level data sets), or Galena (5 yard-level data sets) that are highly susceptible to hop powdery mildew. Twenty-five yard-level data sets were collected from yards planted to cultivar Willamette in Washington. A total of 770 yard-level data sets and 1,606 row-level data sets where $\hat{p} > 0$ were included in the analysis.

Disease incidence was assessed using a cluster sampling design (10), as described in detail previously (33). In 1999, a single transect (row) was sampled from each yard. In 2000, yards were stratified into H strata, where H = number of rows in a yard/20 (rounded-up to the nearest integer), and one transect in each of the strata was selected arbitrarily for sampling. From 2001 to 2007, a transect was sampled from each of the first two strata from the first 40 rows of a yard. Individual yards were sampled multiple times in each season, but transects were chosen arbitrarily on each sampling date. The number of individual yards sampled within a season range from 5 (2006) to 54 (2001).

Ten leaves (n) were sampled arbitrarily from each of the first 75 to 100 (N) plants along the transect, or until the end of the row if the row contained less than the desired number of plants. Each leaf was rated for the incidence (presence or absence) of powdery mildew. Mean $\hat{p}$ was calculated as $\hat{p} = \Sigma x_i/\Sigma n_i$, where x_i is the number of diseased leaves and n_i is the number of leaves sampled in the ith sampling unit.

Distributional analyses. The beta-binomial and binomial distributions were fit to the incidence data using the computer program BBD (12). For binary data collected as a cluster sample, a good fit to the binomial distribution is an indication of a random pattern of diseased plants, whereas a good fit to the beta-binomial distribution is an indication of an aggregated disease pattern (13). A log-likelihood ratio test statistic was calculated to determine whether the data fit the beta-binomial distribution better than the binomial distribution (22).

The degree of aggregation of disease incidence was quantified using the parameter θ of the beta-binomial distribution, which provides a measure of variation in disease incidence per sampling unit (13,15). The index of dispersion (D) was calculated by dividing the observed variance of diseased leaves (v_{obs}) by the theoretical variance for a binomial distribution (v_{bin}) where $v_{\text{obs}} = [\sum(x_i - \hat{p}n_i)^2]/(N-1)$ and $v_{\text{bin}} = n\hat{p}(1-\hat{p})$ and x_i, $\hat{p}$, n, and N are as defined previously. When $\theta = 0$ or $D = 1$, the pattern of diseased plants is random, with aggregation indicated when $D > 1$ or $\theta > 0$ and the degree of aggregation directly proportional to the magnitude of the statistic. D has a chi-square distribution, and can be used to test the null hypothesis of a random distribution of disease incidence with $N - 1$ degrees of freedom (13,15).

Binary power law analyses. The binary power law model (equation 3) was fitted to the observed and binomial variances. Ordinary least squares regression was used to estimate the intercept and slope parameters using the SAS procedure PROC REG (SAS version 9.1, SAS Institute, Cary, NC). In some instances, only a single diseased leaf was found during the sampling at the row or yard level. Taylor (23) referred to discrete data sets with only a single observed individual as "singletons" and suggested excluding these data sets because spatial aggregation cannot be interpreted for one individual. Regression models were fit to all data sets and also to data sets where singletons were removed. Differences in parameter estimates were determined by t tests using Microsoft Excel (Microsoft Corp., Redmond, WA).

A covariance analysis was conducted on the data sets using the GENMOD procedure in SAS to determine the effect of the factors 'cultivar', 'region', and 'year' on the slope and intercept parameters of the binary power law as described previously (33). To conduct the covariance analysis equation 3 was expanded to include the factors (class variables). The continuous variable $\ln[n\hat{p}(1-\hat{p})]$ was included in the model first, and then each of the three factors was added individually as an intercept term and then as an interaction term with the slope. The analyses were conducted on each year, and then a separate analysis was conducted to determine the effect of year on estimates of $\ln(A_x)$ and b. A factor was considered significant if inclusion of the factor as a covariate significantly reduced the sum of square error (SSE) as compared to the null (binary power law) model without the factor. The significance level for the difference between SSE of the models was determined by an F test in which F = (factor SSE/df factor)/(model SSE/df model) and df = degrees of freedom.

TABLE 1. Mean and standard error (in parentheses) of the incidence ($\hat{p}$, the estimate of the probability of disease) of powdery mildew on hop leaves sampled at the yard level from hop yards in Oregon and Washington (Lower Valley, Moxee, and Yakima Indian Reservation [YIR])

			Washington							
	Oregon		Lower Valley		Moxee		YIR		Total	
Year	$\hat{p}$	T[a]	$\hat{p}$	T	$\hat{p}$	T	$\hat{p}$	T	$\hat{p}$	T
2000	0.01 (0.002)	38	0.02 (0.004)	41	0.06 (0.026)	20	0.12 (0.035)	20	0.04 (0.008)	119
2001	0.01 (0.003)	35	0.11 (0.034)	12	0.08 (0.027)	13	0.15 (0.044)	19	0.07 (0.014)	79
2002	0.005 (0.0005)	91	0.08 (0.015)	51	0.01 (0.005)	20	0.08 (0.016)	50	0.04 (0.006)	212
2003	0.03 (0.008)	43	0.05 (0.010)	8	0.01 (0.002)	18	0.04 (0.013)	17	0.03 (0.005)	86
2004	0.03 (0.006)	37	0.03 (0.014)	7	0.12 (0.032)	16	0.01 (0.003)	4	0.05 (0.010)	64
2005	0.01 (0.005)	37	0.05 (0.012)	33	0.12 (0.030)	23	0.18 (0.039)	35	0.09 (0.014)	128
2006	0.004 (0.001)	12	0.16 (0.023)	12	0.02 (0.012)	8	0.06 (0.016)	13	0.07 (0.012)	45
2007	0.02 (0.007)	9	0.07 (0.014)	18	0.02 (0.013)	4	0.09 (0.032)	6	0.06 (0.010)	37
Total	0.01 (0.002)	302	0.06 (0.006)	182	0.06 (0.009)	122	0.11 (0.012)	164	0.05 (0.004)	770

[a] Number of data sets assessed.

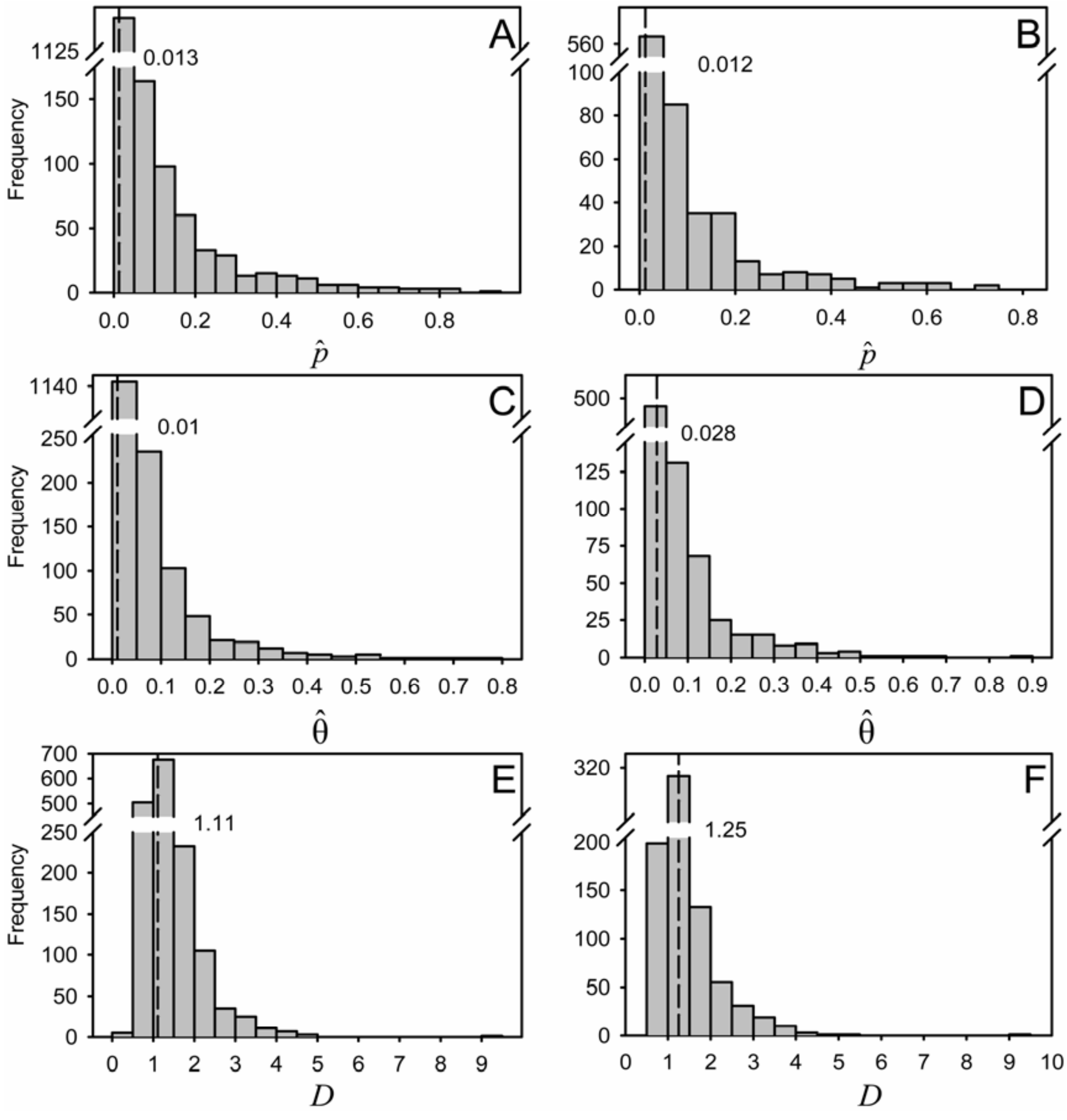

Fig. 1. Frequency distribution of the beta-binomial distribution parameters $\hat{p}$ (**A and B**) and $\hat{\theta}$ (**C and D**), the index of dispersion D (**E and F**) for the incidence of hop powdery mildew on leaves sampled from hop yards from 1999 to 2007 (**A, C, and E**) or 2000 to 2007 (**B, D, and F**). Data sets in **A, C,** and **E** were derived at the row scale, and data sets in **B, D,** and **F** were derived at the yard scale. Vertical dashed lines are the median value for the indicated statistic, with the numerical value given on the figure.

Sample size analyses. The effect of changes in ln(A_x) and b on sample size as determined by fixed and sequential sampling models were quantified for the yard-level analyses from all years combined, as well as 2004 and 2006 since b varied the most between these years. Statistical methods for sample size determination were based on, and fully described in, Madden and Hughes (15) and Madden et al. (17).

Fixed sampling curves were developed to predict the number of sampling units necessary to estimate the incidence of hop powdery mildew on leaves with varying levels of precision. The number of sampling units (N) needed to estimate p with a coefficient of variation of the mean, $C = 0.2$ was calculated by

$$N = \hat{a}\hat{p}^{\hat{b}-2}(1-\hat{p})^{\hat{b}}/C^2 \quad (4)$$

where C = se[$\hat{p}$]/ $\hat{p}$, se $\hat{p}$ = $\sqrt{\hat{a}[\hat{p}(1-\hat{p})]^{\hat{b}}/N}$, $\hat{a} = \hat{A}_x n^{\hat{b}-2}$ and $\hat{A}_x$ and $\hat{b}$ are parameters estimated from the binary power law from all years, 2004, or 2006 (9).

For sequential estimation, the se of $\hat{p}$ was expressed in terms of the binary power law parameters as

$$\sqrt{\hat{a}[\hat{p}(1-\hat{p})]^{\hat{b}}/N} \quad (5)$$

where $\hat{a}$ and $\hat{b}$ are as defined previously. Since heterogeneity was systematically related to disease incidence, stop lines were derived numerically by the formula

$$\gamma_N = T_N^{\hat{b}-2}(nN - T_N)^{\hat{b}} = (C^2/\hat{a})n^{2\hat{b}-2}N^{2\hat{b}-1} \quad (6)$$

A Mathcad (Mathsoft Inc., Cambridge, MA) worksheet developed previously (30) was used to solve equation 6 iteratively for T_N when N = 1 to 500 and C = 0.2.

RESULTS

Disease incidence and distributional analyses. Disease incidence varied among years and regions, with no discernable trends for an overall increase or decrease in disease incidence among row or yard-level data sets (Table 1). The region where the greatest mean disease incidence was observed varied each year,

TABLE 2. Mean and standard error (in parentheses) of the beta-binomial parameter $\hat{\theta}$ for incidence of hop leaves with powdery mildew sampled at the yard-level from hop yards in Oregon and Washington (Lower Valley, Moxee, and Yakima Indian Reservation [YIR])

			Washington							
	Oregon		Lower Valley		Moxee		YIR[a]		Total	
Year	$\hat{\theta}$	T[a]	$\hat{\theta}$	T	$\hat{\theta}$	T	$\hat{\theta}$	T	$\hat{\theta}$	T
2000	0.03 (0.005)	38	0.02 (0.004)	41	0.02 (0.006)	20	0.05 (0.013)	20	0.03 (0.003)	119
2001	0.04 (0.008)	35	0.05 (0.010)	12	0.07 (0.025)	13	0.08 (0.010)	19	0.05 (0.006)	79
2002	0.12 (0.087)	91	0.10 (0.017)	51	0.05 (0.019)	20	0.07 (0.009)	50	0.10 (0.038)	212
2003	0.06 (0.010)	43	0.05 (0.018)	8	0.03 (0.008)	18	0.03 (0.009)	17	0.05 (0.006)	86
2004	0.14 (0.025)	37	0.09 (0.046)	7	0.13 (0.023)	16	0.01 (0.011)	4	0.13 (0.017)	64
2005	0.06 (0.017)	37	0.09 (0.017)	33	0.17 (0.052)	23	0.07 (0.015)	35	0.09 (0.012)	128
2006	0.03 (0.011)	12	0.09 (0.054)	12	0.02 (0.014)	8	0.03 (0.011)	13	0.04 (0.015)	45
2007	0.08 (0.044)	9	0.03 (0.015)	18	0.00 (0.002)	4	0.03 (0.012)	6	0.04 (0.013)	37
Total	0.08 (0.027)	302	0.07 (0.007)	182	0.07 (0.012)	122	0.06 (0.006)	164	0.07 (0.011)	770

[a] T = number of data sets.

TABLE 3. Slope and intercept parameter estimates of the binary power law for the incidence of leaves with powdery mildew from data sets collected from hop yards in Oregon and Washington

		All data sets[b]					Without singletons[b]						t value[c]	
Year	T[a]	ln($\hat{A}_x$)	se	$\hat{b}$	se	R^2	T	ln($\hat{A}_x$)	se	$\hat{b}$	se	R^2	ln($\hat{A}_x$)	$\hat{b}$
Row level														
1999	44	0.342	0.036	1.104	0.020	0.986	42	0.343	0.037	1.116	0.023	0.983	–0.027	–0.383
2000	370	0.266	0.189	1.063	0.007	0.985	315	0.268	0.021	1.065	0.009	0.978	–0.011	–0.151
2001	164	0.458	0.030	1.113	0.013	0.977	150	0.463	0.031	1.122	0.017	0.967	–0.130	–0.420
2002	362	0.529	0.027	1.131	0.010	0.971	302	0.544	0.030	1.146	0.014	0.957	–0.370	–0.892
2003	157	0.502	0.036	1.120	0.013	0.978	138	0.511	0.040	1.127	0.017	0.969	–0.173	–0.334
2004	112	0.719	0.044	1.173	0.018	0.975	100	0.728	0.047	1.184	0.023	0.965	–0.137	0.376
2005	231	0.521	0.031	1.115	0.013	0.971	202	0.520	0.033	1.114	0.017	0.954	0.013	0.061
2006	81	–0.086	0.054	0.931	0.024	0.950	74	–0.104	0.056	0.900	0.029	0.928	0.234	0.809
2007	85	0.162	0.055	1.016	0.026	0.949	79	0.153	0.058	1.004	0.030	0.933	0.106	0.287
All	1,606	0.415	0.012	1.097	0.005	0.970	1,402	0.418	0.013	1.100	0.006	0.957	–0.156	–0.434
Yard level														
2000	119	0.418	0.027	1.092	0.009	0.993	119	0.418	0.027	1.092	0.009	0.993	–	–
2001	79	0.490	0.041	1.100	0.017	0.982	79	0.490	0.041	1.100	0.017	0.982	–	–
2002	212	0.653	0.037	1.146	0.013	0.975	191	0.665	0.040	1.156	0.016	0.967	–0.220	–0.002
2003	86	0.583	0.041	1.134	0.015	0.986	79	0.605	0.045	1.149	0.018	0.981	–0.361	–0.002
2004	64	0.861	0.068	1.175	0.028	0.967	62	0.864	0.071	1.178	0.030	0.962	–0.031	–0.002
2005	128	0.667	0.038	1.139	0.014	0.981	114	0.672	0.041	1.147	0.018	0.973	–0.089	–0.002
2006	45	0.127	0.078	0.983	0.032	0.957	44	0.123	0.079	0.978	0.033	0.954	0.036	–0.001
2007	37	0.166	0.084	0.994	0.036	0.956	36	0.156	0.086	0.984	0.040	0.946	0.083	–0.001
All	770	0.554	0.018	1.117	0.007	0.975	724	0.557	0.019	1.119	0.007	0.969	–0.115	–0.001

[a] T = number of data sets. Row and yard level refers to the scale of the analysis (33).

[b] "Singletons" are defined as data sets where only one diseased leaf was observed among all sampling units assessed in a yard. The parameters $\hat{b}$ and ln($\hat{A}_x$) are slope and intercept estimates, respectively. se = standard error of the mean.

[c] t values are for a test that slope or intercept estimates of the regression for data sets with and without singletons are equal. Values greater than |1.645| indicate significance at P = 0.05.

but was highest in hop yards assessed in Yakima Indian Reservation for 5 of the 8 years. Mean disease incidence was lowest in yards in Oregon for 6 of the 8 years.

For row level data sets, $\hat{p}$ ranged from 0.001 to 0.9 and was positively skewed with mean 0.064 (se = 0.003) and median 0.013 (Fig. 1A). The beta-binomial distribution parameter $\hat{\theta}$ ranged from 0 to 0.758 with mean 0.046 (se = 0.002) and median 0.01, indicating a low degree of aggregation (Fig. 1C). A log-likelihood ratio test indicated that the beta-binomial distribution provided a better fit to the data in 497 of 1,606 data sets (30.9%) than the binomial distribution. Similarly, *D* ranged from 0.4 to 9.17, with mean 1.34 (se = 0.015) and median 1.11 (Fig. 1E), and was significantly greater than 1 in 607 data sets (38.9%).

The distribution of $\hat{p}$, $\hat{\theta}$, and *D* also was positively skewed for yard-level data sets (Tables 1 and 2; Fig. 1B, D, and F). The distribution of $\hat{p}$ ranged from 0.0003 to 0.73 with mean 0.054

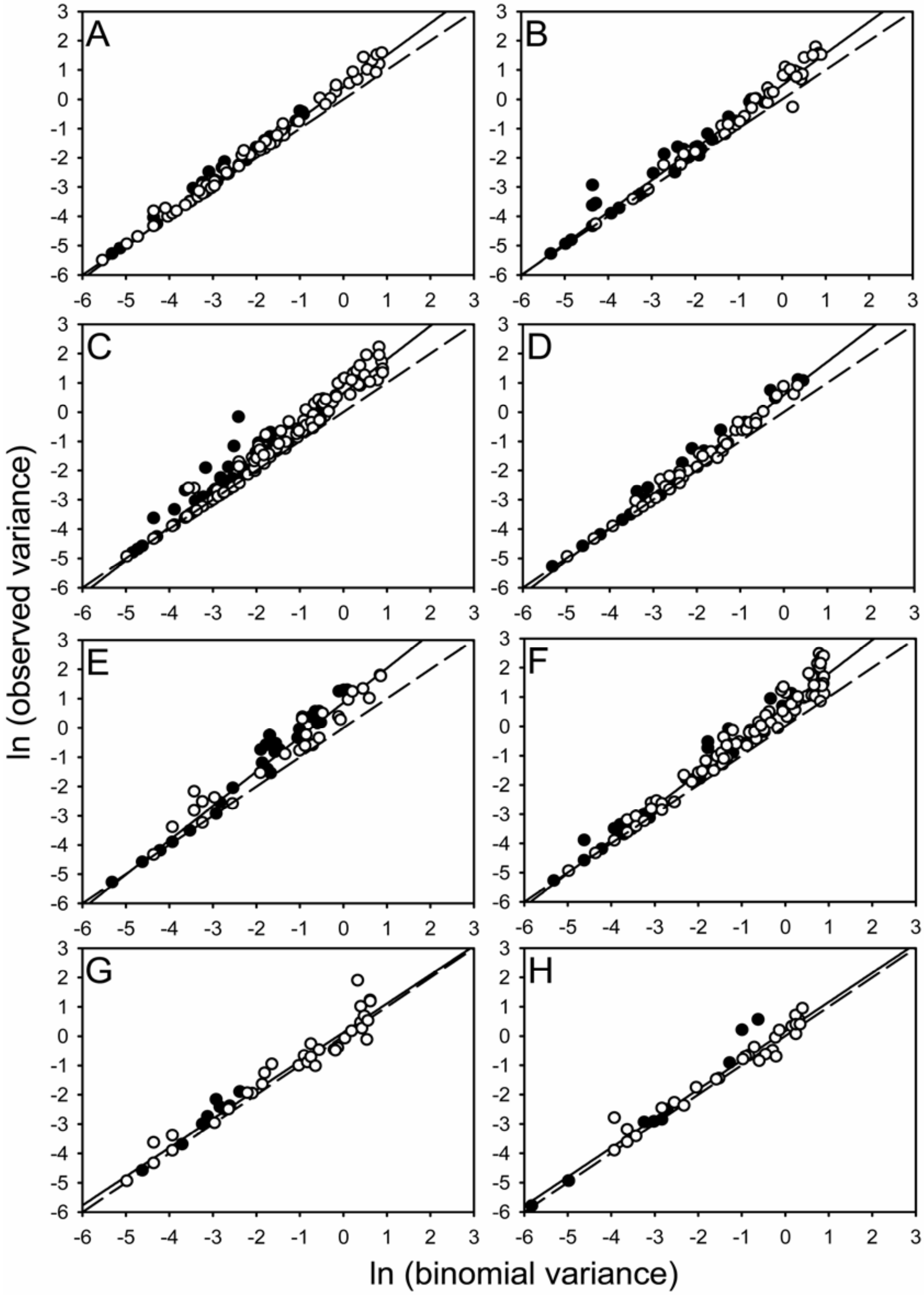

Fig. 2. Relationship between the logarithms of the observed variance and binomial variance for the incidence of hop powdery mildew on leaves sampled from 2000 to 2007 (**A** to **H** in ascending order). The solid line is the least squares regression fit to the data, and the dashed line is the line for a binomial distribution. Slope and intercept parameters are given in Table 4 for yard-level data sets. Data from 1999 are not presented since only one row (transect) was assessed per yard. Solid circles are data sets collected in Oregon, and open circles are data sets collected in Washington.

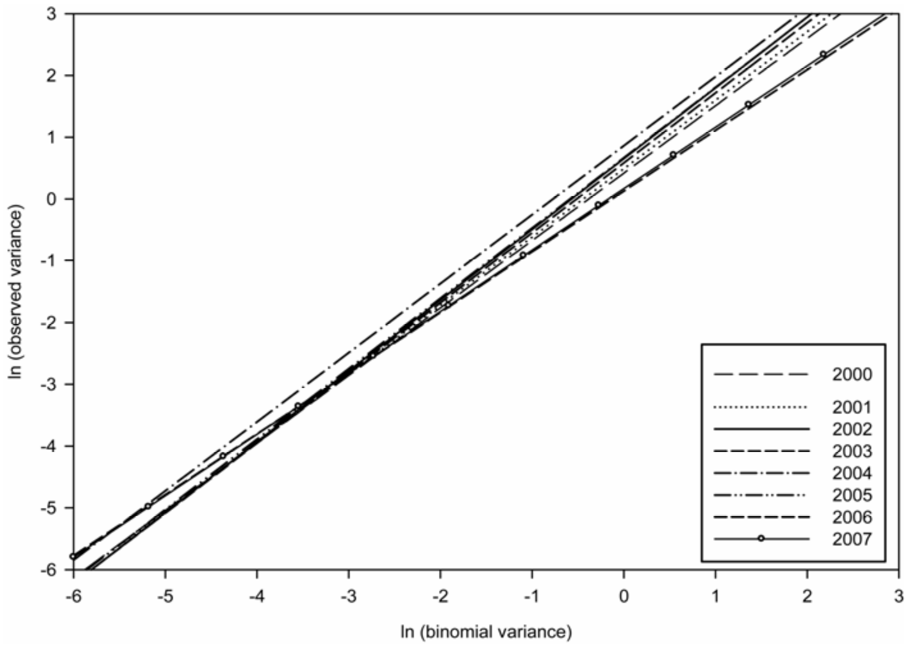

Fig. 3. Relationship between the logarithms of the observed variance and binomial variance for the incidence of hop powdery mildew on leaves sampled from 2000 to 2007. Slope and intercept parameters are given in Table 3 for yard-level data sets.

TABLE 4. Covariance analysis of the effect of cultivar and geographic region on intercept $\ln(\hat{A}_x)$ and slope ($\hat{b}$) parameters of the binary power law for the incidence of hop powdery mildew on leaves sampled at the row level

			$\ln(\hat{A}_x)$				$\hat{b}$			
Factor and year	df model[a]	df factor[a]	SSE[b]	Diff.[c]	*F*[d]	*P*	SSE	Diff.	*F*	*P*
1999										
Power law	42		2.009	–	–	–	2.009	–	–	–
Region	40	2	1.895	0.113	1.200	0.312	1.864	0.145	1.559	0.223
Cultivar	39	3	2.004	0.005	0.031	0.992	1.986	0.023	0.153	0.927
2000										
Power law	368		14.821	–	–	–	14.821	–	–	–
Region	365	3	14.646	0.175	1.450	0.228	14.806	0.015	0.126	0.945
Cultivar	365	3	14.697	0.124	1.027	0.381	14.779	0.042	0.346	0.792
2001										
Power law	162		13.138	–	–	–	13.138	–	–	–
Region	159	3	12.773	0.366	1.517	0.212	13.067	0.072	0.292	0.831
Cultivar	159	3	13.055	0.084	0.341	0.796	13.037	0.101	0.411	0.745
2002										
Power law	360		34.048	–	–	–	34.048	–	–	–
Region	357	3	31.858	2.190	8.182	0.000	33.164	0.885	3.174	0.024
Cultivar	358	2	32.748	1.300	7.106	0.001	33.444	0.605	3.235	0.040
2003										
Power law	155		9.750	–	–	–	9.750	–	–	–
Region	152	3	8.368	1.382	8.370	0.000	9.169	0.581	3.211	0.025
Cultivar	153	2	8.790	0.960	8.356	0.000	9.389	0.361	2.937	0.056
2004										
Power law	110		10.846	–	–	–	10.846	–	–	–
Region	107	3	10.537	0.309	1.046	0.375	10.635	0.211	0.708	0.549
Cultivar	108	2	10.296	0.550	2.885	0.060	10.730	0.116	0.582	0.560
2005										
Power law	229		27.027	–	–	–	27.027	–	–	–
Region	226	3	23.379	3.647	11.753	0.000	26.761	0.265	0.747	0.525
Cultivar	227	2	26.507	0.520	2.226	0.110	26.992	0.034	0.144	0.866
2006										
Power law	79		10.529	–	–	–	10.529	–	–	–
Region	76	3	10.410	0.118	0.288	0.834	9.709	0.820	2.139	0.102
Cultivar	77	2	10.269	0.260	0.974	0.382	10.418	0.111	0.409	0.666
2007										
Power law	83		10.267	–	–	–	10.267	–	–	–
Region	80	3	9.463	0.804	2.266	0.087	9.899	0.368	0.991	0.401
Cultivar	82	1	9.748	0.519	4.368	0.040	10.250	0.018	0.140	0.709

[a] df model = degrees of freedom for the model; df factor = degrees of freedom for factor.

[b] SSE = sum of square error for the covariance model.

[c] Diff. = difference between the SSE of the binary power law model versus the binary power law model with each factor included in the analyses first as an intercept and then as a slope.

[d] Significance level for the difference between SSE of the binary power law model versus binary power law model with each factor as determined by an *F* test, where F = (factor SSE/df factor)/(model SSE/df model).

(se = 0.003) and median 0.012. The parameter $\hat{\theta}$ ranged from 0 to 0.862 with mean 0.046 (se = 0.002) and median 0.028, again indicating a low degree of aggregation of disease incidence. The mean $\hat{\theta}$ varied considerably among regions and years (Table 2), and trends in mean $\hat{\theta}$ over time were not apparent. The region where the greatest value of $\hat{\theta}$ was observed also varied among years, being highest in yards in Oregon for 5 of the 8 years. A log-likelihood ratio test indicated the beta-binomial distribution fit 425 of 770 data sets (55.2%) better than the binomial distribution. D ranged from 0.5 to 9.08 with mean 1.473 (se = 0.026) and median 1.25. D was significantly greater than 1 in 367 data sets (47.7%).

Binary power law analyses. The binary power law provided a good fit to all data sets collected at the row and yard level, with R^2 values ranging from 0.933 to 0.993 (Table 3; Figs. 2 and 3). Comparison of the full data sets and data sets without singletons indicated that estimates of $\ln(A_x)$ and b did not differ significantly ($P = 0.05$). Therefore, herein results are presented only for the full data sets.

At the row level, the intercept parameter $\ln(A_x)$ was >0 for years 1999 to 2005 ($P < 0.0001$) and 2007 ($P = 0.005$), but was not significantly greater than 0 in 2006 ($P = 0.125$). The slope parameter b was significantly greater than 1 for all row-level data sets collected from 1999 to 2005 ($P = 0.05$), indicating that heterogeneity changed systematically with disease incidence. However, in 2006, b was <1 ($P = 0.005$) indicating a more regular pattern (underdispersion) of disease. In 2007, b was not significantly different from 1 ($P = 0.540$), indicating heterogeneity was not systematically related to disease incidence in this year. The factor 'region' significantly affected $\ln(A_x)$ (i.e., height of the line) in 2002, 2003, and 2005, and b in 2002 and 2003 (Table 4). 'Cultivar' had a significant effect on $\ln(A_x)$ in 2002, 2003, and 2007, and b in 2002.

At the yard level, $\ln(A_x)$ was significantly greater than 0 for 2000 to 2005 ($P < 0.0001$). In 2006 and 2007, $\ln(A_x)$ was not significantly different from 0 ($P = 0.112$ and 0.055, respectively), indicating a random pattern of disease incidence (Table 3). The slope parameter b was significantly greater than 1 for 2000 to 2005, but was not significantly different from 1 in 2006 ($P = 0.598$) and 2007 ($P = 0.869$).

Covariance analysis of the binary power law models at the yard scale indicated that the factor cultivar significantly affected $\ln(A_x)$ in 2002, 2003, and 2004. 'Cultivar' also significantly affected b in 2004. The factor 'region' affected $\ln(A_x)$ in 2002, 2003, and 2005 (Table 5; Fig. 4). A covariance analysis of the effect 'year' indicated that this factor had a strong effect on $\ln(A_x)$ in both the row- and yard-level analysis (Table 6). 'Year' did not significantly affect estimates of b at either scale in the covariance analysis.

Sample size analyses. Typical fixed sampling curves were generated from equation 4 depicting that greater numbers of sampling units are needed to determine p (at a constant C) as p decreases (Fig. 5). The height of the sampling curves varied considerably depending on the year (2004 or 2006) or years (pooled data) from which a and b were selected. For $C = 0.2$, the differences between the sampling curves with parameters from 2004 and 2006 were

TABLE 5. Covariance analysis of the effect of cultivar and geographic region on intercept ($\ln[\hat{A}_x]$) and slope ($\hat{b}$) parameters of the binary power law for the incidence of hop powdery mildew on leaves sampled at the yard level

			$\ln(\hat{A}_x)$				$\hat{b}$			
Factor and year	df model[a]	df factor[a]	SSE[b]	Diff.[c]	F[d]	P	SSE	Diff.	F	P
2000										
Power law	117		2.716	–	–	–	2.716	–	–	–
Region	114	3	2.567	0.149	2.203	0.092	2.633	0.083	1.195	0.315
Cultivar	114	3	2.658	0.058	0.835	0.477	2.673	0.043	0.610	0.610
2001										
Power law	77		5.012	–	–	–	5.012	–	–	–
Region	74	3	4.787	0.225	1.159	0.331	4.812	0.200	1.024	0.387
Cultivar	74	3	4.796	0.216	1.108	0.351	4.770	0.242	1.249	0.298
2002										
Power law	210		19.824	–	–	–	19.824	–	–	–
Region	207	3	18.726	1.099	4.048	0.008	19.455	0.370	1.311	0.272
Cultivar	208	2	19.219	0.605	3.274	0.040	19.542	0.283	1.505	0.225
2003										
Power law	84		3.889	–	–	–	3.889	–	–	–
Region	81	3	3.480	0.409	3.173	0.029	3.472	0.417	3.240	0.026
Cultivar	82	2	3.331	0.557	6.859	0.002	3.688	0.200	2.225	0.115
2004										
Power law	62		8.252	–	–	–	8.252	–	–	–
Region	59	3	7.500	0.752	1.971	0.128	3.472	4.780	27.076	0.000
Cultivar	60	2	4.887	3.365	20.652	0.000	3.688	4.564	37.120	0.000
2005										
Power law	126		12.172	–	–	–	12.172	–	–	–
Region	123	3	10.936	1.236	4.635	0.004	11.893	0.280	0.965	0.412
Cultivar	124	2	11.844	0.329	1.722	0.183	12.000	0.173	0.893	0.412
2006										
Power law	43		6.101	–	–	–	6.101	–	–	–
Region	40	3	5.291	0.811	2.043	0.123	5.711	0.390	0.910	0.445
Cultivar	41	2	5.936	0.166	0.572	0.569	5.982	0.119	0.409	0.667
2007										
Power law	35		4.627	–	–	–	4.627	–	–	–
Region	32	3	4.085	0.541	1.413	0.257	4.219	0.408	1.031	0.392
Cultivar	34	1	4.267	0.360	2.866	0.100	4.597	0.029	0.217	0.644

[a] df dev. = degrees of freedom for the model; df factor = degrees of freedom for factor.

[b] SSE = sum of square error for the covariance model.

[c] Diff. = difference between the SSE of the binary power law model versus the binary power law model with each factor included in the analyses first as an intercept and then as a slope.

[d] Significance level for the difference between SSE of the binary power law model versus binary power law model with each factor as determined by an F test, where F = (factor SSE/df factor)/(model SSE/df model).

nearly as great as the difference between the curve from 2004 and assuming a binomial distribution of disease incidence (Fig. 5A). Assuming $p = 0.05$, 0.1, or 0.2, the sampling curve with parameter estimates from 2004 indicated that 99, 53, or 26 sampling units, respectively, would be needed to estimate p with precision $C = 0.2$. With parameter estimates from 2006, 55, 26, or 12 sampling units were required to estimate p at the specified level of precision. Assuming parameter estimates from 2004 represented the true heterogeneity of disease incidence, sampling 55, 26, and 12 sampling units when $p = 0.05$, 0.1, and 0.2 would result in an achieved C of 0.25, 0.27, and 0.28, respectively. Therefore, the specified precision of $C = 0.2$ would not be achieved (Fig. 5C).

Similarly for sequential estimation, the stop lines varied depending on the year from which binary power law parameters were selected (Fig. 5B and D). As expected, the stop lines generated with data from 2004 and 2000 to 2007 were higher than the stop line with a binomial distribution of disease incidence. The stop line with parameters from 2006 was slightly higher than the binomial stop line, and this difference was most pronounced at large N because the binary power law regression line for 2006 did not the cross the binomial line (Figs. 2G and 3).

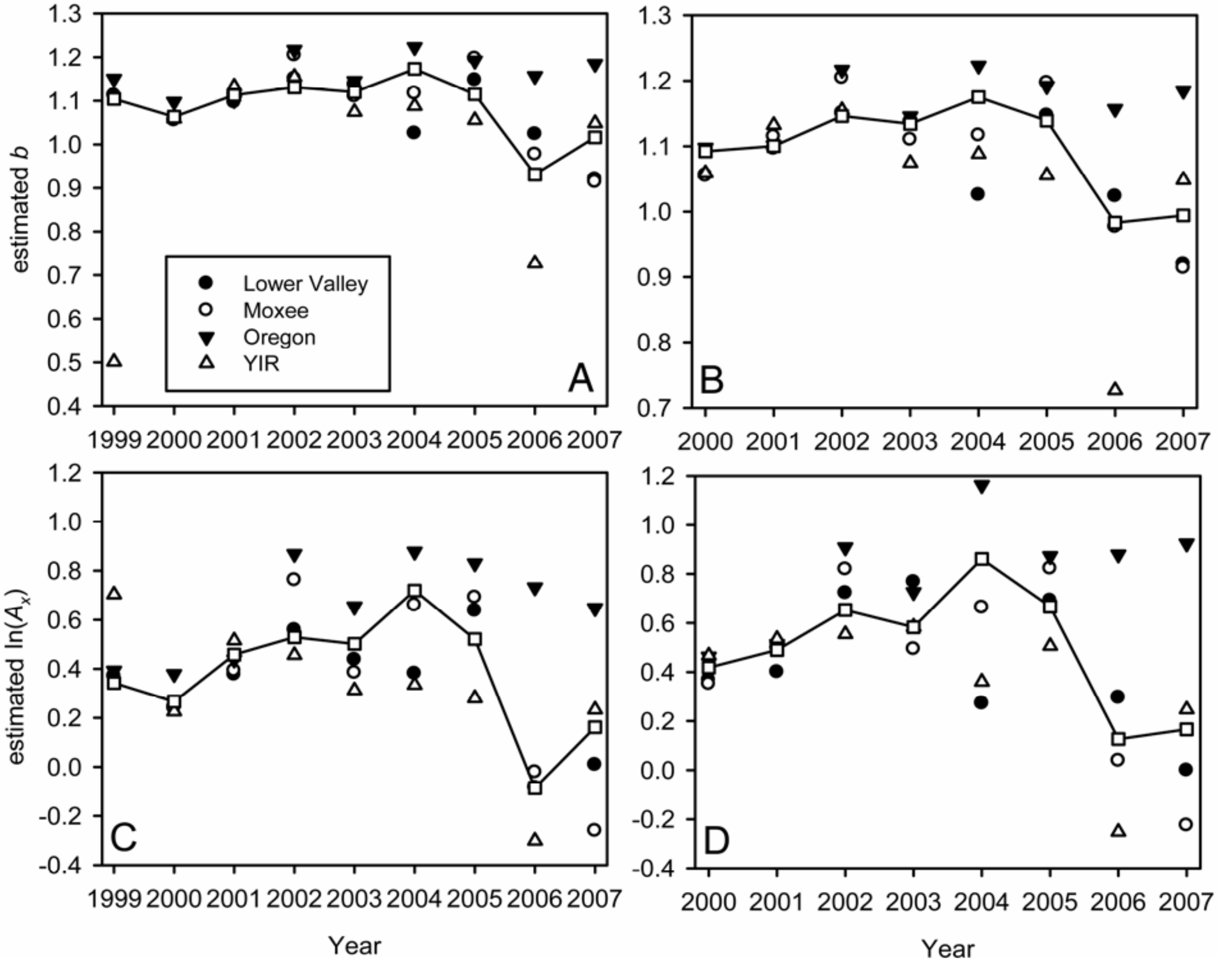

Fig. 4. Variability of mean values of the slope $(\hat{b})$ and intercept $(\ln[\hat{A}_x])$ parameters of the binary power law for the incidence of hop powdery mildew on leaves at the scale of individual rows **(A and C)** or hop yards **(B and D)** over time and four production regions. The open boxes are mean values of parameter estimates for each year. Only five rows were sampled from the Yakima Indian Reservation (YIR) in 1999, and one observation was identified as highly influential in the regression (Cook's D = 2.715). When this observation is removed from the analysis $\hat{b} = 2.038$.

TABLE 6. Covariance analysis of the effect of year on intercept $(\ln[\hat{A}_x])$ and slope $(\hat{b})$ parameters of the binary power law for the incidence of hop powdery mildew on leaves

			$\ln(\hat{A}_x)$				$\hat{b}$			
Factor and scale	df SSE[a]	df factor[a]	SSE[b]	Diff.[c]	F[d]	P	SSE	Diff.	F	P
Row level										
Power law	1,604		157.950	–	–	–	157.950	–	–	–
Year	1,596	8	144.267	13.683	18.921	0.000	156.959	0.991	1.259	0.261
Yard level										
Power law	768		75.440	–	–	–	75.440	–	–	–
Year	761	7	68.290	7.150	11.382	0.000	74.631	0.809	1.178	0.313

[a] df SSE = degrees of freedom for the model sum of squares error; df factor = degrees of freedom for factor.

[b] SSE = sum of squares error.

[c] Diff. = difference between the SSE of the binary power law model versus the binary power law model with year included in the analysis.

[d] Significance level for the difference between SSE of the binary power law model versus binary power law model with year as determined by an F test, where F = (factor SSE/df factor)/(model SSE/df model).

DISCUSSION

The binary power law provides a convenient means to describe and quantify spatial patterns of disease incidence (18) and summarize the spatial dynamics of disease over the course of an epidemic (8). This study provides quantification of the long term spatial and temporal stability of the binary power law parameters in a managed agroecosystem. Similar to previous reports (4,33) which utilized some of the same data, the incidence of powdery mildew on leaves was largely random at the row and yard scale as evidenced by relatively small values of $\hat{\theta}$, D, and parameter estimates of the binary power law.

Taylor and Woiwod (27) provided empirical evidence that variance and population density are systematically linked and spatially stable. Results of the covariance analysis for the 9 years of data in the current study indicated that b also is relatively constant with disease incidence data, similar to other studies with plant diseases (16,30,31). However, in many studies of spatial patterns of plant disease data sets often are collected over a period of 2 to 3 years for practical reasons, such as resource availability. In the case of hop powdery mildew on leaves, if one were to rely on the information gathered during 2006 and 2007, 2 years in which the parameters showed substantial variation compared with other years considered in this study, the binary power law parameter estimates and, perhaps, conclusions on processes associated with the observed patterns could be substantially different than data collected from, say, 2004 and 2005, or 1999 and 2000.

Comparison of b (estimated at the yard level) in 2004 to years 2006 and 2007 by t tests indicates that b was significantly greater in 2004 (t statistic = 4.52 and 3.97, respectively). Differences in b among years were large enough to have practical implications for sample sizes and precision of fixed and sequential sampling, with nearly twice as many sampling units being required to achieve the specified C depending on whether parameter estimates from 2004 or 2006 were used to derive the fixed sampling curves. This difference would result in the greatest differences in sampling costs when p is small (less than 0.1), which typically is where sampling would be needed most to determine the need for a control measure with a polycyclic disease (18). Differences in sample size and costs also would be most pronounced with sequential sampling when disease incidence was low. These results reinforce the need for proper validation of a sampling plan over a range of disease incidence, cultivar, and geographic regions to ensure the sampling plan performs as designed (1).

Similar to other studies with count data (26,29), the intercept parameter of the binary power law did vary significantly among years. Covariance analysis indicated that $\ln(A_x)$ varied among regions and cultivars in 3 of 9 years and 3 of 8 years in the row-

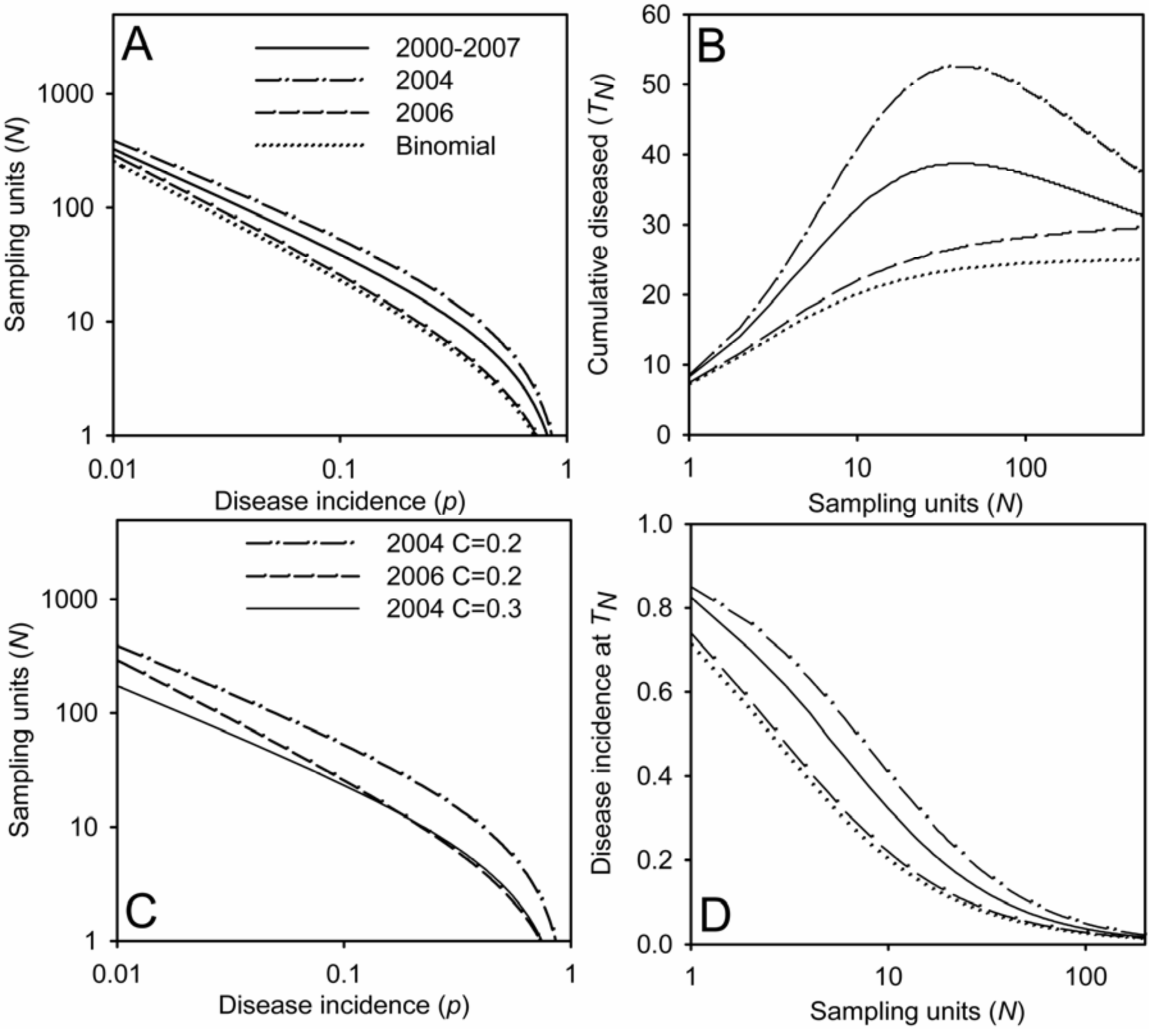

Fig. 5. A, Sample size required to estimate the incidence ($\hat{p}$) of hop leaves with powdery mildew based on parameter estimates of the binary power law derived from yard-level analyses from pooled data from 2000 to 2007 (solid line), 2004 (dot dash), and 2006 (short dash) with coefficient of variation of the mean $C = 0.2$. The sampling curve for a binomial distribution is presented for reference as a dotted line. **B,** Sequential estimation stop lines with binary power law parameter from pooled data from 2000 to 2007, 2004, or 2006 with $C = 0.2$. **C,** Fixed sampling curves with binary power law parameters from 2004 with $C = 0.2$ and 0.3 and 2006 with $C = 0.2$. **D,** Mean disease incidence at critical T_N (the point where the cumulative number of diseased leaves crosses the model T_N curve in **B**) in relation to number of sampling units (N).

and yard-level analyses. The factors that affected parameter estimates did vary somewhat between the row- and yard-level scales. Row-level parameter estimates also tended to have small standard errors, which were expected because of the greater number of data sets available for the analysis. However, inclusion of singleton data sets did not affect parameter estimates at the row or yard level. Such data sets may affect parameter estimates when few data sets are used, but were not important with the sample sizes used in this study.

Results of this study lead to several recommendations for the use of the power law. First, although it is desirable to have multiple years of data to estimate the parameters of the power law, it is recognized that only 2 or 3 years of data may be available. Therefore, the parameters should be estimated from the combined data set. Second, a statistically significant year factor should not be looked at as unfavorable but rather as part of the variability that is expected in a complex system. Third, it would be prudent to reassess sampling plans in agricultural systems periodically over time given that cultivars, disease management programs, and other factors may change. Agricultural systems are often the subject of abrupt changes, such as introductions of new cultivars or fungicide programs, which could impact the spatial dynamic of the pathogen or disease under study.

An interesting question that arises in association with the collection of new data is, should the new spatial data be added to or analyzed separately from older data? As long as the data are collected in the same manner, there would be no statistical reason not to join data sets, thus, the first inclination would be to ultimately join the data sets. However, it is important to carefully scrutinize how a particular subset may differ from the full data set and to note the effect of outliers and high influence observations on parameter estimates derived from the smaller subsets compared to the full data set. This is particularly important since differences in parameter estimates may significantly alter the number of sampling units needed to achieve a desired level of precision (3–5,15). A reasonable approach may be that if a given factor is manageable (e.g., cultivar), then its effect should be included in the overall model if it is found that the effect is not due to highly influential or outlying data points. If the effect is unmanageable (e.g., year), the factor should not be included in the full model and parameters estimates will simply average the effect and lead to larger standard errors of the parameters.

This research demonstrated that although the binary power law parameter tends to be relatively stable, spatial and temporal variability of parameters may have practical consequences for sampling. Collection of data sets collected over multiple geographic locations, years, and a range of disease incidence may be needed to observe the range of a and b in managed agricultural systems. Such data sets also would be robust to rare observations and improve the power of covariance analysis.

ACKNOWLEDGMENTS

Financial support was provided by USDA-ARS CRIS 303-5358-22000-030-00D and 5358-21000-035-00. We thank S. J. Pethybridge and D. Shah for their critical review of an earlier draft of the manuscript and the many individuals that assisted in disease assessments and the hop growers that allowed us access to their farms for sampling. The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

LITERATURE CITED

1. Binns, M. R., Nyrop, J. P., and van der Werf, W. 2000. Sampling and Monitoring in Crop Protection: The Theoretical Basis for Designing Practical Decision Guides. CABI Publishing, Oxon, UK.
2. Dallot, S., Gottwald, T., Labonne, G., and Quiot, J. B. 2003. Spatial pattern analysis of sharka disease (*Plum pox virus* strain M) in peach orchards of southern France. Phytopathology 93:1543-1552.
3. Gent, D. H., Mahaffee, W. F., and Turechek, W. W. 2006. Spatial heterogeneity of the incidence of powdery mildew on hop cones. Plant Dis. 90:1433-1440.
4. Gent, D. H., Turechek, W. W., and Mahaffee, W. F. 2007. Sequential sampling for estimation and classification of the incidence of hop powdery mildew I: Leaf sampling. Plant Dis. 91:1002-1012.
5. Gent, D. H., Turechek, W. W., and Mahaffee, W. F. 2007. Sequential sampling for estimation and classification of the incidence of hop powdery mildew II: Cone sampling. Plant Dis. 91:1013-1020.
6. Hughes, G. 1988. Spatial heterogeneity in crop loss assessment models. Phytopathology 78:883-884.
7. Hughes, G. 1999. Sampling for decision making in crop loss assessment and pest management: Introduction. Phytopathology 89:1080-1083.
8. Hughes, G., and Madden, L. V. 1992. Aggregation and incidence of disease. Plant Pathol. 41:657-660.
9. Hughes, G., and Madden, L. V. 1994. Aggregation and incidence of disease: Some implications for sampling. Aspects Appl. Biol. 37:25-31.
10. Hughes, G., Madden, L. V., and Munkvold, G. P. 1996. Cluster sampling for disease incidence data. Phytopathology 86:132-137.
11. Hughes, G., McRoberts, N., Madden, L. V., and Gottwald, T. R. 1997. Relationships between disease incidence at two levels in a spatial hierarchy. Phytopathology 87:542-550.
12. 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.
13. Madden, L. V., and Hughes, G. 1995. Plant disease incidence: Distribution, heterogeneity, and temporal analysis. Annu. Rev. Phytopathol. 33:529-564.
14. Madden, L. V., and Hughes, G. 1999. An effective sample size for predicting plant disease incidence in a spatial hierarchy. Phytopathology 89:770-781.
15. Madden, L. V., and Hughes, G. 1999. Sampling for plant disease incidence. Phytopathology 89:1088-1113.
16. Madden, L. V., Hughes, G., and Ellis, M. A. 1995. Spatial heterogeneity of the incidence of grape downy mildew. Phytopathology 85:269-275.
17. Madden, L. V., Hughes, G., and Munkvold, G. P. 1996. Plant disease incidence: Inverse sampling, sequential sampling, and confidence intervals when observed mean incidence is zero. Crop Prot. 15:621-632.
18. Madden, L. V., Hughes, G., and van den Bosch, F. 2007. The Study of Plant Disease Epidemics. American Phytopathological Society, St. Paul, MN.
19. Madden, L. V., Turechek, W. W., and Nita, M. 2002. Evaluation of generalized linear mixed models for analyzing disease incidence data obtained in designed experiments. Plant Dis. 86:316-325.
20. Mahaffee, W. F., Thomas, C. S., Turechek, W. W., Ocamb, C. M., Nelson, M. E., Fox, A., and Gubler, W. D. 2003. Responding to an introduced pathogen: *Podosphaera macularis* (hop powdery mildew) in the Pacific Northwest. Published Online. Plant Health Progress doi:10.1094/PHP-2003-1113-07-RV.
21. Patil, G. P., and Stiteler, W. M. 1974. Concepts of aggregation and their quantification: A critical review with some new results and applications. Res. Popul. Ecol. 15:238-254.
22. Paul, S. R., Liang, K. Y., and Self, S. G. 1989. On testing departure from binomial and multinomial assumptions. Biometrics 45:231-236.
23. Taylor, L. R. 1961. Aggregation, variance, and the mean. Nature 189:732-735.
24. Taylor, L. R. 1984. Assessing and interpreting the spatial distributions of insect populations. Annu. Rev. Entomol. 29:321-357.
25. Taylor, L. R., and Taylor, R. A. J. 1977. Aggregation, migration and population mechanics. Nature 265:415-421.
26. Taylor, L. R., Taylor, R. A. J., Woiwod, I. P., and Perry, J. N. 1983. Behavioural dynamics. Nature 303:801-804.
27. Taylor, L. R., and Woiwod, I. P. 1980. Temporal stability as a density-dependent species characteristic. J. Anim. Ecol. 49:209-224.
28. 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.
29. Taylor, L. R., Woiwod, I. P., and Taylor, R. A. J. 1979. The migratory ambit of the hop aphid and its significance in aphid population dynamics. J. Anim. Ecol. 48:955-972.
30. Turechek, W. W., Ellis, M. A., and Madden, L. V. 2001. Sequential sampling for incidence of Phomopsis leaf blight of strawberry. Phytopathology 91:336-347.
31. Turechek, W. W., and Madden, L. V. 1999. Spatial pattern analysis and sequential sampling for the incidence of leaf spot on strawberry in Ohio. Plant Dis. 83:992-1000.

32. Turechek, W. W., and Madden, L. V. 1999. Spatial pattern analysis of strawberry leaf blight in perennial production systems. Phytopathology 89:421-433.
33. Turechek, W. W., and Mahaffee, W. F. 2004. Spatial pattern analysis of hop powdery mildew in the Pacific Northwest: Implications for sampling. Phytopathology 94:1116-1128.
34. Turechek, W. W., Mahaffee, W. F., and Ocamb, C. M. 2001. Development of management strategies for hop powdery mildew in the Pacific Northwest. Published Online. Plant Health Progress doi:10.1094/PHP-2001-0313-01-RS.
35. Xu, X. M., and Madden, L. V. 2002. Incidence and density relationships of powdery mildew on apple. Phytopathology 92:1005-1014.