Letter to the Editor

Comments Regarding the Binary Power Law for Heterogeneity of Disease Incidence

W. W. Turechek, L. V. Madden, D. H. Gent, and X.-M. Xu

First author: United States Department of Agriculture–Agricultural Research Service (USDA-ARS), U.S. Horticultural Research Laboratory, 2001 South Rock Road, Fort Pierce, FL 34945; second author: Department of Plant Pathology, The Ohio State University, Wooster 44691; third author: USDA-ARS, Forage Seed and Cereal Research Unit, Oregon State University, Department of Botany and Plant Pathology, Corvallis 97331; and fourth author: Plant Pathology, East Malling Research, New Road, East Malling, ME19 6BJ, UK.

Accepted for publication 27 July 2011.

ABSTRACT

Turechek, W. W., Madden, L. V., Gent, D. H., and Xu, X.-M. 2011. Comments regarding the binary power law for heterogeneity of disease incidence. Phytopathology 101:1396-1407.

The binary power law (BPL) has been successfully used to characterize heterogeneity (overdispersion or small-scale aggregation) of disease incidence for many plant pathosystems. With the BPL, the log of the observed variance is a linear function of the log of the theoretical variance for a binomial distribution over the range of incidence values, and the estimated scale (κ) and slope (b) parameters provide information on the characteristics of aggregation. When $b = 1$, the interpretation is that the degree of aggregation remains constant over the range of incidence values observed; otherwise, aggregation is variable. In two articles published in this journal in 2009, Gosme and Lucas used their stochastic simulation model, Cascade, to show a multiphasic (split-line) relationship of the variances, with straight-line (linear) relationships on a log-log scale within each phase. In particular, they showed a strong break point in the lines at very low incidence, with b considerably >1 in the first line segment (corresponding to a range of incidence values usually not observed in the field), and b being ≈1 in the next segment (corresponding to the range of incidence values usually observed). We evaluated their findings by utilizing a general spatially explicit stochastic simulator developed by Xu and Ridout in 1998, with a wide range of median dispersal distances for the contact distribution and number of plants in the sampling units (quadrats), and through an assessment of published BPL results. The simulation results showed that the split-line phenomenon can occur, with a break point at incidence values of ≈0.01; however, the split is most obvious for short median dispersal distances and large quadrat sizes. However, values of b in the second phase were almost always >1, and only approached 1 with extremely short median dispersal distances and small quadrat sizes. An appraisal of published results showed no evidence of multiple phases (although the minimum incidence may generally be too high to observe the break), and estimates of b were almost always >1. Thus, it appears that the results from the Cascade simulation model represent a special epidemiological case, corresponding primarily to a roughly nearest-neighbor population-dynamic process. Implications of a multiphasic BPL property may be important and are discussed.

The binary power law (BPL) defines, through a simple power function, the relationship between two variances, the observed variance of a proportion (such as disease incidence) and the theoretical variance of this random variable if it were distributed according to the binomial distribution (18,24). Generally, the relationship is written as

$$s_{obs}^2 = \kappa \cdot (s_{bin}^2)^b \tag{1}$$

where s_{obs}^2 is the observed variance, s_{bin}^2 is the theoretical binomial variance, and κ and b are model parameters. A logarithmic transformation of equation 1 produces a linear relationship:

$$\ln(s_{obs}^2) = \ln(\kappa) + b \cdot \ln(s_{bin}^2) \tag{2}$$

with a slope of b and an intercept of ln(κ), which allows the power law's parameters to be estimated with simple linear regression (28). This is directly analogous to Taylor's power law for unbounded counts (39), which relates the observed variance (v_{obs}) to the population mean (m), because the variance is equal to the mean for the Poisson distribution (39). A plot of $\ln(s_{obs}^2)$ versus $\ln(s_{bin}^2)$ for a set of observations—either from different times in the same epidemic or from single times from a collection of epidemics—nearly always results in a straight line. The fit of equation 2 to observed data (18,24) and to data from stochastic simulators (47,48) is often very good, with R^2 values typically >0.9 and with estimated slope values typically >1 and <2.

The value of b is a fundamental parameter in characterizing the spatial pattern of incidence at a small scale (at the scale of the sampling unit or smaller), which is manifested by overdispersion or extra-binomial variation in disease incidence among the sampling units (24). When $b = 1$, the magnitude of overdispersion does not change with incidence, and the value of κ then represents the constant (on average) index of dispersion (i.e., $D = s_{obs}^2 / s_{bin}^2 = \kappa$). When b does not equal 1, the index of dispersion changes systematically with disease incidence and $D = s_{obs}^2 / s_{bin}^2 = \kappa \cdot (s_{bin}^2)^{(b-1)}$ (24). The relationship between the observed and theoretical variance, in general, and the value of b, in particular, are of critical importance in several applications, including (i) the development of sequential and nonsequential sampling protocols both for estimation purposes and for decision making (10,11, 26,40), (ii) the proper weighting of incidence values in generalized linear mixed models for analysis of treatment effects

Corresponding author: W. W. Turechek
E-mail address: William.Turechek@ars.usda.gov

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.

doi:10.1094/PHYTO-04-11-0100

(29), (iii) the characterization of the relationship between incidence at different scales in a spatial hierarchy (e.g., leaflet, leaf, plant) (21,25,43), (iv) the selection of a variance stabilizing transformation for analysis of variance (ANOVA)-type data analysis (18), and (v) the description of spatial dynamics of disease incidence during epidemics (28) (Table 1).

No particular single mechanism has been proposed for the BPL, and it is likely that several mechanisms based on environmental and epidemiological stochasticity are responsible for the realized relationship between the observed and theoretical variances that is commonly found. Stochastic simulation models have been useful for addressing epidemiological effects, such as infection rates and dispersal gradients, on the small-scale patterns of disease incidence (13,14,45,47). Xu and Ridout (48) showed that the BPL, as represented by a version of equation 2 (with $b > 1$), arises naturally over a wide range of epidemiological conditions. More recently, Gosme and Lucas (13) developed a stochastic simulation model, called "Cascade", and found relationships between the observed and theoretical variances that have not been observed yet with real-world data sets, leading to speculation about the processes that generated these new results. This letter addresses the issues raised about the BPL by Gosme and Lucas (13).

Cascade is a model designed to simulate disease spread in a spatial hierarchy (13,14). The overall structure of the model is complex but a description of it can be found in the original set of publications found in the July 2009 issue of *Phytopathology*. In the first publication reporting on Cascade by Gosme and Lucas (13), a set of simulations was performed, and a plot of the logarithms of the observed (i.e., simulated) versus the theoretical variances did not result in the characteristic simple linear relationship across all disease incidence values; rather, a relation that consisted of two or more phases or segments was found, with a straight line within each segment (on the log scale) connected at break points (or switch points). The authors differentiated four distinct phases in the BPL relationship; this can be seen in their Figure 4 (13). The four phases are described by the authors as follows: (i) "[T]he beginning of the epidemic, the simulated variance increased more than the theoretical variance resulting in a slope greater than 1"; (ii) "[T]he second phase, the slope was equal to one, indicating a constant level of aggregation"; [note: in later communications with Dr. Gosme, she indicated that "*b* was not (mathematically) equal to one, but it was close to one" (*personal communication*)]; (iii) "The third phase corresponded to a phase where the logarithm of the simulated variance decreased at almost the same rate as the logarithm of the binomial variance, resulting in a slope only slightly greater than one"; and (iv) "The fourth segment corresponded to a phase when the simulated variance decreased more quickly than the theoretical variance." In general, the largest phase was the second one, or the second and third phases combined, comprising a large range in disease incidence values. The fourth phase may not be seen unless incidence is very large (e.g., >0.98), or may look the same as the first phase. In many cases, therefore, the four phases are reduced to two (segments one and two/three) (14). The authors suggest that this discovery of a bi- or multiphasic relationship could not have been made without simulation because experimental data often have too few observations, particularly at the beginning and end of the epidemic where disease incidence is at its lowest and highest, respectively, and that real-world data have too much uncertainty in their estimated incidence values to detect this full theoretical relationship. In their following article (14), the authors expanded upon their results and ran a number of additional simulations to support their original findings. Some of their additional analyses were based on the assumption that *b* was reasonably close to 1 in phase two/three, so that $b = 1$ could be used as an approximation (e.g., Table 2 in literature citation 14).

The results are genuinely interesting and may have important theoretical and practical implications. Given the importance of the BPL for the reasons summarized above, it is essential to explore the likelihood that the results generated by the Cascade model can be found with observed data and to explore the epidemiological conditions that could produce their results. In particular, attention should be focused on (i) the multiphasic (break-point) relationship between the observed and theoretical variances on a log scale and (ii) the slope in their second and third phases being equal to or close to one. In this letter, we summarize the estimates of *b* (and a transformation of κ) found for a wide range of studies published in the literature and perform additional simulations with the model by Xu and Ridout (47) to address these issues. Our results below show that, although a multiphasic relationship can occur, as predicted by Gosme and Lucas (13), the slope of the line segment for the range of disease incidence values commonly observed in empirical data will be >1 under most circumstances.

MATERIALS AND METHODS

Basic notation and model formulations. The expected probability of a plant or plant unit (e.g., leaf) being diseased is given by *p*. There are *N* sampling units and *n* plants or plant units in each sampling unit; *n* can be considered the size of the sampling unit. The expected number (count) of diseased plants per sampling unit is *np*. An estimate of *p* is the proportion of plants or plant units that are diseased ($\overline{X}$, where each unit is recorded as $X = 0$ or $X = 1$). For convenience here, we do not distinguish between the true and estimated *p* in the equations.

There are (at least) four common (and interchangeable) expressions of the BPL (equation 1), depending on whether the variance of the counts or variance of the proportions across the *N* sampling units is used in the formulation, or whether *n* is written as part of the theoretical variance or absorbed into the κ scale parameter. The exponent *b* is unaffected by all of these formulations but the scale parameter is directly affected; thus, a different symbol is used in place of κ for each model. Details are given in the Appendix.

Simulation study. Simulations were performed using a two-dimensional stochastic spatial contact model (47), where the distance that spores are dispersed follows a half-Cauchy distribution with median dispersal distance μ (37). Full details of the simulation model can be found in Xu and Ridout (47). Epidemics were started with one or nine initially infected leaves. For epidemics initiated with a single infection, the initial infection was located at the center of a 200-by-200 simulation grid of plants. Each plant had a total of 30 susceptible units (leaves) and, thus, this model simulated epidemic development at two hierarchical levels (within and between plants). For epidemics initiated with nine infected leaves, the leaves were regularly spaced in the grid to obtain the maximum separation distance. Simulation conditions were chosen here to generate results for very small disease incidence values in order to explore the results found by Gosme and Lucas (13,14). Five median dispersal distances values (μ) were used: 0.25, 0.5, 1, 2, and 4. Twenty simulations were run for each dispersal distance on a daily time step and an infection rate of 0.4 per day was assumed. Only a single epidemic failed to establish in one run when $\mu = 0.25$. Four different quadrat sizes were used to sample the central 160-by-160 grid: 2 by 2, 4 by 4, 5 by 5, and 8 by 8 (n = 4, 16, 25, and 64, respectively). All the plants in the central 160-by-160 grid were assessed daily for disease and a plant was considered diseased if ≥1 of its 30 susceptible units/leaves were diseased. Depending on the dispersal distance, the model was run for a certain time period of 102 to 172 days to ensure that final disease incidence was >0.95 for all the simulations, for most of which it was >0.98.

For each simulation run, three different models were fitted to the quadrat (sampling unit) data across all times. Each model was fitted separately to the data for each of the 20 simulation runs for each dispersal parameter value. The first model was the log transformation of the usual (symmetric) BPL defined in equation A2 of the Appendix:

$$\ln(s^2_{obs}) = \ln(A_p) + b \cdot \ln[p(1-p)/n] \tag{3}$$

where the parameters $\ln(A_p)$ and b are the intercept and slope, respectively, of a straight line on a log-log scale. The second model was the log transformation of the asymmetric binary power law (ABPL) defined in equation A5 and explained in the Appendix:

$$\ln(s^2_{obs}) = \ln(a_1) + b_1 \cdot \ln(p) + b_2 \cdot \ln(1-p) \tag{4}$$

where $\ln(a_1)$, b_1, and b_2 are the parameters. The third model, labeled a split-line (SL) or broken-stick model, is written as

$$\ln(s^2_{obs}) = (\ln(A_{p3}) + b_3 t) \cdot I[t \le t_*] + (\ln(A_{p4}) + b_4 t) \cdot I[t > t_*] \tag{5}$$

where t is the log of the binomial variance (i.e., $t = \ln(p(1-p)/n)$), t_* is a constant ($t_* = \ln(p_*(1-p_*)/n)$, and $I[\cdot]$ is an indicator function taking on the value of 1 when the condition is true and 0 otherwise. Equation 5 defines two straight-line segments on a log-log scale that meet at a join-point binomial variance (t_*) determined by the data. The parameters b_3 and b_4 are the slopes of the two segments and $\ln(A_{p3})$ and $\ln(A_{p4})$ are the intercepts. The parameter p_* is the join-point (break-point or switch-point) incidence, calculated from a back-transformation of t_*, between the two different line segments (35). In order for the two lines to meet at an incidence of p_*, the intercept for the first line segment [$\ln(A_{p3})$] is constrained to be an exact function of the other parameters; thus, there are four parameters in equation 5 (b_3, b_4, $\ln(A_{p4})$, and p_*).

Equations 3 and 4 were fitted using ordinary linear least squares regression, and equation 5 was fitted with nonlinear least squares regression. All equations were fit using Genstat (VSNI Ltd., England). It was assumed that there was an additive residual term in each model (not shown for convenience) with a mean of 0 and a variance of σ^2.

Evaluation of published data sets. A search of the literature was conducted to identify a representative set of studies that utilized the BPL as a descriptor of overdispersion (small-scale pattern) in epidemics. Because authors do not necessarily cite the original article by Hughes and Madden (18) or one of the standard reviews or syntheses (27,28), it was not possible to retrieve all articles that used this model. Nevertheless, the articles obtained reflect a wide diversity of pathosystems and experimental or survey methods. The list of the studies identified and various characteristics associated with each of the studies is shown in Table 1. The BPL results from the studies shown in Table 1 were assembled and the scale parameters were transformed to the common value a (Appendix). When multiple power law results were provided, only the summary or overall fit to the data are given in Table 1 if those results were provided. Otherwise, all results are provided.

For the analysis here, we excluded studies that involved artificial inoculation (e.g., citation 22); in other words, only naturally occurring epidemics were considered. Furthermore, we only

TABLE 1. Plant pathology-based studies identified from a search of the literature that utilized the binary power law (equation 1) and various characteristics associated with each of the studies (the power-law scale parameter was standardized to a common a for each study [equation A3])

Ref.[a]	Disease	Crop	Organism	Mode of dispersal	T[b]	n	a	b
7[c]	Apple scab	Apple	Fungus	Aerial	100	12	0.2542	1.12
5	Fire blight	Apple	Bacterium	Insect	211	4	0.3990	1.10
46	Powdery mildew	Apple	Fungus	Aerial	62	4	0.2751	0.95
36	*Alfalfa mosaic virus* (AlMV)	Bean	Virus	Insect	24	5	0.3183	1.05
36	*Cucumber mosaic virus* (CMV)	Bean	Virus	Insect	24	5	0.3302	1.07
36	AlMV and CMV	Bean	Virus	Insect	24	5	0.3229	1.05
38	Black spot	Citrus	Fungus	Rain	303	5	0.6010	1.06
16[d]	Citrus canker	Citrus	Bacterium	Rain	321	9	0.4804	1.21
2[e]	Citrus sudden death	Citrus	Virus	Graft	98	4	0.3765	1.07
2[e]	Citrus sudden death	Citrus	Virus	Graft	98	16	0.1758	1.20

(*continued on next page*)

[a] References for the study are provided in the Literature Cited section. Only studies that identified the sampling unit (quadrat) size (n) are included. When there were several intercepts and slopes reported for the same pathosystem but with equal quadrat sizes (e.g., from different locations or years), the average is shown here. Different versions of the estimated scale parameter ("intercept" for linear log-log model) were presented in different publications. We converted all of these to the a parameter shown here (equation A3) based on the value of n and the functions given in the Appendix for a, A_p, and A_x.

[b] T = number of data sets used to fit the power law model.

[c] Parameter a was calculated from A_p, not A_x as reported in text, based on the binary power law model fit provided in their Figure 3.

[d] Parameter a was calculated from A_p, not A_x as reported in text, based on the binary power law model fit provided in their Figure 1.

[e] Data are reported for quadrat sizes of 4 (2 by 2), 16 (4 by 4), and 64 (8 by 8) trees. The causal pathogen has not been established definitely but is known to be graft transmissible. An insect vector is considered likely. Parameter a was calculated from A_p based on the binary power law model fit provided in their Figure 4.

[f] Data are reported for quadrat sizes of 4 (2 by 2) and 16 (4 by 4) trees. The causal pathogen has not been established definitively but is known to be graft transmissible. An insect vector is considered likely. Parameter a was calculated from A_p based on the binary power law model fit provided in their Figure 2.

[g] Data are reported for leprosis-diseased plants from quadrat sizes of 4 (2 by 2), 9 (3 by 3), 16 (4 by 4), and 25 (5 by 5) trees.

[h] Actually mite transmissible.

[i] Data are reported for quadrat sizes of 4 (2 by 2) and 16 (4 by 4) trees. Parameter a was calculated from A_p based on the binary power law model fit provided in their Figure 2.

[j] Data are from analyses conducted at the row level (T = 1,606) and yard level (T = 770) as reported by Gent et al. (12).

[k] Data are from analyses conducted at the row level (T = 578) and yard level (T = 198) as reported by Turechek and Mahaffee (44).

[l] Estimate of n was obtained by averaging n from the quadrats reported in each of the five plots (P1 to P5). Parameter a was calculated from A_p based on the formulation provided in the text and their Figure 3.

[m] Data are reported for quadrat sizes of 4 (2 by 2), 9 (3 by 3), and 16 (4 by 4) trees. Parameter a was calculated from A_p, not A_x as reported in text, based on the binary power law model fit provided in their Figure 1.

[n] Parameter estimates are provided separately for analysis of different sampling approaches for leaves and leaflets.

[o] Unpublished data collected by W. W. Turechek from a 0.4-ha strawberry field located at the United States Department of Agriculture–Agricultural Research Service Beltsville Area Research Center in Beltsville, MD.

[p] TEV = T*obacco etch virus*; TvmV = *Tobacco vein mottling virus*.

[q] Unpublished data collected by W. W. Turechek, S. Adkins, and P. D. Roberts from a 1-ha watermelon field located at the University of Florida SW Research and Education Center in Immokalee, FL.

[r] Combined the 2003 and 2004 data sets and calculated the average n_h to recalculate parameters (data provided by M. Gosme).

utilized studies where the sampling unit size (n) was identified (or was identified after contacting the authors), so that a bivariate analysis of the two parameters of the usual BPL could be performed (e.g., citation 34 was excluded because the value of n was not provided). In some cases, it was discovered that the incorrect formulation (Appendix) or the incorrect logarithmic transformation was specified in a publication to the point where it was difficult to correctly interpret the parameters (e.g., citation 33). This was determined by plotting the power-law function as written in a given publication, and then comparing this with the original plot in the article. Appropriate adjustments were made when the error could be identified; otherwise, the study was dropped from the analysis. Regression analyses were conducted to determine whether the power law parameters varied predictably relative to each other and to n; a one-way ANOVA was used to determine if the parameters were affected by the common factors listed in Table 1 (crop, causal pathogen type, and pathogen dispersal mechanism). Analyses were conducted using Minitab (v. 15; Minitab Inc., State College, PA) and TableCurve 2D (v. 5.0; Systat Software Inc., San Jose, CA).

RESULTS

Simulation study. As found in previous studies with field observations and simulations, there was a strong relationship between the observed variance and the theoretical variance for a binomial distribution on a log-log scale (Figs. 1 and 2). The relationships were clearly linear over different ranges of the theoretical variance (and, hence, different ranges of incidence), although a break point could be seen in some graphs. The break point was at a very low incidence ($0.001 < P < 0.01$ for most conditions). The break point incidence, p_*, increased as sampling unit (quadrat) size increased (Tables 2 and 3; Fig. 2) and the SL property was only very apparent with more local dispersal (i.e., small values of μ) (Fig. 1) and larger sampling units (Fig. 2).

In Figure 1, the natural logarithm of the observed variance of disease incidence among the 1,024 quadrats (based on 5 by 5 individuals for each quadrat) at each sampling time was plotted against the corresponding theoretical binomial variance for all simulation runs at a given median dispersal distance. Note (as mentioned in Materials and Methods) that the BPL (equation 3), ABPL (equation 4), and the SL (equation 5) models were fitted to each individual simulated epidemic separately. From these, the averages of the model parameters and other fit statistics were calculated (Tables 2 and 3). In general, the SL provided the best fit, particularly for the smaller dispersal distances ($\mu < 2$), whereas the ABPL provided better fit for larger dispersal distances. However, it should be noted that the SL model failed to fit many data sets (Table 4), especially for smaller quadrat size and large dispersal distance. In other words, convergence often could not be achieved for those simulation conditions, probably because there was less evidence for two line segments in those cases. For individual runs, the improvement of fitting an SL or ABPL model over a BPL was marginal (Tables 2 and 3) and is unlikely to be apparent visually when plotted for a single simulation run.

For data where the SL model (equation 5) provided a good fit, the initial slope (b_3) was greater than the secondary slope (b_4) at small dispersal distance. This can be partially understood by

TABLE 1. (*continued from previous page*)

Ref.[a]	Disease	Crop	Organism	Mode of dispersal	T[b]	n	a	b
2[e]	Citrus sudden death	Citrus	Virus	Graft	98	64	0.2446	1.54
4[f]	Citrus sudden death	Citrus	Virus	Graft	55	4	0.3372	1.05
4[f]	Citrus sudden death	Citrus	Virus	Graft	55	16	0.1553	1.17
3[g]	*Citrus leprosis virus*	Citrus	Virus	Insect[h]	161	4	0.4894	1.13
3[g]	*Citrus leprosis virus*	Citrus	Virus	Insect[h]	161	9	0.3360	1.23
3[g]	*Citrus leprosis virus*	Citrus	Virus	Insect[h]	161	16	0.3210	1.34
3[g]	*Citrus leprosis Virus*	Citrus	Virus	Insect[h]	161	25	0.2773	1.40
4[i]	*Citrus tristeza virus*	Citrus	Virus	Insect	49	4	0.4871	1.15
4[i]	*Citrus tristeza virus*	Citrus	Virus	Insect	49	16	0.1678	1.25
21	*Citrus tristeza virus*	Citrus	Virus	Insect	54	4	0.3220	1.05
17	*Citrus tristeza virus*	Citrus	Virus	Insect	17	4	0.3853	1.07
27	Downy mildew	Grape	Fungus	Aerial	100	15	0.2546	1.30
20	Eutypa dieback	Grape	Fungus	Aerial	22	9	0.1480	0.97
24	*Maize chlorotic dwarf virus*	Maize	Virus	Insect	7	100	0.0912	1.45
10	Powdery mildew	Hop	Fungus	Aerial	104	10	0.2545	1.14
12[j]	Powdery mildew	Hop	Fungus	Aerial	1,606	10	0.1893	1.10
12[j]	Powdery mildew	Hop	Fungus	Aerial	770	10	0.2278	1.12
9	Powdery mildew	Hop	Fungus	Aerial	201	25	0.1470	1.22
44[k]	Powdery mildew	Hop	Fungus	Aerial	578	10	0.1738	1.09
44[k]	Powdery mildew	Hop	Fungus	Aerial	198	10	0.1984	1.10
30	Verticillium wilt	Olive	Fungus	Soil	29	2	0.6179	1.06
30	Verticillium wilt	Olive	Fungus	Soil	29	4	0.2923	1.02
30	Verticillium wilt	Olive	Fungus	Soil	29	8	0.2131	1.12
30	Verticillium wilt	Olive	Fungus	Soil	29	16	0.1340	1.19
32[l]	Bacterial blight	Onion	Bacterium	Rain	30	65.2	1.0105	1.48
8[m]	Plum pox	Peach	Virus	Insect	35	4	0.5648	1.22
8[m]	Plum pox	Peach	Virus	Insect	35	9	0.4610	1.34
8[m]	Plum pox	Peach	Virus	Insect	34	16	0.3972	1.48
23	Pear scab	Pear	Fungus	Rain	59	4	0.4794	1.15
19	Armillaria root rot	Spruce	Fungus	Soil	4	4	0.4383	1.06
42[n]	Leaf blight	Strawberry	Fungus	Rain	121	15	0.2697	1.18
42[n]	Leaf blight	Strawberry	Fungus	Rain	121	5	0.2874	1.07
42[n]	Leaf blight	Strawberry	Fungus	Rain	15	15	0.2247	1.12
42[n]	Leaf blight	Strawberry	Fungus	Rain	8	15	0.2231	1.11
41	Leaf spot	Strawberry	Fungus	Rain	59	15	0.4515	1.23
n/a[o]	Angular leaf spot	Strawberry	Bacterium	Rain	63	5.87	0.7558	1.33
18[p]	TEV & TVMV	Tobacco	Virus	Insect	188	40–60	0.1200	1.21
n/a[q]	*Squash vein yellowing virus*	Watermelon	Virus	Insect	18	4	0.7985	1.07
n/a[q]	*Cucurbit leaf crumple virus*	Watermelon	Virus	Insect	18	4	0.5381	1.03
15[r]	Take-all	Wheat	Fungus	Soil	36	20.59	0.2811	1.18

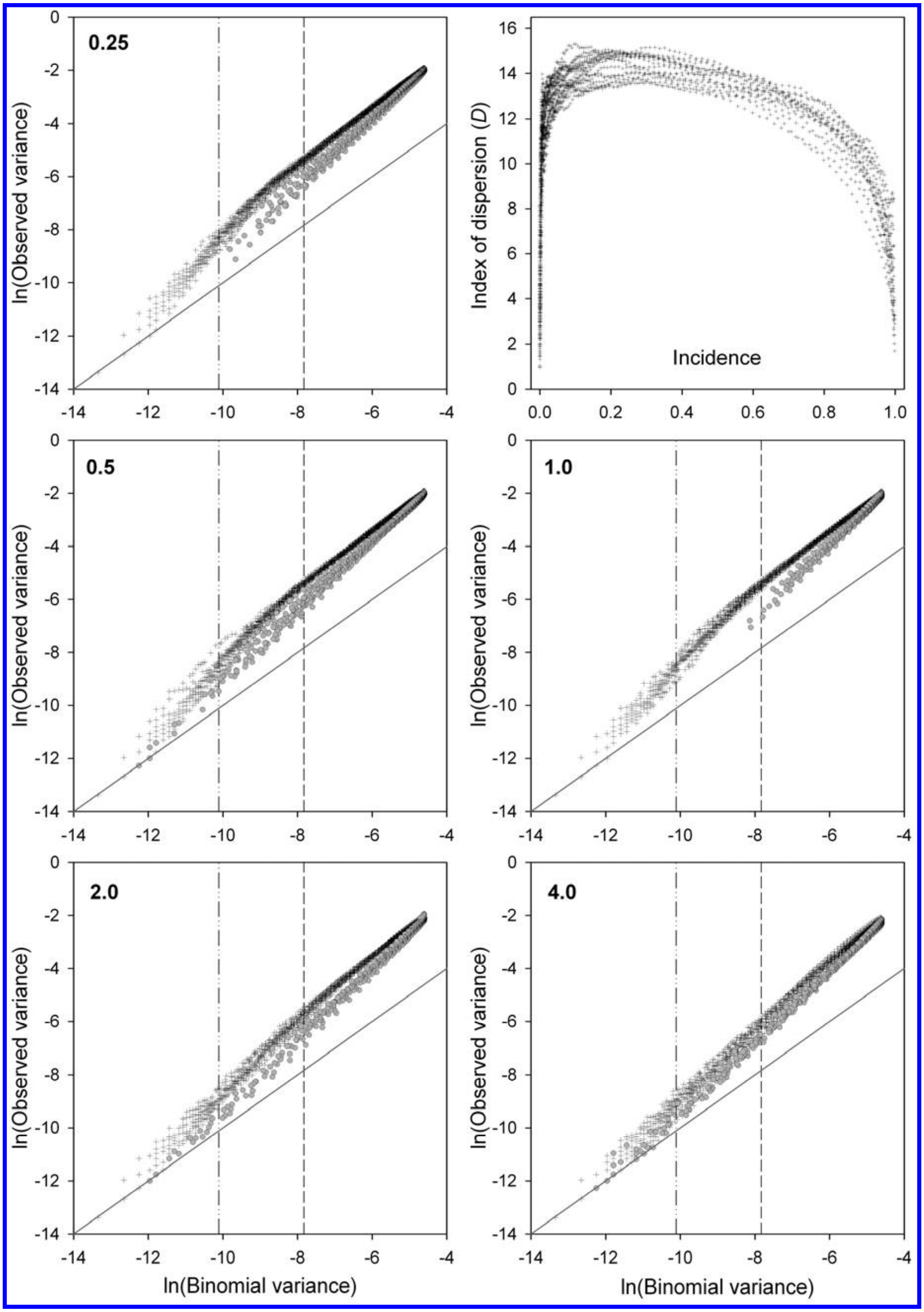

Fig. 1. Relationship between the logarithms of the observed variance and theoretical binomial variance from epidemic data generated from simulations of a two-dimensional stochastic spatial contact model where the distance that spores were dispersed followed a half-Cauchy distribution with median dispersal distance μ. Five dispersal distances were used (i.e., 0.25, 0.5, 1.0, 2.0, and 4.0). Twenty simulations were run for each dispersal distance on a daily time step and an infection rate of 0.4 per day was assumed. Epidemics were started with one initially infected leaf. Incidence was assessed daily and only data collected from the 5-by-5 quadrat is shown. The model was run until final incidence for most simulations was >0.95. Gray circles represent incidence >0.5 and black crosses represent incidence ≤0.5. The gray line represents the binomial line (i.e., when the observed variance equals the binomial variance). For interpretation of the horizontal scale, p = 0.001 corresponds to a binomial variance (with n = 25 here) of –10.11 and p = 0.01 corresponds to a binomial variance of –7.83. Top right graph: relationship between the ratio of the observed and theoretical variances (D) and disease incidence.

visualizing a plot s^2_{obs} / s^2_{bin} against *p,* where a sharp increase in the ratio is observed at the start of the epidemic (Fig. 1). However, the difference between slopes for the two line segments decreased as the dispersal distance, μ, increased (Tables 2 and 3). For large dispersal distances and small quadrat sizes, b_4 was often greater than b_3. Thus, b_3 does not have to be greater than b_4. In the majority (≈95%) of cases, b_4 was significantly >1 according to a *t* test, and the value of b_4 generally increased as the Cauchy parameter increased (Fig. 3). This is actually an indication of reduced overdispersion (small-scale aggregation) with a fixed intercept because, for a given value of ln(A_{p4}), smaller values for the slope parameter indicates greater aggregation (because $s^2_{obs} / s^2_{bin} = A_p \cdot (s^2_{bin})^{(b-1)}$, but s^2_{bin} < 1). With our simulations, b_4 was closest to 1, primarily for small μ and small *n*. Thus, the fact that b_4 = 1 for certain phases of the BPL in the simulations generated by Cascade is an indication of extreme local dispersal (small μ) or a very small quadrat size for a given intercept. In our simulations, b_4 was almost always >1.0 and, in >50% of cases, >1.1 (Fig. 3).

Evaluation of published data sets. In the representative studies listed in Table 1, the estimate of *b* for the fit of the BPL (equation 2) was numerically >1 in all but two cases; 50% were >1.175 and 90% were >1.05. Because disease incidence is determined with measurement error in actual studies, it is likely that the true values of *b* are somewhat larger than those determined in these studies. That is, it is well known that measurement

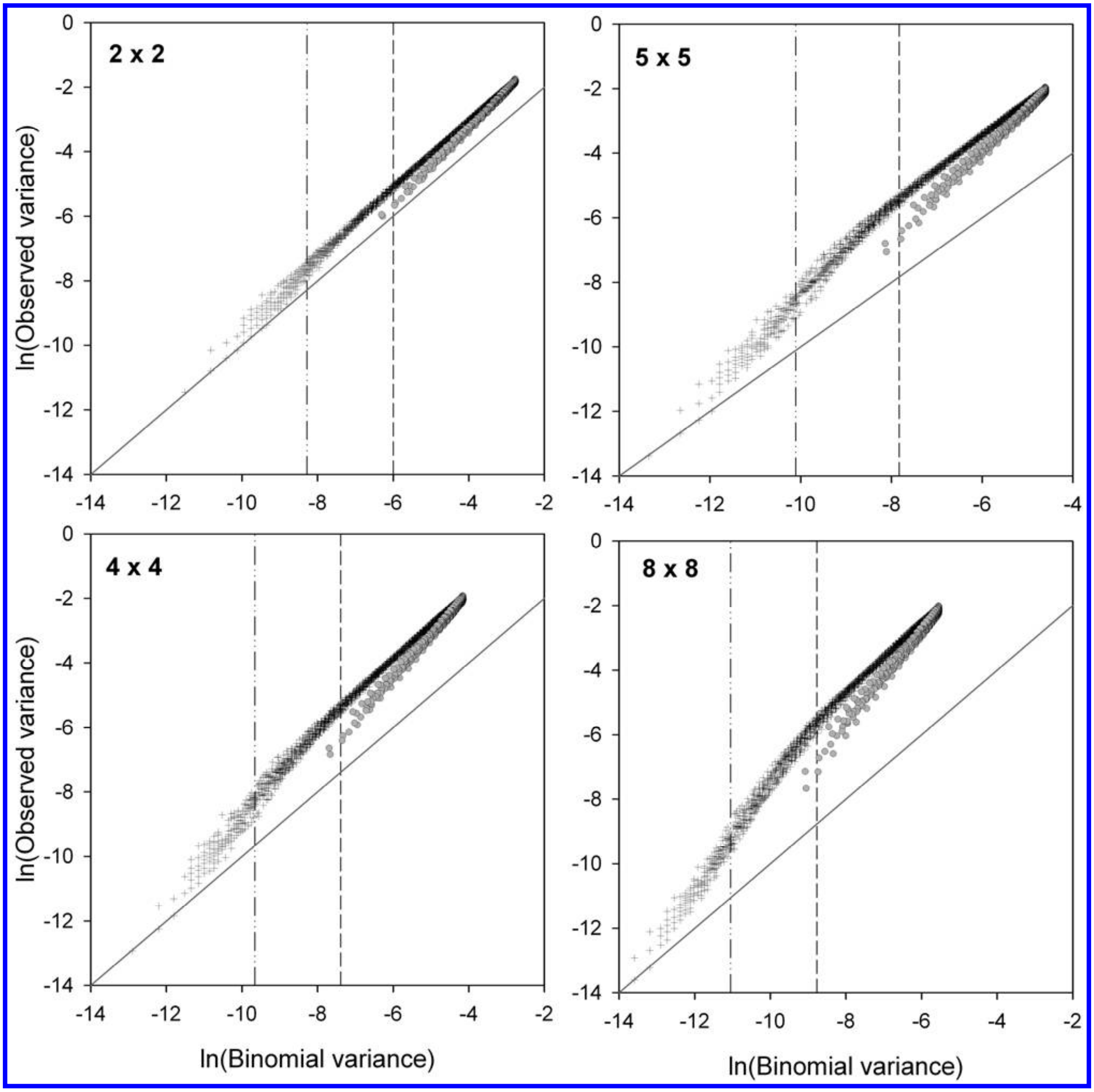

Fig. 2. Relationship between the logarithms of the observed variance and theoretical binomial variance from epidemic data generated from simulations of a two-dimensional stochastic spatial contact model where the distance that spores were dispersed followed a half-Cauchy distribution with median dispersal distance μ = 1. Four different quadrat sizes were used to sample the central 160-by-160 grid: 2 by 2, 4 by 4, 5 by 5, and 8 by 8. Twenty simulations were run for each dispersal distance on a daily time step and an infection rate of 0.4 per day was assumed. Epidemics were started with one initially infected leaf. The model was run until final incidence for most simulations was >0.95. Incidence was assessed daily. Gray circles represent incidence >0.5 and black crosses represent incidence ≤0.5. The gray line represents the binomial line (i.e., when the observed variance equals the binomial variance). For ease of comparison, *p* = 0.01 occurs at the binomial values –6.00 (*n* = 4), –7.39 (*n* = 16), –7.83 (*n* = 25), and –8.77 (*n* = 64); and *p* = 0.001 occurs at the binomial values –8.28 (*n* = 4), –9.66 (*n* = 16), –10.11 (*n* = 25), and –11.05 (*n* = 64).

error results in somewhat negatively biased estimates of the slope in a linear regression analysis (31). Inspection of the graphs from the articles (for the cases where graphs were shown) indicated no departure from linearity of the observed and theoretical variances on a log-log scale (*unpublished*). For most data sets, the smallest value of p in these studies was ≈0.0003 to 0.01 and only a limited number of data points had incidence <0.03 for any given data set. Thus, the first line segment identified originally by Gosme and Lucas (13), and confirmed by our simulations above, would generally not be observed with these datasets. In some data sets—namely Gottwald et al. (16) and Gent et al. (12)—the minimum p was on the order of 0.0001, and no evidence of two (or more) phases of the power law relation was seen in the graphs.

There was a very general tendency for the scale parameter (in the a formulation of equation A3 in the Appendix) to increase with b based on a regression analysis (Fig. 4A), although the relationship was weak. Except for one outlier, the a parameter decreased with the inverse of $\underline{n}$ (Fig. 4B). The slope parameter increased in proportion with $\sqrt{n}$ (Fig. 4C). The general increase in b with n is consistent with the results of simulations performed as part of this study, as well as in other studies (48). These results do not provide evidence to support the assertion of the slope tending toward the value 1 during the so-called second and third phases of epidemics.

There were significant relationships between plant species and pathogen type and the BPL parameters according to ANOVA (Table 5). Epidemics caused by bacteria had a significantly higher values for a than those caused by fungi or viruses, whereas epidemics caused by fungi had a significantly smaller b than those caused by either bacteria or viruses. There were significant differences associated with the crop for both parameters but there was no clear pattern in their differences, with the exception of the onion bacterial-blight outlier having a significantly higher value of a than all the other crop species except watermelon, and a value for b that was significantly higher than all other crop species except for maize, peach, tobacco, and wheat. The mode of dispersal, as characterized in this study, had a significant effect on a but not b. Mean separation using Fisher's least significant difference produced three overlapping groups for a. The rain-dispersed, insect-transmissible, and soilborne pathogens formed the upper group (i.e., the largest values for a); the insect-transmissible, soilborne, and graft-transmissible pathogens formed the middle group; and the soilborne, graft-transmissible, and aerially dispersed pathogens formed the bottom group.

DISCUSSION

The BPL has found many applications for characterizing overdispersion of disease incidence data in epidemics since its introduction in 1992 (18,27,28 [chapter 9]). Our interest in this study was focused on investigating two results of Gosme and Lucas (13,14): (i) a bi- or multiphasic relationship between the log of the observed variance and log of the theoretical binomial variance, with straight line segments found for each phase, and a break point (switch point) separating each line segment; and (ii) the slope of the BPL on a log-log scale being close to the value of 1 in what Gosme and Lucas (13) call the second and third phases of the relationship (corresponding to the range of incidence values typically observed in field studies). With respect to the breakpoint issue, an inspection of the literature did not reveal examples of two or more line segments in binary power-law graphs (log-log scale) for a wide range of published data sets. Using the very general stochastic spatio-temporal model of Xu and Ridout (47), we ran an independent set of simulations over a wide range of conditions (i.e., varying pathogen dispersal distance, initial disease level, and quadrat [sampling unit] size) to determine what

TABLE 2. Average slope and scale parameter estimates and average percentage of variance accounted for by the binary power law (BPL; equation 3), asymmetric binary power law (ABPL; equation 4), and split-line (SL; equation 5) models, and the break point incidence for the SL model for simulated epidemics initiated with one initial infection[a]

	BPL			ABPL				SL				
n[b]	b	a	R^2	b_1	b_2	a_1	R^2	p_*	b_3	b_4	a_4	R^2
Cauchy = 0.25												
4	1.06	0.764	99.6	1.07	1.16	0.835	**99.8**	0.0013	1.32	1.04	0.711	99.7
16	1.16	0.806	98.9	1.17	1.32	0.932	**99.4**	0.0024	1.50	1.08	0.635	**99.4**
25	1.21	0.815	98.8	1.22	1.37	0.980	99.1	0.0031	1.56	1.09	0.626	**99.5**
64	1.32	0.971	98.6	1.34	1.50	1.162	98.9	0.0073	1.62	1.13	0.589	**99.4**
Cauchy = 0.50												
4	1.08	0.773	99.3	1.09	1.18	0.844	**99.5**	0.0032	1.24	1.04	0.703	**99.5**
16	1.18	0.817	98.9	1.19	1.34	0.923	99.3	0.0031	1.53	1.08	0.635	**99.5**
25	1.23	0.870	98.6	1.24	1.40	0.990	98.9	0.0050	1.56	1.10	0.625	**99.3**
64	1.34	1.028	98.4	1.36	1.53	1.174	98.7	0.0095	1.64	1.13	0.595	**99.3**
Cauchy = 1.0												
4	1.08	0.736	99.7	1.09	1.24	0.844	**99.8**	0.0055	1.21	1.02	0.661	**99.8**
16	1.18	0.770	99.0	1.21	1.46	0.970	99.4	0.0045	1.49	1.06	0.560	**99.7**
25	1.22	0.829	98.8	1.26	1.54	1.062	99.2	0.0053	1.54	1.06	0.548	**99.6**
64	1.33	0.969	98.5	1.37	1.71	1.297	99.0	0.0079	1.66	1.09	0.542	**99.6**
Cauchy = 2.0												
4	1.08	0.721	99.7	1.08	1.18	0.763	**99.9**	0.0164	1.19	1.05	0.640	99.7
16	1.20	0.750	99.2	1.21	1.36	0.827	**99.6**	0.0159	1.40	1.10	0.589	99.5
25	1.26	0.814	99.3	1.28	1.43	0.914	**99.6**	0.0182	1.43	1.14	0.589	**99.6**
64	1.36	0.918	99.1	1.37	1.53	1.062	99.4	0.0180	1.56	1.16	0.569	**99.6**
Cauchy = 4.0												
4	1.11	0.632	**97.7**	1.11	1.18	0.670	**97.7**	0.0141	0.93	1.11	0.685	**97.7**
16	1.25	0.634	98.6	1.26	1.40	0.719	**98.7**	0.0417	1.18	1.16	0.514	**98.7**
25	1.29	0.987	98.1	1.29	1.49	0.763	**98.4**	0.0337	1.27	1.20	0.581	98.3
64	1.37	0.829	97.9	1.38	1.51	0.905	**98.2**	0.0326	1.35	1.22	0.575	98.1

[a] Intercept term in equations 3 to 5 was converted to the a parameter in the table based on the interrelations of scale parameters described in the Appendix. Note that a_4 was derived from $\ln(A_{p4})$. The initial infection was located at the center of a 200-by-200 simulation grid of plants. R^2 values in bold type indicate the best-fitting model for a given dispersal distance (μ) and quadrat size (n).

[b] Four different quadrat sizes were used to sample the central 160-by-160 grid: 2 by 2, 4 by 4, 5 by 5, and 8 by 8. Simulations were performed using a two-dimensional stochastic spatial contact model, where the distance that spores are dispersed follows a half-Cauchy distribution with median dispersal distance μ (47). Twenty simulations were run for each dispersal distance on a daily time step and an infection rate of 0.4 per day was assumed (note: the epidemic failed to establish in one run when μ = 0.25).

conditions could lead to multiple line segments with different slopes (log-log scale) with the BPL. The results of our simulations revealed that, under some circumstances, a break point can, indeed, be identified, with two line segments and different slopes being found; these results can help to explain the results generated by Cascade and explain why the two segments are not found with most observed data sets, as discussed below. With respect to the second point, it is very unusual for $b = 1$ with published data sets (Table 1), although our simulations show some (extreme) conditions—very local dispersal coupled with small sampling quadrat size—where b may approach 1. Thus, it appears that the results generated by Cascade (13,14) apply only to epidemics with a very specific set of characteristics, as discussed below.

In our set of simulations, a two-line-segment relationship with a break point was clearly evident in generated data corresponding to small dispersal distances (μ), consistent with a small length scale of the contact distribution (28). This was based on the goodness of fit of the SL model compared with the usual BPL, and the differences in magnitude between b_3 and b_4 of equation 5. A break point was less evident in generated data corresponding to larger dispersal distances or, if a break point was found, the differences in slopes between the two line segments (b_3 and b_4) were very small. The evidence for a break point was also stronger for large rather than small quadrat size (n). Based on our results here, it appears that the simulations run by Gosme and Lucas (13,14) are characteristic of epidemics with overall small dispersal distances, of the nearest-neighbor type (28), with something akin to transmission characteristics of soilborne diseases. This was acknowledged in their article, where take-all disease of wheat was used as an example of the dispersal process typical of the simulations that were run in Cascade (13).

In defining how the infection process is implemented in Cascade, Gosme and Lucas (13) write "…the model relies on the

TABLE 3. Average slope and scale parameter estimates and average percentage of variance accounted for by the binary power law (BPL; equation 3), asymmetric binary power law (ABPL; equation 4), and split-line (SL; equation 5) models, and the break point incidence for the SL model for simulated epidemics initiated with nine initial infections[a]

	BPL			ABPL				SL				
n[b]	b	a	R^2	b_1	b_2	a_1	R^2	p_*	b_3	b_4	a_4	R^2
Cauchy = 0.25												
4	1.10	0.737	99.7	1.11	1.16	0.795	99.8	0.0065	1.29	1.04	0.649	**99.9**
16	1.21	0.715	99.1	1.22	1.31	0.795	99.3	0.0048	1.63	1.09	0.542	**99.7**
25	1.31	0.813	99.1	1.32	1.39	0.896	99.2	0.0148	1.57	1.11	0.515	**99.8**
64	1.41	0.832	99.1	1.41	1.45	0.887	99.2	0.0149	1.69	1.19	0.512	**99.8**
Cauchy = 0.50												
4	1.10	0.730	99.7	1.11	1.17	0.787	**99.8**	0.0056	1.33	1.05	0.647	**99.8**
16	1.21	0.694	99.1	1.22	1.32	0.779	99.4	0.0053	1.56	1.11	0.539	**99.6**
25	1.31	0.789	99.2	1.32	1.39	0.861	99.3	0.0144	1.54	1.14	0.522	**99.7**
64	1.41	0.832	99.2	1.42	1.46	0.878	99.2	0.0162	1.67	1.20	0.511	**99.8**
Cauchy = 1.0												
4	1.11	0.720	**98.4**	1.10	1.12	0.719	**98.4**	0.0326	1.17	1.03	0.620	**98.4**
16	1.24	0.699	96.4	1.23	1.25	0.698	96.5	0.0153	1.41	1.11	0.523	**96.7**
25	1.31	0.781	96.3	1.31	1.30	0.756	96.4	0.0264	1.46	1.12	0.504	**96.6**
64	1.41	0.815	96.1	1.41	1.37	0.771	96.3	0.0306	1.58	1.18	0.483	**96.5**
Cauchy = 2.0												
4	1.11	0.638	**99.9**	1.10	1.11	0.638	**99.9**	0.0471	1.13	1.07	0.587	**99.9**
16	1.24	0.614	98.9	1.23	1.23	0.600	**99.0**	0.0272	1.38	1.15	0.497	**99.0**
25	1.31	0.666	99.3	1.31	1.28	0.638	99.4	0.0282	1.45	1.18	0.488	**99.5**
64	1.42	0.694	98.6	1.42	1.35	0.664	**98.9**	0.0252	1.59	1.24	0.488	**98.9**
Cauchy = 4.0												
4	1.08	0.498	**96.1**	1.09	1.08	0.497	**96.1**	0.0212	1.00	1.09	0.506	**96.1**
16	1.21	0.405	94.5	1.21	1.19	0.399	**94.6**	0.0104	1.19	1.21	0.397	94.5
25	1.28	0.415	94.5	1.28	1.25	0.415	**94.6**	0.0316	1.18	1.24	0.386	94.5
64	1.38	0.423	95.5	1.40	1.31	0.423	**96.0**	0.0497	1.29	1.31	0.376	95.6

[a] Intercept term in equations 3 to 5 was converted to the a parameter in the table based on the interrelations of scale parameters described in the Appendix. Note that a_4 was derived from $\ln(A_{p4})$. The nine initially infected leaves were regularly spaced in the grid to obtain the maximum distance apart from each other. R^2 values in bold type indicate the best-fitting model for a given dispersal distance (μ) and quadrat size (n).

[b] Four different quadrat sizes were used to sample the central 160-by-160 grid: 2 by 2, 4 by 4, 5 by 5, and 8 by 8. Simulations were performed using a two-dimensional stochastic spatial contact model, where the distance that spores are dispersed follows a half-Cauchy distribution with median dispersal distance μ (47). Twenty simulations were run for each dispersal distance on a daily time step and an infection rate of 0.4 per day was assumed (note: the epidemic failed to establish in one run when μ = 0.25).

TABLE 4. Number of individual simulation runs (out of 20 unless indicated otherwise) that cannot be fitted by a split-line (SL) model (equation 5) but can be well fitted by the linear version of the binary power law (BPL, equation 3) and asymmetric binary power law (ABPL, equation 4) models

	One initial infection, Cauchy parameter (μ)[b]						Nine initial infections, Cauchy parameter (μ)[b]					
Quadrat size[a]	0.25[c]	0.50	1.00	2.00	4.00	Total	0.25	0.50	1.00	2.00	4.00	Total
4	9	8	6	14	9	46	3	2	6	8	8	27
16	2	4	0	8	8	22	0	0	1	1	13	15
25	4	1	0	2	6	13	0	0	0	1	12	13
64	2	2	0	3	4	11	0	0	0	1	7	8
Total	17	15	6	27	27	92	3	2	7	11	40	63

[a] Four different quadrat sizes were used to sample the central 160-by-160 grid: 2 by 2, 4 by 4, 5 by 5, and 8 by 8.

[b] Simulations were performed using a two-dimensional stochastic spatial contact model, where the distance that spores were dispersed followed a half-Cauchy distribution with median dispersal distance μ. Twenty simulations were run for each dispersal distance on a daily time step and an infection rate of 0.4 per day was assumed (note: the epidemic failed to establish in one run when μ = 0.25).

[c] Only 19 runs were used.

assumption that the host's spatial structure is sufficiently strong so that interactions between holons from two different groups are negligible compared with interactions between holons of the same group." In their model, which is not spatially explicit, infections between holons (= units) at one level (e.g., leaves) can only occur within previously infected units of the level above (e.g., plants) and, therefore, "long-distance" infections only occur at the highest level and must occur at a lower rate than infections at the lower levels (i.e., the dispersal distances must be short. Even though the infection rate between units at each scale was varied or investigated in their article, their results routinely produced the SL property, which was taken as evidence that the phenomenon is a (hidden or cryptic) characteristic of many epidemics. In contrast, our simulation results, where the spore dispersal function was explicitly represented, showed that the SL property for the BPL on a log-log scale is primarily a characteristic of epidemics with short dispersal distances. Again, however, such relationships were not apparent in published empirical studies conducted with pathogens that possess short dispersal distances (15,30).

It is noteworthy that the break point occurs at a very low level of incidence, both with the Cascade model and in our simulations, in the vicinity of $P \approx 0.001$ to 0.03 (for most situations). It also occurs at very high incidences with the Cascade model (their phase 4). The larger slope in the first phase (when it occurs) could have multiple causes. For the Taylor power law (for unbounded counts), compared with the overall relationship, unusual or artificial results can occur at extremely low densities, where the vast majority of quadrats are "empty" (i.e., 0 individuals of interest), and a very small number of quadrats have one or a small number of individuals (39). At such low densities, concepts of aggregation and heterogeneity may even be difficult to apply. Similar artifacts occur with the BPL in the case when only one diseased individual is observed among all N sampling units, yielding $\ln(s^2_{obs}) = \ln(s^2_{bin})$ and a potentially high leverage point in regression models.

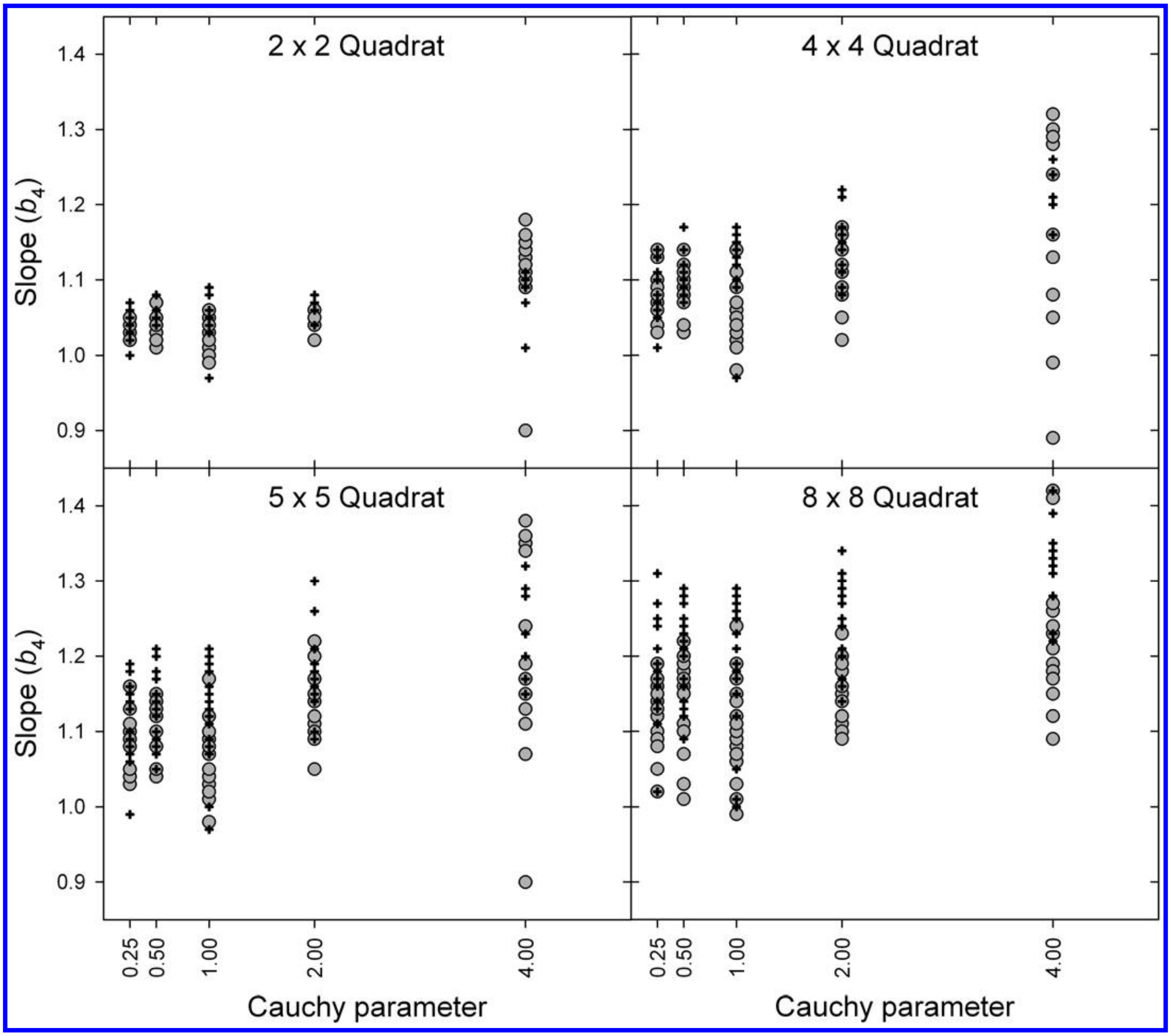

Fig. 3. Slope (b_4) of the second line in the split-line (SL) model fitted to describe the relationship between the logarithms of the observed variance and theoretical binomial variance from epidemic data generated from simulations of a two-dimensional stochastic spatial contact model where the distance that spores were dispersed followed a half-Cauchy distribution with five median dispersal distances μ = 0.25, 0.5, 1, 2, and 4. Simulations were run for each dispersal distance on a daily time step and an infection rate of 0.4 per day was assumed. The model was run until final incidence for most simulations was >0.95. Four different quadrat sizes were used to sample the central 160-by-160 grid: 2 by 2, 4 by 4, 5 by 5, and 8 by 8. The SL model failed to fit a varying number of data sets (Table 4). Gray circles and black cross represent simulations with one and nine initial infected plants, respectively.

Furthermore, in a theoretical assessment of population demographics (birth, death, immigration, and emigration) on the (unbounded) Taylor power law, Anderson et al. (1) showed that the slope can vary with density for different demographic situations, and the slope at very low density can be either higher or lower than the overall slope.

Gosme and Lucas rightly argued that the break-point discovery for overdispersion of disease incidence (or any binary variable) could not have been made without stochastic simulation modeling, because experimental data often have too few (if any) observations, particularly at the beginning and end of the epidemic, where disease incidence is at its lowest and highest. Sampling and measurement error at very low and high incidence also will result in estimates of the mean and variance with low relative precision when the number of sampling units is not large (28). Few investigators monitoring single epidemics over time or many epidemics at a single time can commit resources to collect data on ≥1,000 sampling units to obtain precise information on incidence and overdispersion near the extremes of the incidence scale. The minimum value of incidence in the data sets summarized in Table 1 was typically >0.005, or there were few observations below this value to be able to visualize or model a break point. However, Gottwald et al. (16) show results for 321 data sets with minimum incidence values ≈0.001 (6 data sets with $P < 0.001$ and 95 with $P < 0.01$), and they did not observe the SL property. Thus, it remains uncertain how common an SL phenomenon is with actual epidemic data if sufficient data were available to quantify the process; however, we predict that it could be found when there is close to a nearest-neighbor dispersal process (small μ). The practical question arises, then, whether the SL property needs to be accounted for in developing sampling models for either incidence estimation or decision making (6,28), or in conducting other analyses dependent on the degree of overdispersion. It seems reasonable to answer "yes" if the SL property can be identified in observed data, the SL model fits the data better than the BPL model, and the objective is to sample or perform other analyses at very low or high levels of incidence (near 0 or 100%). Sensitivity analyses will have to be performed to determine the error introduced in these applications if the SL is ignored and analytical efforts focus on incidence near 0 or 100%.

When the SL relationship occurs, it is clear that the ABPL (equation 4) and SL (equation 5) models are better choices for representing the power law relationship than the simpler BPL model (equation 3). However, the SL model must be fitted by nonlinear estimation methods and cannot always be fitted successfully (when the data do not support two line segments). The ABPL model, first introduced by Hughes and Madden (18) in their original article on the BPL, may be a viable alternative to the SL model, because the former can be fitted using linear least squares. The ABPL generally provided a better fit to the simulation data than did the usual BPL model. However, as mentioned above, the improved fit to the data by the ABPL and SL models is often marginal relative to the BPL model, so the gain tends to be minimal and the improved fit is unlikely to be apparent visually when the data are plotted. As demonstrated above, a wide range of incidence values, including many points with values <0.01 (as deliberately chosen with our simulations) would be needed to clearly distinguish among these model fits. Thus, it is not surprising that the SL relationship for the variances on a log-log scale is not apparent for observed data sets.

The property of the slope being ≈1 in what Gosme and Lucas (13) refer to as the second and third phases of the BPL log-log relationship (b_4 in equations 5) was not generally seen in our simulations. The property also was not evident in the published data sets we evaluated, both those described in Table 1 and those not included because n was not given (*unpublished*). The slope for the second/third phase was clearly >1 under most circumstances in our simulations, and only came very close to 1 for small μ and small n. In fact, the range of b_4 values (equation 5) in our simulations was very similar to the values of estimated b (equation 3, or one of the equivalent formulations in the Appendix) in Table 1 for observed data sets. This was expected because the BPL (equation 3) is usually fitted to data with incidence values >0.005 or 0.01. Previous simulation results by Xu and Ridout (47,48) also resulted in b values >1 and generally <2, with many values between 1.05 and 1.5. Even values of b of 1.05 or 1.10 can have substantial effects on the index of dispersion (D) at a given mean

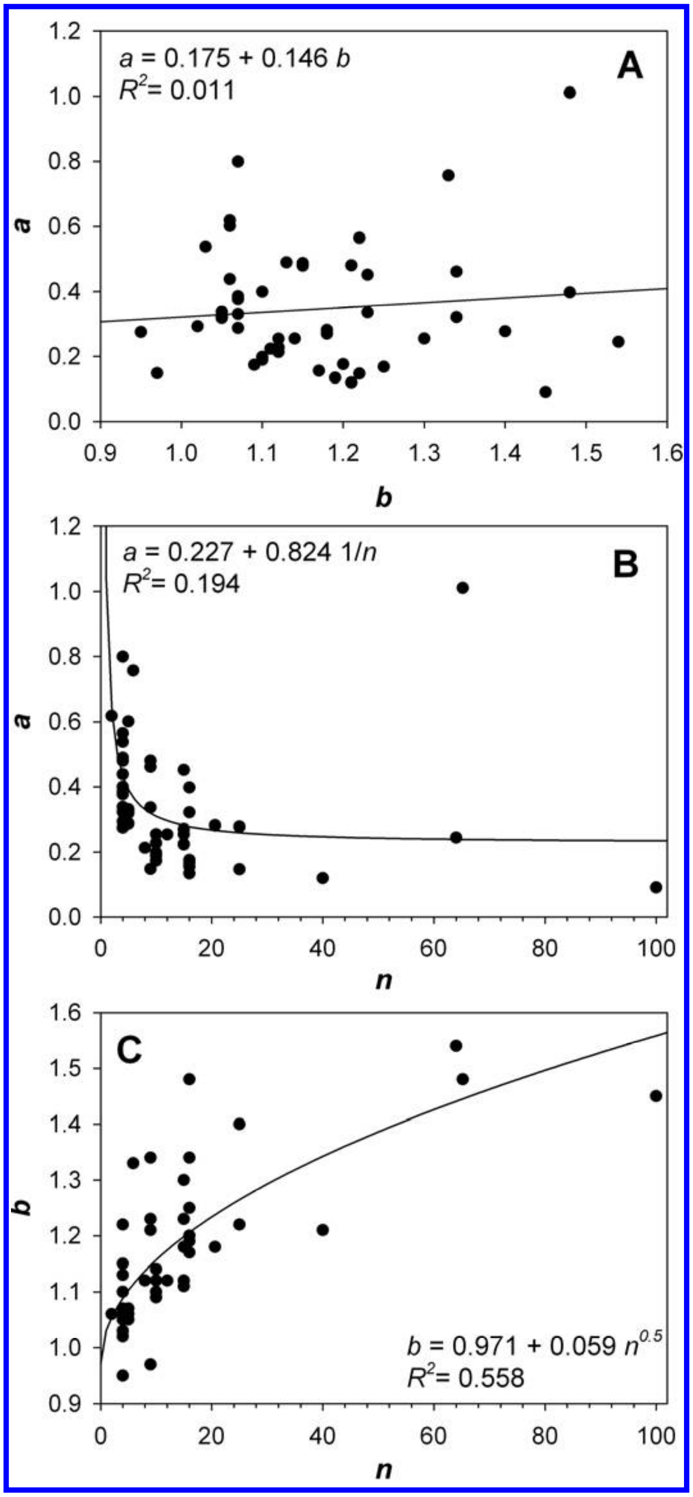

Fig. 4. Relationships between **A,** the scale parameter (a) and slope (b) parameters of the binary power law (see equation A3 and equations 1 and 2); **B,** a and the sampling unit size (n); and **C,** b and n from the plant pathology-based studies listed in Table 1. Solid lines are model fits to the data; equation and fit statistics are given in the individual panels.

TABLE 5. Result from a one-way analysis of variance where the effects of plant species (crop), pathogen type (organism), and main dispersal mechanism on the binary power law parameters *a* and *b* reported in Table 1 were examined

		a			*b*		
Factor	df	MS	*F*	*P*	MS	*F*	*P*
Crop	14	0.0735	3.88	0.0006	0.0308	2.32	0.0220
Organism	2	0.2370	9.13	0.0004	0.0690	4.28	0.0196
Dispersal	4	0.1026	3.60	0.0125	0.0208	1.15	0.3440

incidence. Because the lowest incidence of disease in most of the published studies assembled here was greater than what is needed to recognize a break point between phases one and two/three, results from the published studies would generally correspond to the proposed phases two/three of Gosme and Lucas (13); b from the usual BPL (equation 3) would then directly correspond to b_4 from the SL model (equations 5 and 6b). Given the magnitude of b_4 values in this study, and the fact that this parameter generally increased with increasing n, we conclude that slope in phase two/three will rarely equal or be very close to 1.

Although the stochastic simulation model of Xu and Ridout (47) is more general than the Cascade model, especially in terms of the dispersal function (contact distribution) that is allowable, there is a limit to how far one can go in using one model to refute (or support) another model. However, the results from the Xu and Ridout model—both here and in previous publications (47,48)—agree with the results from observed data sets for the range of incidence values typically found with disease in the field. The simulations in the current study, and those by Gosme and Lucas (13,14), were concerned with the effects of stochastic population-dynamic (demographic and dispersal) factors and some sampling methods (i.e., size of sampling units) on binary power-law-based characterizations of overdispersion over time in epidemics. However, the BPL (equations 1 and 2, or any of the formulations in the Appendix), with a single slope on a log-log scale, also provides a good description of overdispersion for data points coming from different epidemics; that is, where each data point is a variance for a different field (with different incidence values) (Table 1). Several of the data sets in Table 1 correspond to this situation. For observed data sets, stochastic environmental effects influence the overdispersion results as much as population-dynamic effects. Using the collected data, we were unable to identify any biological factors that had a consistent effect on the BPL parameters. Hughes and Madden (18), in their original article and in subsequent publications (24), did not prescribe any particular mechanism to the BPL because of the likelihood that several mechanisms work together to produce the repeatable results.

In conclusion, the SL results of Gosme and Lucas (13), with a break point at a low incidence, were confirmed using a more general simulator. However, the small values of the slope they found in the second/third phase likely reflect extreme short-range contact distributions and are not typical of most observed data. Accounting for the break-point relationship in data sets where they are apparent can be accomplished by fitting an SL or asymmetric form of the BPL, although the improved fit of these models may be marginal relative to the usual BPL.

APPENDIX

With a binomial random variable, the variance of the counts is given by $np(1 - p)$, and the variance of the proportions is given by $[p(1 - p)]/n$; with the estimate of p, one substitutes these expressions for s_{bin}^2 in equation 1 to obtain

$$s_{obs}^2 = V_x = A_x[np(1-p)]^b \tag{A1}$$

$$s_{obs}^2 = V_p = A_p[p(1-p)/n]^b \tag{A2}$$

The observed variance in these equations (left side) must be based on the number of diseased individuals per sampling unit in equation A1 (V_x) and the proportion of diseased individuals per sampling unit in equation A2 (V_p). When n is fixed for each sampling unit, one can write $[p(1-p)/n]^b$ as $n^{-b}[p(1-p)]^b$, and equation A2 can be written as

$$s_{obs}^2 = V_p = a[p(1-p)]^b \tag{A3}$$

where $a = A_p n^{-b}$. Equation A3 was the form used originally by Hughes and Madden (18), although equations A1 and A2 have been more commonly used in subsequent years. Here, we prefer the versions based on proportions (equations A2 and A3), and we generally follow the nomenclature of Turechek and Madden (42) for the scale parameter.

With fixed n, one can also write $[np(1-p)]^b$ of equation A1 as $n^b[p(1-p)]^b$, and move n into the scale parameter for the variance of counts:

$$s_{obs}^2 = V_x = a_x[p(1-p)]^b \tag{A4}$$

where $a_x = A_x n^b$. Because $V_x = n^2 V_p$, by definition, one can write $a = a_x n^{-2}$, $A_p = A_x n^{2b-2}$, and $a = A_x n^{b-2}$. These conversions are essential to compare results from different studies (when authors use different formulations of the BPL) and to use formulas for calculating sample sizes and estimates of overdispersion (28).

All of the equations in the Appendix describe a symmetrical relation between the observed variance and p (i.e., the same variance is specified for p and $1 - p$). An asymmetrical relation can be defined by a generalization of equation A3 (or A4). For instance, consider the right-hand side of equation A3, which can be written as $ap^b(1-p)^b$. If the exponents for p and $1 - p$ are different (b_1 and b_2, respectively), then a more complex BPL can be written as

$$s_{obs}^2 = V_p = a_1 p^{b_1}(1-p)^{b_2} \tag{A5}$$

which allows for different variances at p and $1 - p$ (24).

ACKNOWLEDGMENTS

Financial support for W. Turechek was provided in part by the United States Department of Agriculture–Agricultural Research Service (USDA-ARS) CRIS project 6618-22000-026-00D. Financial support for L. Madden was provided in part by USDA Special Grant 2010-34493-21087. Financial support for D. Gent was provided in part by USDA-ARS CRIS project 5358-21000-040-00D. Financial support for X.-M. Xu was provided in part by the Chinese Natural Science Foundation project number 30928016.

LITERATURE CITED

1. Anderson, R. M., Gordon, D. M., Crawley, M. J., and Hassell, M. P. 1982. Variability in the abundance of animal and plant species. Nature 296:245-248.
2. Bassanezi, R. B., Bergamin, A., Amorim, L., Gimenes-Fernandes, N., Gottwald, T. R., and Bove, J. M. 2003. Spatial and temporal analyses of citrus sudden death as a tool to generate hypotheses concerning its etiology. Phytopathology 93:502-512.
3. Bassanezi, R. B., and Laranjeira, F. F. 2007. Spatial patterns of leprosis

and its mite vector in commercial citrus groves in Brazil. Plant Pathol. 56:97-106.

4. Batista, L., Bassanezi, R. B., and Laranjeira, F. F. 2008. Comparative epidemiology of citrus tristeza in Cuba and citrus sudden death in Brazil. Trop. Plant Pathol. 33:348-355.
5. Biggs, A. R., Turechek, W. W., and Gottwald, T. R. 2008. Analysis of fire blight shoot infection epidemics on apple. Plant Dis. 92:1349-1356.
6. Binns, M. R., Nyrop, J. P., and van der Werf, W. 2000. Sampling and Monitoring in Crop Protection: The Theoretical Basis for Developing Practical Decision Guides. CABI Publishing, Oxon, UK.
7. Carisse, O., Meloche, C., Boivin, G., and Jobin, T. 2009. Action thresholds for summer fungicide sprays and sequential classification of apple scab incidence. Plant Dis. 93:490-498.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. Gosme, M., and Lucas, P. 2009. Cascade: An epidemiological model to simulate the disease spread and aggregation across multiple scales in a spatial hierarchy. Phytopathology 99:823-832.
14. Gosme, M., and Lucas, P. 2009. Disease spread across multiple scales in a spatial hierarchy: Effect of host spatial structure and of inoculum quantity and distribution. Phytopathology 99:833-839.
15. Gosme, M., Willocquet, L., and Lucas, P. 2007. Size, shape and intensity of aggregation of take-all disease during natural epidemics in second wheat crops. Plant Pathol. 56:87-96.
16. Gottwald, T. R., Bassanezi, R. B., Amorim, L., and Bergamin-Filho, A. 2007. Spatial pattern analysis of citrus canker-infected plantings in Sao Paulo, Brazil, and augmentation of infection elicited by the Asian leafminer. Phytopathology 97:674-683.
17. Hughes, G., and Gottwald, T. R. 1999. Survey methods for assessment of citrus tristeza virus incidence when *Toxoptera citricida* is the predominant vector. Phytopathology 89:487-494.
18. Hughes, G., and Madden, L. V. 1992. Aggregation and incidence of disease. Plant Pathol. 41:657-660.
19. Hughes, G., and Madden, L. V. 1998. Comment—using spatial and temporal patterns of Armillaria root disease to formulate management recommendations for Ontario's black spruce (*Picea mariana*) seed orchards. Can. J. For. Res. 28:154-158.
20. Hughes, G., Madden, L. V., and Munkvold, G. P. 1996. Cluster sampling for disease incidence data. Phytopathology 86:132-137.
21. 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.
22. Humeau, L., Roumagnac, P., Picard, Y., Robene-Soustrade, I., Chiroleu, F., Gagnevin, L., and Pruvost, O. 2006. Quantitative and molecular epidemiology of bacterial blight of onion in seed production fields. Phytopathology 96:1345-1354.
23. Li, B. H., Yang, J. R., Li, B. D., and Xu, X. -M. 2007. Incidence-density relationship of pear scab (*Venturia nashicola*) on fruits and leaves. Plant Pathol. 56:120-127.
24. Madden, L. V., and Hughes, G. 1995. Plant disease incidence: Distributions, heterogeneity, and temporal analysis. Annu. Rev. Phytopathol. 33:529-564.
25. Madden, L. V., and Hughes, G. 1999. An effective sample size for predicting plant disease incidence in a spatial hierarchy. Phytopathology 89:770-781.
26. Madden, L. V., and Hughes, G. 1999. Sampling for plant disease incidence. Phytopathology 89:1088-1103.
27. Madden, L. V., Hughes, G., and Ellis, M. A. 1995. Spatial heterogeneity of the incidence of grape downy mildew. Phytopathology 85:269-275.
28. Madden, L. V., Hughes, G., and van den Bosch, F. 2007. The Study of Plant Disease Epidemics. American Phytopathological Society, St. Paul, MN.
29. Madden, L. V., Turechek, W. W., and Nita, M. 2002. Evaluation of generalized linear mixed models for analyzing disease incidence data obtained from designed experiments. Plant Dis. 86:316-325.
30. Navas-Cortés, J. A., Landa, B. B., Mercado-Blanco, J., Trapero-Casas, J. L., Rodríguez-Jurado, D., and Jiménez-Díaz, R. M. 2008. Spatiotemporal analysis of spread of infections by *Verticillium dahliae* pathotypes within a high tree density olive orchard in southern Spain. Phytopathology 98:167-180.
31. Neter, J., Wasserman, W., and Kutner, M. H. 1983. Applied Linear Regression Models. Richard D. Irwin, Inc., Press, Howewood, IL.
32. Roumagnac, P., Pruvost, O., Chiroleu, F., and Hughes, G. 2004. Spatial and temporal analyses of bacterial blight of onion caused by *Xanthomonas axonopodis* pv. *allii*. Phytopathology 94:138-146.
33. Ruiz, L., Janssen, D., Martin, G., Velasco, L., Segundo, E., and Cuadrado, I. M. 2006. Analysis of the temporal and spatial disease progress of *Bemisia tabaci*-transmitted *Cucurbit yellow stunting disorder virus* and *Cucumber vein yellowing virus* in cucumber. Plant Pathol. 55:264-275.
34. Savary, S., Castilla, N. P., and Willocquet, L. 2001. Analysis of the spatiotemporal structure of rice sheath blight epidemics in a farmer's field. Plant Pathol. 50:53-68.
35. Schabenberger, O., and Pierce, F. J. 2002. Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press LLC, Boca Raton, FL.
36. Shah, D. A., Dillard, H. R., and Nault, B. A. 2005. Sampling for the incidence of aphid-transmitted viruses in snap bean. Phytopathology 95:1405-1411.
37. Shaw, M. W. 1995. Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. Proc. R. Soc. B 259:243-248.
38. Sposito, M. B., Amorim, L., Bassanezi, R. B., Filho, A. B., and Hau, B. 2008. Spatial pattern of black spot incidence within citrus trees related to disease severity and pathogen dispersal. Plant Pathol. 57:103-108.
39. Taylor, L. R. 1984. Assessing and interpreting the spatial distributions of insect populations. Annu. Rev. Entomol. 29:321-357.
40. 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.
41. 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.
42. Turechek, W. W., and Madden, L. V. 1999. Spatial pattern analysis of strawberry leaf blight in perennial production systems. Phytopathology 89:421-433.
43. Turechek, W. W., and Madden, L. V. 2003. A generalized linear modeling approach for characterizing disease incidence in a spatial hierarchy. Phytopathology 93:458-466.
44. 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.
45. Willocquet, L., and Savary, S. 2004. An epidemiological simulation model with three scales of spatial hierarchy. Phytopathology 94:883-891.
46. Xu, X.-M., and Madden, L. V. 2002. Incidence and density relationships of powdery mildew on apple. Phytopathology 92:1005-1014.
47. Xu, X.-M, and Ridout, M. S. 1998. Effects of initial epidemic conditions, sporulation rate, and spore dispersal gradient on the spatio-temporal dynamics of plant disease epidemics. Phytopathology 88:1000-1012.
48. Xu, X.-M, and Ridout, M. S. 2000. Effects of quadrat size and shape, initial epidemic conditions, and spore dispersal gradient on spatial statistics of plant disease epidemics. Phytopathology 90:738-750.