Letter to the Editor

A Distribution-Free Method for Estimating the Effect of Aggregated Plant Damage on Crop Yield

Francis J. Ferrandino

Assistant scientist, the Connecticut Agricultural Experiment Station, Department of Plant Pathology and Ecology, P. O. Box 1106, New Haven 06504.

Accepted for publication 22 September 1989 (submitted for electronic processing).

Many plant pathogens and herbivorous insects can cause spatially aggregated damage to plants. Recently, Hughes (4,5) pointed out that the spatial variability of crop damage due to harmful organisms can have an important effect on the prediction of crop yield and that this effect is often ignored. The magnitude and direction of the bias introduced into the estimation of yield by spatial aggregation will depend on both the shape of the curve for observed yield versus severity of attack, and the frequency distribution of damage to plants. To demonstrate this bias in the prediction of yield, Hughes (5) expressed yield from a field with irregularly distributed damage in terms of a weighted average over the expected yield from each plant. In these calculations a truncated negative binomial function was chosen to describe the distribution of severity of damage, and the yield versus severity relation was given by a one-parameter function, which allowed for either a positive or a negative second derivative of yield with respect to severity.

Hughes (5) showed that when the second derivative of yield with respect to severity is positive, then the overall average yield in a field with a constant amount of damage increases with increasing aggregation of damage. However, when this second derivative is negative, a field with a clustered distribution of damage tends to yield less than its homogeneously damaged counterpart. These two shapes of the yield versus severity of attack curve have been termed Type I (positive second derivative) and Type II (negative second derivative), respectively (Fig. 1; 8). Johnson (6) suggested that, in general, Type I curves are characteristic of pests that affect the radiation use efficiency, RUE, and that Type II curves may be associated with pests that cause foliar damage to crops with moderate to high LAI and, thus, directly affect intercepted radiation. The shape of the yield versus severity of attack curve may also, however, depend on the way in which the attack on the host plant is quantified and averaged over time (10). For example, in a study on the effect of several *Meloidogyne* species of nematodes on the yield of flue-cured tobacco (2), a Type I yield-severity relation was observed when nematode population density was used to describe severity of attack. However, when a direct measure of damage to plants (root-gall index) was used to quantify severity attack, the resultant yield-severity relation was either linear or a Type II curve.

A similar methodology was used by Noe and Barker (9) to predict the effect of aggregation of nematodes (*Meloidogyne incognita*) on yield loss. They used pathogen population density to characterize severity of attack. Yield was assumed to linearly decrease with the log of the population density and, once again, the negative binomial function was used to describe the distribution of the pathogen. Predicted yield was found to increase when the nematode population was spatially aggregated because the second derivative of the logarithmic relation between yield and population density is positive. Of course, this effect was largest for small values of the parameter k in the negative binomial distribution (3), which is a measure of dispersion in the population.

To implement the above approach, one needs to know, a priori, the frequency distribution for different levels of plant damage throughout a field. When spatial aggregation is important, the negative binomial distribution has been applied successfully to describe damage to plants for many observed epidemics (11). There are, however, some difficulties in applying this unbounded and discrete distribution to disease data. When applied to a bounded variable such as fractional disease severity, the distribution must be truncated and renormalized (5). When applied to population density (9), which is estimated by counting the number of individuals within a certain sample volume, the fitted value of k for the distribution is very sensitive to the size of the chosen sampling unit.

The purpose of this letter is to present an approximate method that provides an easily calculated yet reasonably accurate estimate of expected yield when damage due to disease is distributed nonhomogeneously. This method uses data directly and is not dependent on a fitted distribution function to describe the observed variability of damage to plants.

METHODS

I assume that the final yield of a damaged plant with severity of attack, s, can be expressed as a fraction, $y(s)$, of the yield produced by a healthy plant ($s = 0$), and that s is divided into L discrete categories s_i for $1 < i < L$. Within a field containing a total of N plants, let n_i be the number of plants with severity of attack s_i. The mean severity of attack, $\bar{s}$, is then defined by:

$$\bar{s} = [\sum_{i=1} n_i \cdot s_i] / N. \quad (1)$$

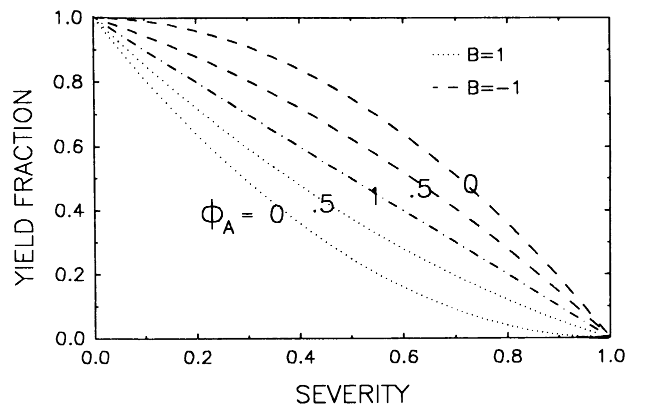

Fig. 1. Effect of severity (s) and aggregation on fractional yield (y). The predictions for $\bar{y}$ using Eq. 8 are plotted versus $\bar{s}$ for $B = +1$ (dotted line, Type I) and $B = -1$ (dashed line, Type II) for $\phi_A = 0$, 0.5, and 1.

A plant with severity s_i will produce fractional yield, $y(s_i)$, and the mean fractional yield per plant for the entire field $\bar{y}$ can be expressed in terms of a weighted average of $y(s_i)$:

$$\bar{s} = [\sum_{i=1}^{L} n_i \cdot y\,(s_i)] \;/\; N. \tag{2}$$

Expanding $y(s_i)$ in a Taylor series about $y(\bar{s})$ one obtains:

$$\bar{y} = [\sum_{i=1}^{L} n_i \cdot [(y(\bar{s}) + y'\,(\bar{s}) \cdot (s_i - \bar{s}) + y''\,(\bar{s}) \cdot (s_i - \bar{s})^2/2 + \ldots] \;/\; N, \tag{3}$$

in which $y'(s)$ and $y''(s)$ represent the first and second derivatives, respectively, of y with respect to s. By the definition of $\bar{s}$, the second term in this expansion, which is linear in $(s_i - \bar{s})$, will sum to zero. Eq. 3 is the discrete analog of the result presented by Allen (1) for continuous distributions.

As a first approximation, consider only the first two nonzero terms of Eq. 3 (1), so that:

$$\bar{y} = y(\bar{s}) + y''\,(\bar{s}) \cdot (MSD(s))/2\,, \tag{4}$$

where $MSD(s)$ is the mean square deviation from the mean disease severity defined by:

$$MSD\,(s) = [\sum_{i=1}^{L} (s_i - \bar{s})^2 \cdot n_i] \;/\; N. \tag{5}$$

$MSD(s)$ equals 0 when damage is homogeneous and $MSD(s)$ increases with increasing aggregation. When severity is bounded (e.g., the maximum value of fractional foliar damage is unity), there is also an upper bound on the value of $MSD(s)$. For later convenience, I define a coefficient of aggregation ϕ_A

$$\phi_A = MSD(s)/[\bar{s} \cdot (s(MAX) - \bar{s})], \tag{6}$$

where the denominator corresponds to the maximum possible $MSD(s)$. Using the above relation Eq. 4 can be rewritten as:

$$\bar{y} = y(\bar{s}) + y''(\bar{s}) \cdot \phi_A \cdot [\bar{s} \cdot (s(MAX) - \bar{s})]/2\,. \tag{4'}$$

The parameter MSD is by definition positive-definite and, thus, the effect of aggregation of damaged on the average yield is totally determined by the sign of the second derivative of the yield-severity relation. When the yield-severity relation is of Type I, the second term in both Eq. 4 and Eq. 4′ is positive and aggregation of damage increases yield. When the relation is of Type II, aggregation decreases yield, in qualitative agreement with Hughes (5). In addition, Eq. 4 provides a prediction for the magnitude of the effect of aggregation on yield in terms of the product of the second derivative of yield with respect to severity and the calculated value of $MSD(s)$.

THEORETICAL AND EMPIRICAL EXAMPLES

In general, Eq. 3 can be used to evaluate the effect of aggregation of damage on expected yield to any desired degree of accuracy. The number of terms that must be included in the expansion will depend on the particular case in question. Often, however, the approximation provided by Eq. 4 is sufficient, as the following examples will illustrate.

As a theoretical example, assume that the curvature of the yield-severity relation is constant, which is equivalent to an assumed quadratic relation between percent yield and severity of damage s. Letting the fractional yield equal one at $s = 0$ and zero at $s = 1$, gives:

$$y(s) = 1 - s - B \cdot s \cdot (1 - s), \tag{7}$$

where B is a parameter related to the curvature of the yield versus damage relation. Eq. 7 can be combined with Eq. 4′ to yield:

$$\bar{y} = 1 - \bar{s} - B \cdot (1 - \phi_A) \cdot \bar{s} \cdot (1 - \bar{s}). \tag{8}$$

The variables characteristic of a field with nonhomogeneously distributed damage ($\bar{y}$, $\bar{s}$, and ϕ_A) can be directly measured and, thus, the parameter B can be obtained by nonlinear regression if there are several values for these variables. In general, aggregation of damage reduces the curvature of the yield-severity relation and both the Type I ($B = 1$) and Type II ($B = -1$) relations give the same yield prediction when $\phi_A = 1$ (Fig. 1).

An example of yield loss for potato plants defoliated by late blight. Often severity of attack is obtained, in a field plot, as the mean of many observations over time and space, while the final yield is bulked. If in addition to the mean damage, the variation about this mean is also taken into account, some estimate of the effect of aggregation on the bulk yield may be obtained. As an example of this procedure, the dry tuber yield from 180 1.5- × 1.5-m quadrats from a potato field are plotted (open circles) against mean fractional defoliation caused by late blight, D, in

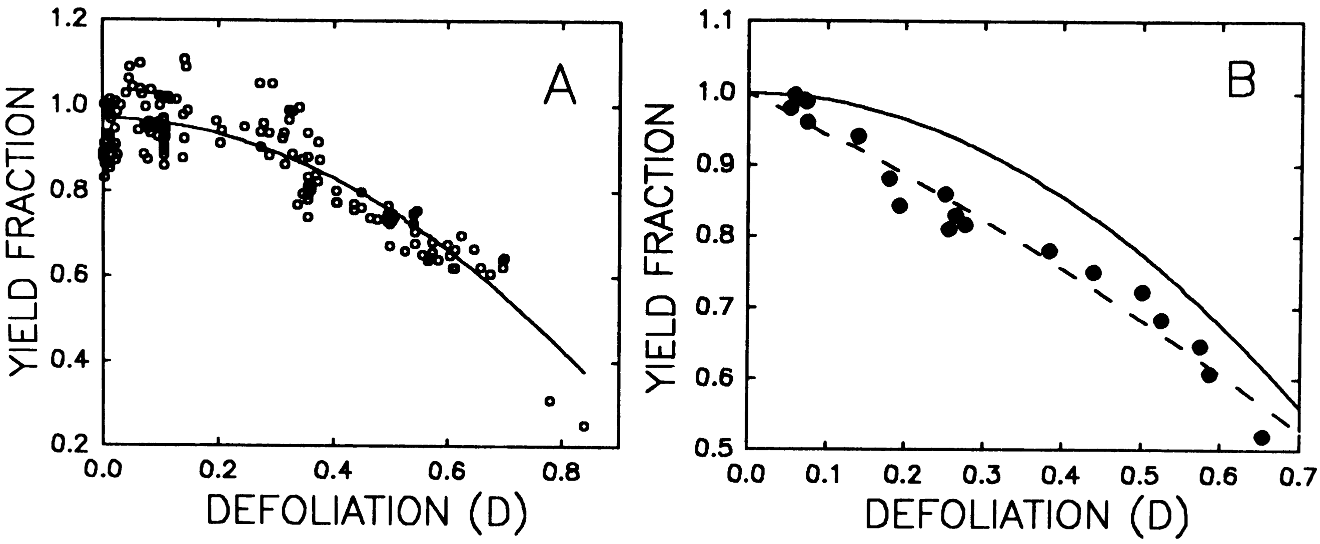

Fig. 2. Fractional dry tuber yield versus fractional defoliation of potato plants due to late blight for 1.5- × 1.5-m quadrats (A) (open circles) and 4.5- × 4.5-m bulked quadrats (B) (closed circles). Also shown are yield-severity predictions (Eq. 9) resulting from regressions I and II of Table 1 (solid line and dashed line, respectively).

Figure 2A (Ferrandino, *unpublished*). D is defined by the equation:

$$D = 1 - HAD/HAD(MAX),$$

where HAD is the healthy area duration (12) and $HAD(MAX)$ is the maximum observed value of HAD. This data set was fit to a two-parameter equation analogous to Eq. 7:

$$\bar{y}(s) = 1 - A \cdot D + B \cdot D^2. \quad (9)$$

The nonlinear regression fit to Eq. 9 (solid line) is also plotted in Figure 2A and B and the parameters for the fit are shown in Table 1 (Regression I). To simulate bulking, the mean yield y from 20 blocks containing nine contiguous quadrats from the same data set are plotted in Figure 2B (solid circles). The bulked data were also fitted by nonlinear regression to Eq. 9 (Table 1: Regression II) and the fit is plotted in Figure 2B (dashed line). Because of the aggregation of damage, bulking resulted in a lessening of curvature for the yield versus severity relation (Fig. 2B) and a corresponding decrease in the absolute value of the parameter B in the regression (Table 1). The effects of aggregation of damage on the bulked yields can be estimated by combining Eqs. 4 and 9 to yield:

$$\bar{y} = 1 - A \cdot D + B \cdot (D^2 + MSD(D)), \quad (10)$$

where D and $MSD(D)$ are defined in analogy with the expressions for s (Eqs. 1 and 5, respectively). Eq. 10 can be rewritten in terms of ϕ_A (Eq. 6):

$$\bar{y} = 1 - (A - \phi_A \cdot B) \cdot D + B \cdot (1 - \phi_A) \cdot D^2, \quad (11)$$

TABLE 1. Results of nonlinear regression using Eq. 9 for the late blight data shown in Fig. 2

	Data form	df[a]	Parameters[b] A	B	r^2[c]
I	Single quadrat	178	0.02 ± 0.02	−0.82 ± 0.03	0.74
II	Bulked (9 quadrats)	18	0.32 ± 0.02	−0.48 ± 0.04	0.95

[a] Degrees of freedom in the regression fit.
[b] Values shown are the estimated regression parameters ± the standard error for Eq. 9.
[c] Coefficient of determination.

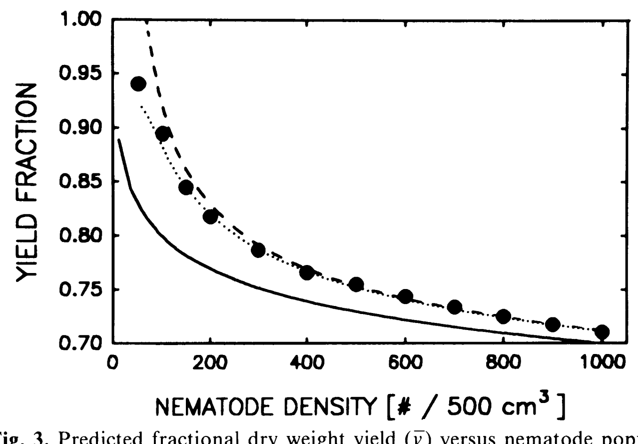

Fig. 3. Predicted fractional dry weight yield ($\bar{y}$) versus nematode population density [m; number/(500 cm^3)] for tobacco plants infested with nematodes. Closed circles: calculations of Noe and Barker (1985); Solid line: homogeneously distributed damage [$y(m) = 1 - 0.1 \log 10(m)$]; Dashed line: first two terms of Eq. 13; Dotted line: Eq. 13 including the third order term.

where $s(MAX)$ has been set equal to 1. For $\phi_A = 0$, damage is distributed homogeneously and Eq. 11 becomes identical in form to Eq. 9. As ϕ_A increases, the contribution of the D^2 term in Eq. 11 decreases and the yield-severity relation becomes more nearly linear.

The yield predictions of Noe and Barker (9) for tobacco are shown as a function of mean nematode density in Figure 3. They assumed that the normalized yield y was a linearly decreasing function of the log of the mean nematode density m expressed as the number of nematodes per 500 cubic centimeters of soil [$y(m) = 1 - 0.1 \log 10(m)$]. They also assumed that the population distribution was given by a negative binomial with the k parameter set equal to $0.0018 \cdot m$. For this case, $y''(\bar{s})/2 = (46.06 \cdot m^2) - 1$ and the variance of the negative binomial distribution is given by $\sigma \cdot m^2 = m(m + k)/k$ (2) and is equal to $MSD(m)$. For this example, the general formula (Eq. 3) can be written as:

$$y = y\,(m) + \frac{(m+k)}{46.06 \cdot k \cdot m} - \frac{(m+k) \cdot (2m+k)}{138.18 \cdot k^2 \cdot m^2} + \ldots \quad (12)$$

which can be simplified with the relation $k = 0.0018 \cdot m$ (8) to yield:

$$\bar{y} = y(m) + 12.08/m - 4{,}479/m^2 + \ldots \quad (13)$$

Once again the estimate provided by Eq. 4, which includes only the first two terms of Eqs. 12 and 13, is reasonably good over most of the range of mean nematode densities (Fig. 3). For $m < 250$, the last included term in Eqs. 12 and 13, which is third order in deviations from the mean population density, becomes important (Fig. 3).

The estimation of yield using Eq. 3 involves one or two calculations depending on whether or not the third order term is included. By way of contrast, the results of the Noe and Barker study involved a summation over as many as 6,000 terms for each individual yield estimation.

DISCUSSION

To implement Eq. 3 to estimate yield one needs to know the functional form, $y(s)$, of the yield-severity relation for a single plant. In addition, this relation must be well behaved so that the second and higher order derivatives with respect to s can be estimated. The simplification of Eq. 3 to second order (Eq. 4) is contingent on the additional restriction that the higher order terms in the expansion remain small. The method can be applied to more general forms of the yield-severity relation; however, each application must be properly handled to make sure that singularities and other complications are taken into account. For example, the model can be applied to the yield-severity relation proposed by Madden et al (7), as long as samples are restricted to cases for which damage exceeds the threshold level, below which no yield loss occurs.

The foregoing method of using the observed variance of severity of damage in Eq. 4 to estimate yield when plant damage is spatially aggregated, is mathematically simple and in many cases sufficiently accurate. The method also has the advantage of directly using data as opposed to using a fitted population distribution function to describe the variability of damage. Despite several restrictions on the applicability of the method, it should be useful for estimating the effect of spatial aggregation of plant damage due to disease, insects or nematodes on final yield.

LITERATURE CITED

1. Allen, J. C. 1988. Averaging functions in a variable environment: A second-order approximation method. Environ. Entomol. 17:621-625.
2. Barker, K. R., Todd, F. A., Shane, W. W., and Nelson, L. A. 1981. Interrelationships of *Meloidogyne* species with flue-cured tobacco. J. Nematol. 13:67-78.
3. Fisher, L. 1974. Probability and statistics. Pages 1181-1236 in: Handbook of Applied Mathematics. C. E. Pearson, ed. Van Nostrand

Reinhold, New York.
4. Hughes, G. 1988. Models of crop growth. Nature 332:16.
5. Hughes, G. 1988. Spatial heterogeneity in crop loss assessment models. Phytopathology 78:883-884.
6. Johnson, K. B., 1987. Defoliation, disease, and growth: A reply. Phytopathology 77:1495-1497.
7. Madden, L. P., Pennypacker, S. P., Antle, C. E., and Kingsolver, C. H. 1981. A loss model for crops. Phytopathology 71:685-689.
8. Mumford, J. D., and Norton, G. A. 1987. Economics of integrated pest control. Pages 191-200 in: Crop Loss Assessment and Pest Management. P. S. Teng, ed. The American Phytopathological Society, St. Paul, MN.
9. Noe, J. P., and Barker, K. R. 1985. Overestimation of yield loss of tobacco caused by the aggregated spatial pattern of *Meloidogyne incognita*. J. Nematol. 17:245-251.
10. Romig, R. W., and Calpouzos, L. 1970. The relationship between stem rust and loss in yield of spring wheat. Phytopathology 60:1801-1805.
11. Waggoner, P. E., and Rich, S. 1981. Lesion distribution, multiple infection, and the logistic increase in plant disease. Proc. Natl. Acad. Sci. USA 78:3292-3295.
12. Waggoner, P. E., and Berger, R. D. 1987. Defoliation, disease, and growth. Phytopathology 77:393-398.