

Probability Range in Damage Predictions as Related to Sampling Decisions¹

H. FERRIS²

Abstract: The risk involved in basing a nematode management decision on predicted crop loss is related to the uncertainty in the crop damage function and error in measuring nematode population density. The sampling intensity necessary to measure a nematode population with specified precision varies with population density. Since the density is unknown prior to sampling, optimum sampling intensity for a management decision is calculated for the economic threshold population level associated with the management cost. Population densities below the threshold are measured with greater precision than required; those above the threshold are less precisely measured, but invoke management. The approach described provides resolution to sampling strategies and allows assessment of the risk associated with the management decision.

Key words: crop loss, management decisions, sampling intensity, risk analysis, economic thresholds.

Variability in the spatial distribution of nematodes, error in population density estimates, and the resultant uncertainties associated with nematode damage functions have been examined elsewhere (4). I discussed earlier how the probability level associated with the damage estimate is a function of two probability levels, that of the damage function confidence interval and that of the population estimate (4). These uncertainties result from variability in nematode distribution, biology, host-parasite interactions, and seasonal effects. Here I explore the problem of determining the optimum sample number to arrive at a damage prediction with a specified probability level. This paper involves the development of conceptual tools and provides a forum for their exploration. The validity of the tools depends upon the underlying perceptions and conclusions, which are presented to allow objective evaluation.

Rationale: The first requirement for nematode damage prediction is an appropriate damage function. Confidence bands about a damage function can be calculated from the cumulative deviation from the function of observations on which the function is based (12). The damage function and confidence bands used in this paper can be summarized by the relationship $y =$

$m + (1 - m)z^{P-T} \pm t_{\alpha} \cdot s_y$, which is adapted from the basic damage function of Seinhorst (11). In this relationship, y is the relative yield, m is a minimum relative yield at high nematode population densities, P is the nematode population density, T is the tolerance level below which damage is not observed, and z is a slope parameter of the damage function. Further, $s_y = s_{y \cdot x} \sqrt{(1/n + x^2/\Sigma x^2)}$ where x is the deviation of each predictive independent variable from the mean of such variables, Σx^2 is the sum of all such deviations, and t_{α} is Student's t value for the required probability level. The value $s_{y \cdot x}$ is $(\Sigma d^2_{y \cdot x})/(n - 2)$ where d is the deviation of an observed and predicted point and n is the number of points on which the model is based (12). The α value is the proportion of predictions which will not fall within the confidence bands (Fig. 1). Some questions may be raised as to the validity of calculating the confidence bands using techniques essentially derived from linear regression approaches; however, the principles are not affected by the mechanics of the algebra.

The second requirement for the nematode damage prediction is an estimate of the nematode population density for which the prediction is to be made. This estimate will have an associated error component (4,7,10). Determination of a confidence interval for the population estimate requires knowledge of the standard error of this estimate. Since nematode population distributions can usually be described by the negative binomial distribution (6,9), it is possible to calculate the expected standard

Received for publication 2 September 1983.

¹ This study was supported in part by USDA CSRS Grant 59-2064-1-1-748-0.

² Nematologist, Department of Nematology, University of California, Riverside, CA 92521.

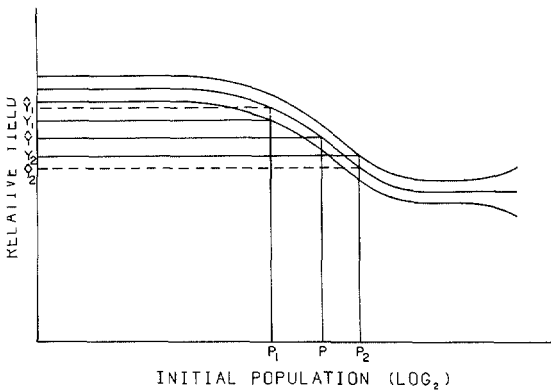


FIG. 1. Relationship between yield and nematode density indicating confidence bands associated with the damage function, economic decision value (\hat{Y}), economic threshold population (P), predictive population interval (P_1 to P_2), and economic decision interval (Y_1 to Y_2).

TABLE 1. Taylor's power function parameters relating variance to mean for populations of *Meloidogyne incognita* in small plot studies.

Data set #	No. of observations	a*	b	r ²
1	195	1.0	1.732	0.79
2	101	1.0	1.728	0.94
3	106	1.0	1.790	0.94
4	54	1.0	1.883	0.98
5	110	1.0	1.765	0.81
6	237	1.0	1.740	0.78
7	88	1.0	1.731	0.72
8	74	1.0	1.860	0.91
\bar{x}		1.0	1.779	

* Parameter forced to 1.0 in estimation process.

error if the mean (μ) and dispersion parameter (k) of the distribution are known (2). A problem with this approach is that the dispersion parameter (k) is defined as a function of the mean density and of the variance of the population, which must be known in advance. Since both mean and variance vary with time, prediction of k may be difficult. Another estimator of the error associated with a population estimate can be obtained through Taylor's power law (13,14) which states that in aggregated populations the variance of the population estimate is an exponential function of the mean. That is, $s^2 = ax^b$ where s^2 is the variance, x is the mean population estimate, a is a coefficient influenced by sample size (2,13,14), and b is a species-specific index of aggregation (2,16). This variance to mean relationship can be linearized through log transformation: $\log s^2 = \log a + b \log x$, allowing determination of the values a and b by linear regression. At small sample numbers the variance to mean ratio may be variable; however, it stabilizes rapidly as the number of samples used to calculate the variance and mean is increased. The stability of the parameters of Taylor's power law is illustrated by values of b obtained from a series of data sets collected from small plots (Table 1). The a value in these observations was close to 1, so it was held constant at 1 by constrain-

ing the regression line through the origin for the log transformed data. This allows more clear assessment of variability associated with the estimate of b .

Since both methods discussed for estimating the variance are functions of the population mean, prediction of the variance before sampling involves prior knowledge of this mean. Basic to the applied use of nematode damage functions is the requirement for predicting expected yield or yield loss for a crop. The nematode sampling process is undertaken to estimate this yield loss, and determination of the nematode population density is merely an intermediate step in this process. Further, basic economics requires that a particular management procedure not be invoked unless the value of the loss to the nematodes is greater than the cost of the management procedure (3). Hence, in an applied sense, the use of damage functions requires that the population be sampled with sufficient intensity to measure it with an acceptable probability if it is at a density which will invoke the management decision. This economic threshold population density (P) can be determined by solving the damage function $y = m + (1 - m)z^{P-T}$ for P ,

$$P = ((\log(y - m)) - \log(1 - m)) / (\log z) - T \quad (i)$$

The variance associated with estimates at this population density can be calculated

from Taylor's power law ($s^2 = aP^b$) or from the negative binomial if k is known.

Karandinos (8) rederived some published formulae for the number of samples (n) required to provide estimates of the density of an aggregated population within a specified range of the true mean with a specified confidence level:

$$n = Z_{\alpha/2}^2 \left(\frac{1}{\mu} + \frac{1}{k} \right) / D^2$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ point for the standard normal variate, μ is the population mean, k is the dispersion parameter of the negative binomial distribution, and D is half the length of the acceptable confidence interval expressed as a proportion of the mean. This estimate for n was derived from the general formula $n = (Z_{\alpha/2}/D)^2 \cdot \sigma^2 / \mu^2$ where σ^2 is the variance. The standard variate ($Z_{\alpha/2}$) can be replaced by Student's t variate and σ^2 substituted by its estimator $a\bar{x}^b$ (15) to yield an equation (16) for the number of samples necessary to predict \bar{x} :

$$n = (t_{\alpha/2}/D)^2 \cdot a \cdot \bar{x}^{(b-2)} \quad (\text{ii})$$

A characteristic of an aggregated population is that the variance is greater than the mean (2). From Taylor's power law, when $a = 1$ and $b = 1$ the variance is equal to the mean and the population distribution would be random or Poisson. When $a = 1$ and $b > 1$ the variance is greater than the mean and will increase at a greater rate than the mean as the mean increases. Consequently, if the measured population is actually lower than the economic threshold population, it will have been measured with greater precision than was required for the decision. If it is greater than the economic threshold, it will have been measured with less than the required precision but will invoke the management decision.

It is useful to explore the contributions of the various components of the sampling precision equation (ii) to the sampling intensity determined. The component $a\bar{x}^{(b-2)}$ has a relatively small contribution (5 or less) at intermediate population densities and when $a = 1$ and $b \approx 2.0$. The probability component ($t_{\alpha/2}$) varies with sample number, with the probability level,

and with the number of samples, but is usually around 2.0 so that its squared component approximates 4.0. The acceptable deviation about the population estimate (D), however, may have a major impact on the required sampling intensity. Thus, if it is desired to estimate the population density within 10%, as perhaps in a population or biomass study, the impact of this component squared is to increase the required sampling intensity 100-fold ($1/D^2 = 100$). For crop loss prediction, however, it is desirable to predict the expected loss within a certain range (say 10%). Since the damage per nematode decreases with increasing nematode population density (11) the population interval associated with a 10% damage range reflects a value of D much greater than 10% in relation to the population mean. Hence, the impact of the magnitude of D is minimized by evaluating sampling intensity based upon an acceptable range in the damage prediction.

Optimum sampling intensity: Through the outlined logic, the number of samples required to estimate a population density with a specified level of precision will vary directly with the density. Conversely, the precision associated with a population estimate at a specified sampling intensity will decrease with increasing density. When sampling a field to determine the nematode population density for management decisions, the population density and its variance are not known in advance. The minimum sampling intensity is that number of samples (n) which would measure the population with acceptable precision if the population density were at the economic threshold level (P). Thus, from equations (i) and (ii):

$$n = (t_{\alpha/2}/D)^2 \cdot a \cdot P^{(b-2)} \quad (\text{iii})$$

Since the optimum sampling intensity depends upon the economic threshold population, it is specific to the cost of the management alternative and the nature of the damage function for the nematode species. Thus, the sampling intensity varies with the magnitude of the investment contemplated.

Relationship between confidence range of prediction and number of samples: For risk analysis of the management decision, it is necessary to know the number of samples which will place the yield prediction within

a specified interval. If a management decision is to be made at a predicted yield (\hat{y}) (Fig. 1), the number of samples should be determined to place the yield prediction within a small interval around \hat{y} ; for example, $\hat{y} \pm 10\%$. Projecting that interval from the y axis indicates that the population must be estimated in the range P_1 to P_2 . The probability associated with this estimate will be a function of the $t_{\alpha 1}$ value associated with the confidence bands on the damage function and the $t_{\alpha 2}$ value in the equation yielding the number of samples required to estimate the population density within a certain range. The probability levels for the damage function bands and for the population estimate must be larger than required for the yield interval estimate, or solution of the sampling intensity equation is impossible.

The upper bound of the damage interval is projected onto the lower confidence band of the damage function, while the lower bound of the interval is projected onto the upper confidence band. This procedure may be verified by widening the damage interval, resulting in a corresponding increase in the population estimate interval. If the damage interval bounds are projected onto the opposite confidence bands, widening of the damage interval decreases the population estimate interval. Note that in the extreme case, if absolute prediction of the damage level is required and the damage function is known with perfection, a single line would be projected from the y -axis to the damage function and onto the x -axis requiring measurement of the population without deviation ($D = 0$). From equation (iii) this would give an infinite number of samples. Projections from the y -axis are constrained by the damage function. If the upper bound of the damage interval is too high to project onto the lower confidence band of the damage function, the appropriate sampling intensity is determined from the lower bound projection. Similarly, if the lower bound projection is too low to intersect the upper confidence band, the appropriate sampling intensity is determined from the upper band projection. An approximation for the estimated projection of the missing bound may be provided by balancing the population interval about the economic threshold population on a logarithmic scale,

whereby $\log P_2 = \log P + (\log P - \log P_1)$. If the economic decision value (that crop value caused by the economic threshold population density) is less than the value attributed to the minimum yield (m), the economic threshold is infinite and sampling is unnecessary; the management option is rejected since the investment cannot be recouped from yield improvement.

It is possible to calculate P_1 and P_2 from y_1 and y_2 (Fig. 1) as follows:

$$y_1 = \hat{y}_1 - t_{\alpha 1} s_{\hat{y}}$$

$$\hat{y}_1 = m + (1 - m)z^{P_1 - T} \quad \text{for } P_1 > T,$$

else $\hat{y}_1 = 1$

then, $y_1 = m + (1 - m)z^{P_1 - T} - t_{\alpha 1} s_{\hat{y}} \quad \text{for } P_1 > T,$

else $y_1 = 1 - t_{\alpha 1} s_{\hat{y}}$.

Solving for P_1 :

$$P_1 = ((\log((\hat{y}_1 - m + t_{\alpha 1} s_{\hat{y}}) / (1 - m))) / \log z) - T \quad \text{for } P_1 > T,$$

else $P_1 = 0$.

All terms in the latter expression are known except for

$$s_{\hat{y}} = s_{y \cdot x} \sqrt{\frac{1}{n} + \frac{(\bar{x} - P_1)^2}{\sum x^2}}$$

where \bar{x} is the mean population density and n the number of samples involved in derivation of the damage function. Since P_1 is embedded in this term, a direct solution for P_1 is impossible and an iterative approach is used. Initially P is substituted for P_1 and the equation is solved for P_1 . The new value of P_1 is substituted and the calculation repeated until the difference between repeated values of P_1 approaches zero. A similar process can be used to calculate P_2 .

The values obtained for P_1 and P_2 are used to generate $D = (P_2 - P_1)/P$, that is a density range expressed as a proportion of the mean. By substituting the value of D into equation (iii), the number of nematode density measurements necessary to establish the damage prediction within the interval $\hat{y} \pm 10\%$ is derived. The probability associated with this interval is a multiplicative function of the confidence level

TABLE 2. Number of samples necessary to predict specified economic decision intervals, with a probability of 0.90 of the prediction falling within the interval, for Taylor's population parameters $a = 1.0$, $b = 1.779$, and the damage function $y = 0.56 + 0.44 \times 0.997^{(P-20)}$. Potential crop value = \$100.

Management cost (\$)	% interval	Number of samples
10	0.0	4
20	0.0	5
30	0.0	5
10	5.0	3
20	5.0	4
30	5.0	4
10	10.0	3
20	10.0	3
30	10.0	3

associated with the damage function ($1 - \alpha_1$) and that associated with calculation of the number of samples ($1 - \alpha_2$).

The t_{α_2} value used in the probability level calculation is a function of the number of samples, which is not yet known. Once again, an iterative process may be used:

- (1) reformulate equation (iii) to predict $\hat{D} = \sqrt{(t_{\alpha_2}^2 \cdot a \cdot P^{(b-2)})/n}$;
- (2) solve this equation repeatedly for increasing values of n from $n = 1$;
- (3) when the difference between \hat{D} and D is at a minimum, the appropriate number of samples has been determined.

DISCUSSION

The rationale developed here provides the basis for a sampling decision model incorporating the variance of the sample estimate and the variability associated with the damage function. If the predicted crop loss is less than the economic decision interval, management is unnecessary; if it is above the economic decision interval, management is justified. If the predicted crop loss is within the decision interval, a subjective decision must be made based upon the grower's economic status or risk aversion/risk acceptance level. The boundaries of the economic decision interval are specified in the management decision process. It is possible to calculate the number of samples necessary to provide a specified range of an economic decision value at a

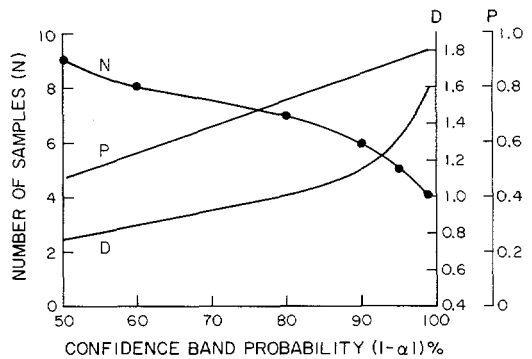


FIG. 2. Influence of damage function confidence band probability levels on the number of samples (N) necessary to predict an economic decision interval of fixed size, the associated probability (P) of the nematode density confidence interval, and the proportional nematode density interval (D) being identified.

known probability level, provided that parameters of the damage function and nematode distribution are known (Table 2). The required number of samples decreases as the economic decision value approaches minimum yield. Here the predictive population interval is greatest, decreasing the required sampling intensity.

The following examples are included to illustrate the dynamic influence and interdependent nature of the parameters of the population distribution and of the damage function in determining sampling intensity and predictive precision relative to economic thresholds and decision values. The probability (P) associated with the economic decision interval increases as confidence bands on the damage function are widened by raising the associated probability level. Since the effect is to widen the proportional predictive population interval (D), the economic decision interval can be determined with fewer soil samples (Fig. 2). The number of samples necessary to achieve a specified interval about an economic decision value is influenced by the magnitude of the economic decision value because the slope of the damage function is variable (Fig. 3). The influence of Taylor's b value (or other dispersion parameter) is to increase the number of samples for a required economic decision interval as aggregation of the population increases (Fig. 3).

As the complexities of nematode soil sampling for management decision pur-

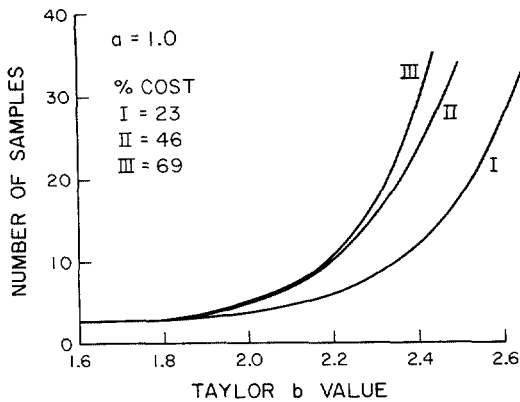


FIG. 3. Influence of Taylor's b value on the number of samples necessary to measure a 10% economic decision interval for economic decision values 23, 46, and 69% of the potential value loss due to nematodes.

poses are explored, a theoretical basis for determining sampling intensity and the associated precision is evolving. It is emphasized that data in this paper are used primarily to provide examples; however, they originate from field studies by the author. The parameters of Taylor's power function were derived from small-plot studies and are expected to be larger in field-scale investigations. The concepts presented here provide a basis for computer-delivered predictions of damage and for analysis of the risk associated with their acceptance. They do not replace or detract from the already well-established principles for sampling for nematodes and other soil organisms (1,5), but merely provide resolution to those principles. It is still necessary to stratify fields according to edaphic and cultural conditions, but further necessary to determine damage function and population distribution parameters for the conditions of individual strata. The recommended maximum stratum size of 5–10 acres should probably remain unchanged, and distribution parameters should be measured accordingly.

LITERATURE CITED

1. Barker, K. R., and C. L. Campbell. 1981. Sampling nematode populations. Pp. 451–474 in B. M. Zuckerman and R. A. Rohde, eds. *Plant parasitic nematodes*, vol. 3. New York: Academic Press.
2. Elliott, J. M. 1977. Some methods for statistical analysis of samples of benthic invertebrates. *Freshwater Biology Association Scientific Publication No.* 25.
3. Ferris, H. 1978. Nematode economic thresholds: Derivation, requirements, and theoretical considerations. *Journal of Nematology* 10:341–350.
4. Ferris, H. 1984. Nematode damage functions: The problems of experimental and sampling error. *Journal of Nematology* 16:1–9.
5. Ferris, H., P. B. Goodell, and M. V. McKenry. 1981. Sampling for nematodes. *California Agriculture* 35:13–15.
6. Goodell, P. B., and H. Ferris. 1980. Plant-parasitic nematode distributions in an alfalfa field. *Journal of Nematology* 12:136–141.
7. Goodell, P. B., and H. Ferris. 1981. Sample optimization for five plant-parasitic nematodes in an alfalfa field. *Journal of Nematology* 13:304–313.
8. Karandinos, M. G. 1976. Optimum sample size and comments on some published formulae. *Bulletin of the Entomological Society of America* 22:417–421.
9. McSorley, R. 1982. Simulated sampling strategies for nematode distribution according to a negative binomial model. *Journal of Nematology* 14:517–522.
10. McSorley, R., and J. L. Parrado. 1982. Estimating relative error in nematode numbers from single soil samples composed of multiple cores. *Journal of Nematology* 14:522–529.
11. Seinhorst, J. W. 1965. The relation between nematode density and damage to plants. *Nematologia* 11:137–154.
12. Snedecor, G. W., and W. G. Cochran. 1967. *Statistical methods*. Ames: Iowa State University Press.
13. Taylor, L. R. 1961. Aggregation, variance and the mean. *Nature* 189:732–735.
14. Taylor, L. R. 1971. Aggregation as a species characteristic. Pp. 357–372 in G. P. Patil, E. C. Pielou, and W. E. Waters, eds. *Statistical ecology*, vol. 1. University Park: Pennsylvania State University Press.
15. Taylor, L. R., I. P. Woiwood, and J. N. Perry. 1978. The density-dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology* 47:383–406.
16. Wilson, L. T., and P. M. Room. 1982. The relative efficiency and reliability of three methods for sampling arthropods in Australian cotton fields. *Journal of the Australian Entomological Society* 21:175–181.