JOURNAL OF NEMATOLOGY
VOLUME 16 JANUARY 1984 NUMBER 1

Journal of Nematology 16(1):1–9. 1984.

Nematode Damage Functions: The Problems of Experimental and Sampling Error[1]

H. FERRIS[2]

Abstract: The development and use of pest damage functions involves measurement and experimental errors associated with cultural, environmental, and distributional factors. Damage predictions are more valuable if considered with associated probability. Collapsing population densities into a geometric series of population classes allows a pseudo-replication removal of experimental and sampling error in damage function development. Recognition of the nature of sampling error for aggregated populations allows assessment of probability associated with the population estimate. The product of the probabilities incorporated in the damage function and in the population estimate provides a basis for risk analysis of the yield loss prediction and the ensuing management decision.

Key words: crop loss, management decisions, confidence intervals, risk analysis, economic thresholds.

A knowledge of the relationship between crop yield or value and numbers of plant-parasitic nematodes is fundamental to the application of quantitative rationale to nematode management decisions. A theoretical basis for damage functions has been developed (8), and the influence of environmental factors has been discussed extensively (2). The basic approach to developing nematode crop damage functions is to establish different nematode population densities, either in field plots or microplots, and to grow a crop at each density. Yield and yield loss can then be related to initial nematode population densities.

METHODOLOGY

In one method, the soil is fumigated and known numbers of nematodes are introduced to establish the range of initial population densities. The independent variable (population density) is controlled, and a replicated experiment can be conducted. Because the initial population densities are known, there is essentially no sampling error in their measurement. Similarly, since the experiment can be replicated, it is possible to measure and remove experimental error in yield by averaging across replicates for each initial population density level. A disadvantage of this approach, however, is the uncertainty involved in equating initial population densities achieved by inoculation to field-measured population densities. The uncertainty includes concern about inoculum viability, spatial distribution, and population age structure. Furthermore, the preparation of such experiments is costly.

Another approach is to manipulate nematode population densities in a field site by multiple cropping and rotation patterns, or with pesticides. Such measures achieve a more natural population distribution and age structure. However, the initial population density must be assessed by a sampling process, hence another source of variation is introduced into the data set. Because of the sampling error involved in measuring the population, the independent variable is not precisely controlled. Consequently, both the nematode population and the crop yield are subject to sources of error (Fig. 1). The yield estimate includes experimental error generated through location effects, variable culture and cultivation of the crop within the field, and measurement error. The nematode

---

Received for publication 31 January 1983.

[1] This study was supported in part by USDA-SEA Grant 70-59-2064-1-2-207-2 and USDA CSRS Grant 59-2064-1-1-748-0. These thoughts were stimulated by correspondence with Dr. Peter M. Burrows, Experimental Statistics Unit, Clemson University, Clemson, SC. I thank Larry W. Duncan for much helpful discussion.

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

The JOURNAL OF NEMATOLOGY for October (15:475–654) was issued 16 December 1983.

population estimate includes sampling and assessment error. Since there is only one observation at each population density, we lose the ability to remove experimental error by averaging across replications at the same population density. Variability in the data set may mask relationships and result in low r^2 values.

An analytical approach: This paper discusses approaches to minimizing the sampling and experimental error components of damage functions developed in field plots. The basic premise is that treatment averages from replicated experiments reflect minimal experimental error. Nematode density estimates from soil samples incorporate error reflective of the variance associated with their aggregated distribution. As with many pest species, the aggregated distribution of nematodes is frequently adequately described in mathematical terms by the negative binomial distribution (4,5). This distribution is defined by two parameters, the mean population density (μ) or its sample estimate ($\bar{x}$), and the index of dispersion (k). The probability of assessing a population density x in a sample is

$$P(x) = \left(1 + \frac{\mu}{k}\right)^{-k} \frac{(k + x - 1)!}{x!(k - 1)!} \left(\frac{\mu}{\mu + k}\right)^{x}.$$

A characteristic of such distributions is that the variance is greater than the mean, hence the variance and sampling error increase with population density. A sample providing a high population assessment reflects a mean population density with a large variance.

Seinhorst (8) adapted Nicholson's (6) competition theory to demonstrate that the damage per nematode is less at high than at low population densities. The assumption herein is that it is reasonable to partition the continuously varying set of independent variables from a damage-function data set into discrete groupings which could be considered replicates of each other. Densities within each class are considered essentially equal in effect on plant growth, but they reflect an associated unmeasured sampling error. Following the rationale of increased variance and decreased importance per nematode with increased population density, I select a convenient geometric series of density classes, the $\log_2$ series which is frequently used to display nematode damage functions. For each density class i, the mean density $\bar{x}_i$ is calculated for all population values between $2^{(i-0.5)}$ and $2^{(i+0.5)}$. Similarly, the mean yield, $\bar{y}_i$, is calculated for the corresponding yield measurements. Exceptions to this are that density class $i = 0$ consists of all values with $x = 0$, and class $i = 1$ consists of all values $0 < x < 2^{1.5}$. Note that the arithmetic range incorporated in each density class increases with x_i.

This approach reduces variability within a data set of paired observations. First, the variance associated with assessment of the nematode population density, a "horizontal variance," is stabilized by averaging values within population size classes which are biologically appropriate and justified. Each density estimate is considered a representative of one of these classes. Second, experimental error or "vertical variance" is stabilized by using the mean yield associated with all population estimates which fall within a density class as being representative of the influence of that class. By fitting a model to such a modified data set, we have reduced known sources of error and, hence, have more confidence in the goodness-of-fit assessment of the damage function. There is a fundamental problem with attempting to fit a model to the means of nematode density classes. Since the design of the experiment is not readily balanced, there will inevitably be different numbers of observations within each class. By using the means of each class, equal weight is given to one observation as to the mean of 20 observations. This situation is unsatisfactory, although the problem may be minimal if there is a large number of plots. An alternative is to include the density class mean and average yield in the data set as many times as there were observations in that range. Thus each observation is equally weighted for measurement of the goodness-of-fit of the model to the error-corrected data set.

The method described is especially effective with a large data set. Five experiments over 4 years have produced results on the response of tomatoes to varying levels of *Meloidogyne incognita* (Kofoid &

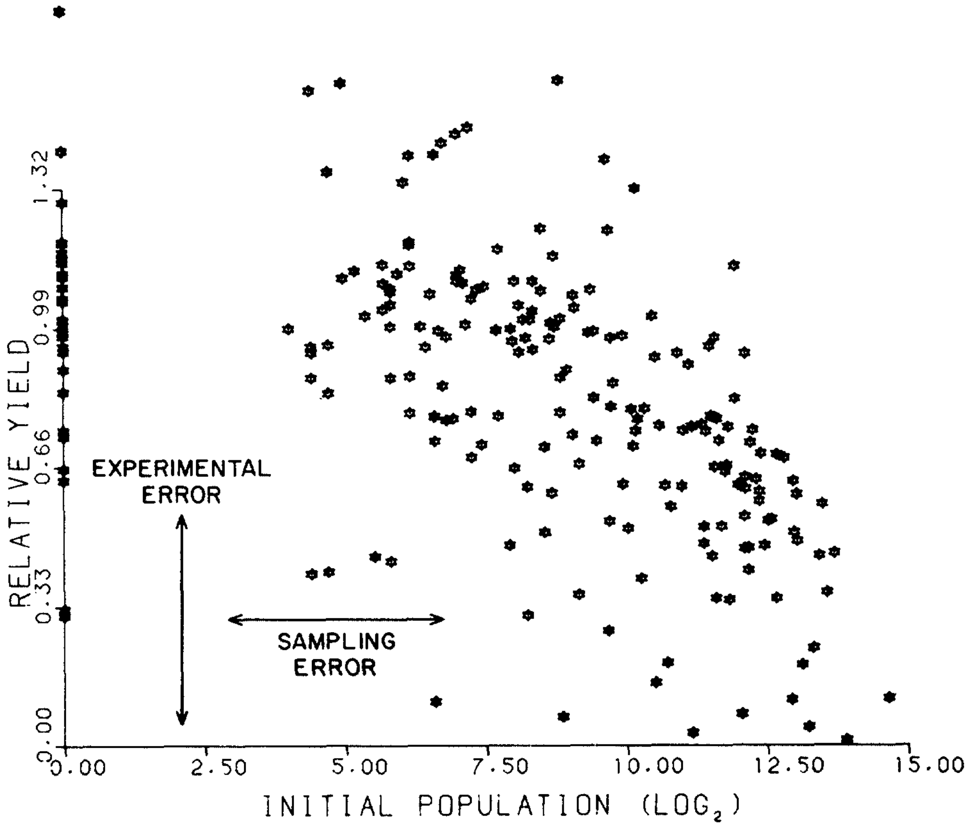

FIG. 1. Components of variation in relative yield measurements of tomatoes in response to varying population densities of *Meloidogyne incognita*. Four-year data set corrected for seasonal effects.

White) Chitwood infection (unpubl. data). Population densities of second-stage juveniles were measured prior to planting and corrected for extraction efficiency. If each year's data are standardized by dividing through by the maximum yield for that year (Fig. 1), a nematode damage function can be fitted to this multiple data set by the method described (Fig. 2). The high r^2 value obtained by grouping observations is produced at the expense of reducing the number of degrees of freedom to two less than the number of density classes. As a consequence of weighting the density class averages by the number of observations (n_i) constituting each average, the minimum yield estimate obtained using a curve-fitting alogrithm (3) may appear different from that represented by the data (Fig. 2). The n_i values are recorded above each point in the graph. Frequently, however, the best estimate of minimum yield may be a few observations at high nematode density, and the unweighted data set may better reflect this parameter (Fig. 3).

Variability in yield prediction: The yield value predicted by the damage function is a mean value for plants subjected to a particular nematode population density. In fact, there is a population of such yield estimates, normally distributed about the damage function for any nematode population density. Assume that an experiment is conducted to develop a damage function for a crop–nematode combination in a series of 100-m^2 plots in a field; assume also that there are several replicates of each nematode density class. For each density class, there will be a series of yield observations equal to the number of replications. The mean of these yield observations is the best estimate of the damage function value for the density class and reflects the removal of experimental error already discussed. If a whole field was divided into 100-m^2 plots, and if yields for each density

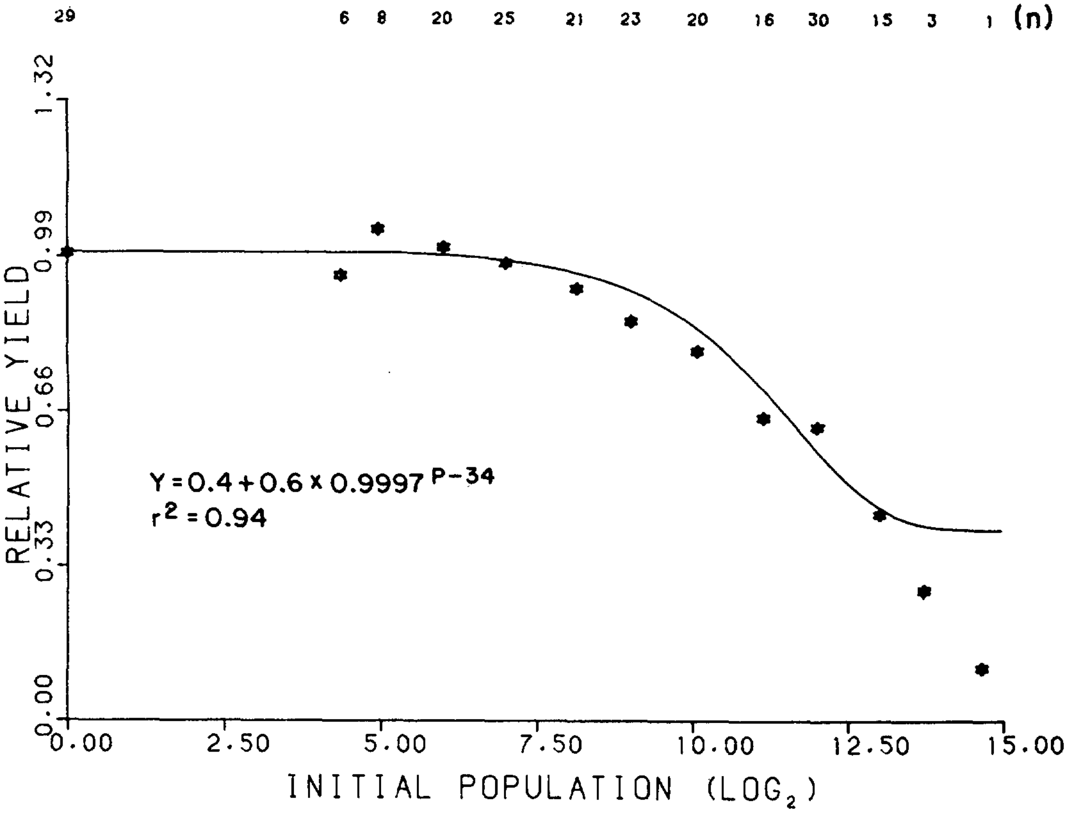

FIG. 2. Least-squares fit of the model $y = m + (1 - m)z^{P-T}$ to a 4-year data set of tomato yields averaged across *Meloidogyne incognita* density classes. Data set weighted for number of observations (n) per density class.

class from these plots were averaged together in groups equal to the number of replications in the experiment, a family of observed mean yields and densities would be obtained. The family of mean yields would be normally distributed around the damage function prediction, but more tightly grouped than the original plot observations on which they were based. The average of this family of mean yields would be the best possible estimate from that field of the yield at the average of the mean nematode population densities. The variation associated with the family of means is a useful measure of the expected variation in using the damage function as a predictor of yield in a whole field. If the damage function were to be used as a predictor of yield in an individual plot from that field, the expected variation would be much greater.

The proportion of yield estimates which will occur within a given distance from the predictive line for damage is calculable if the standard deviation among the family of estimates is known. Consequently, it is possible to apply confidence bands around the damage function which reflect the normal distribution of the means of individual estimates (Fig. 4). Confidence bands of this type are narrowest at the point reflecting the means of the population density and yield estimates. This is the region of the regression about which there is the most information. As the deviation of population estimates from this mean increases, so confidence in the prediction decreases. Actually, the confidence bands in Fig. 4 represent confidence limits on a prediction of the 4-year mean yield for a nematode population density. Predictions for individual years would have wider confidence bands for the same probability level. The confidence bands are calculated from the standard error of the y (yield) values pooled across all x (population) values, weighted

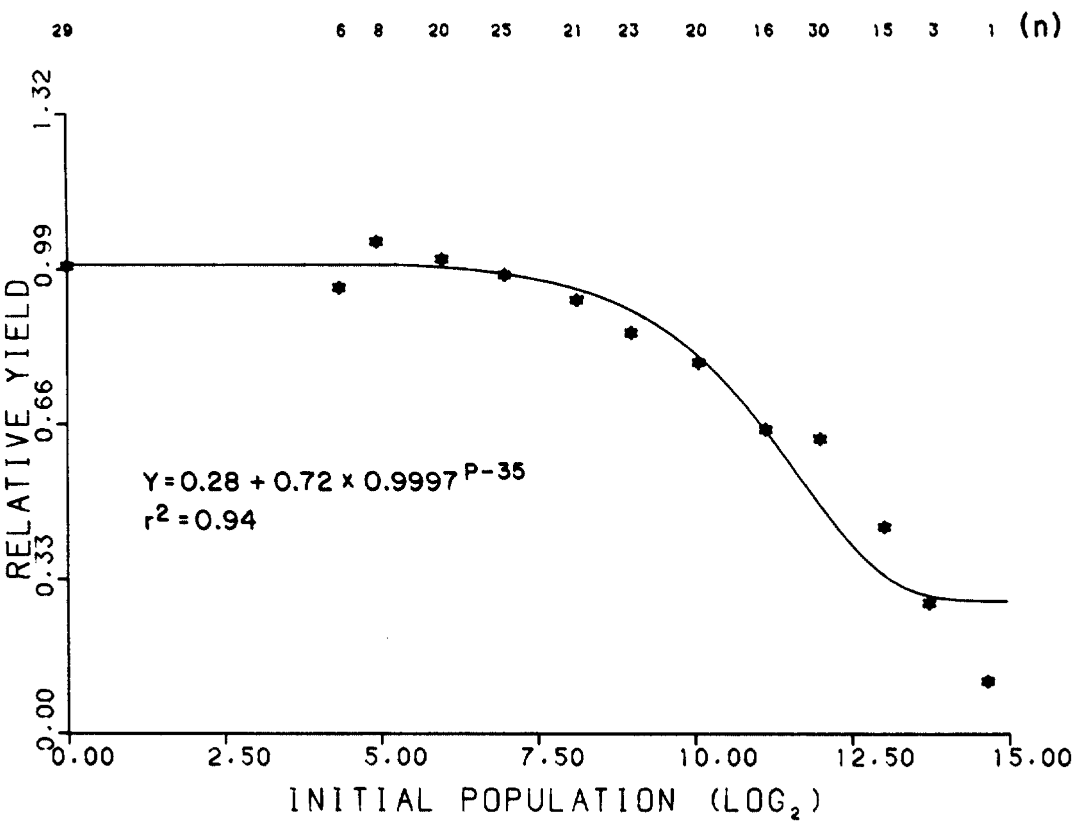

FIG. 3. Least-squares fit of the model $y = m + (1 - m)z^{P-T}$ to a 4-year data set of tomato yields averaged across *Meloidogyne incognita* density classes. Data set not weighted for number of observations (n) per class.

by the number of observations and distance from the mean population estimate for the damage function and by Student's t value at the chosen probability level. An extra term is incorporated in the estimator for predictions of individual plot yield values (10).

Variability in the sample estimate: Another problem in yield prediction using a damage function derived from an independently collected data set is variation in the nematode density estimate (sampling error) for which the prediction is being made. A family of possible population estimates is distributed about the mean sample density. Projection of the nematode density estimate confidence intervals on the damage function and its confidence intervals produces an expected yield interval (Fig. 5). Since the damage function is independently derived, the associated probability is the product of the respective probability levels associated with the confidence intervals of the damage function and the population estimate. If the nematode population is estimated to lie within a given range at $P = 0.5$, and the damage function confidence belt is generated at $P = 0.8$, the yield estimate interval is known at $P = 0.4$ (i.e., 0.5×0.8). As the probabilities associated with each of these independent assessments approach 100%, so the loss in precision associated with their multiplication is reduced. If the variances associated with the damage function and population estimate remain at their current size, the probability associated with the yield estimate can only be improved by widening the range of the confidence band. The only way to improve the precision of the yield estimate is to narrow the confidence bands on the damage function and population estimate without reducing probability.

The index of dispersion of the statistical distribution which describes the nematode distribution influences sampling error.

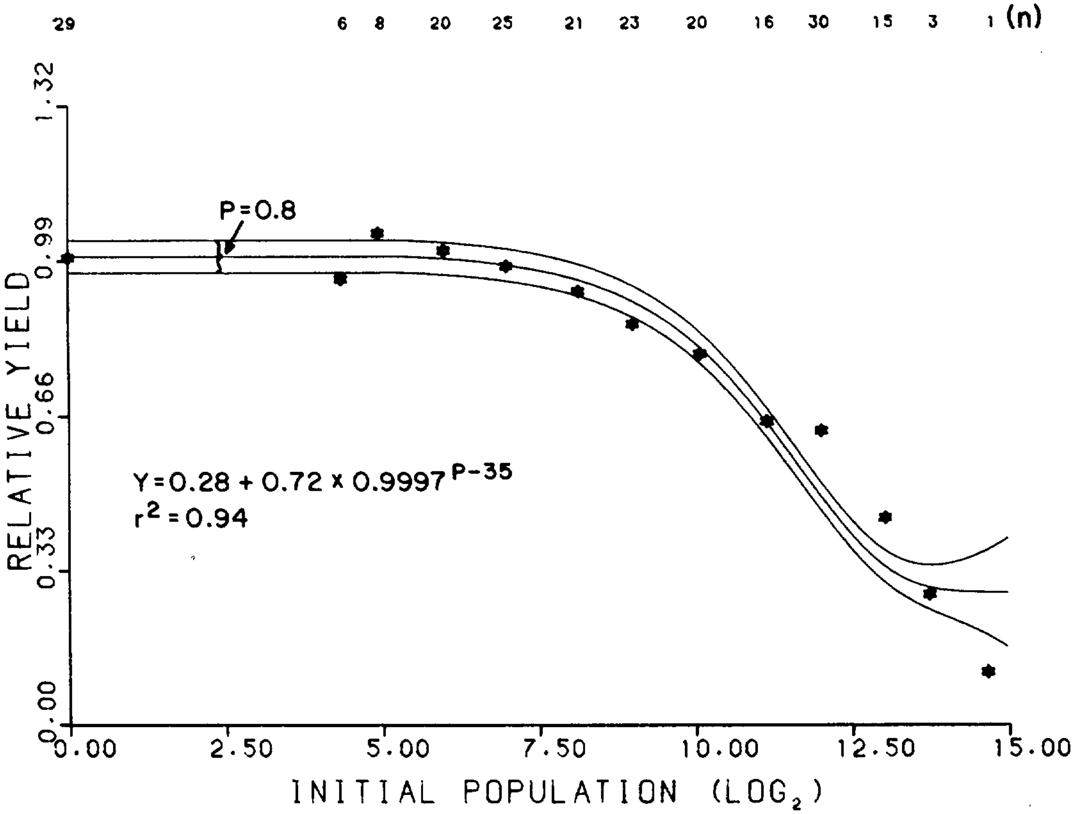

FIG. 4. Confidence interval (80%) for 4-year-average yields for tomatoes in response to *Meloidogyne incognita.*

Since such indices may change with time during and between growing seasons, time of sampling and sampling intensity affect sampling error. The magnitude of the error can be measured by taking repeated samples from the same area. Since the population is aggregated, the density classes are not symmetrical about the mean. Consequently, the confidence interval about the mean assessment is not symmetrical; the variance is higher on the upper side. An approximation is provided by log transformation to attempt to normalize the distribution (1). Then the mean of the transformed counts ($\bar{q}$) is $[\Sigma \log(x + 1)]/n$, and the transformed confidence limits are

$$\bar{q} \pm t\sqrt{(s^2_q)/n}$$

or in arithmetic terms:

$$2^{\bar{q}} \times 2^{t}(\sqrt{(s^2_q)/n}) \quad \text{and} \quad 2^{\bar{q}} \div 2^{t}(\sqrt{(s^2_q)/n})$$

where s^2 is the variance of the transformed x values and $\log_2$ is used. The convenience of using $\log_2$ is that the transformed confidence interval can be plotted on the same scale as the population density of the damage function (Fig. 5). The $\log_2$ scale is established simply as a geometric series of 2, or by the transformation

$$\log_2 x = (\log_{10} x)/(\log_{10} 2).$$

If, for example, a population density estimate of 1,000 is based on soil samples containing 600 and 1,400 nematodes, the respective $\log_2$ values are 9.97, 9.23, 10.45. The variance on the transformed scale is 0.778 and the *t* value at 1 df for 80% probability is 3.078. Hence, the 80% confidence interval on a $\log_2$ scale is 9.97 ± 1.92 (Fig. 5). In arithmetic terms, the limits are 265 and 3,795, obviously not centered on the arithmetic density mean of 1,000.

Influence of nematode distribution: The dispersion parameter of the negative binomial statistical description of nematode distribution influences the precision of nematode density estimates. The distri-

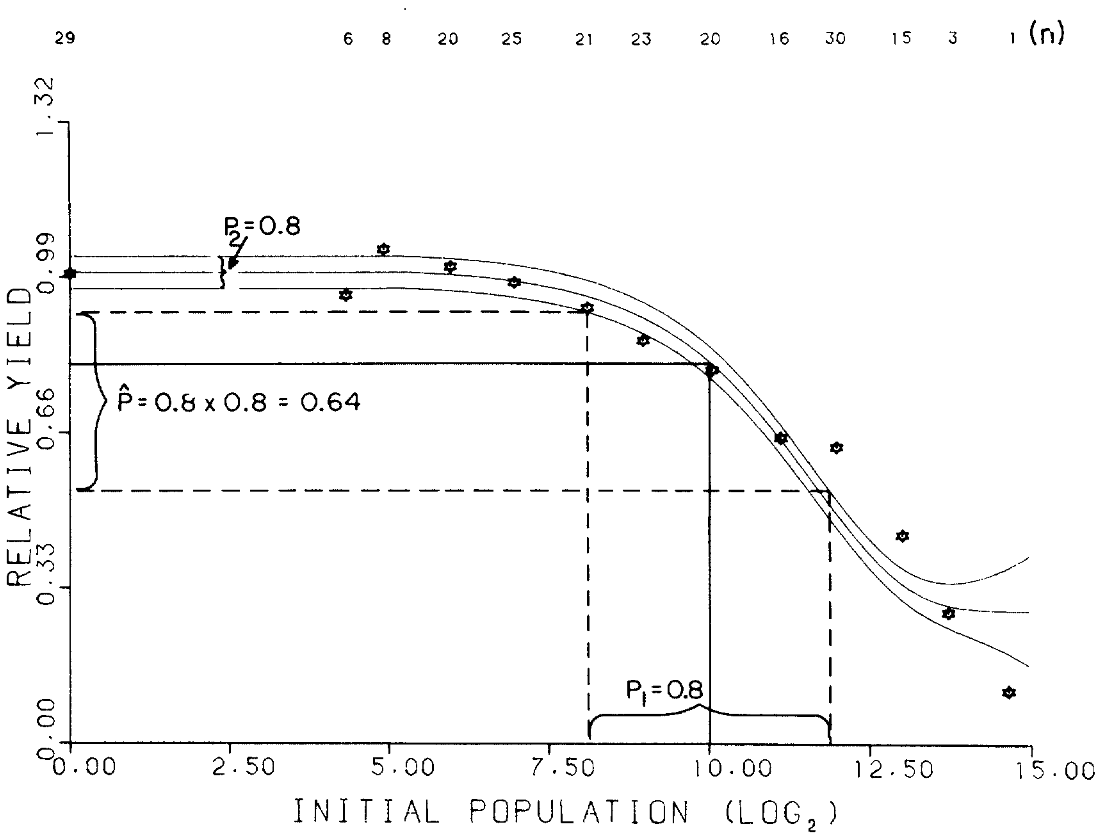

FIG. 5. Probability range for yield-loss prediction for tomatoes in response to *Meloidogyne incognita*. Estimate encompasses variability associated with sampling error and damage function determination.

bution is skewed, with a high frequency of low population density observations and a low frequency of extremely high population density observations. The applicability of this distribution is a consequence of the feeding habits, food distribution, and behavioral and reproductive patterns of nematodes and other organisms. The arithmetic mean—that is, the total number of organisms observed divided by the number of observational units (samples)—may seem intuitively to be an unsatisfying measure of population density for an aggregated population. The process of nematode sampling, where a sample consists of a composite of soil cores, is equivalent to obtaining an arithmetic mean estimate of the population. By compositing cores, individual population counts occurring in each core are added. Further, if the field is represented by more than one composite sample, standard practice would be to average across samples as a measure of the population density. Consequently, the estimate is the arithmetic mean of a negative binomial distribution, with a greater number of observations on one side of the mean than on the other.

The effect of using an average nematode population density is to underestimate yield (7,9). The descriptive parameters, μ and k, of the negative binomial distribution can be derived from sample assessments of the population providing estimates of the mean and variance (1). An interesting approach to the analysis of sampling efficiency is to simulate nematode distribution across a field and to predict total field yield. A field may be divided into a finite number of sample units; as an extreme example, say the total number of 2.5-cm-d soil cores which could be removed from the field. The parameters of distribution of the population, as described by the negative binomial, allow calculation of the proportion of cores in the field representing specific density classes of nematodes. If a nematode damage function for that crop and nematode combination is known, the yield loss incurred by the fraction of the crop growing in each core volume of soil can be calculated. By summing the remaining fractional yields across the whole field, the actual yield is determined. Computer simulation studies reveal slightly inflated damage estimates from the arithmetic mean of nematode population density. The magnitude of the inflation varies with the parameters of the damage function and of the distribution. The effect is to promote conservative crop management decisions, which should foster user confidence.

DISCUSSION

The variability associated with predictions of expected yield loss caused by plant-parasitic nematodes is an important component of the pest management decision process. Prediction of expected yield loss at a specified probability level allows risk analysis of the management decision. Further research and consideration of the variability associated with population assessment, and the impact of the descriptive parameters of nematode distribution, is necessary. It is interesting to speculate on the impact of an "old infestation" versus a "new infestation" on sampling error. A new infestation is defined as a new introduction or focus in a field, whereas an old infestation has been established sufficiently to express the full potential of its biological distribution. Unless there is an obvious reason for stratification of a new infestation field, sampling errors may be greater than expected. Vertical gradients in the dispersion parameters of population distribution probably occur. Sampling errors may vary with depth, underscoring the need for consideration of the location of the population likely to influence the proposed crop prior to preplant sampling. Measurement of a surface nematode population density for yield-loss prediction is meaningless if overwintering survival occurs at greater depths. Such scenarios can be simulated if relevant dispersion parameters are identified.

In multiple-species infestations, it is important to recognize that dispersion parameters are species specific. A single sampling plan is unlikely to minimize sampling error for all species, and confidence in predictions will vary among species. This complicates the assessment of yield-loss probability for multispecies models. The estimate with the lowest probability becomes the limiting factor if all species present have equal damage potential. Differences in pathogenicity and abundance of the individual species would confound the probability component of the yield-loss assessment.

The predictive capability of damage functions for different growing seasons is important. Examples used in this paper (Figs. 1–4) reflect the relatively stable conditions of irrigated agriculture in the southwestern United States. Wider confidence bands would be expected on 4-year averages for rain-fed agricultural systems. Also the use of relative yields, corrected for seasonal variation and productive potential of the field, removes much of the error which would be associated with predicting actual yields. Validation of damage functions across years and geographic regions, including measurement of associated variance, is an important aspect of the science of crop loss.

In summary, the purpose of this paper is to encourage the development of nematode damage functions. Data sets from small field plots are amenable to adjustment for experimental and sampling error which reveal the relationship between crop yield and nematode population density. Measurement of the variability associated with damage functions and population estimates will promote refinement of technique in both areas. It also allows statement of probability levels associated with yield-loss predictions.

LITERATURE CITED

1. Elliott, J. M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Freshwater Biol. Assoc. Sci. Publ. No. 25.

2. Ferris, H. 1980. Nematology—Status and prospects: Practical implementation of quantitative approaches to nematology. J. Nematol. 12:164–170.

3. Ferris, H., W. D. Turner, and L. W. Duncan. 1981. An algorithm for fitting Seinhorst curves to the relationship between plant growth and preplant nematode densities. J. Nematol. 13:300–304.

4. Goodell, P. B., and H. Ferris. 1980. Plant-parasitic nematode distributions in an alfalfa field. J. Nematol. 12:136–141.

5. McSorley, R., and J. L. Parrado. 1982. Estimating relative error in nematode numbers from single soil samples composed of multiple cores. J. Nematol. 14:522–529.

6. Nicholson, A. J. 1933. The balance of animal populations. J. Anim. Ecol. 2:132–178.

7. Noe, J. P., and K. R. Barker. 1983. Overestimation of yield loss due to patchy distribution of *Meloidogyne incognita*. Amer. Phytopathol. Soc. (Abstr.).

8. Seinhorst, J. W. 1965. The relation between population density and damage to plants. Nematologica 11:137–154.

9. Seinhorst, J. W. 1973. The relation between nematode distribution in a field and loss in yield at different average nematode densities. Nematologica 19:421–427.

10. Snedecor, G. W., and W. G. Cochran. 1967. Statistical methods. Ames: Iowa State University Press.