

Simulated Sampling Strategies for Nematodes Distributed According to a Negative Binomial Model¹

R. MCSORLEY²

Abstract: A FORTRAN computer program was developed to simulate nematode soil sampling strategies consisting of various numbers of samples per field, with each sample consisting of various numbers of soil cores. The program assumes that the nematode species involved fit a negative binomial distribution. Required input data are estimates of the mean and k values, the number of samples per field and cores per sample in the strategy to be investigated, and the number of times the simulation is to be replicated. Output consists of simulated values of the relative deviation from the mean and standard error to mean ratio, both averaged over all replications. The program was used to compare 150 simulated sampling strategies for *Meloidogyne incognita*, involving all combinations of two mean values (2.0 and 10.0 larvae/10 cm² soil), three k values (1.35, 0.544, and 0.294), five different numbers of samples per field (1, 2, 4, 10, 20), and five different numbers of cores per sample (1, 2, 4, 10, 20). Simulations resulting from different mean values were similar, but best results were obtained with higher k values and 20 cores per sample. Relatively few 20-core samples were needed to obtain average deviations from the mean of 20–25%. *Key words:* computer simulation, sampling error, spatial distribution, *Meloidogyne incognita*.
Journal of Nematology 14(4):517-522. 1982.

The accuracy of a sampling plan is critical to the operation of a nematode diagnostic laboratory, yet recommendations on the numbers of samples and number of soil cores per sample to be collected per field vary greatly (1). Relatively little information is available about how to construct an accurate and efficient sampling plan, but available studies (5,8) suggest that obtaining accurate estimates of field populations requires considerable effort.

A recent study (7) has examined the relative errors involved in estimating soil populations from a single composite soil sample composed of multiple cores from fields of various sizes. In some cases, estimates of field populations within acceptable error limits could not be obtained without collecting very large numbers of cores. However, it may be possible to obtain more accurate estimates by taking several replicated composite samples from a field with a lower number of cores per composite sample.

Goodell and Ferris (5) have developed a method for comparing the relative accuracy of different sampling schemes consisting of different numbers of composite samples and cores per sample. Their method involves searching a large data base of nematode counts from an alfalfa field (4) for simulated nematode counts in various sample and core combinations. The rela-

tive accuracy of several sampling strategies were compared by calculating DEV, the deviation of the estimate from the field mean (of the large data base), expressed as a percentage of the field mean. Results were then used to optimize sampling plans to achieve maximum efficiency.

This paper outlines a FORTRAN computer program developed to extend the work of Goodell and Ferris (5) to the more general situation in which the underlying spatial distribution is the negative binomial. Instead of obtaining simulated counts from a stored data base, this program uses simulation from a negative binomial distribution to obtain counts of nematodes per soil core. Provided that a nematode population fits a negative binomial distribution model, a sampling plan in terms of number of samples and cores per sample can be developed by using the appropriate mean and k values. The relative efficiency of several sampling strategies can be compared in terms of relative deviations from the mean and standard error to mean ratios.

COMPUTER PROGRAM STRUCTURE

Required input variables are the mean (\bar{x}) and k value for the appropriate negative binomial distribution. Using these values, the individual terms of the negative binomial distribution are calculated from generalized formulae developed from Elliott (2). The probability of a zero value (P_0) is calculated from the formula:

Received for publication 6 April 1982.
¹Florida Agricultural Experiment Stations Journal Series No. 3758.

²University of Florida Agricultural Research and Education Center, 18905 S. W. 280 Street, Homestead, FL 33031.

$$P_0 = \left(1 + \frac{\bar{x}}{k}\right)^{-k} \quad (1)$$

while the probability of each successive i^{th} term (P_i) is given by:

$$P_i = \left(\frac{k+(i-1)}{i}\right) \left(\frac{\bar{x}}{\bar{x}+k}\right) P_{(i-1)} \quad (2)$$

The cumulative probability (CMP_{*i*}) is also calculated for each term, and calculations continue until the first 1,000 terms are computed or a cumulative probability of 0.999995 is reached.

Once the terms of the appropriate negative binomial distribution have been calculated, the sampling loops are run for a given sampling strategy. Required input variables for each such problem are the number of samples (IS) and number of cores per sample (IC) to be collected under the given strategy, the number of times the simulated strategy is to be replicated (IR), and a seed for the FORTRAN pseudo-random number generator (S).

A Monte Carlo simulation technique (3) was used for generating simulated values for the nematode counts in each core collected. A pseudo-random number is matched to the corresponding point on the cumulative probability distribution and the corresponding discrete count is used in the simulation.

The interaction of the various sampling loops is illustrated (Fig. 1). Each simulated core count is accumulated until IC is reached, at which time a mean count for the sample is calculated. Means of successive samples are computed in a similar manner until the total number of samples (IS) for the strategy is reached. At this point, a mean count (\bar{y}) for the strategy is computed, and DEV, the relative deviation from the theoretical mean, is calculated, where

$$\text{DEV} = \left| \frac{\bar{x} - \bar{y}}{\bar{x}} \right| \quad (3)$$

The procedure is repeated for each replication, after which AMDEV, an average value of DEV, is calculated. Standard error to mean ratios are also computed for each replication, and an average value, ASMDEV, is computed for all replications. Thus, for a sampling strategy of four com-

posite samples of 20 cores each, replicated 10 times, simulated values for 800 cores of soil would be generated and included in the computations. A second strategy can be compared with the first by entering a second set of IC, IS, IR, and S values. Output for each problem consists of AMDEV, the average deviation from the field mean, and ASMDEV, the average standard error to mean ratio.

SIMULATION RESULTS

Simulations were run using data from a previous study (7) on counts of *Meloidogyne incognita* (Kofoid & White) Chitwood larvae. Appropriate k values were 1.35 for fields of ca. 0.5 ha in size, 0.544 for ca. 1.0 ha and 0.294 for fields of ca. 1.5 ha (7). With each k value, mean counts of 2.0 and 10.0 larvae per 10 cm³ of soil were used. Thus, simulations were performed for a total of six different combinations of \bar{x} and k values. For each combination, simulations of 25 different sampling strategies were run. These strategies consisted of all combinations of five different numbers of composite samples per scheme (1, 2, 4, 10, 20) and five numbers of cores per composite sample (1, 2, 4, 10, 20). All combinations were replicated 20 times. Results of the simulations are shown for the $\bar{x} = 2$ cases for the 20 cores per sample (Fig. 2A) and the 10 cores per sample (Fig. 2B) strategies. Theoretical curves of the form $y = ax^b$ were fit to the simulated points for each k value. For a given k value, it is apparent that more composite samples of 10 cores/sample (Fig. 2B) are needed to maintain the same relative deviation (AMDEV) than in the corresponding 20 cores/sample case (Fig. 2A). In general, the smaller the field, the greater the k value will be (7). For a given number of cores per sample, fewer samples per field are needed to maintain a given level of error with a larger k value compared to a smaller one.

Simulated values for the average relative deviation from the mean (AMDEV) and the average of the standard error to mean ratios (ASMDEV) were similar for a given case; thus, only AMDEV values are shown (Fig. 2).

In comparing the $\bar{x} = 2$ cases (Fig. 2)

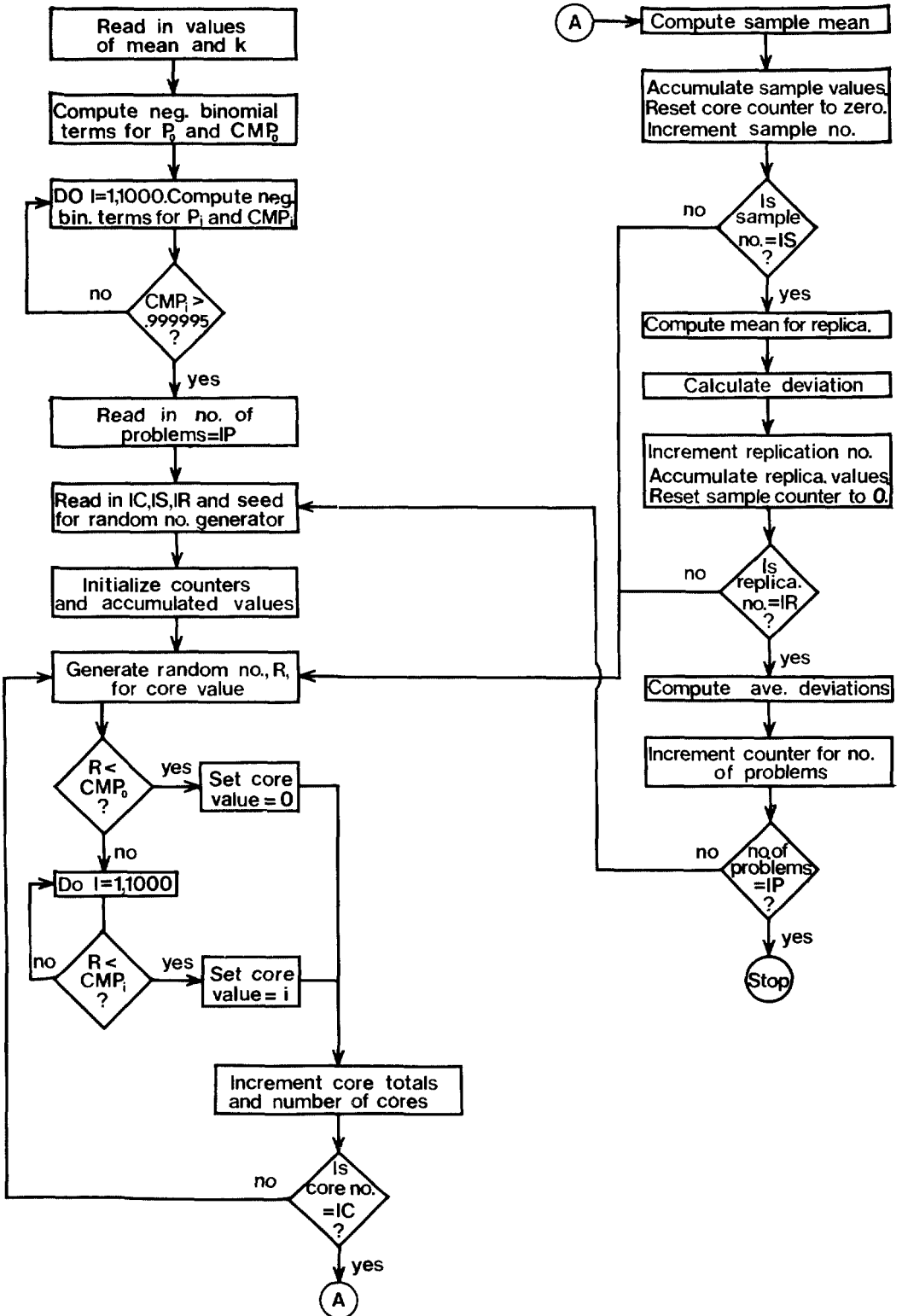


Fig. 1. Flow chart of program for generating terms of a negative binomial distribution and simulating corresponding nematode counts per core, sample, and replication. IS = number of samples; IC = number of cores; IR = number of replications; P_i = probability of the i^{th} term; CMP_i = cumulative probability including the i^{th} term.

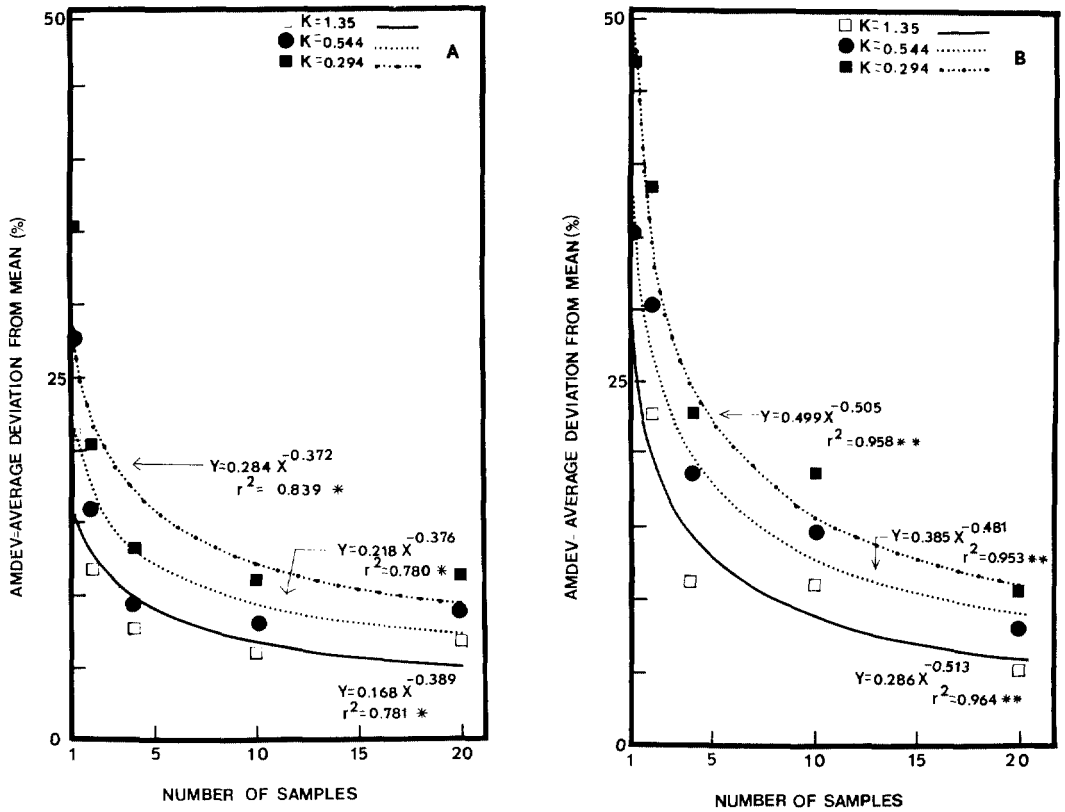


Fig. 2. Relationship between average of percent deviation from field mean (AMDEV) and number of samples per field for mean = 2.0 and various k values. Simulated values indicated by points; curves represent best fit to simulated points. A) 20 core per sample. B) 10 cores per sample.

to the $\bar{x} = 10$ cases (not shown) in the simulation runs, the mean deviations (AMDEV) obtained for the $\bar{x} = 10$ case were only slightly lower than those obtained for the $\bar{x} = 2$ case. Largest deviations are generally obtained for the smallest mean values. Thus, in developing a sampling plan for a field where no previous estimate of the mean has been made, the most conservative sampling strategy is to assume a very low mean value for the computer simulation. One must also assume that the spatial distribution of the nematode fits a negative binomial model, and an appropriate k value for the nematode species and field size involved should be selected (7). Transformation of mean values depend on the fitted negative binomial distributions. Actual raw counts are normally used in the goodness of fit tests (4,7), and the mean values used here also represent actual counts per 10 cm³ of soil.

DISCUSSION

It is impossible to develop meaningful error estimates for nematological sampling plans without knowledge of the spatial distribution of the nematode species involved. Much additional research into the statistical distribution of plant parasitic nematodes is needed to determine if there are consistent patterns for a given species, crop, or geographical region. Use of this particular computer program assumes knowledge of the mean and k parameters of the appropriate negative binomial distribution for the nematode to be studied. The program also assumes a random sampling scheme in the field. Similar results would be anticipated for other patterns, and so the division of the field into strips for sampling is recommended (5). In this study, multiple samples of multiple cores refers to repeated samples from the same

field. Dividing a field into portions and taking a separate sample from each portion can give different results and transposes to the single sample/multiple cores case (7). In general, subdividing a field into smaller units will require more sampling effort, but can provide more detailed information if spot treatment of only a portion of the field is feasible. The advantage of computer simulation in developing sampling plans is apparent. For example, to test a strategy of 20 samples of 20 cores per sample, replicated 20 times, count data on 8,000 cores of soil would be needed.

Various statistics have been used to evaluate the magnitude of error terms in sampling studies. Goodell and Ferris (5) and the present study used DEV, as defined by equation 3. This is a particularly useful term when comparing the deviations obtained by two simulated sampling strategies, since it can compare the single sample/multiple core case to the multiple sample/multiple core case. The standard error to mean ratio (E) is widely used in entomology, where it is called relative variation when expressed as a percent (9). Sampling error can also be expressed in terms of percentage confidence limits of the mean, and various formulae are available for computing such terms (2,6,10).

Because it is a measure of precision and therefore requires at least two samples for its computation, the standard error to mean ratio, E , is not as versatile as DEV in simulation studies such as the present study or that of Goodell and Ferris (5). The limitation is that it is difficult to compare the single sample/multiple core case to the multiple sample/multiple core case using E . Nevertheless, E is the more meaningful term, statistically. In the present study, ASMDEV, an estimate of E obtained for multiple sample/multiple core cases, was relatively close to the corresponding values of AMDEV.

For the single sample/multiple core case, E must be calculated over the cores involved in the single sample by the formula used elsewhere (7,10):

$$n = \frac{1}{E^2} \left(\frac{1}{\bar{x}} + \frac{1}{k} \right) \quad (4)$$

For $n = 20$ cores per sample, $\bar{x} = 2$, and k

$= 1.35$, equation 4 simplifies to $E = 0.249$. In the simulation of the one sample of 20 cores case for these mean and k values, AMDEV = 0.213, relatively close to the calculated value of E .

If information on the spatial distribution of a given nematode is available, then the procedures described here and elsewhere (5) can be used to develop sampling schemes. Because these methods involve simulation, error estimates obtained by them are stochastic and therefore a number of replications should be run to assure that the estimates are reasonable. Calculation of actual error terms is possible only for the single sample/multiple core case using equation 4. This calculation can be compared with the simulation results to assure that sufficient replications are being used in the simulation.

Before a sampling program is established, it is apparent that consideration should be given to the amount of error that can be tolerated. The method by which error terms are to be calculated should be understood, since discrepancies or misunderstandings can lead to differences in the numbers of samples to be collected.

LITERATURE CITED

1. Barker, K. R., and C. J. Nusbbaum. 1971. Diagnostic and advisory programs. Pp. 231-301 in B. M. Zuckerman, W. F. Mai, and R. A. Rhode, eds. Plant parasitic nematodes. Vol. I. Morphology, anatomy, taxonomy, and ecology. New York: Academic Press.
2. Elliott, J. M. 1979. Some methods for the statistical analysis of samples of benthic invertebrates. Freshwater Biological Association Scientific Publication. No. 25. Windermere, Eng.
3. Giffin, W. C. 1971. Introduction to operations engineering. Homewood, Illinois: Richard D. Irwin, Inc.
4. Goodell, P., and H. Ferris. 1980. Plant parasitic nematode distribution in an alfalfa field. *J. Nematol.* 12:136-141.
5. Goodell, P. B., and H. Ferris. 1981. Sample optimization for five plant-parasitic nematodes in an alfalfa field. *J. Nematol.* 13:304-313.
6. Karandinos, M. G. 1976. Optimum sample size and comments on some published formulae. *Bull. Entomol. Soc. Amer.* 22:417-421.
7. 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.
8. Proctor, J. R., and C. F. Marks. 1975. The determination of normalizing transformations for nematode count data from soil samples and of

efficient sampling schemes. *Nematologica* 20:395-406.

9. Ruesink, W. G. 1980. Introduction to sampling theory. Pp. 61-78 in M. Kogan and D. C. Herzog, eds. *Sampling methods in soybean entomology*. New

York: Springer-Verlag.

10. Southwood, T. R. E. 1978. *Ecological methods with particular reference to the study of insect populations*. New York: Halsted Press.