Improving the Accuracy of Sampling Field Plots for Plant-Parasitic Nematodes[1]

L. J. FRANCL[2]

Abstract: The validity of nematode data from field experiments depends largely on how well samples represent the nematode population. Data from an intensive sampling of three field plots before and after spring cultivation were used to compare eight simulated sampling schemes. Average deviation from the plot mean ranged from 10% to 34% before cultivation and from 7% to 16% after cultivation. Samples taken from only the plant row erred most before cultivation but were comparable to other schemes after cultivation. Several schemes achieved a 25% deviation or less in 90% of the sample simulations. Sampling a nematode population usually involves subsampling a composite bulk sample, however, and this increases error by an estimable amount. A random sample with 35 cores and four random subsamples estimated mean plot densities within 25% with probabilities ranging from 0.77 to 0.85. The probability of a sample–subsample combination coming within a specified percent error of the true mean can be extended cautiously to any field mean and variance more-or-less independent of species and area using formulae presented herein. The most economical method of increasing sample accuracy was to increase the number of soil cores.
Key words: Heterodera glycines, soybean cyst nematode, subsampling.

Research on management of plant-parasitic nematodes often uses field experimentation to generate data. The validity of these data depends largely on how well the samples estimated the nematode populations. Sampling for nematodes usually involves taking the bulk soil sample and subsampling the composite bulk sample. The accuracy of the bulk sample in estimating mean nematode population density depends on the number of soil cores collected and on how closely core spatial coordinates represent the three-dimensional nematode population dispersion. Mixing the bulk sample at best randomizes nematodes so that variance among subsample counts equals the mean (3,6,8). Bulk sampling and subsampling both contribute to the uncertainty of mean estimation and to total unexplained variation (2). My objectives were to compare eight schemes for sampling field plots for the soybean cyst nematode (SCN), *Heterodera glycines* Ichinohe, and to calculate the probable accuracy of a population density estimate as a function of the number of cores, number of subsamples, mean nematode population density, and population spatial variation.

MATERIALS AND METHODS

Methods for selecting field plots and collecting single-core samples in a lattice pattern were discussed previously (4). Random and systematic samples were drawn with replacement from the data base of each plot and combined to simulate a multiple-core bulk sample. The mean of the multiple-core sample represented the sample result. Random samples (Fig. 1A) consisting of 10, 20, or 40 cores were iterated 100 times per plot with grid coordinates randomly selected on a uniform distribution as implemented in SAS (registered trademark of SAS Institute Inc., Cary, North Carolina). Manually drawn systematic samples used all possible nonrepetitive starting points. The diagonal scheme (Fig. 1B) consisting of 25 cores was repeated 10 times per plot. The basic zig-zag scheme had four variations (Fig. 1C, D). Two zig-zag schemes sampled only the plant row with either 12 or 25 cores and with eight or four repetitions per plot, respectively. A third zig-zag scheme covered the entire study area with 25 cores, repeated 10 times per plot. The final zig-zag scheme covered the area enclosed by the plant rows with 25 cores and was repeated six times per plot. Error rates for the sampling scheme were calculated as percent absolute deviation (PDEV, equation 1) from the grand mean of the plot (5). The mean PDEV and

percentiles for the scheme iterations were determined by PROC UNIVARIATE in SAS.

Percent absolute deviation for jth iteration of scheme i in plot k (PDEV)
$$= (|x_{ij} - \bar{x}_k| \times 100)/\bar{x}_k \qquad (1)$$

Subsampling the field sample: Randomness of subsample counts was tested by linear regression analysis on data from an SCN field experiment. The data set, collected in 1982 and 1983, consisted of 578 nonzero samples, each with three subsamples. Bulk samples were mixed by hand, and 100-g subsamples were eluted.

If subsampling was random, the subsample mean would equal the subsample variance and the slope (b_1) should not differ significantly from one when variance is regressed on the mean (13).

The accuracy of a nematode density estimate is dependent upon the total variation among sample cores and subsamples. An estimate of total variation of the density estimate was given by equation 2, similar in form to a random effects model (11) except for the inclusion of a correction factor for finite subsample observations. An adequate number of representative soil cores thus is essential to reduce the effect of σ^2_1 which is higher than the mean in spatially aggregated populations (12). Furthermore, the total number of nematodes collected in the first stage of sampling provides requisite information for mean density estimation to the second stage. If subsampling nematodes is a random process, then the expected value of σ^2_2 equals the area's mean density, μ.

$$Var(\hat{\mu}) = \sigma^2_1/n_1 + \sigma^2_2/n_2 \times (1 - n_2/n_1) \qquad (2)$$

Where:

$Var(\hat{\mu})$ = variance in estimate of μ;
σ^2_1 = spatial variance of field population;
σ^2_2 = variance among subsamples;
n_1 = number of soil cores;
n_2 = number of subsamples (in core units);
$1 - n_2/n_1$ = correction factor.

Sample reliability is defined here as the probability that a mean density estimate
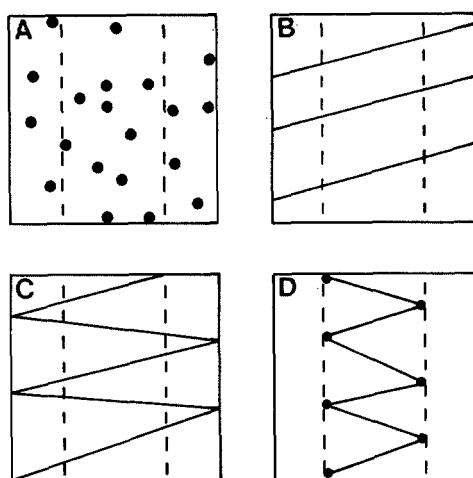

Fig. 1. Sample scheme patterns. A) Random, 10, 20, or 40 cores. B) Diagonal, 25 cores. C) Zig-zag, 25 cores from area between two rows (1 m wide) or 25 cores over entire width (2 m). D) Zig-zag, 12 or 25 cores taken only from plant rows.

will be within specified bounds of the true mean. An error of ± 25% from the mean allows for the detection of population halving or doubling (12), sufficient for most studies on nematode management. Reliability of sampling was found by calculating a Z deviate (equation 3) and determining area under the normal curve between Z and −Z. Subsampling from 100 40-core random samples was simulated by generation of 1,200 pseudorandom variates in SAS to check the validity of equations 2 and 3. (A BASIC program that extends probability calculations to other means and variances is available from the author.)

$$Z = \% \text{ error} \times \mu/\sqrt{Var(\hat{\mu})} \qquad (3)$$

An estimate of the time required to sample and subsample a single field plot was deduced from personal experience. It is recognized that there are differences among laboratory extraction techniques for various nematode species (1).

RESULTS AND DISCUSSION

Field sampling: Deviation percentages for eight sampling schemes were averaged across plots within sampling time (Fig. 2). Systematic samples taken only from the plant rows had a mean deviation of 33–34% before cultivation. This is a consequence of taking samples from a stratum having an average 26% higher population
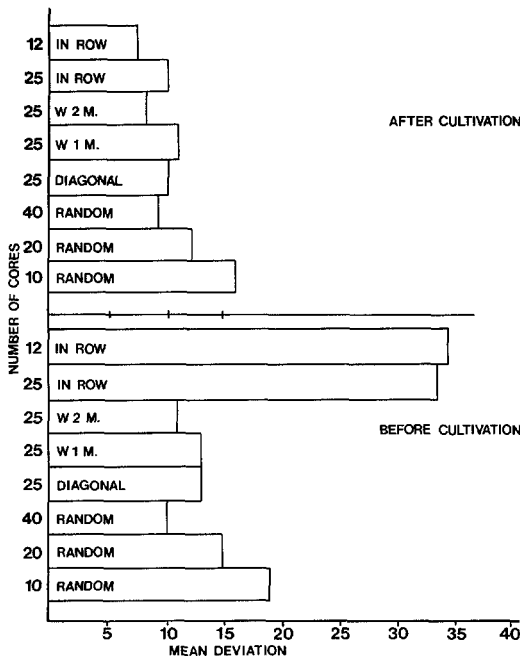
FIG. 2. Mean percent deviation from the average number of *Heterodera glycines* cysts per soil core in field lots calculated for eight sampling schemes before and after primary cultivation.

population density present in plant row strata. In-row samples of 25 and 12 cores before cultivation had an average deviation of 11% and 18% of the row stratum mean, respectively. Bias in the estimation of mean plot density could be avoided by collecting samples from between the rows as well as in the rows. Average deviation from the whole-plot mean decreased from 19% for 10 randomly collected cores to 10% for 40 randomly collected cores. The rate of improvement in sampling accuracy showed arithmetic returns as the number of cores doubled from 10 to 20 and from 20 to 40. Increased numbers of cores therefore yielded decreased returns to labor. Before cultivation, the zig-zag covering the entire study area had the lowest average PDEV of the systematic schemes and was comparable to the 40-core random sample. All sampling schemes were much closer in average PDEV after cultivation.

Numerous plots often must be sampled in field experiments; therefore the number of large errors that can be expected to occur is an important consideration. Reporting only the mean error rate can be misleading because there is a distribution of values around the mean. The 90th percentile, for example, indicates that interval of PDEV which occurred in 90% of the samples with the remaining 10% of the samples having a larger PDEV. The distributions of PDEV from sample iterations are compared to formal probabilistic statements in Table 1 (7). Most percentile estimates are close to expected results from probability calculations. Before cultivation, the 2-m-wide zig-zag sampling scheme consisting of

density than the whole-plot means (4). An in-row sample consisting of 20 or more cores is the current recommendation for nematode field plot research (1). This practice would result in an adequate representation of the population density at the beginning of the season after plowing but would overestimate the density of SCN at the end of the season, represented here by the first sampling time. In-row sampling schemes provided good estimates of the

TABLE 1. Percent deviation of sampling schemes evaluated at the 90th percentile and expected percent deviation at $\alpha = 0.90$.

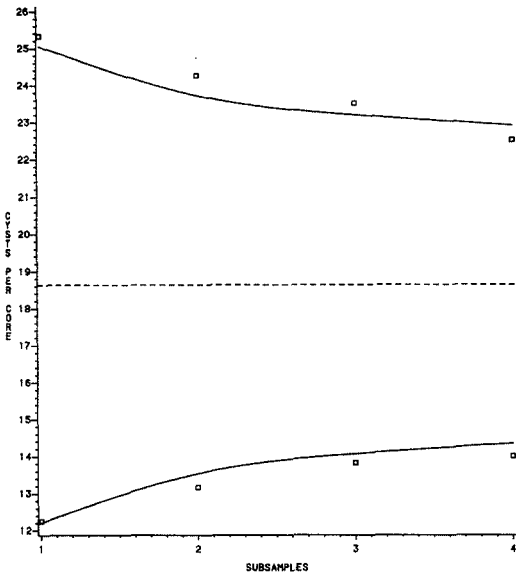| Sampling schemes (cores)	Before cultivation			After cultivation		
	Iterations	Observed error	Expected error	Iterations	Observed error	Expected error
Random (10)	300	38	43	200	35	35
Random (20)	300	28	31	200	25	25
Random (40)	300	22	22	200	19	18
Diagonal (25)	30	28	27	20	24	22
Zig-zags						
1 m wide (25)	18	32	27	12	26	22
2 m wide (25)	30	23	27	20	24	22
In row (25)	12	52	27	8	13	22
In row (12)	24	73	40	16	25	32

FIG. 3. Observed errors (squares) at 10th and 90th percentiles from a computer simulation and predicted (solid lines) 80% confidence intervals from equations 2 and 3 for sampling plot 1 prior to cultivation with 40 cores and 1–4 subsamples. Plot mean is marked with broken line, and the upper and lower bounds at 25% error at 23.3 and 14.0 cysts per core, respectively.

25 cores gave less than the expected error under the assumption of random sampling. In-row sampling schemes erred most before cultivation but produced better than expected results after cultivation. This lat-

ter result is possibly due to the method of cultivation.

Subsampling the field sample: Subsampling a composite bulk sample is an efficient way to reduce the time required to collect data but it adds to the total error expected from a sample. The only exception to this rule occurs when the dispersion of the sampled characteristic in the bulk sample is uniform with zero variance. Therefore, the sampling schemes in the previous section should be considered relative to one another rather than as indications of what is achievable in practice.

Subsample variance was regressed on subsample mean by year for 1982 and 1983 to test the hypothesis of random nematode dispersion in the bulk sample. The slope (b_1) was not different from one in 1983, but in 1982 it was significantly greater than one ($P < 0.01$). Subsamples with large mean numbers of cysts had higher than expected variance in 1982 under the hypothesis of randomness. This result indicates that randomization is the best consequence to be expected from mixing a bulk sample and that higher variances can occur. Subsampling error is confounded with errors in handling and counting. Further discussion will assume that subsamples follow a random (Poisson) process.

The validity of equations 2 and 3 for estimating $Var(\hat{\mu})$ and sampling reliability was tested by comparing the predicted

TABLE 2. Probabilities that a sample result will be within 25% of true mean for different sample and subsample sizes. Calculated for plot 2 where means were 10.2 and 12.5 cysts per core and variances were 77 and 70 before and after cultivation, respectively.

Subsam-ples (n_2)	Cores in sample (n_2)								
	20	25	30	35	40	45	50	55	60
Before cultivation									
1	0.51	0.52	0.53	0.54	0.54	0.54	0.55	0.55	0.55
2	0.62	0.64	0.65	0.66	0.67	0.68	0.68	0.69	0.69
3	0.67	0.70	0.72	0.73	0.74	0.75	0.76	0.76	0.77
4	0.70	0.73	0.76	0.77	0.78	0.79	<u>0.80</u>	0.81	0.82
5	0.73	0.76	0.78	<u>0.80</u>	0.81	0.82	0.83	0.84	0.85
6	0.74	0.78	<u>0.80</u>	<u>0.82</u>	0.83	0.84	0.86	0.86	0.87
After cultivation									
1	0.57	0.58	0.59	0.59	0.60	0.60	0.60	0.60	0.60
2	0.70	0.71	0.72	0.73	0.74	0.74	0.75	0.75	0.76
3	0.76	0.78	0.79	<u>0.80</u>	0.81	0.82	0.82	0.83	0.83
4	<u>0.80</u>	0.82	0.84	<u>0.85</u>	0.86	0.86	0.87	0.87	0.88
5	<u>0.82</u>	0.84	0.86	0.87	0.88	0.89	0.90	0.90	0.91
6	0.84	0.86	0.88	0.89	0.90	0.91	0.92	0.92	0.93

TABLE 3. Relative cost efficiency of processing nematode samples that met the criterion for 0.80 probability of a result within 25% of the true mean.

	Number of sample cores	Number of subsamples	Relative efficiency
Before cultivation			
Plot 1	66	2	1.00
	37	3	0.95
	31	4	0.80
	28	5	0.68
Plot 2	121	3	0.86
	48	4	1.00
	36	5	0.89
Plot 3	41	3	1.00
	27	4	0.88
	22	5	0.75
After cultivation			
Plot 1	44	3	1.00
	24	4	0.91
	19	5	0.78
Plot 2	33	3	1.00
	21	4	0.86
	17	5	0.74

Total cost (in minutes) = $0.25 \times n_1 + 9.5 \times n_2 + 5.0$ where n_1 = number of soil cores and n_2 = number of subsamples. Relative cost efficiency of optimal combination = 1.00.

range of 40-core random samples at several probability limits with the observed range of sample simulations between comparable percentiles. Predicted and observed confidence limits were in close agreement (Fig. 3). Skewness was detected in the observed results due to aggregation of the field population, but this was not an important source of error. Equation 2 casts subsample size in soil core units because subsample volume or weight, not subsample number per se, is the determinant of variance (3).

The reliability of sampling within 25% of the mean of plot 2 before and after cultivation using 20–60 soil cores and 1–6 subsamples is presented in Table 2. This plot had the highest and lowest sampling reliability over the course of the experiment. The choice of an acceptable level of reliability and error rate is up to the individual investigator. A minimal probability for success in the 0.75–0.80 range seems both reasonable and achievable for SCN if a 25% deviation is acceptable. More than one subsample is necessary for improved sampling accuracy. Many more samples would have to be collected to reach a 10% error rate. A 0.50 probability of a sample

result within 10% of the mean of plot 2 after cultivation would have required 65 cores and five subsamples.

Three subsamples and 45 cores was an economic compromise in most plots (Table 3). However, a conservative approach when extracting SCN or other nematodes from soil with a four-cone semiautomatic elutriator would be to have all four cones process a single sample and to eliminate variability among cones as a source of error. Taking four subsamples therefore leads to choosing 35 cores as an acceptable level of sampling intensity. Probabilities ranged from 0.77 to 0.85, and the relative cost efficiency was at worst 80% of the most efficient combination of cores and subsamples (Tables 2, 3).

Sample reliability is independent of nematode species. Area and formulae presented here should prove useful for any soil characteristic, organic or inorganic, provided that the sample represents the spatial dispersion of the organism or property and that subsampling the composite is a random process (11). Stratification of large areas therefore remains a sound recommendation (5). The utility of equations 2 and 3 is illustrated by reanalyzing the data of Proctor and Marks (10). They stated that five 40-core samples, each with five subsamples, were needed to estimate the mean density ± 20% of *Pratylenchus penetrans* (Cobb) with a probability of 0.95 and with a total expenditure of 7 hours per plot. Nematodes were extracted from 25 g and a 20% aliquant of the final volume was counted. A subsample in this case was about $\frac{1}{10}$ of a soil core (10). Calculations presented in this paper estimated a probability of 0.95 at ± 20% and 0.99 if the interval of acceptance is widened to ± 25%. The time expended per subsample would increase by perhaps 20 minutes if the entire final volume were counted, but the reduction in total sample variance is significant. A 40-core sample with four subsamples (two soil core equivalents) could have estimated the population ± 25% with a probability of 0.95 and might have been accomplished within 1.5 hours.

Nematode population density and spatial dispersion usually are unknown before field experiments begin, and it is not practical to run a uniformity trial prior to every

experiment. As more reports on spatial analysis for specific plant–nematode situations are published (4,9), precise scientific methodologies should emerge for sampling plant-parasitic nematodes in field plots.

Literature Cited

1. Barker, K R., chairman. 1978. Determining nematode population responses to control agents. Pp. 114–125 *in* E. Zehr, ed. Methods for evaluating plant fungicides, nematicides, and bactericides. St. Paul: American Phytopathological Society.

2. Ferris, H. 1984. Nematode damage functions: The problems of experimental and sampling error. Journal of Nematology 16:1–9.

3. Fidler, J. H., B. M. Church, and J. F. Southey. 1959. Field sampling and laboratory examination of cereal root eelworm cysts. Plant Pathology 8:27–34.

4. Francl, L. J. 1986. Spatial analysis of *Heterodera glycines* populations in field plots. Journal of Nematology 18:183–189.

5. Goodell, P. B., and H. Ferris. 1981. Sample optimization for five plant-parasitic nematodes in an alfalfa field. Journal of Nematology 13:304–313.

6. Jones, F. G. W., and R. A. Kempton. 1979. Population dynamics, population models and integrated control. Ch. 18 *in* J. F. Southey, ed. Plant nematology. London: Ministry of Agriculture, Fisheries and Food.

7. Karandinos, M. G. 1976. Optimum sample size and comments on some published formulae. Bulletin of the Entomological Society of America 22:417–421.

8. Muller, J. 1983. On the problem of estimating population densities of *Heterodera schachtii* by soil sampling and extraction. I. Estimation of the population density in mixed soil samples. Nachrichtenblatt des Deutschen Pflanzenschutzdienstes 35:132–136.

9. Noe, J. P., and C. L. Campbell. 1985. Spatial pattern analysis of plant-parasitic nematodes. Journal of Nematology 16:86–93.

10. Proctor, J. R., and C. F. Marks. 1974. The determination of normalizing transformations for nematode count data from soil samples and of efficient sampling schemes. Nematologica 20:395–406.

11. Reed, J. F., and J. A. Rigney. 1947. Soil sampling from fields of uniform and nonuniform appearance and soil types. Agronomy Journal 39:26–40.

12. Southwood, T. R. E. 1978. Ecological methods with particular reference to the study of insect populations, 2nd ed. Cambridge, England: University Printing House.

13. Stuart, A. 1956. The efficiencies of tests of randomness against normal regression. Journal of the American Statistical Association 51:285–287.