Factors influencing the precision of soil seed bank estimates[1]

D. L. BENOIT[2], N. C. KENKEL[3], AND P. B. CAVERS
Department of Plant Sciences, University of Western Ontario, London, Ont., Canada N6A 5B7

Received February 28, 1989

BENOIT, D. L., KENKEL, N. C., and CAVERS, P. B. 1989. Factors influencing the precision of soil seed bank estimates. Can. J. Bot. **67**: 2833–2840.

The dimension of soil augers needed to sample a seed bank of *Chenopodium* spp. (lamb's-quarters) was determined by randomly sampling a 1.35-ha area within a cornfield in Oxford County, Ontario. Sampling units of three different auger sizes (1.9, 2.7, and 3.3 cm in diameter) were collected. On a per volume basis, there were no significant differences between the three sizes of auger in estimating the number of lamb's-quarters seeds in the soil. Three sampling methods, systematic, stratified random, and cluster, were compared with random sampling in their capacity to minimize the sampling variance. Soil cores of 1.9 cm diameter and 15 cm deep were taken systematically at 3.5-m intervals to form a 32 × 32 matrix. Repeated sampling within the matrix using Monte Carlo techniques indicated that the estimate of sampling variance decreased with increasing sample size, regardless of the sampling method used. No fewer than 60 sampling units should be collected to quantify the seed bank of an abundant weed such as lamb's-quarters. The estimates of sampling variance of systematic and cluster sampling were clearly influenced by the sampling interval and the cluster's shape, respectively. This was attributed to the underlying seed distribution of lamb's-quarters in the soil that was clustered with patterns of high and low seed density parallel to corn rows. There were no significant differences between the estimate of sampling variance of random and stratified random sampling with a fixed sample size of 64 units.

BENOIT, D. L., KENKEL, N. C., et CAVERS, P. B. 1989. Factors influencing the precision of soil seed bank estimates. Can. J. Bot. **67** : 2833–2840.

Trois tarières de grosseur différente (1,9, 2,7 et 3,3 cm de diamètre) ont été utilisées dans un champ de maïs de 1,35 ha afin de déterminer le volume de sol nécessaire pour échantillonner une banque de graines de *Chenopodium* dans le sol. Il n'y a aucune différence significative entre les estimés du nombre de graines dénombré par volume de sol et récolté avec les trois tarières. Une étude a été initiée afin de comparer la précision des estimés de moyenne de différentes tailles d'échantillons obtenus suivant des méthodes d'échantillonnage aléatoire, systématique, stratifié et en grappe. Des échantillons de sol (1,9 cm de diamètre et 15 cm de profondeur) ont été ramassés à des intervalles de 3,5 m formant ainsi une matrice de 32 × 32. Une technique de Monte Carlo a été utilisée pour effectuer les simulations des différentes méthodes d'échantillonnage sur cette matrice. Une simulation de l'échantillonnage aléatoire a démontré que la variance de la moyenne diminue au fur et à mesure que la taille de l'échantillonnage augmente. Cette courbe indique qu'un minimum de 60 échantillons est nécessaire pour décrire une banque de graines dans le sol. La disposition des grappes et des intervalles de méthodes d'échantillonnage en grappe et systématique ont influencé la variance de la moyenne à cause de la distribution spatiale des graines de *Chenopodium* dans le sol. En effet, ces graines étaient groupées formant des surfaces de fortes et de faibles densités situées parallèlement aux rangs de maïs. Il n'y avait aucune différence significative entre la variance de la moyenne mesurée par un échantillonnage aléatoire ou stratifié pour une taille totale de 64 échantillons de sol récoltés.

Introduction

The use of soil sampling to estimate the numbers of weed seeds in the soil is both time and labour intensive, regardless of the techniques used (Kropáč 1966; Malone 1967; Feast and Roberts 1973; Thorsen and Crabtree 1977; Fay and Olson 1978; Standifer 1980). The limiting factor in seed bank studies is ultimately the total volume of soil that can be sampled and processed. The dimension and number of sampling units (or soil cores) are important aspects of the sampling methods.

There is no standard sampling unit used in seed bank studies. Some investigators (Numata *et al.* 1964; Hayashi and Numata 1971) used a species – soil volume curve to estimate the minimum soil volume required to adequately describe the content of a seed bank. This method is analogous to the species – area curve of the phytosociologists, and as such, its main objective is to characterize the species composition of the seed bank. Other investigators (Dospekhov and Chekryzhov 1972; Tulikov *et al.* 1981) examined the total soil weight required to characterize a seed bank.

In most seed bank studies, the sampling cost, the available resources (time, space, and labour), and the sampling tool have dictated an arbitrarily chosen but reasonable sample size. The optimal sample size (total number of sampling units) required in seed bank studies has been investigated specifically by Champness (1949), Rabotnov (1958), Goyeau and Fablet (1982), Barralis *et al.* (1986), and Lopez *et al.* (1988). The general consensus is that a large number of small sampling units is more appropriate than a small number of large sampling units (Kropáč 1966). This finding concurs with other studies, particularly on benthic invertebrates (Elliott 1977) where similar problems occur.

Numerous descriptions of seed bank populations have been done in studies of agricultural and natural vegetation, and these results have been reviewed by Roberts (1981) and Leck *et al.* (1989). Seed populations in the soil are frequently and erroneously assumed to be homogeneous and normally distributed. The problem in describing the seed distribution in soil is associated with its inherent heterogeneity. Seeds often are shed close to the parent plant. This leads to strong departures from randomness in the seed distribution of populations on and in the soil (Major and Pyott 1966). Although the most abundant species often have a normal distribution, the less abundant ones usually have a Poisson or aggregated distribution (Goyeau and Fablet 1982).

[1]Agriculture Canada contribution No. 335/89.

[2]Present address: Agriculture Canada, Research Station, St-Jean-sur-Richelieu, Que., Canada J3B 6Z8.

[3]Present address: Department of Botany, University of Manitoba, Winnipeg, Man., Canada R3T 2N2.

Printed in Canada / Imprimé au Canada

This study was designed to show how the methods of sampling seed banks affect the precision of the estimate of the mean when the underlying distribution of the seed population is unknown. The seed population of an abundant weed, in this case lamb's-quarters (*Chenopodium* spp.), was chosen for investigation. The objectives were (*i*) to determine which of three auger sizes provides the most precise estimate of the mean for the seed population of *Chenopodium* spp. in cultivated soil; (*ii*) to describe the distribution of *Chenopodium* seeds over a large area (1.35 ha); (*iii*) to determine the minimum number of sampling units needed to provide an acceptable estimate of the mean density and sampling variance of a seed bank population of *Chenopodium* spp.; and (*iv*) to examine the influence of various sampling methods on the precision of the estimate of the mean. The four sampling methods used in this study were random, systematic, stratified random, and cluster.

Materials and methods

Study area

A 9.7-ha cornfield near Mount Elgin in Dereham Township, Oxford County, Ontario (42°56′N, 80°46′W), was selected for this study. The field is level and has a clay loam soil with good drainage. In this study, the field excluded the border rows of corn (headlands). An area of 1.35 ha (108.5 × 108.5 m) was randomly chosen within the field and identified as the sampling site. All soil cores were taken to a depth of 15 cm to include most of the plow layer of the cultivated soil.

Seed extraction method

Seeds were extracted from the soil cores using a modified Malone's (1967) technique. Each soil core was soaked for a minimum of 30 min in a solution of sodium hexametaphosphate (50 g/L) and sodium bicarbonate (25 g/L). This suspension was poured over a set of sieves, the upper one being a No. 10 Canadian Standard (2.0-mm mesh opening) and the lower one a No. 20 Canadian Standard (0.85-mm mesh opening). The material was washed through these sieves by a fine spray of water. The debris collected in the lower sieve was transferred onto a Whatman No. 1 filter paper and left to dry for 24 h. Once dried, all seeds of *Chenopodium* spp. were separated by hand and the number of undamaged seeds were recorded. Although most extracted seeds were probably those of common lamb's-quarters (*Chenopodium album* L.), it was impossible to exclude the possibility that seeds of other *Chenopodium* species were also present, since the pericarp of the retrieved seeds was often eroded by the seed extraction technique. Hence, we refer to the seeds as *Chenopodium* spp.

Sampling procedures

Auger size

Three sizes of soil auger (Table 1) were compared for their efficiency in sampling seeds from the soil. The amount of work required to extract seeds from the large volume of soil collected by the largest auger prevented the taking of many soil cores. As a result, the total volume of soil collected with each size of auger was uniform, and the total number of sampling units taken with each size of auger varied (Table 1).

The sampling units for each auger size were taken at random on the sampling site. Sampling was carried out on 5 November 1982 when most of the cornfield remained to be harvested.

Sample size and sampling methods

Soil cores 1.9 cm in diameter and 15 cm deep were systematically taken at 3.5-m intervals throughout the sampling site on 5−9 August 1982. This sampling procedure created a total of 1024 sampling units arranged in a 32 × 32 matrix. All soil cores, numbered and bagged individually, were stored immediately in a constant-temperature room (6.3 ± 0.8°C) until they could be processed.

The cores from the 32 × 32 matrix were used as a seed bank "population" of *Chenopodium* spp. with known parameters, the assumption being that all sampling units in the matrix are immediately adjacent to each other and form a contiguous population. The columns of the matrix were aligned with the corn rows in the field.

FORTRAN programs (designed by N.C.K.) were used to simulate sampling carried out on a seed population in the soil. These programs were based on the accepted definition of each sampling method. Simple random sampling is a method where each possible sampling unit has an equal (or known) probability of being selected, and the random selection of such units provides us with unbiased estimates of population means and sampling variance (Cochran 1977). With systematic sampling, a complete description of the units (or individuals) and their arrangement in the population is required. The first unit is drawn at random from the population, and every *i*th unit is selected until the desired sample size has been obtained (Sampford 1962). With stratified random sampling, a population is first divided into subpopulations or strata, which may or may not be of equal size. Within each stratum, a sample is selected randomly and independently (Elliott 1977). Cluster sampling is similar to simple random sampling whereby groups of units are selected randomly from the population. These groups can also be called clusters or primary units and are composed of secondary units. With cluster sampling, all secondary units are sampled (Stuart 1976).

A sampling event in the FORTRAN program is defined as a sequence of draws of a sampling unit (or soil core) from the population (or matrix) where every member was given an equal chance of being drawn. At each draw, sampling with replacement was used, and the sequence of draws is terminated when the desired sample size (or total number of sampling units) is obtained. The desired sample size was a multiple of 2 in order for the simulation program to operate within the 32 × 32 matrix.

Precision estimates were measured by the estimate of sampling variance of the mean ($S^2_{\bar{x}}$) and obtained through a Monte Carlo technique. The sampling variance of the mean ($S^2_{\bar{x}}$) measures, for all samples, the dispersion of the sample means ($\bar{X}_i$) from the true population mean (μ). The Monte Carlo technique consists of calculating the mean for each random sampling event, thereby providing the sampling variance of the mean.

There were 400 sampling events on the population to calculate each Monte Carlo estimate of sampling variance (henceforth termed MC sampling variance) for each sample size studied with each sampling method under investigation. The general formula for the MC sampling variance of the random sampling method is given as

$$\text{MC } S^2_{\bar{x}} = \frac{\sum_{i=1}^{400} (X_{i.} - X_{..})^2}{400}$$

where

$$X_{i.} = \frac{\sum_{j=1}^{m} X_{ij}}{m} \qquad (i = 1,\ldots 400;\ j = 1,\ldots m)$$

$$X_{..} = \frac{\sum_{i=1}^{m} X_i}{400}$$

and m is the total number of sampling units in a sampling event.

Statistical analysis

Auger size

The comparison of auger sizes was made on an equal soil volume basis (100 cm^3). A square root transformation of the non-normal data was made prior to analysis. Four soil cores from harvested areas of the sampling site with impeded drainage had compacted soil in the auger and significantly greater numbers of seeds for 100 cm^3 of soil.

TABLE 1. Characteristics of three auger sizes

Category of auger	Diameter of auger (cm)	Volume (cm^3)[a]	Total no. of sampling units	Total volume sampled (cm^3)
Small	1.9	42.5	60	2552
Medium	2.7	85.9	30	2576
Large	3.3	128.3	20	2566

[a] $V = \pi r^2 h$.

TABLE 2. Number of *Chenopodium* spp. seeds per 100 cm^3 of soil for different auger sizes after transformation $(x - 0.5)^{1/2}$

Category of auger	Sample size (n)	No. of *Chenopodium* spp. seeds per 100 cm^3 soil		Sampling variance ($S^2_{\bar{x}}$)[b]
		$\bar{X} \pm SD$[a]	Range	
Small	58	2.6*a*±1.1	0.7−5.1	0.020*a*
Medium	28	2.9*a*±0.8	1.7−4.4	0.026*a*
Large	20	2.9*a*±0.8	1.7−4.6	0.032*a*

[a] Values followed by the same letter are not significantly different at the 0.01 level based on Scheffé's test.

[b] Values followed by the same letter are not significantly different at the 0.01 level based on the pairwise F-test.

Consequently, these soil cores were excluded from the analysis. An analysis of variance was carried out to detect differences between the estimates of means made from samples by the different auger sizes (SAS Institute Inc. 1982). Pairwise F-tests were done to detect differences in the sampling variance of the mean from different auger sizes.

Population distribution

Before a comparison of the different sampling technique can be made, it is necessary to describe the sampled population. The goodness of fit to a Poisson distribution of the sampled population of *Chenopodium* seeds was examined with a χ^2 test. A t-test was also used to test the population for departure from randomness using the variance to mean ratio (Kershaw 1973). The homogeneity of row and column totals for the number of *Chenopodium* seeds was checked for departure from equidistribution by the χ^2 test.

Sample size

Sample size obtained by a random sampling method and ranging between 4 and 512 units was tested. A natural logarithm transformation was used to demonstrate a linear relationship between sample size and both the MC sampling variance and population sampling variance using a regression program (Orloci and Kenkel 1985). Student's t-tests were used to verify the equality of regression coefficients and the equality of elevation of both regression lines (Zar 1974).

Multiple pairwise comparisons of the MC sampling variance for sample sizes between 45 and 120 units were conducted using the F-test (F = MC $S^2_{\bar{x}_1}$ / MC $S^2_{\bar{x}_2}$). As the number of items to be compared increases, so does the chance of making a type I error; the probability of such an error is α. If the level of probability of any such error is to be maintained, the probability of a type I error for any of all possible pairs of comparisons (α') must be so small that their summation does not exceed the desired α (Sokal and Rohlf 1981). The α' for 12 pairwise comparisons was calculated as 0.001 using the formula

$$\alpha = \frac{\alpha' k(k-1)}{2}$$

where α is the desired probability level, α' is the probability level of any pairwise comparisons, and k is the number of items to be compared. Using the α' as the probability of rejection, the critical F value to be used for the F-test was calculated as $F_{[(2),\,0.001,\,399,\,399]} = 1.365$ using the FPROBS program (Orloci and Kenkel 1985).

Multiple pairwise comparisons of the the MC sampling variance of the four sampling methods for a given total number of sampling units (n = 64) were also made using the F-test described above.

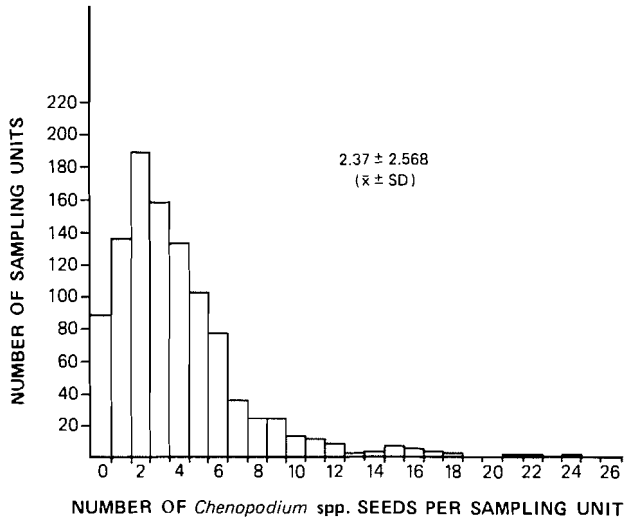

FIG. 1. Distribution of *Chenopodium* spp. seeds per sampling unit for a 1.18 ha area (n = 1024 units).

Results

Auger size

No significant differences were detected between auger sizes in their estimation of the mean number of *Chenopodium* spp. seeds per 100 cm^3 of soil (Table 2). Even though sampling variance decreased with decreasing auger sizes, the precision was improved, though not significantly, by taking a larger number of samples with the smaller auger (Table 2).

Population distribution

The pattern of the number of *Chenopodium* seeds per sampling unit is illustrated in Fig. 1. The distribution is skewed to the right, and a larger than expected proportion of sampling units with high seed numbers was observed. The χ^2 test revealed a strong departure from randomness, since the goodness of fit to a Poisson distribution was rejected (χ^2 = 943.2, $P < 0.001$). The variance to mean ratio was greater than 1, thereby initiating a clustered distribution (σ^2/μ = 2.78, t = 40.3, $P < 0.001$). The hypothesis of homogeneity of the total number of *Chenopodium* spp. seeds for the columns and rows of the matrix was rejected for both totals (χ^2_{row} = 58.47, $P < 0.001$; χ^2_{column} = 868.67, $P < 0.001$). The divergence of row totals for the matrix, taken across the corn rows, is not very pronounced (Fig. 2A), but the differences between column totals are particularly evident (Fig. 2B). These show differences in seed number between corn rows.

Sample size

As the total number of sampling units increased, the MC sampling variance decreased, following an exponential decay curve (Fig. 3). This relation is linear, as it was observed by plotting $\ln(S^2_{\bar{x}})$ versus $\ln(n)$ for the population (Fig. 4). The regression coefficients and the elevations of the regression lines of the population sampling variance ($\ln y = 1.885 - 0.999 \ln x$) were not significantly different (t_b = 1.46, $t_{[0.05(2)60]}$ = 2.000; $t_{elevation}$ = 0.63, $t_{[0.05(2)61]}$ = 1.995). There

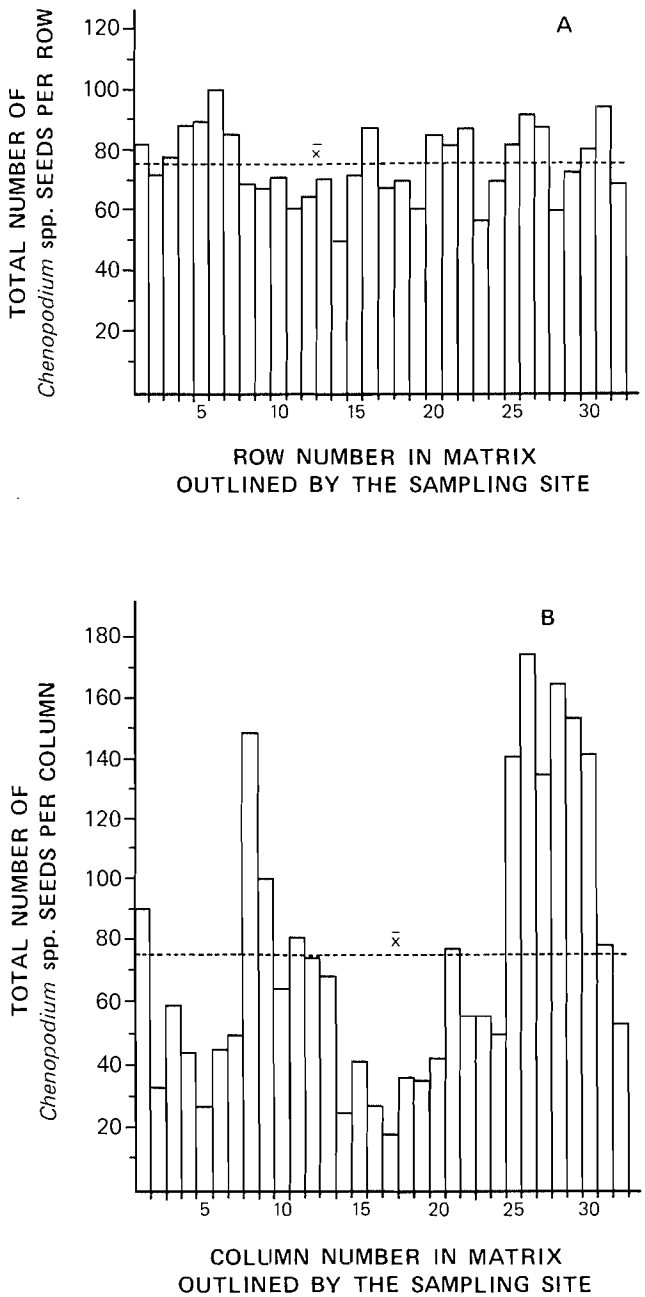

FIG. 2. Total number of *Chenopodium* spp. seeds per row and column of the matrix outlined by the sampling site. The mean of the total number of *Chenopodium* spp. seeds per row or column is 75.9. (A) Rows of matrix run perpendicular to corn rows in the sampling site. (B) Columns of matrix run parallel to corn rows in the sampling site.

were significant differences ($P < 0.05$) in the MC sampling variance between sample size of 60 and 45 units or less and 75 units or more (Table 3).

Sampling methods

With simple random sampling, as the total number of sampling units increases, the MC sampling variance decreases, following an exponential decay curve (Fig. 3). With systematic sampling, as the total number of sampling units increases, the MC sampling variance decreases (Fig. 5A). For any sample size (total number of sampling units), the MC sampling variance can fluctuate depending on the configuration of the sampling interval used, but such fluctuations are less pronounced as the sample sizes increases (Fig. 5A).

TABLE 3. Comparison of the Monte Carlo estimates of sampling variances (MC $S^2_{\bar{x}}$) for different sample sizes

Sample size	MC $S^{2a}_{\bar{x}}$
40	0.170*a*
45	0.165*a*
50	0.146*ac*
55	0.137*ac*
60	0.116*bc*
65	0.098*bd*
70	0.091*bde*
75	0.075*def*
80	0.063*f*
90	0.067*ef*
100	0.072*ef*
120	0.058*f*

[a]Values followed by different letters are significantly different at the 0.001 level based on the test of equality of variance for 12 pairwise comparisons. The rejection level (α') for each comparison is $F_{[(2)0.001,399,399]} = 1.365$.

With stratified random sampling, as the total number of sampling units increases, the MC sampling variance decreases (Fig. 5B). For a fixed total number of sampling units, extensive stratification within the matrix does not lead to a decrease in the MC sampling variance (Fig. 5C). The orientation of the strata has a slight influence on the precision of the mean. Subdivision of the matrix leading to vertically oriented strata (i units $> j$ units) results in smaller MC sampling variance and vice versa (Fig. 5B). This trend attenuates as the total number of sampling units increases.

With cluster sampling, as the total number of sampling units increases, the MC sampling variance fluctuates (Fig. 5D). However, for a given total number of sampling units collected, the MC sampling variance decreases as the number of clusters increases (Fig. 5E). The orientation of the clusters has also a strong influence on the precision of the mean. Sampling with horizontally shaped clusters (i units $< j$ units) resulted in a substantial decrease in the MC sampling variance (Fig. 5D). This trend is consistent regardless of either the total number of sampling units or the number of clusters sampled.

For a total of 64 sampling units collected, the comparison of all four sampling methods indicated that cluster sampling is significantly different ($P < 0.001$) from all other sampling methods (Table 4).

Discussion

Auger size

For the same total volume of soil sampled, the smallest auger permitted the collection of a greater number of sampling units. This is important because a larger sample size can provide a more precise estimate of the mean (Kershaw 1973; Elliott 1977). This result was also confirmed by Bigwood and Inouye (1988). Tulikov *et al.* (1981) warned that small soil samples (< 100 g) significantly overestimated the total number of seeds in the soil ($P < 0.01$). This was not confirmed by our results (Table 2).

Population distribution

There was evidence (χ^2 test, variance to mean ratio) for a

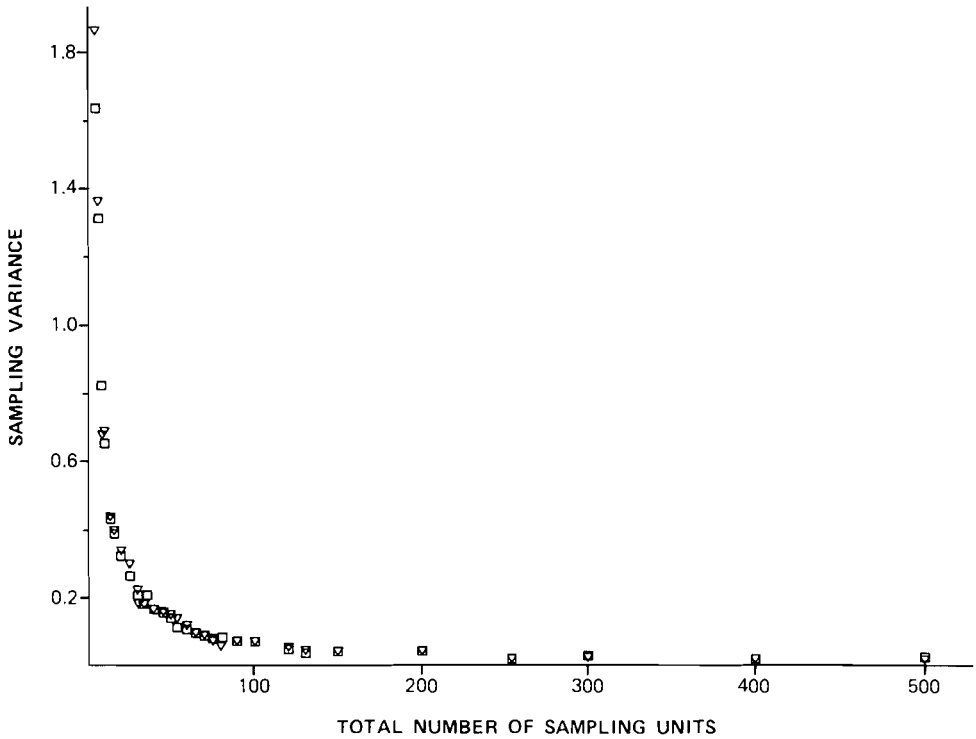

FIG. 3. The effect of different sample sizes on the population sampling variance (□) and the Monte Carlo estimate of sampling variance (▽) when a simple random sampling method is used.

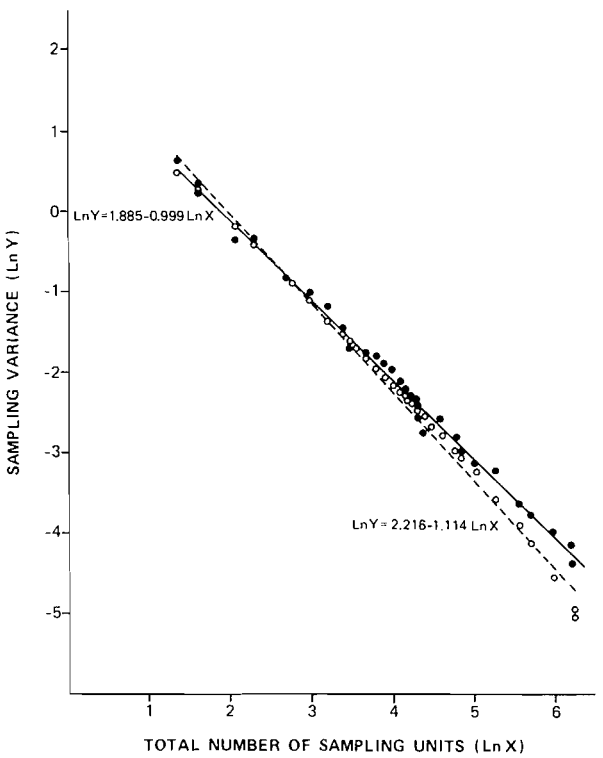

FIG. 4. Linear regressions of the transformed Monte Carlo estimate of sampling variance (●) and the transformed population sampling variance (○) on the transformed sample size.

TABLE 4. Comparisoon of the Monte Carlo estimate of sampling variance and the theoretical sampling variance for different sampling methods of a given sample size (n = 64)

Sampling method	Monte Carlo sampling variance[a]
Random	0.096*a*
Systematic	
(4 × 4)	0.088*a*
(8 × 2)	0.027*b*
(2 × 8)	0.094*a*
Stratified random (32 strata)	
(16 × 2)	0.080*a*
(2 × 16)	0.101*a*
(8 × 4)	0.086*a*
(4 × 8)	0.075*a*
Cluster (2 clusters)	
(16 × 2)	1.062*c*
(2 × 16)	0.249*e*
(8 × 4)	0.912*cd*
(4 × 8)	0.769*d*

[a]Values followed by different letters are significantly different at the 0.001 level based on the test of equality of variance for 12 pairwise comparisons. The rejection level (α') for each comparison is $F_{[(2)0.001,399,399]} = 1.365$.

strong clumping pattern (or clustered distribution) of seeds in the soil. This could be caused in part by the seed dispersal pattern of the parent plants. In weedy species such as *Chenopodium* spp., the seeds are often shed close to the parent plant, thereby creating a clumping pattern of seeds in or on the soil (Major and Pyott 1966).

Upon closer examination, large differences between column totals of the matrix were observed in the number of *Chenopodium* seeds. Since the columns in the matrix corresponded to corn rows in the field, the seed pattern in the sampling site is

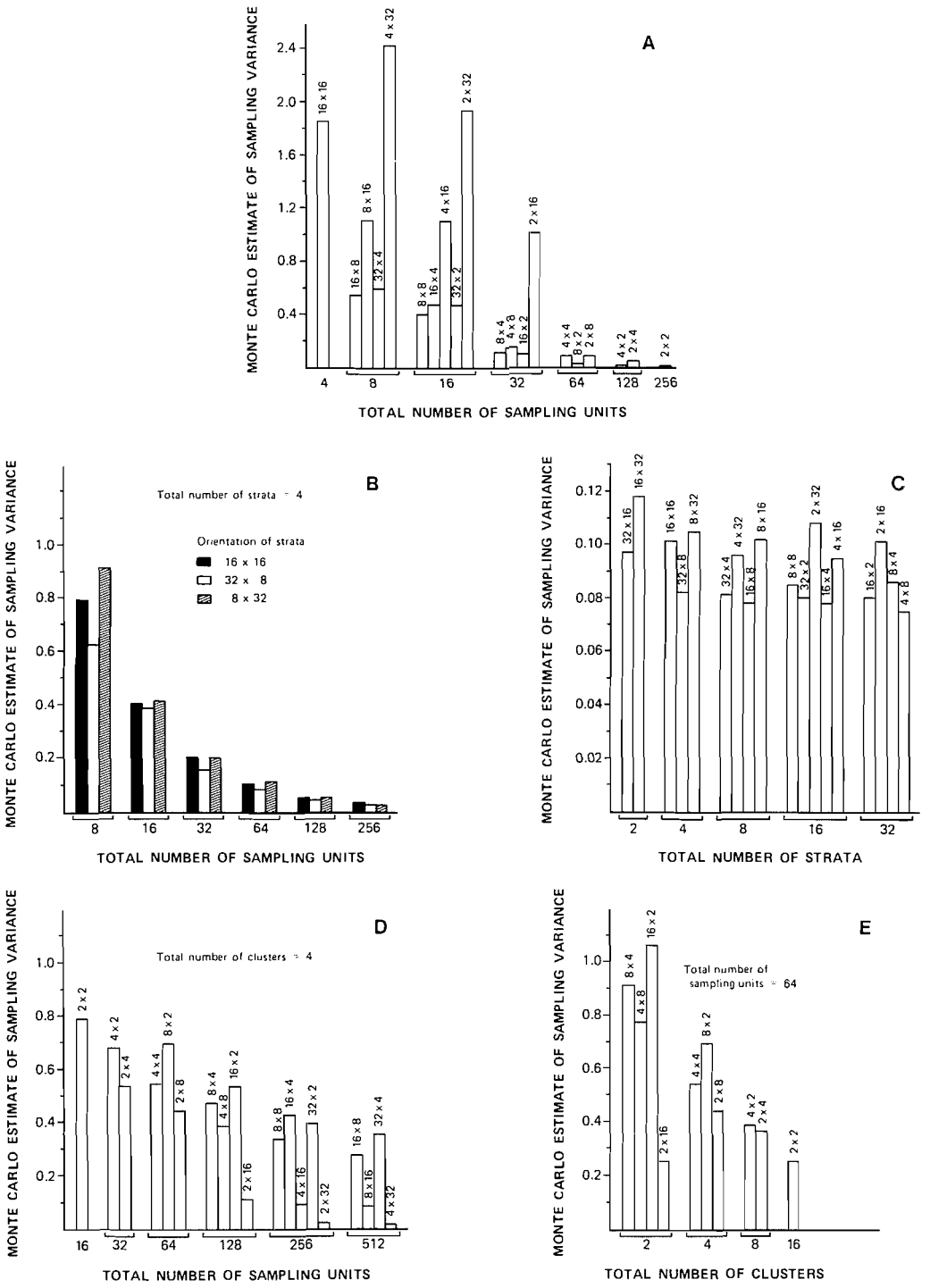

FIG. 5. The effect of sample size, stratification, and clustering on the Monte Carlo estimate of sampling variance using the different sampling methods. (A) Effect of increased sample size using the systematic sampling method. The numbers above each bar represent the sampling intervals. The first number represents the ith row interval and the second number, the jth column inverval. (B) Effect of increased total number of sampling units using stratified random sampling. The matrix was divided into four strata. The first number of the orientation of the strata refers to the number of i rows of the matrix in a stratum and the second number, to the number of j columns of the matrix in a stratum. (C) Effect of increased stratification within the matrix for a total of 64 sampling units using the stratified random sampling method. The numbers above each bar represent the orientation of the strata. The first value refers to the number of i rows of the matrix in a stratum and the second value, to the number of j columns of the matrix in a stratum. (D) Effect of increased total number of sampling units using the cluster sampling method. Four clusters were sampled within the matrix. The numbers above each bar represent the cluster shape used for sampling. The first value refers to the number of i rows of the matrix in each cluster and the second value, to the number of j columns of the matrix in each cluster. (E) Effect of increasing the number of the clusters for a total of 64 sampling units sampled using the cluster sampling method. The numbers above each bar represent the cluster shape. The first value refers to the number of i rows of the matrix in each cluster and the second value, to the number of j columns of the matrix in each cluster.

one of groups of corn rows with high numbers of *Chenopodium* seeds and groups of corn rows with low seed numbers. These large variations were not apparent across corn rows.

Heterogeneity in the seed bank between crop rows may be explained by the movement of machinery along the same axis of the field each year. This results in dispersal of the weed seeds along the crop rows rather than across them. Eventually, areas that consistently had higher weed populations may have developed large seed banks in the soil. Areas that had either good weed control or a microenvironment favouring germination or rapid decay of weed seeds may have ended up with relatively small seed banks.

Sample size

The regression lines of the population sampling variance and the MC sampling variance coincided, thereby confirming that the Monte Carlo estimate did come from the same population. This linear regression indicated that doubling the sample size reduced the sampling variance by half (or doubled the precision of the estimate of the mean). The inflection point of the curve was defined as the level beyond which the sampling effort required to substantially decrease the sampling variance increased dramatically. Up to a total of 75 sampling units, greater precision is obtained by increasing the number of sampling units; however, beyond this sample size, the reduction of the sampling variance (or gain in precision of the mean) does not compensate for the substantial increase in sampling effort.

Goyeau and Fablet (1982) reported that if the distribution is expected to be aggregated, then the sample size should be greater than 100. They also found that if the expected mean seed density per sampling unit ranged between 1 and 5, then a sample size ranging between 100 and 200 was needed to estimate the mean seed density with 20% precision ($\alpha = 0.01$). This was further confirmed by Lopez *et al.* (1988). Morin and Wojewedka (1985) suggested 90 cores to estimate the seed bank of heavy infestations of *Digitaria sanguinalis* (L.) Scop. with 20% precision ($\alpha = 0.05$). Similarly, Barralis *et al.* (1986) found that 90 cores gave precision estimates varying between 20 and 70% when the mean number of seeds per core varied between 0.1 and 5 and recommended a sample size of 100 units for abundant species.

The distribution of *Chenopodium* spp. seeds over the study site was both clustered and high ($\bar{x} \pm SD = 2.37 \pm 2.568$) and ranged between 0 and 20 seeds/unit. Both of these attributes suggest that a sample size between 100 and 200 would be needed (Goyeau and Fablet 1982; Lopez *et al.* 1988). However, because of the extra labour needed to collect and process additional sampling units without an appreciable decrease in the sampling variance (Table 3), we suggest that a sampling size ranging between 60 and 75 small sampling units (1.9 cm in diameter) should be sampled to describe the seed bank of abundant species such as *Chenopodium* spp. This is in agreement with Rabotnov (1958) but lower than recommendations by Barralis *et al.* (1986) and Lopez *et al.* (1988).

Sampling methods

The seed distribution in the soil over the area surveyed exhibited banks of high and low seed densities parallel to the corn rows. Randomly located clusters within the surveyed area would reflect this pattern, accentuating the variation between clusters. Being in direct contradiction with the clustering principle, which aims at increasing within-cluster variance and decreasing among-cluster variance (Stuart 1976), a substantial loss in precision results. Indeed, the MC sampling variances for the cluster sampling method were significantly different ($P < 0.001$) from those for all other sampling methods tested.

Similarly, the placement of systematic units in our sampling site resulted in poor precision when units were taken, so only a few columns (or corn rows) were sampled. This problem was eliminated when units were taken at equal distances in both directions (i units = j units) or when sampling was more frequent across corn rows than along them. However, this systematic sampling method is not recommended, since an unbiased estimate of the sampling variance is not available (Sampford 1962).

Thus, either random or stratified random sampling could be used to study the seed banks of weeds in the soil. The multiple pairwise comparisons of sampling methods indicated clearly that systematic and stratified random sampling methods were as good as random sampling. However, stratified random sampling is more advantageous in cultivated fields, since stratification along crop rows can provide separate estimates of the seed bank for different areas of a field where heterogeneity is suspected, based on known physical, biological, or environmental characteristics of the area.

Acknowledgements

This research was conducted while D. L. Benoit was on leave from Agriculture Canada Research Station, Saint-Jean-sur-Richelieu, Québec. We thank Mr. John Tucsok for letting us work on his farm, Peter Andreae for designing the soil washing machine, Steve Cruise and Tim Ellis for assisting with the sampling, David Hill and Andrea Patch for washing and cleaning soil samples, and Claire Ménard and Ghyslaine Brodeur for typing the manuscript. Financial support from the Ontario Pesticide Advisory Committee and the Natural Sciences and Engineering Research Council of Canada through an operating grant to P. B. Cavers is gratefully acknowledged.

BARRALIS, G., CHADOEUF, R., and GOUET, J.-P. 1986. Essai de détermination de la taille de l'échantillon pour l'étude du potentiel semencier d'un sol. Weed Res. **26**: 291–297.

BIGWOOD, D. W., and INOUYE, D. W. 1988. Spatial pattern analysis of seed banks: an improved method and optimized sampling. Ecology, **69**: 497–507.

CHAMPNESS, S. S. 1949. Notes on the buried seed populations beneath different types of ley in their seeding year. J. Ecol. **37**: 51–66.

COCHRAN, W. G. 1977. Sampling techniques. 3rd ed. John Wiley & Sons, Inc., New York.

DOSPEKHOV, B. A., and CHEKRYZHOV, A. D. 1972. Counting weed seeds in the soil by the method of small samples. Izv. Timiryazev. Skh. Akad. **2**: 213–215.

ELLIOTT, J. M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. 2nd ed. Freshwater Biol. Assoc. Sci. Publ. No. 25.

FAY, P. K., and OLSON, W. A. 1978. Techniques for separating weed seed from soil. Weed Sci. **26**: 530–533.

FEAST, P. M., and ROBERTS, H. A. 1973. Note on the estimation of viable weed seeds in soil samples. Weed Res. **13**: 110–113.

GOYEAU, H., and FABLET, G. 1982. Étude du stock de semences de mauvaises herbes dans le sol: le problème de l'échantillonnage. Agronomie (Paris), **2**: 545–552.

HAYASHI, I., and NUMATA, M. 1971. Viable buried-seed population in the *Misanthus* and *Zoysia* type grasslands in Japan — Ecological studies on the buried-seed population in the soil related to plant succession. VI. Jpn. J. Ecol. **20**: 243–252.

KERSHAW, K. A. 1973. Quantitative and dynamic plant ecology. 2nd ed. Edward Arnold (Publishers) Ltd., London.

KROPÁČ, Z. 1966. Estimation of weed seeds in arable soil. Pedobiologia, **6**: 105–128.

LECK, M. A., PARKER, V. T., and SIMPSON, R. L. (*Editors*). 1989. Ecology of soil seed banks. Academic Press Inc., San Diego, CA.

LOPEZ, C., ABRAMOVSKY, P., VERDIER, J. L., and MAMAROT, J. 1988. Estimation du stock semencier dans le cadre d'un essai étudiant l'influence de systèmes culturaux sur l'évolution de la flore adventice. Weed Res. **28**: 215–221.

MAJOR, J., and PYOTT, W. T. 1966. Buried viable seeds in two California bunchgrass sites and their bearing on the definition of a flora. Vegetatio, **13**: 253–282.

MALONE, C. R. 1967. A rapid method for enumeration of viable seeds in soil. Weeds, **15**: 381–382.

MORIN, C., and WOJEWEDKA, A. 1985. Evaluation du potentiel semencier d'un sol. *In* Septième Colloque International sur l'Écologie, la Biologie et la Systématique des Mauvaises Herbes, 9–11 octobre 1984. COLUMA-EWRS, Paris. pp. 55–62.

NUMATA, M., HAYASHI, I., KOMURA, T., and OKI, K. 1964. Ecological studies on the buried-seed population in the soil as related to plant succession. I. Jpn. J. Ecol. **14**: 207–215.

ORLOCI, L., and KENKEL, N. C. 1985. Introduction to data analysis. International Co-operative Publishing House, Fairland, MD.

RABOTNOV, T. A. 1958. On methods of studying the supply of living seeds in soils of meadows. Bot. Zh. (Leningr.), **43**: 1572–1581.

ROBERTS, H. A. 1981. Seed banks in soils. Adv. Appl. Ecol. **6**: 1–55.

SAMPFORD, M. R. 1962. An introduction to sampling theory. Oliver and Boyd Ltd., Edinburgh, U.K.

SAS INSTITUTE INC. 1982. SAS user's guide: statistics. SAS Institute Inc., Cary, NC.

SOKAL, R. R., and ROHLF, F. J. 1981. Biometry. 2nd ed. W. H. Freeman and Company, San Francisco, CA.

STANDIFER, L. C. 1980. A technique for estimating weed seed populations in cultivated soil. Weed Sci. **28**: 134–138.

STUART, A. 1976. Basic ideas of scientific sampling. 2nd ed. Charles Griffin and Company Ltd., London.

THORSEN, J. A., and CRABTREE, G. 1977. Washing equipment for separating weed seed from soil. Weed Sci. **25**: 41–42.

TULIKOV, A. M., FROLOVA, T. N., and CHIBOTAR, FV. V. 1981. A comparative evaluation of sample sizes for calculating the number of weed seeds in soil. Izv. Timiryazev. Skh. Akad. **5**: 190–193.

ZAR, J. H. 1974. Biostatistical analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.