Landscape and Urban Planning, 20 (1991) 277–283
Elsevier Science Publishers B.V., Amsterdam

Approach to determine effective sampling size for urban street tree survey

Wen-Quan Sun and Nina L. Bassuk

Department of Floriculture and Ornamental Horticulture, 20 Plant Science Building, Cornell University, Ithaca, NY 14853, USA

(Accepted 22 March 1991)

ABSTRACT

Sun, W.-Q. and Bassuk, N.L., 1991. Approach to determine effective sampling size for urban street tree survey. *Landscape Urban Plann.*, 20: 277–283.

The collection of data on the current condition of street trees is the first step in developing an urban street tree planning and maintenance program. The objective of this study was to establish a recommendation of sampling size to facilitate street tree surveys through computer simulation. The hypothetical tree populations used for sampling simulation were set up according to published literature. Simulation results concluded that sample quality was improved little after a certain sampling size, but it declined significantly at a smaller size. With the assumption that a relative deviation from the true value (X) within $\pm 10\%$ (i.e. $X \pm 10\% \cdot X$) was permitted for estimating the percentage of species in a street tree population, the sampling sizes for urban street tree surveys were recommended on the basis of simulation results. The accuracy of surveys with recommended sampling sizes was estimated by using the street tree population of the City of Ithaca, New York. It was confirmed that recommended sampling size provided the estimates with relative deviations of approximately 10% for major species.

INTRODUCTION

The development of methods for inventory of the street tree population is the first issue in the street tree planning process (Amir and Misgav, 1990). The inventory collects information on street trees about species, location and their current condition (size, age, damage, disease, insect pests, etc.). Such information works as a baseline for long-term planning of the street tree system (e.g. developing goals, implementing policies, selecting tree species, etc.). A street tree inventory is also needed for budgeting and day-to-day maintenance. Without regularly updated tree data, it is difficult to develop a tree management plan (Jim, 1986). To gather street tree data, several street tree projects used complete inventories (Green, 1984; McPherson and Rowntree, 1989). Complete surveys, however, are costly and for many cities it is almost impossible to regularly carry out such works because of limited funding. A sampling survey is a cheaper and more efficient method. It can also provide sufficient and accurate data for the purposes of planning and management, if properly conducted.

A review of the literature found that methods of street tree sampling surveys used by researchers differed greatly, and included random sampling and stratified random sampling, with sampling units from single trees, blockface to town and sampling sizes from 1 to 92% of the population sizes (Wray and Mize, 1985; Jim, 1986; Denne, 1987; Talarchek, 1987; Wong et al., 1988). Various techniques of street tree inventories have been reviewed by McBride and Nowak (1989). However, much less attention has been paid to the sampling size, one of the most influential factors affecting the accuracy of the survey. The present study

0169-2046/91/$03.50

aimed to determine appropriate sampling sizes for street tree surveys through computer simulation.

SAMPLING SIMULATION METHOD

Hypothetical street tree populations

Through the literature survey, we found that street tree populations usually were composed of less than 50 species (excluding the rare species) and that more often a few species (less than ten) made up a great proportion of the entire population (Bassuk and Jaenson, 1990). In the three hypothetical populations (Fig. 1), it was assumed that each hypothetical population was composed of 50 tree species. Population A has the highest species diversity, and population C has the lowest diversity. In population A, every species is evenly presented

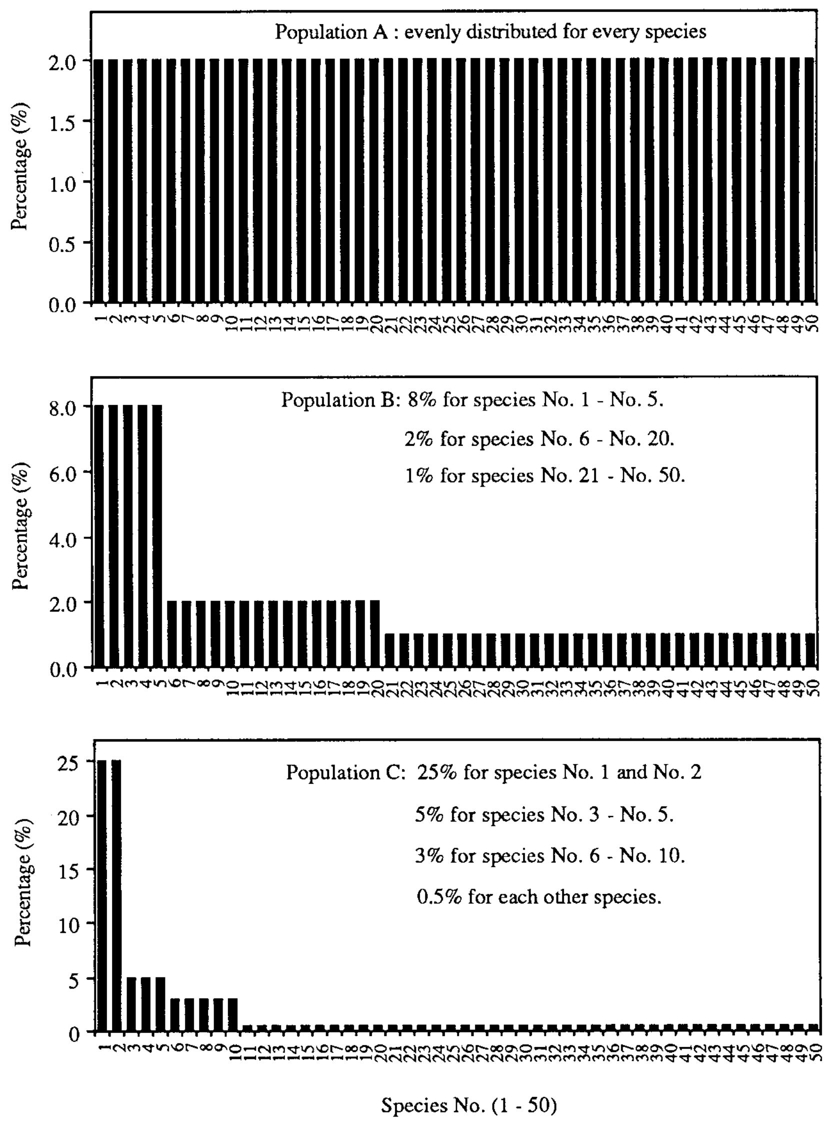

Fig. 1. The species compositions in three hypothetical street tree populations.

(2% of the population). In population B, five major species make up 40% of the population, 15 species make up 30% and 30 minor species make up another 30%. In population C, two species represent 50% of street trees. The top ten species make up 80% of the population and 40 other species only account for 20% (Fig. 1). Most urban street tree populations are like populations B and C.

Population size is another important factor affecting the accuracy of the sampling inventory. Four population sizes, 1000, 2000, 5000 and 10 000 trees were then used in the sampling simulation. Although the street tree populations of most medium to large cities are larger than 10 000 trees (McPherson and Rowntree, 1989), urban tree populations are well segregated by man-made zones and districts. In different zones or districts, street tree subpopulations usually show remarkable differences (species diversity, tree conditions, etc.). Therefore, stratified inventory techniques were used more often in street tree surveys. The range of 1000–10 000 trees should be more appropriate for this situation and for small cities.

Sampling simulation

A data base of hypothetical populations was set up on a AT & T-PC6300 personal computer by using the software package "Minitab" (Ryan et al., 1985). The 50 species were coded for with integers from 1 to 50, and the data base contained the number of codes corresponding to the number of trees of a particular species that the code represented. For example, the data base of the 2000-tree population A contained each code 40 times. For each sampling size, ten samples were randomly drawn without replacement. Sampling size was gradually increased from 5 to 40%. The codes in samples were sorted and samples were compared with their population.

Internal sample quality indicators

A sample should be representative of the three population from which it is drawn. Two internal indicators were used to monitor sample quality in this study. One was the average percentage of species missed in ten samples. It indicated how many species in the population were missed in the samples. Another indicator was the coefficient of variation (CV)[a] of species diversity index (SDI) in samples. (CV (%) is calculated as the standard deviation of SDI divided by the mean SDI.) SDI is an index which describes the state of diversity in a tree population, and can be calculated by the following formula

$$\mathrm{SDI}=\frac{\sum N_j\cdot(\sum N_j-1)}{\sum N_j(N_j-1)}$$

where N_j is the number of trees in the jth species and $\sum N_j$ is the number of total trees in a population or a sample. If the frequency distribution for every species in a sample is sufficiently similar to that in the population, the sample and the population would have very close SDI values. Therefore, the CV (%) of SDI in ten samples shows the variation from sample to sample in estimating the proportion of individual species in a population. A valid sampling size should be able to ensure the variation from sample to sample within a certain degree.

These two indicators are the functions of sampling size. The higher accuracy of survey required, the larger the sampling size needed. Before sample size can be recommended, it is necessary to set an accuracy level which could be widely acceptable. Assume that a relative deviation from true value (X) within $\pm 10\%$ (i.e. $X\pm 10\%\cdot X$) is permitted for estimating the percentage of species in a street tree population, then SDI variation in samples needs to be controlled to within $\pm 10\%$, and the CV

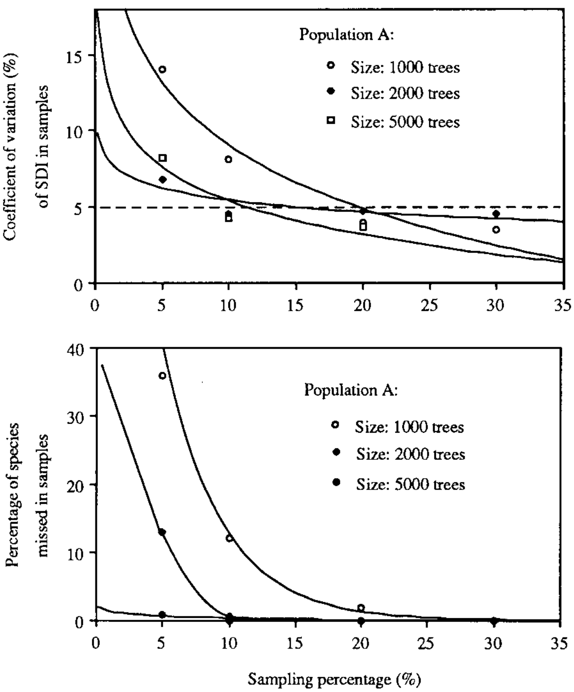

Fig. 2. The coefficient of variation (%) of SDI and the average percentage of species missed in ten samples as a function of population size and sample size in population A.

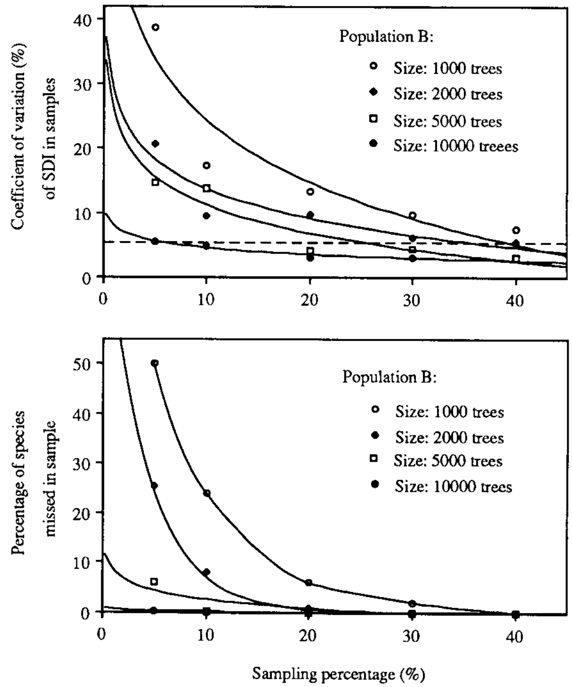

Fig. 3. The coefficient of variation (%) of SDI and the average percentage of species missed in ten samples as a function of population size and sample size in population B.

should be no more than 5%[1]. A limit for species missing is also set at 5%.

SAMPLING SIZES AND SAMPLE-TO-SAMPLE VARIATION

SDI of three hypothetical populations A, B and C were on average 51.5, 24.7 and 7.3, respectively, with very slight decreases as population size increased. The variation of SDI in ten samples primarily depended on sampling size and population size. As sampling size was increased, the CV of SDI and percent species missing in samples declined sharply at first and then slowly (Figs. 2–4). The pattern revealed that sample quality could not be improved much if sampling size was increased past a certain point, but that sample quality declined significantly at a smaller sampling size.

In population A, the CV of SDI and percentage of species missing in samples were less than 5% when sampling size was larger than 20% for populations of 1000 or 2000 trees and 10% for populations of 5000 trees (Fig. 2). In population B, the percentage of species missing was less than 5% at sample sizes of 25%, 12%, 5% and 5% for populations of 1000, 2000, 5000 and 10 000 trees, respectively; the CV, however, did not decline to 5% until sample size was increased to 40%, 35%, 25%, and 7%, respectively (Fig. 3). To satisfy the 5% limits of CV and species missing in population C, sample size had to be 45%, 35%, 20% and 10%, respectively (Fig. 4). Compared with population B, larger sample sizes are required for population C, particularly when the survey is also concerned with minor species. Based on these simulation results, the recommended

[1] $SDI=\Sigma N_j(\Sigma N_j-1)/\Sigma N_j(N_j-1)$. If the N_j changes to M_j, the numerator $\Sigma N_j(\Sigma N_j-1)$ remains unchanged because $\Sigma N_j=\Sigma M_j$, while the denominator $\Sigma N_j(N_j-1)$ will vary. In the extreme case, a 10% departure of N_j will induce an approximate 20% variation of SDI. Therefore it is reasonable to set a 10% limit for SDI deviation. Assume that SDI in samples is normally distributed, then a CV of 5% will give a 95% confidence interval of SDI within ±10%, because the Z value equals 1.96 at $a=0.05$.

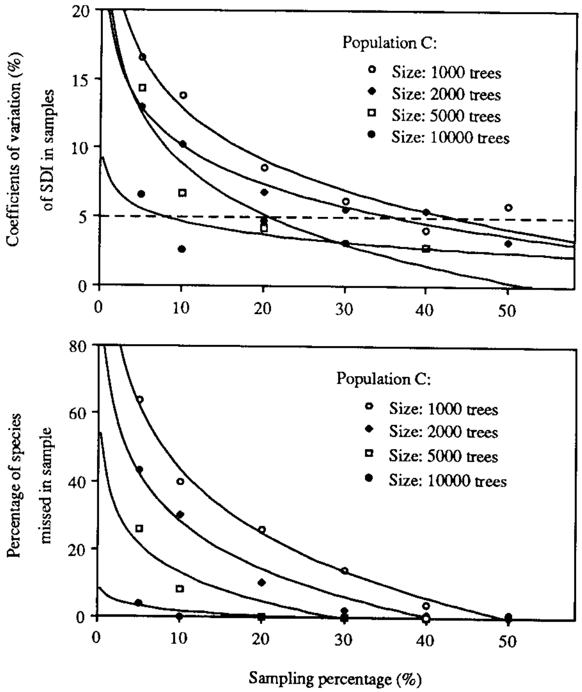

Fig. 4. The coefficient of variation (%) of SDI and the average percentage of species missed in ten samples as a function of population size and sample size in population C.

TABLE 1

Recommended sample size for street tree survey (% of population)

Population size (trees)	Species diversity[1]		
	High SDI > 35	Middle SDI 10–35	Low SDI < 10
< 2000	20	40	50
2000–5000	10	30	35
5001–10 000	5	20	25
> 10 000	5	10	15

[1]Species diversity (high, middle or low) corresponds to the three hypothetical populations A, B and C, respectively.

sampling sizes for urban street trees surveys are given in Table 1.

CASE STUDY

To examine the accuracy of sampling inventories with the recommendation in Table 1, street trees of the City of Ithaca, NY were used in the case study. In 1987, the Shade Tree Advisory Committee (STAC) in the City of Ithaca conducted a complete street tree survey (Bassuk and Jaenson, 1988). The STAC survey showed that the city had in total 5541 street trees, which belonged to 103 species. However, the top five species accounted for 72.2% and the top 12 species for 86.2% of the total population. The population diversity was very low, with an SDI of 7.6. From Table 1, a sample size of 25% (1385 trees) was suggested.

A data base of all the street trees in the City of Ithaca was established by the methods described earlier. Then, three random and independent samples with 25% size were generated by computer, and the sample-estimated percentages and their relative deviations (relative deviation (%) is calculated as |sample-estimated % − actual %|/actual %) from the population for 17 major species (accounting for 90.4% of the population) were calculated. The relative deviations were less than 10% for the top five species, of which each species made up more than 5% of the population. To obtain the 95% confidence intervals for 17 major species, another 25 random and independent samples were generated. The 95% confidence intervals of the sample-estimated percentage for those species are presented in Table 2. The comparison between the actual proportion in the population and 95% confidence intervals confirmed that a sampling size of 25% as adequate for major species in the street tree population of the City of Ithaca, NY. The relative deviations using the 95% confidence interval limits were also around 10% for the top five species.

APPLICATION OF SAMPLING SIZE RECOMMENDATION

Table 1 provided guidance on the size of samples. To use Table 1, however, the size and species diversity of the population need to be roughly guessed. An estimate of within ±20–30% should be accurate enough for the total

TABLE 2

The actual and sample-estimated percentages for 17 major species (90.39%) in the street tree population, Ithaca, NY

Species	Actual %[1]	Estimated %[2]
Acer platanoides	33.14	31.24–34.46
Acer saccharum	19.06	17.37–21.03
Gleditsia triacanthos	8.81	7.50–10.14
Acer saccharinum	5.99	4.97–6.70
Acer rubrum	5.23	4.76–6.12
Malus spp.	2.56	1.52–3.67
Platanus × acerifolia	2.31	1.80–3.33
Ginkgo biloba	2.24	1.49–3.09
Fraxinus pennsylvania	1.88	1.17–2.41
Tilia cordata	1.79	0.90–2.59
Quercus rubra	1.71	1.37–2.25
Pyrus calleryana	1.59	1.10–2.08
Hibiscus syriacus	0.99	0.67–1.41
Taxus spp.	0.96	0.69–1.37
Acer negundo	0.72	0.24–1.04
Quercus palustris	0.71	0.34–1.25
Zelkova serrata	0.70	0.33–1.10

[1]Data from a complete survey (Bassuk and Jaenson, 1988).
[2]Estimated % are the 95% confidence intervals based on 25 random and independent samples from sampling simulation with a sample size of 25%.

number of trees. The species diversity, however, is difficult to estimate. One can compare the real tree population with three hypothetical populations in Fig. 1 and determine which hypothetical population is closer to the real population. With a little careful consideration, it would not be difficult to choose a proper sample size. When the population is larger than the maximum size used in this study, (say, 20 000 or 30 000 trees) a smaller sample size could be used.

Although this recommendation was based on an assumption of 50 tree species in a population, it could almost be applied to any street tree populations. In some cities, the number of street tree species may be several times more than 50, but the majority of those species make a relatively small contribution to the whole population. It seems to be always the case that a relatively small number of species forms a large proportion of the population (McPherson and Rowntree, 1989).

Under the condition of lower species diversity, it can be seen from Table 1 that the larger sample size is required. For some surveys, however, those species forming a very small proportion of the population can be ignored and a smaller sample size may be used. Although it is unlikely that the street tree population has a similar degree of species diversity to the hypothetical population A, the definition of a street tree may be varied (Lohmann, 1988). Some urban surveys probably include not just the shade trees along streets, but also other trees in residential areas or home yards. This will tremendously increase species diversity. In this situation, it is reasonable to assume that the concerned population has high diversity.

The accuracy of a sampling inventory of street trees not only depends on sample size, but also on the sample unit used. Sampling using single trees as the sampling unit is more difficult to carry out. Researchers would prefer using area zone (site) as sample unit. An area zone can be a street, block-face (block perimeter along one street), or a certain area unit which divides the whole area into equal plot sites. For instance, Talarchek (1987) used a computer to generate random *X* and *Y* coordinates on a 1:24 000 map of the City of New Orleans, and then selected the block-face nearest the coordinates for sampling. Sampling with area zones, however, creates another problem, the variation of tree density from site to site. To improve zone sampling quality, stratified random sampling and small area units like a block-face (rather than a large unit, such as street and district) could be used. We believe, however, that the sample sizes recommended in Table 1 will also be useful whatever sampling method is adopted. Using the recommended sampling size instead of a 100% tree survey, makes regular inventories possible for an urban street tree planning and management program.

REFERENCES

Amir, S. and Misgav, A., 1990. A framework for street tree planning in urban areas in Israel. Landscape Urban Plann., 19: 203–212.

Bassuk, N.L. and Jaenson, R.J., 1988. Ithaca street tree survey, City of Ithaca, New York. Urban Horticulture Institute, Cornell University, Ithaca, NY.

Bassuk, N.L. and Jaenson, R.J., 1990. Urban tree species diversity — sampling and recommendations for improvement, In: Abstracts of Contributed Papers for the 23rd International Horticultural Congress, Firenze, Italy, 1: 724.

Denne, M.P., 1987. The tree resource of churchyards of Gwynedd, Wales, Arboric. J., 11: 33–52.

Green, T.L., 1984. Maintaining and preserving wood parks. J. Arboric., 10: 193–197.

Jim, C.Y., 1986. Street trees in high density urban Hong Kong. J. Arboric., 12: 257–263.

Lohmann, G., 1988. How valuable are the street trees of Lambeth Borough? Arboric. J., 12: 1–16.

McBride, J.R. and Nowak, D.J., 1989. Urban park tree inventories. Arboric. J., 13: 345–361.

McPherson, E.G. and Rowntree, R.A., 1989. Using structural measures to compare twenty-two US street tree populations. Landscape J., 8: 13–23.

Ryan, B.F., Joiner, B.L. and Ryan, T.A., 1985. Minitab. PWS-KENT, Boston.

Talarchek, G.M., 1987. Indicators of urban forest condition in New Orleans. J. Arboric., 13: 217–224.

Wong, T.W., Good, J.E.G. and Denne, M.P., 1988. Tree root damage to pavements and kerbs in the city of Manchester. Arboric. J., 12: 17–34.

Wray, P.H. and Mize, C.W., 1985. Species adapted for street-tree environments in Iowa. J. Arboric., 11: 249–252.