Technical Note

Number of Samples Required for Estimating Herbaceous Biomass

Michio Tsutsumi,[1] *Shiro Itano,*[2] *and Masae Shiyomi*[3]

Authors are [1]*Researcher, Ecosystem Functions Research Team, National Institute of Livestock and Grassland Science, Nasushiobara, Tochigi, Japan 3292793;* [2]*Chief, Grassland Dynamics Monitoring Laboratory, National Institute of Livestock and Grassland Science, Nasushiobara, Tochigi, Japan 3292793; and* [3]*Director, Ibaraki Study Center, The University of the Air, Mito, Ibaraki, Japan 0056.*

Abstract

Although the precision of herbaceous biomass estimation depends on the sample number, the spatial heterogeneity of the biomass, and sampling procedures, the magnitudes of the influences on the precision have not been clarified. We simulated virtual plant communities based on the gamma distribution to clarify the relationships between the precision of estimating herbaceous biomass and the number of samples, sampling density, spatial heterogeneity of the biomass, and sampling procedures. Using only two parameters, the gamma distribution can approximate the frequency distribution of herbage mass with varying heterogeneity. Our simulations demonstrated that the number of samples is a more influential factor than sampling density on the precision of the herbaceous biomass estimation. Moreover, our simulations confirmed that biomass heterogeneity strongly affected the precision and quantified the magnitude of the influence. When we estimated biomass with random sampling and a 50×50 cm quadrat and accepted estimation error of $\pm$ 10% of the mean for a confidence interval of 95%, the numbers of samples needed were 200, 77, and 9 for very, intermediate, and less heterogeneous grasslands, respectively. Similarly, when we estimated biomass with a ranked set sampling (RSS), then 24, 15, and 4 samples were needed in very, intermediate and less heterogeneous grasslands, respectively. We came to two conclusions: 1) In less heterogeneous grasslands, good precision of estimation can be obtained with a small number of samples, and it is useful to employ RSS. The cutting method, as well as nondestructive methods, will be practical; and 2) estimation for heterogeneous grassland requires a large number of samples, and it is not so useful to employ RSS. For that reason, more research is needed on nondestructive methods.

Resumen

Aunque la precisión de la estimación de la biomasa herbácea depende de número de muestras, la heterogeneidad espacial de la biomasa y los procedimientos de muestro, su magnitud de influencia sobre la precisión no ha sido clarificada. Simulamos comunidades vegetales virtuales en base a la distribución gama para clarificar las relaciones entre la precisión de la estimación de la biomasa herbácea y el número de muestras, densidad de muestreo, heterogeneidad espacial de la biomasa y los procedimientos de muestreo. Usando solo dos parámetros, la distribución gama puede aproximar la distribución de frecuencia de la biomasa herbácea con una heterogeneidad variable. Nuestras simulaciones demostraron que el número de muestras influye más en la precisión de la estimación de la biomasa herbácea que la densidad de muestreo. Más aún, nuestras simulaciones confirmaron que la heterogeneidad de la biomasa afecta fuertemente la precisión y cuantificó la magnitud de la influencia. Cuando estimamos la biomasa con un muestreo aleatorio usando un cuadrante de 50×50 cm y una error de $\pm$ 10% de la media para establecer un intervalo de confianza al 95%, los tamaños de muestras necesitados fueron 200, 77, y 9 para pastizales con heterogeneidad alta, intermedia y baja respectivamente. Cuando estimamos la biomasa con el muestreo de grupos clasificados (RSS), entonces se necesitaron 24, 15, y 4 muestras para la heterogeneidad alta, intermedia y baja del pastizal. Concluimos que: 1) en pastizales menos heterogéneos, se puede obtener una buena precisión con un pequeño número de muestras y fue útil emplear el RSS, el método de corte, así como métodos no destructivos, serían prácticos; y 2) la estimación en pastizales heterogéneos requiere de un gran número de muestras y no es tan útil emplear el RSS. Por esa razón se necesita más investigación en los métodos no destructivos.

Key Words: estimating plant biomass, gamma distribution, ranked set sampling, random sampling, sampling density, spatial heterogeneity

INTRODUCTION

Plant biomass measurement is a fundamental procedure for grassland management and grassland field studies. The precision of herbaceous biomass estimation depends not only on the sample number (Iwasaki 1976), but also on the spatial heterogeneity of the biomass (Iwasaki 1976) and sampling procedures (Cobby et al. 1985). Sampling density was also inferred to affect the precision of biomass estimation (Kayama 1961). However, the magnitudes of the influences of the sample number, the spatial heterogeneity of the biomass, sampling procedures, and sampling density on the precision have not been clarified. Field experiments examining a number of grasslands with different conditions are necessary to clarify this problem; however, such experiments are only marginally practical. Consequently use of a computer simulation is an appropriate method for resolving the issue.

Correspondence: Michio Tsutsumi, Ecosystem Functions Research Team, National Institute of Livestock and Grassland Science, Nasushiobara, Tochigi, Japan 3292793. Email: mcot@affrc.go.jp

Manuscript received 4 July 2006; manuscript accepted 17 March 2007.

Grassland vegetation shows spatial heterogeneity (Clarke et al. 1995; Cid and Brizuela 1998; Tsutsumi et al. 2000). Spatial heterogeneity in grassland vegetation might be generated and maintained through the activities of grazing animals, herbage plants, and soil microbes and their interactions (Bullock 1996; Hirata 2000; Ogura et al. 2002). Several studies have proposed models for describing spatial heterogeneity in grassland plant communities (Remington et al. 1992, 1994; Shiyomi et al. 2000; Barthram et al. 2005). Shiyomi et al. (1983, 1984) demonstrated that the frequency distribution of herbage mass with varying heterogeneity could be approximated using the gamma distribution, a statistical model composed of two parameters (see also Shiyomi et al. 1998; Tsutsumi et al. 2002). The gamma model is rather simple, thereby easing the evaluation of parameters. The Weibull distribution can approximate biomass heterogeneity (Remington et al. 1992, 1994) as well as the gamma distribution; however, it requires three parameters, and estimation of the parameters is more difficult.

There are many procedures for selecting a sampling area, which can be roughly classified into two types, i.e., systematic and random sampling. Systematic sampling is easier to carry out in the field than random, while the values obtained by systematic sampling are likely to be less accurate than those from random (Greig-Smith 1983). On the other hand, McIntyre (1952) has proposed another procedure, ranked set sampling (RSS). The RSS procedure is as follows: 1) randomly select sets of n quadrats, 2) nondestructively rank the quadrats based on the biomass within each set of n quadrats, and 3) for the first set conduct sampling in the highest ranked quadrat, for the second set conduct sampling in the second highest ranked quadrat, and so on, until in the nth set the nth ranked quadrat is sampled. Cobby et al. (1985) reported that use of RSS increased the precision of estimating biomass compared to random sampling from the results of field experiments on grass and grass-clover swards.

We simulated virtual plant communities based on the gamma distribution to clarify the relationships between the precision of estimating herbaceous biomass and number of samples, sampling density, spatial heterogeneity of the biomass, and sampling procedures. Based on those simulations, we can present the theoretically required number of samples for precise estimation under different conditions.

METHODS

Gamma Model

The gamma distribution is given by the following equation for $w \geq 0$ and μ, $p > 0$:

$$f(w) = \frac{w^{p-1}p^p}{\Gamma(p)\mu^p}\exp\left(\frac{-pw}{\mu}\right). \qquad [1]$$

In this equation w denotes the biomass per unit area (quadrat), μ is the mean of biomass per quadrat, p is an index representing the spatial heterogeneity (shape of frequency distribution) of biomass over the grassland, and $\Gamma(p)$ is a gamma function of p (Shiyomi et al. 1983, 1998). Here let σ^2 be the variance among quadrats,

$$p = \frac{\mu^2}{\sigma^2}. \qquad [2]$$

The parameter p is equivalent to the reciprocal of the square of coefficient of variation. As p increases, spatial heterogeneity of the biomass decreases (Shiyomi et al. 1983). The p can be easily approximated with non- or less destructive survey methods proposed by Shiyomi (1991), Itano et al. (2006), and Tsutsumi and Itano (2005). The methods of Shiyomi (1991) and Itano et al. (2006) are carried out by visual observation, and Tsutsumi and Itano's (2005) method uses some nondestructive measure of herbage biomass such as an electronic capacitance probe. In actual data observed on grazed pastures with a 50×50 cm quadrat, p takes a value of about 1–30 (Shiyomi et al. 1984, 1998; Tsutsumi et al. 2000, 2002; Hirata et al. 2002; Ogura et al. 2002). In a well-managed mowed pasture, p will be rather high. Examples of applying the gamma, the normal and the lognormal distributions to observed frequency distribution of herbage biomass are shown in Figure 1.

Though many studies may implicitly assume that samples follow a normal distribution, we can not apply normal distribution to herbage biomass in grassland. In a less heterogeneous grassland (p takes a high value), the biomass can be approximated with the normal distribution (which can be considered as a shape of the gamma distribution) as shown in Figure 1A, while in a highly heterogeneous grassland the frequency distribution of the biomass displays a "long tail" as shown in Figure 1B. If we assume that samples follow the normal distribution, we will have to accept that some samples take negative values (see Fig. 1B), particularly on high heterogeneity of the biomass, because of the symmetrical shape of the normal distribution. On the other hand, a lognormal distribution also could approximate the biomass heterogeneity in the case of Figure 1. However, the gamma distribution has more biological meaning than the lognormal distribution (Shiyomi et al. 1983).

Simulations Using Virtual Communities

According to the assumption that the gamma model describes biomass heterogeneity, biomass was estimated for each of three small (2 500 m^2) and three large (12 500 m^2) virtual pastures. We created six virtual communities as follows. Three sets of 10 000 (for small areas) and three sets of 50 000 (for large areas) gamma random numbers were generated. We designated each set as a virtual community, and each number as the biomass within a given 50×50 cm quadrat in the community. For the six communities, the average biomass per quadrat (μ) was 50 g DM; p was set to 2, 5, and 50 for each of the two levels of areas (Fig. 2). We assumed $p = 2$, 5, and 50 for very heterogeneous grassland, intermediate heterogeneous grassland, and less heterogeneous grassland, respectively, on a scale of 0.25 m^2.

In the six virtual communities, we conducted simulations to estimate biomass with different sampling procedures and numbers of samples. The sampling procedures were random sampling and RSS. Each simulation was iterated 1 000 times

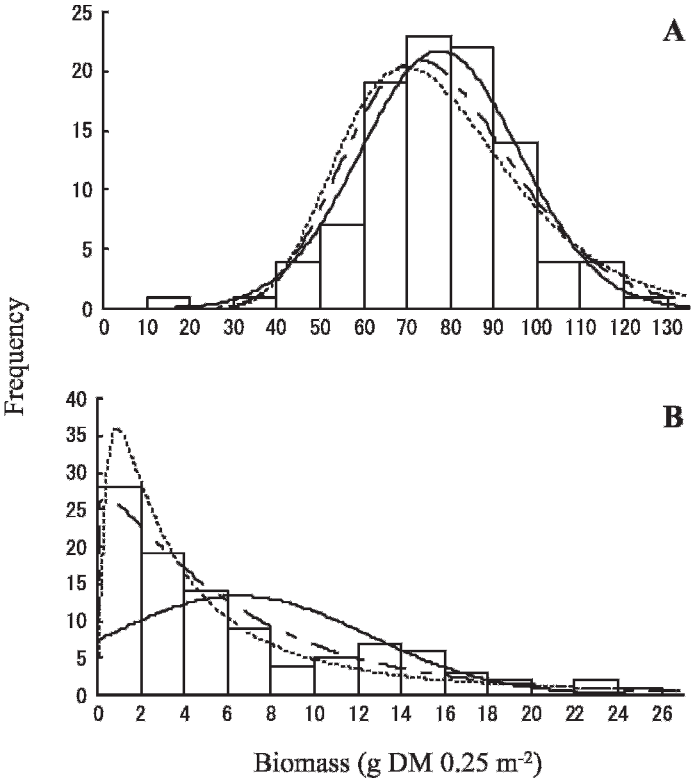

Figure 1. Examples of observed frequency distribution of herbage biomass and application of the normal, lognormal, and gamma distributions in **(A)** less heterogeneous and **(B)** more heterogeneous grasslands. Bar, solid line, dotted line, and broken line indicate observed frequency distribution, normal distribution, lognormal distribution, and gamma distribution, respectively. The parameter values and goodness-of-fit of the three distributions are as follows: **A,** $\mu = 77.6$, $p = 17.84$; goodness-of-fit of the normal, lognormal, and gamma distributions are $P = 0.56$ (Shapilo-Wilk W test), $P = 0.07$ (Kolmogrov-Smirnov Lilliefors test), and $P = 0.25$ (Cramer–von Mises W test), respectively. **B,** $\mu = 6.49$, $p = 1.186$; goodness-of-fit of the normal, lognormal, and gamma distributions are $P < 0.0001$ (Shapilo-Wilk W test), $P = 0.15$ (Kolmogrov-Smirnov Lilliefors test), and $P = 0.25$ (Cramer–von Mises W test), respectively. This figure was redrawn with the data from **(A)** Tsutsumi et al. (2002) and **(B)** Tsutsumi et al. (2000).

using the macro function of Excel 2003® (Microsoft, Redmond, WA).

To evaluate the precision of estimation in the simulations, we computed the margin of estimation error, which was obtained as the difference between an estimated value and the true value, such as 50. Then we counted frequencies for which estimation was precise when acceptable error limits were selected to be $\pm 5\%$, $\pm 10\%$, and $\pm 20\%$ of the mean. In addition, we computed the root mean square error of estimated values to the true value as average estimation error.

RESULTS

Figure 3 shows the probability that the margin of estimation error was within $\pm 10\%$ of the mean and root mean square error in the case of $p = 2$ with random sampling in small and large areas as examples. Comparing the results of the small and large areas, the precision of the estimated values with equal numbers of samples was very similar (i.e., no effects of sampling density were apparent on the precision of estimation). Similar results were obtained in simulations with $p = 5$, $p = 50$, and with RSS. Hereafter we present only the results of simulations conducted for small areas.

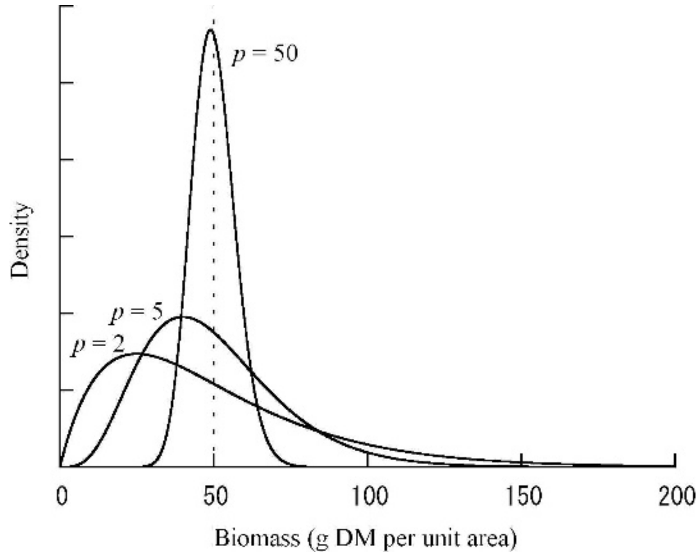

Figure 2. Probability of the density distribution of herbaceous biomass in very heterogeneous ($p = 2$), intermediate ($p = 5$), and less heterogeneous ($p = 50$) grasslands. Average biomass values (g DM per unit area) were all 50.

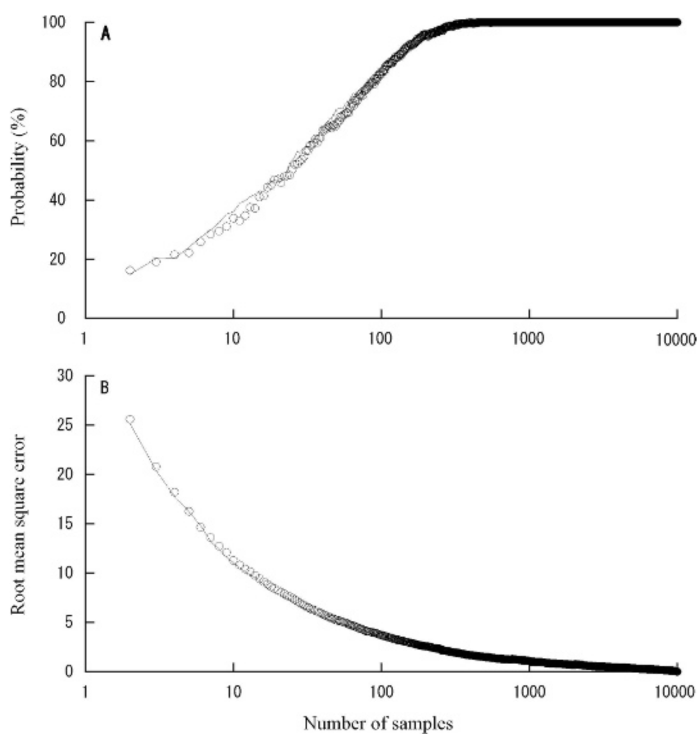

Figure 3. A, Probability that the margin of estimation error ranged within $\pm 10\%$ of the mean; **B,** Root mean square error of the estimated values to the true value (i.e., 50) in the case of $p = 2$ with random sampling in small (circle) and large areas (solid line).

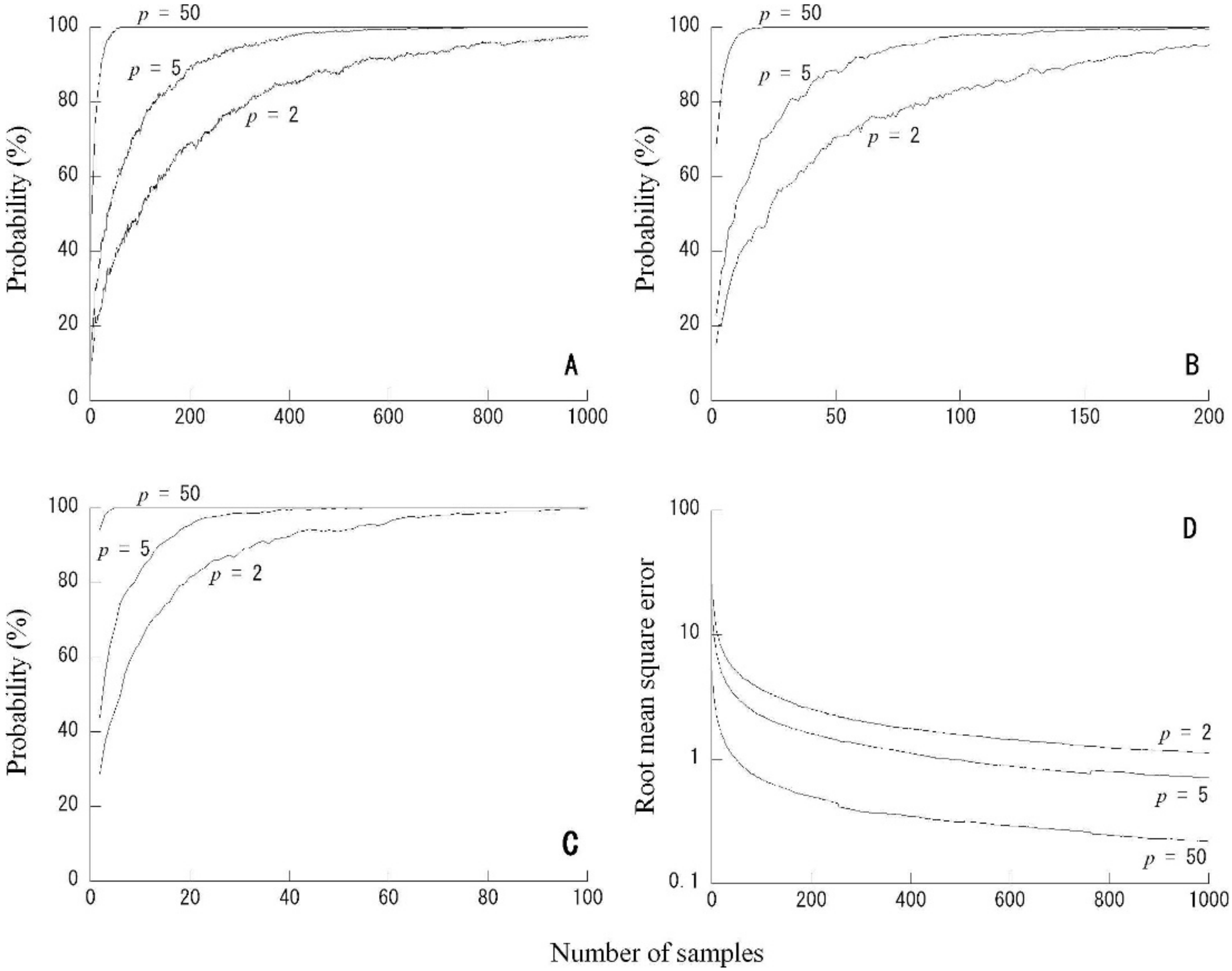

Figure 4. Results of simulations with random sampling in a small area. **A,** Relationship between number of samples and probability that the margin of estimation error was within ± 5% of the mean. **B,** Relationship between number of samples and probability that the margin of estimation error was within ± 10% of the mean. **C,** Relationship between number of samples and probability that the margin of estimation error was within ± 20% of the mean. **D,** Relationship between number of samples and root mean square error of the estimated values to the true value (i.e., 50).

Results of the simulations with random sampling and with RSS are shown in Figures 4 and 5, respectively. Table 1 lists the necessary number of samples for estimation at a confidence interval of 95%, when acceptable error limits were selected to be ± 5%, ± 10%, and ± 20% of the mean. The margins of estimation error became larger as *p* decreased for the same number of samples (Figs. 4 and 5). Similarly, the necessary number of samples became large as *p* decreased at the same acceptable error limit (Figs. 4 and 5; Table 1). In all cases the margins of estimation error that were generated in the simulations with RSS were lower than those observed in equivalent simulations with random sampling.

DISCUSSION

Our simulations demonstrated that the number of samples is a more influential factor than sampling density on the precision of the herbaceous biomass estimation (Fig. 3). For this reason researchers can determine an appropriate number of samples irrespective of the grassland area. Though Kayama (1961) suggested that sampling density affected the precision in the empirical study, our simulations disputed his suggestion.

The results of our simulations confirmed that biomass heterogeneity strongly affected the precision of biomass estimation (Figs. 4 and 5; Table 1), as suggested by Iwasaki (1976). Additionally, this paper has quantified the magnitude of the influence; this was significant new information. For grasslands with low heterogeneity, such as a mown pasture, we require fewer than 10 samples if we accept an estimation error of ± 10% of the mean (Table 1). Moreover, adoption of RSS decreases the required number of samples. For grasslands with intermediate heterogeneity, we require 77 or more samples with random sampling, even if we accept an estimation error of ± 10% of the mean, whereas we need only 15 samples with

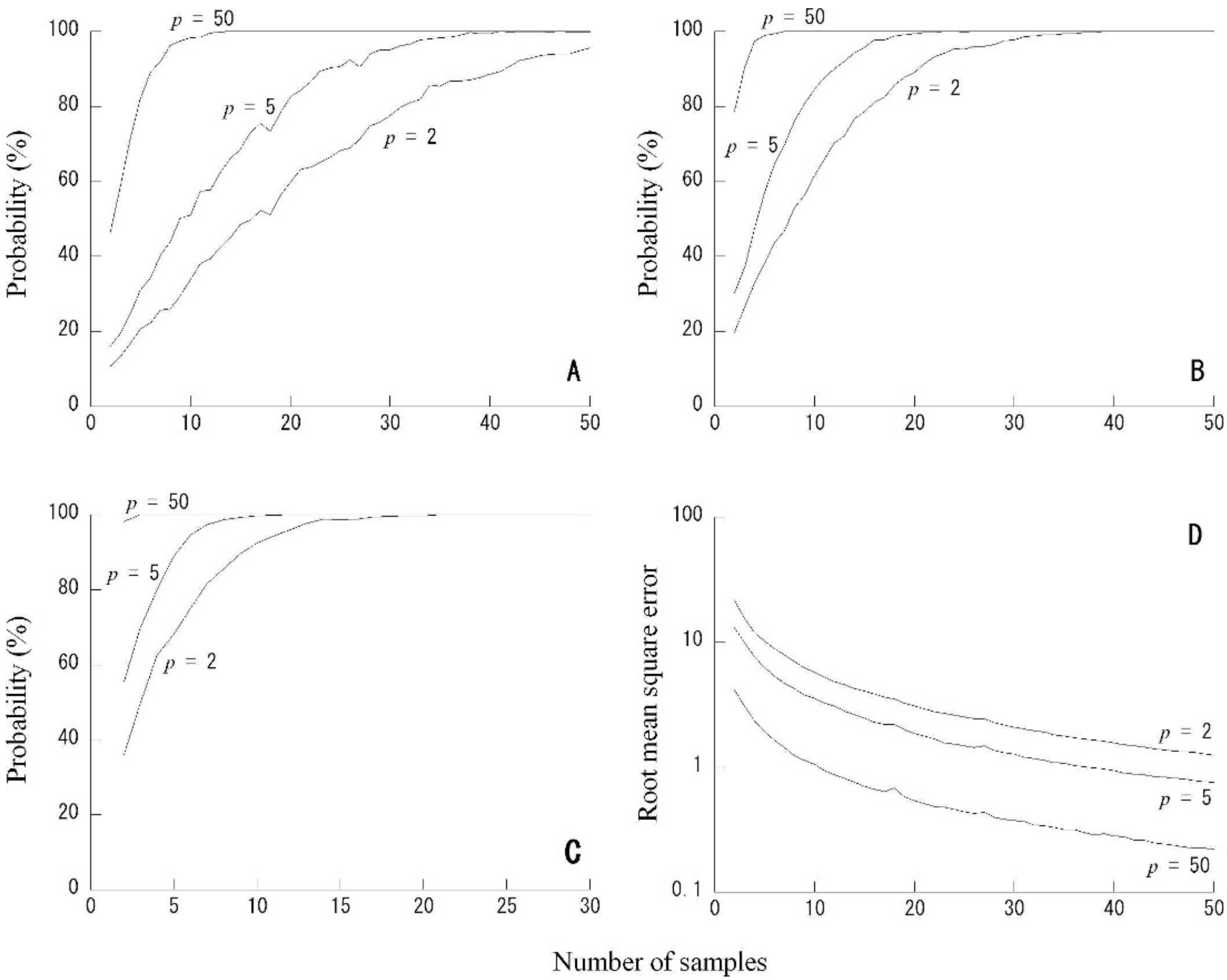

Figure 5. Results of the simulations with ranked set sampling in a small area. **A,** Relationship between number of samples and probability that the margin of estimation error was within ± 5% of the mean. **B,** Relationship between number of samples and probability that the margin of estimation error was within ± 10% of the mean. **C,** Relationship between number of samples and probability that the margin of estimation error was within ± 20% of the mean. **D,** Relationship between number of samples and root mean square error of the estimated values to the true value (i.e., 50).

Table 1. Necessary number of samples for herbaceous biomass estimation at a confidence interval of 95%, when acceptable error limits were assumed to be ± 5%, ± 10%, and ± 20% of the mean.

Biomass heterogeneity	Acceptable error limit	Random sampling	RSS[1]
Very heterogeneous ($p = 2$)	± 5%	766	50
	± 10%	200	24
	± 20%	54	12
Intermediate ($p = 5$)	± 5%	328	29
	± 10%	77	15
	± 20%	20	7
Less heterogeneous ($p = 50$)	± 5%	31	8
	± 10%	9	4
	± 20%	3	2

[1]Ranked set sampling.

RSS. On the other hand, for very heterogeneous patchy grassland, even if we accept an estimation error of ± 10% of the mean, we require 200 random samples and 24 with RSS. RSS is superior to random sampling at the same number of samples (Figs. 4 and 5). However, the RSS procedure is laborious, especially when sampling is conducted at numerous (i.e., more than 10) points because RSS requires ranking n^2 quadrats before sampling n quadrats.

In the present study, we assumed a 50×50 cm quadrat for sampling. Several researchers have discussed the appropriate sampling unit size (Kayama 1961; Soplin et al. 1975; Iwasaki 1976). Shiyomi (1987) showed that biomass heterogeneity, as indicated with p-values, is lower for larger quadrats. Accordingly, if we adopt a larger quadrat than 50×50 cm, the necessary number of samples will be smaller than those indicated in the present study. However, in such a case, the total sampling area might be even larger, as indicated by the previous studies (Kayama 1961; Iwasaki 1976; Shiyomi 1987).

Recently geostatistical analysis has been used for evaluating sampling schemes in various fields (Webster and Oliver 2000). Geostatistical analysis could be used for evaluating the number of samples required for estimating herbaceous biomass. Our analysis is applicable when random sampling is employed (in RSS we assumed to select samples for ranking randomly). If systematic sampling, such as along a diagonal, is employed, geostatistical analysis will be necessary and useful. In such a case, detailed information on spatial distribution of herbage biomass in the pasture of interest will be required. On the other hand, the heterogeneity of herbage biomass appears in various spatial scales, that is, patch size of the biomass varies within and among pasture(s), so that it is difficult to evaluate the number of samples required for estimating the biomass based on the patch size.

Based on the results of the simulations, we concluded the following:

1. In less heterogeneous grasslands, good precision of estimation can be obtained with a small number of samples, and it is useful to employ RSS. The cutting method, as well as nondestructive methods, will be practical.

2. Estimation for heterogeneous grassland requires a large number of samples, and it is not so useful to employ RSS. For that reason, more research is needed on nondestructive methods.

LITERATURE CITED

Barthram, G. T., E. I. Duff, D. A. Elston, J. H. Griffiths, T. G. Common, and C. A. Marriott. 2005. Frequency distributions of sward height under sheep grazing. *Grass and Forage Science* 60:4–16.

Bullock, J. M. 1996. Plant competition and population dynamics. *In:* J. Hodgson and A. W. Illius [eds.]. The ecology and management of grazing systems. Wallingford, UK: CAB International. p. 69–100.

Cid, M. S., and M. A. Brizuela. 1998. Heterogeneity in tall fescue pastures created and sustained by cattle grazing. *Journal of Range Management* 51:644–649.

Clarke, J. L., D. Welch, and I. J. Gordon. 1995. The influence of vegetation pattern on the grazing of heather moorland by red deer and sheep: II. The impact on heather. *Journal of Applied Ecology* 32:177–186.

Cobby, J. M., M. S. Ridout, P. J. Bassett, and R. V. Large. 1985. An investigation into the use of ranked set sampling on grass and grass-clover swards. *Grass and Forage Science* 40:257–263.

Greig-Smith, P. 1983. Quantitative plant ecology. 3rd ed. Oxford, UK: Blackwell Scientific Publications. p. 19–53.

Hirata, M. 2000. Quantifying spatial heterogeneity in herbage mass and consumption in pastures. *Journal of Range Management* 53:315–321.

Hirata, M., R. Sato, and S. Ogura. 2002. Effects of progressive grazing of a pasture on the spatial distributions of herbage mass and utilization by cattle: a preliminary study. *Ecological Research* 17:381–393.

Itano, S., M. Tsutsumi, S. Sakanoue, K. Nakagami, and M. Shiyomi. 2006. Variant of estimation method of aboveground plant biomass in grassland with the gamma model. 2. Proposal of a modified method with visual observation and tests in *Zoysia* pastures. *Grassland Science* 52:53–59.

Iwasaki, M. 1976. Influence of size and shape of quadrate upon distribution pattern of grass weight and number of samples necessary for estimation of grass weight on grassland. *Bulletin of National Grassland Research Institute* 8:11–26 (in Japanese with English summary).

Kayama, R. 1961. The size of quadrat to the distribution type of the coverage and yield of pasture grass. *Journal of Japanese Society of Grassland Science* 7:21–27 (in Japanese with English summary).

McIntyre, G. 1952. A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research* 3:385–390.

Ogura, S., T. Nakahara, and M. Hirata. 2002. Spatial distribution of herbage consumption by cattle in a bahiagrass (*Paspalum notatum* Flügge) pasture monitored two-dimensionally. *Grassland Science* 48:317–325.

Remington, K. K., C. D. Bonham, and R. M. Reich. 1992. Blue grama-buffalograss responses to grazing: a Weibull distribution. *Journal of Range Management* 45:272–276.

Remington, K. K., C. D. Bonham, and R. M. Reich. 1994. Modeling the distribution of *Agropyron cristatum* biomass in a grazed pasture using the Weibull distribution. *Journal of Japanese Society of Grassland Science* 40:190–197.

Shiyomi, M. 1987. Characteristics of herbage biomass estimated by the quadrat method. *Journal of Japanese Society of Grassland Science* 32:307–312 (in Japanese with English summary).

Shiyomi, M. 1991. Method for the estimation of herbaceous biomass in grazed pasture by visual observation. *Journal of Japanese Society of Grassland Science* 37:231–239.

Shiyomi, M., T. Akiyama, and S. Takahashi. 1983. A spatial pattern model of plant biomass in grazing pasture I. *Journal of Japanese Society of Grassland Science* 28:373–382.

Shiyomi, M., T. Akiyama, and S. Takahashi. 1984. A spatial pattern model of plant biomass in grazing pasture II. *Journal of Japanese Society of Grassland Science* 30:40–48.

Shiyomi, M., M. Okada, S. Takahashi, and Y. Tang. 1998. Spatial pattern changes in aboveground plant biomass in a grazing pasture. *Ecological Research* 13:313–322.

Shiyomi, M., S. Takahashi, and J. Yoshimura. 2000. A measure for spatial heterogeneity of a grassland vegetation based on the beta-binomial distribution. *Journal of Vegetation Science* 11:627–632.

Soplin, H., H. Gross, and J. Rawlings. 1975. Optimum size of sampling unit to estimate coastal bermudagrass yield. *Agronomy Journal* 67:533–537.

Tsutsumi, M., and S. Itano. 2005. Variant of estimation method of above-ground plant biomass in grassland with gamma model. 1. Use of an electronic capacitance probe. *Grassland Science* 51:275–279.

Tsutsumi, M., M. Shiyomi, H. Hayashi, S. Takahashi, and K. Sugawara. 2000. Small-scale spatial heterogeneity of live-shoot biomass of plant species composing sown grassland communities. *Grassland Science* 46:209–216.

Tsutsumi, M., M. Shiyomi, S. Sato, and K. Sugawara. 2002. Use of gamma distribution in aboveground biomass of plant species in grazing pasture. *Grassland Science* 47:615–618.

Webster, R., and M. Oliver. 2000. Geostatistics for environmental scientists. Chichester, UK: John Wiley and Sons. p. 23–34.