MARDI Res. J. 24(2) (1996): 127–134

A three-stick sampling technique for data collection in direct-seeded rice experiments

(Teknik pensampelan dengan tiga batang kayu untuk pengumpulan data dalam ujikaji padi tabur terus)

B. H. Yap*, M. Azmi* and P. M. Chang*

Key words: unbiased, three-stick sampling technique, wooden frame, random, systematic, haphazard, quadrat, triquadrat

Abstrak

Ujikaji untuk menilai teknik baru pensampelan dengan menggunakan tiga batang kayu bagi menentukan lokasi unit pensampelan telah dijalankan. Unit pensampelan akhir yang diambil data terdiri daripada kuadrat yang berukuran 25 cm x 25 cm dan 15 cm x 15 cm. Empat pelan pensampelan dan empat kadar biji benih telah dinilai. Pembolehubah yang diambil data ialah jumlah anak pokok padi/rumpai (peringkat awal), bilangan pokok padi, jumlah benah hijau dan benah belakang putih, jumlah musuh serangga semula jadi dan bilangan rumpai. Analisis varian dijalankan bagi setiap pembolehubah dan setiap saiz kuadrat. Hasil daripada ujikaji menunjukkan bahawa pelan pensampelan sistematik dan rawak yang menggabungkan penggunaan prosidur pensampelan tiga batang kayu dan rangka kayu berukuran 5 m x 5 m lebih sesuai dan menjamin cara pengumpulan data yang lebih benar. Varian untuk petak utama, subpetak dan pensampelan bagi kuadrat 15 cm x 15 cm lebih kecil daripada nilai bagi kuadrat 25 cm x 25 cm kecuali bilangan serangga. Cara untuk memudahkan pensampelan itu dengan penggunaan trikuadrat dan sebatang kayu segi empat sepanjang 5 m sebagai pengganti rangka 5 m x 5 m dicadangkan. Pokok padi dan rumpai di petak didapati berkelompok dan tidak bercorak rawak atau seragam seperti yang dikehendaki. Keadaan ini khusus disebabkan oleh penyediaan tanah yang kurang memuaskan oleh petani.

Abstract

An experiment to evaluate the innovative use of a three-stick sampling procedure to determine the locations of sampling units was conducted. The final sampling units from which data were obtained, comprised quadrats of sizes 25 cm x 25 cm and 15 cm x 15 cm. Four sampling plans and four seed rates were evaluated. The data collected were for the variables total paddy/weed seedling counts (early stage), number of paddy plants, total number of green leafhoppers and white-back planthoppers, total number of predators and weed counts. An analysis of variance was done on each of the variables for each quadrat size. The findings revealed that the systematic and random sampling plans incorporating the use of the three-stick sampling procedure and a 5 m x 5 m wooden frame were practical and would ensure a more objective approach to data collection. The main plot, sub-plot and sampling variances for the 15 cm x 15 cm quadrats were much smaller than the values for 25 cm x 25 cm quadrats, except insect counts. Ways

*MARDI Research Station, Seberang Perai, P.O. Box 203, 13200 Kepala Batas, Seberang Perai, Penang, Malaysia
Authors' full names: Yap Beng Ho, Azmi Man and Chang Poon Min

to simplify the sampling procedure using a triquadrat and a four-sided 5 m pole instead of the 5 m x 5 m wooden frame are suggested. The paddy and weed seedlings in the plots were found to be clumped or clustered as opposed to the random or uniform patterns desired. This discrepancy was mainly due to the unsatisfactory land preparations carried out by the farmers.

Introduction

The planting of rice by direct seeding has been adopted on a very wide scale in all the major rice granary areas. This is mainly due to the shortage of labour, especially in the agricultural sector in Malaysia. Most rice experiments in MARDI are, therefore, done on direct-seeded rice.

In the direct-seeded rice environment, the individual paddy plants or hills are not easily identifiable and the distances between them can vary considerably. Hence, it is tedious to separate the plants unarbitrarily for data collection. In addition, there are no distinct planting rows as found in a transplanted rice experiment which can be used in the preliminary sampling plan.

Quadrats comprising a square frame of standard dimensions of 25 cm x 25 cm or smaller, have been used for locating sampling 'points' for data collection in direct-seeded rice. It is usually done in a 'pseudo-random' manner, i.e. the quadrat is thrown and the area enclosed by the quadrat where it lands, is sampled. The throwing of the quadrats will be influenced by the person's visual and other preconceived notions of the area to be sampled, and will thus vary subjectively. Information obtained in this manner is open to abuse and the amount of bias recorded is difficult to assess. An alternative is to randomly locate and demarcate several sampling 'points' within the treatment plots before seeding (Gomez 1972). However, this method can cause unfair attention to be given to these demarcated 'points' either consciously or unconsciously. There are, therefore, defects in the existing sampling practices.

Field studies of statistical sampling procedures are rare, especially for the direct-seeded rice experiments. Since one of the most important components of success in an experiment is the quality of the data collected (and the practicability of the proposed sampling procedure), this study aims to find practical sampling procedures which can minimize bias in locating sampling 'points' for the collection of data in direct-seeded rice experiments. A method using three sticks, and four sampling procedures using two quadrat sizes were evaluated.

Materials and methods

The experiment on sampling methods using three sticks was done on a farmer's field located in Block K3, IADP Jabatan Pertanian, Bukit Merah, Seberang Perai. Land preparation and levelling were done by the farmer. A randomized complete block design with four seedling rates of 40, 60, 80, and 100 kg seeds/ha in four replicated blocks was laid out. The pre-germinated seeds of variety MR 84 were sown on 5 April 1994. The fertilizers applied were 90 kg N, 45 kg P_2O_5 and 30 kg K_2O per hectare.

Using two quadrat sizes (15 cm x 15 cm and 25 cm x 25 cm), four sampling procedures evaluated were systematic (exact), systematic (step method), random (table) and haphazard (throwing).

How to locate the sampling strips and sampling units

In *Figure 1,* the typical (seed rate) treatment plot (6 m x 6 m) with a movable 5 m x 5 m wooden frame placed 0.5 m from the edges of the plot demarcates the boundary of the sampling area. However, the 3 m x 3 m sub-plot at the centre of each plot was reserved for the yield assessment. For this experiment, the wooden frame was made up of eight pieces

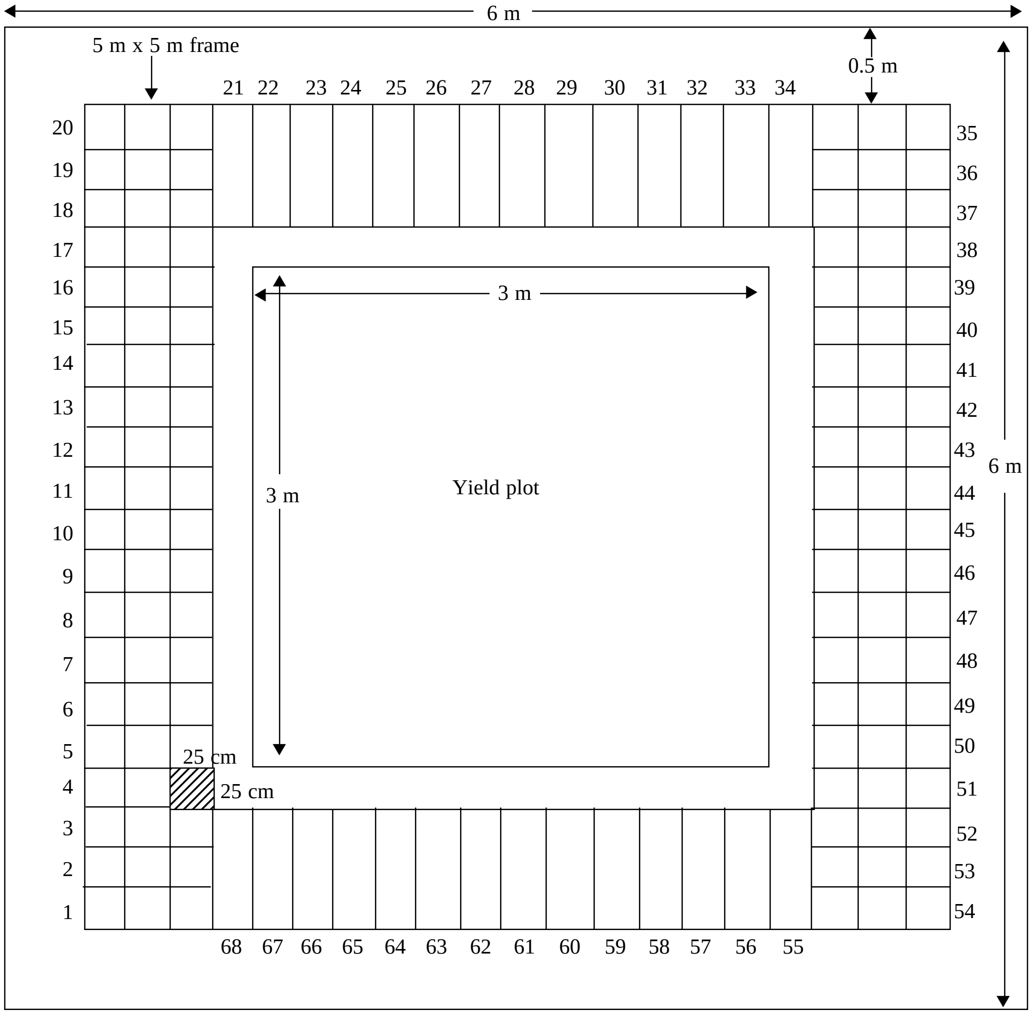

Figure 1. Treatment plot showing the 5 m x 5 m frame with the 68 first-stage sampling strips and the final-stage sampling units (25 cm x 25 cm)

of 2.5 cm x 2.5 cm wood, each 2.5 m long. Each length of wood was in the shape of so that it could be easily put together to form an upright 5 m x 5 m square frame inside the 6 m x 6 m plot. The eight pieces of wood were marked 1 to 10, 11 to 20 and so on up to 68, at 25 cm intervals. These numbers on the wooden frame were the locations for the 68 (25 cm x 75 cm) first-stage sampling strips. The formula for calculating the number of strips is 8(2L – 4QRW – 1.5) where L = plot length (m) and QRW = quardrow width (m). The final sampling units consisted of three 25 cm x 25 cm quadrats in each of the first-stage sampling strips.

For the systematic exact and step sampling plans, a random starting ‘point’ (number) was alloted to each replicate. In this case, a number between 1 and 17 was randomly chosen. This number was then added to 17 to give the next selected strip number and so on. By using a sampling interval of 17, a total of four sampling strips were selected. These strips were used for all the four seed rates in one replicate. For the next replicate, another random number between 1 and 17 was used and so on

(Gomez 1972). Having selected the numbers, the strips could be easily located in the plot by the numbers on the 5 m x 5 m wooden frame. The three sticks all of equal length (75 cm) were marked at 25, 50 or 75 cm respectively. For each selected strip, one of the three sticks would be randomly picked to identify one of the three 25 cm x 25 cm quadrats for data recordings. If the stick with the 50 cm marking was picked, this stick was then placed on the strip (number) perpendicularly to the wooden frame. The other two sticks were then used to position the 25 cm x 25 cm wooden quadrat in the flooded plot, 25 cm from the sampling frame, i.e. if the stick with the 50 cm marking was picked. Data were first recorded for the 15 cm x 15 cm quadrat by placing it at the centre of the 25 cm x 25 cm quadrat. Then the data for the surrounding area outside the 15 cm x 15 cm quadrat were carefully recorded. These values, added to the value for the 15 cm x 15 cm quadrat, made up the value for the 25 cm x 25 cm quadrat. The three sticks were then 'reshuffled', and another stick was randomly picked to identify the next quadrat for data collection in the next strip along the wooden frame and so on. In the systematic step method, each strip was located by walking along the side of the wooden frame starting from the right-hand corner of the plot and moving in a clockwise direction. The number of 'standard' walking steps to take to locate the first strip would be equal to the randomly selected number (between 1 and 17) used in the systematic exact method. The strip (number) on the wooden frame nearest to the last step was selected and the three sticks were again used to pick the quadrat for data collection in the same way as for the systematic method. The next strip would be located by taking another 17 steps from the first strip and so on. The systematic step approach was studied to see whether this simpler and practical method could be used and would not give significant statistical differences compared with the random (table) method. The strips located by the systematic step method would generally be different from those 'fixed' by the systematic exact method. It might not be possible to avoid some bias in the step method but it was hoped that it would be small, and this simpler method should be more acceptable to the rice scientists. Systematic sampling was equivalent to a kind of partial stratification and the sample estimate would be somewhat more precise than a fully random sample (Yates 1965).

In the random (table) method, four strips (numbers) were selected out of the 68 strips on the wooden frame using the random number table. A different set of four numbers were used for each replicate. The three sticks were again used to locate the quadrats in the strips for the data collection.

In the haphazard (throw) method, the 25 cm x 25 cm quadrat was thrown from each side of the plot (with the thrower's back facing the plot) and where it landed, the area within it was recorded. The three sticks were not used.

Data collection

It was originally planned to collect data from a sample of eight quadrats per plot. However, after the joint paddy/weed seedling counts in two replicates, it was decided that the other two replicates would be used for the recording of counts of only paddy plants the following week, when the paddy plants could be differentiated from the weed plants. From the two sets of results, we might be able to get an estimate of the extent of weed infestations. From the observed variations of the counts and the time required to collect the data from each plot for the four sampling methods studied, it was decided that a sample of four instead of eight quadrats per plot would be used and data obtained from all the four replicates. The additional data recorded were for the number of green leafhoppers and white-back planthoppers, predators and weed counts.

Statistical analyses

An analysis of variance was done on each of the variables recorded. The four seed rates were regarded as the main plot treatments and the four sampling methods (systematic exact, step method, random and haphazard), were taken as the sub-plot treatments in all the analyses. The analyses were done separately for each quadrat size. This was partly to avoid the use of an extrapolation factor for the conversions of the 15 cm x 15 cm quadrats (Q15) to make them comparable with the 25 cm x 25 cm (Q25) quadrats in a combined analysis. More importantly, a separate estimate of the sampling variances for the two quadrat sizes was required.

Results and discussion

Results of the analyses of variance for the five recorded variables are given in *Table 1*. Mean squares for the three components of error, main plot, pooled sub-plot and the pooled quadrat sampling, are also given.

From the analysis of the original data (not transformed), scatter plots of means versus variances from the seed rates x sampling method components of variance for each quadrat size found that variances tend to increase with the mean values. The square root transformation of the data before analysis can reduce this dependence, but generally found to be not satisfactory. For count variables whose distributions are skewed, the averages of counts can be used instead, as they tend to be normally distributed with increasing sample size. As this paper is concerned only with the evaluation of sampling methods and their statistical properties for some commonly collected data in direct-seeded rice experiments, the discussion of results is focused on them only. The analysis of variance which is quite robust to certain amount of deviations from normality, is used as a convenient computational tool to give approximate tests of effects and to obtain the necessary statistics required for this study.

The effect of higher seed rates on the paddy/weed and paddy counts was significant at the 5% and 10% probability levels respectively for Q15. In both the variables, the F-values for seed rates decreased and were not significant when the bigger quadrats Q25 were used. This could be attributed to the large increase in the magnitude of the main plot error when the bigger quadrats Q25 were used, compared with Q15. For the final weed counts, the main plot error also increased with the quadrat size *(Table 1)*. However, the effects of quadrat size on the main plot errors were minimal for the pest and predator counts. The F-value for the green leafhopper and white-back planthopper counts in the different seed rates was significant at the 5% probability level for Q25.

The F-values for effects due to sampling methods were not significant at $p = 0.05$ for all the variables *(Table 1)*. Generally, the differences in the mean values of the variables estimated by the four sampling methods were not large for both Q15 and Q25 *(Table 2)*. However, further comparisons with random sampling method as the unbiased standard, showed that the estimates obtained using the systematic exact sampling method were generally closer to the random method estimates than those produced by the haphazard and systematic step methods.

However, since the characters measured varied from point to point in the plot, the sampling procedure adopted should involve counts of randomly, or at least, objectively selected portions of the population to avoid serious bias (Finney 1946). It is also important to note that the same sampling method, quadrat and sample size must be used in the recording of a character for all treatment plots.

The values of the standard deviations for the four sampling methods also did not differ widely except for weed counts where random sampling for Q25, gave much lower values than the other three sampling methods *(Table 2)*. The weed species

Table 1. Analysis of variance

Source of variation	Paddy/weed seedlings					Paddy seedlings					GWBPH (counts)					Predators				Weed counts			
	df.#	Q15		Q25		df.	Q15		Q25		df.	Q15		Q25		Q15		Q25		Q15		Q25	
		MS	FR	MS	FR		MS	FR	MS	FR		MS	FR	MS	FR	MS	FR	MS	FR	MS	FR	MS	FR
Block (B)	1	50.8	3.8ns	0.1	<1ns	1	29.1	6.4	36.1	<1ns	3	125.3	3.4ns	114.0	2.6ns	4.40	4.4ns	1.31	6ns	20.80	2.3ns	131.60	6.0*
Seed rate (S)	3	267.1	20.1*	848.6	6.9ns	3	30.3	6.7ns	171.7	2.9ns	3	55.4	1.5ns	231.7	5.3*	1.10	1.1ns	0.80	<1ns	9.80	1.1ns	27.10	1.2ns
B x S (main plot error)	3	13.3	–	123.6	–	3	4.5	–	59.9	–	9	37.2	–	43.4	–	1.00	–	0.80	–	8.90	–	21.90	–
Sampling method (M)	3	27.2	2.3ns	54.8	1.5ns	3	17.3	3.0ns	0.8	<1ns	3	28.8	<1ns	6.0	<1ns	1.10	2.2ns	1.50	2.8ns	4.70	2.5ns	10.40	1.9ns
M x S	9	16.3	1.4ns	23.4	<1ns	9	3.9	<1ns	12.9	1.2ns	9	30.0	<1ns	24.0	<1ns	0.20	<1ns	0.50	1.0ns	2.00	1.0ns	5.20	<1ns
(M x B) & (M x B x S) (pooled sub-plot error)	12	4.4	–	37.6	–	12	5.8	–	10.9	–	36	35.5	–	28.8	–	0.47	–	0.53	–	1.91	–	5.58	–
Quadrat sampling variation (pooled)	224	11.7	–	20.2	–	96	3.8	–	9.4	–	192	32.6	–	33.1	–	0.90	–	0.78	–	2.03	–	4.86	–
Overall mean		6.2		10.4			3.4		7.2			10.4		11.1		0.77		0.78		1.21		1.95	

MS = mean square
FR = F-ratio
*$p <$ = 0.05
**$p <$ = 0.01
ns = not significant at $p = 0.05$
GWBPH = green leafthoppers and white-back planthoppers
Q15 = 15 cm x 15 cm quadrat
Q25 = 25 cm x 25 cm quadrat

#For paddy & weed seedlings, Rep I & III and 8 quadrats/plot were used
For paddy seedlings (subsequently), Rep II & IV and 4 quadrats/plot were used
For the remaining 3 characters in table, 4 replicates and 4 quadrats/plot were used

Table 2. Four sampling methods, two quadrat sizes, means and standard deviations

Sampling method	Paddy/weed seedlings @ 21 DAS		Paddy seedlings @ 30 DAS		GWBPH @ 60 DAS		Predators @ 60 DAS		Weed counts @ 50 DAS	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
15 cm x 15 cm										
Syst. exact	6.0	4.0	3.2	2.3	10.2	5.7	0.95	1.13	1.19	1.32
Syst. step	5.7	3.2	2.7	1.8	11.1	7.0	0.73	0.90	0.98	1.33
Random	7.1	4.7	3.3	2.0	9.5	5.1	0.66	0.86	1.06	1.49
Haphazard	5.9	3.3	4.4	2.7	10.6	5.5	0.73	0.78	1.59	2.14
25 cm x 25 cm										
Syst. exact	10.3	5.9	7.2	4.6	11.4	6.0	0.78	0.98	2.00	3.36
Syst. step	9.5	5.3	6.9	3.4	11.4	6.2	0.47	0.71	2.08	2.77
Random	11.7	5.4	7.3	2.7	10.9	5.8	0.67	0.80	1.39	1.61
Haphazard	10.2	6.0	7.2	4.6	10.8	5.9	0.80	0.93	2.34	2.92

GWBPH = green leafhoppers & white-back planthoppers

encountered in the experiment consisted of mainly *Leptochloa chinensis* and *Monochoria vaginalis*. The number of *L. chinensis* plants visible might be small but they were usually interconnected by runners.

Increase in the quadrat size from Q15 to Q25 also significantly increased the sub-plot error and the quadrat sampling variances of the three variables, paddy/weed, paddy and weed counts *(Table 1)*. Such an effect generally indicated that the spatial distributions of paddy and weed seedlings in the plots were not regular or random. Except for the predator counts, the variance to mean ratios were much greater than one, also indicating that the spatial distributions of these variables were inclined to the negative binomial distributions (Elliott 1977). Poor land preparation and unevenness in the experimental area were often responsible for the patchy distributions of paddy and weed seedlings in direct-seeded rice experiments. Various quadrat-variance methods could be used to study the effect of varying quadrat sizes for the detection of the underlying population patterns of plant and insect species (Ludwig and Reynolds 1988).

Conclusion

This study found that the sampling method using a 5 m x 5 m wooden frame together with the objective three-stick procedure to determine the final sampling units in the direct-seeded rice experimental plots was applicable. However, to make it more practical, several modifications are suggested. The four corners for the 5 m x 5 m wooden frame in all the plots should be pegged prior to the date of data collection. The 5 m x 5 m wooden frame should be replaced by one four-sided wooden pole of 5 m where the four sides of the pole (marked accordingly) are used to represent the four sides of the 5 m x 5 m wooden frame. It is easier to move the single 5 m pole from one side to another in a plot and from plot to plot, to locate the sampling strips identified by the numbers selected from 1 to 68. The three sticks can also be replaced by three coloured glass marbles and the 25 cm x 25 cm quadrat by a bigger triquadrat (25 cm x 75 cm) in which each of the three 25 cm x 25 cm quadrats is coloured like one of the glass marbles.

The determination of a quadrat in a selected strip is then decided by the colour of the marble drawn. To locate a sampling strip, the triquadrat is placed at the selected strip number on the 5 m pole which is then laid down on one side of the plot, using the pegs for alignment, and taking recording from the quadrat which has the same colour as the marble.

Of the four sampling procedures evaluated, the random method is ideal as it is unbiased and the actual field operations of locating the sampling units can be made similar to the systematic sampling procedure by arranging the randomly chosen strips (numbers) consecutively for data collection. The main disadvantage is the extra effort of choosing four instead of just one random number for the systematic procedure. The haphazard procedure is not reliable because subjective bias can render any results obtained questionable. This study also found that the haphazard procedure was inferior to the systematic exact sampling procedure. The time taken for each of the four sampling procedures was, however, not very much different from one another.

Acknowledgements
We would like to thank Wan Roslan W. Saari for implementing the experiment and helping to collect the data in the field. We are also thankful to Ismail Sharif, Hashim Che Mat and Shaharuddin Saari for helping us in the pest and weed counts.

References

Elliott, J. M. (1977). Some methods for the statistical analysis of samples of benthic invertebrates. *Freshwater Biological Assoc. Sc. Pub. No.* ***25*** L.A.: The Ferry House

Finney, D. J. (1946). Field sampling for the estimation of wireworm populations. *Biometrics* ***2(1):*** 1–6

Gomez, K. A. (1972). *Techniques for field experiments with rice* Los Banos, Laguna, Philippines: IRRI

Ludwig, J. A. and Reynolds, J. F. (1988). *Statistical ecology A primer on methods and computing* New York: John Wiley & Sons

Yates, F. (1965). *Sampling methods for censuses and surveys* 3rd ed. London: Charles Griffin & Co. Ltd.

Accepted for publication on