

Increasing Plot Length Reduces Experimental Error of On-Farm Tests

Stewart B. Wuest, Baird C. Miller, J. Richard Alldredge, Stephen O. Guy, Russ S. Karow, Roger J. Veseth, and Donald J. Wysocki

Research Question

The use of side-by-side, combine-width experimental plots in farmer-conducted on-farm tests is increasing. Information on how plot length affects the success of these tests is lacking. It is also desirable to know what level of precision to expect from on-farm tests performed in highly variable dryland cereal production areas. This study investigated the relationship between the length of combine-wide, side-by-side plots and experimental error under the dryland grain production conditions of the Pacific Northwest.

Literature Summary

Researchers studying the performance of on-farm tests have found that the randomized complete block design used in the Midwest produces results comparable in precision to research station small plot experiments. These designs use 1200 to 1300 ft long plots 20 to 40 ft wide, and primarily involve corn or soybean. The performance of on-farm tests under dryland small grain production conditions has not been evaluated. There has also been little research to date that can be used to recommend minimum, maximum, or optimum plot lengths.

Study Description

Fourteen uniformity trials were harvested in commercial wheat and barley fields in Washington, Idaho, and Oregon. A uniformity trial measures natural variability between plots by harvesting plots where no treatments have been placed. Side-by-side strips 1500 ft long were harvested in 250 ft segments to allow recombination of the data into plots of different lengths. The grain yield data (bu/acre) were analyzed to determine variance between pairs of side-by-side plots.

Ten of fourteen sites had a variance of < 5 at plot lengths of 1500 ft, and were classified as low variance. The remaining four sites ranged from 6 to 22 (high variance). Figure 1 shows least significant differences at $\alpha = 0.05$ (LSD 0.05) for an individual on-farm test with two treatments and four replications based on average variances for the low and high variance groups.

Applied Questions

How long should on-farm test plots be?

In most fields there was a large decrease in variance as plot length increased. Therefore, on-farm tests will produce more reliable results as plot length increases from 250 to 750 ft or more. In some very uniform fields even short plots will have acceptably low variability, but in every field measured, variability decreased as plot length increased to 1500 ft. To ensure the best results, we recommend that plots be as long as is practical.

Full scientific article from which this summary was written begins on page 211 of this issue.

How does replication affect precision?

The variability encountered in field experiments makes replication the key to a successful test. Figure 1 shows how LSD (0.05) decreases when replications are added. A low LSD is important because it allows detection of smaller differences between the performance of the treatments, or if there is no difference, it allows a high confidence that the treatments do not perform differently.

On-farm tests can provide valuable information to farmers and researchers. Small differences can be detected with a high degree of confidence in most fields with four or more replications of 1000 ft or longer side-by-side plots.

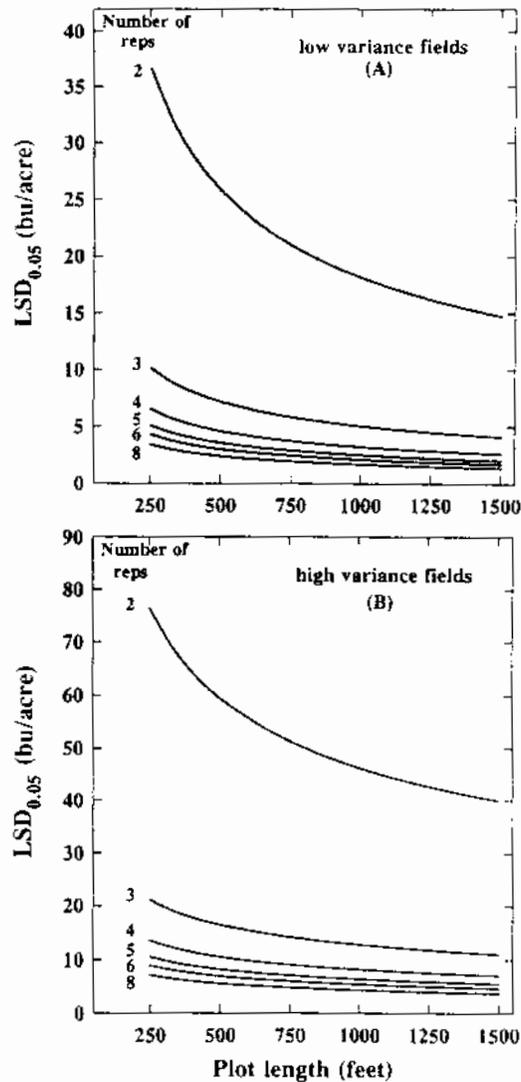


Fig. 1. The effect of increasing plot length on LSD values at $\alpha = 0.05$ is shown for experiments with different numbers of replications (reps). (A) is based upon data from the 10 fields with low variance and (B) from the remaining four fields with high variance. Note that the scales on the vertical axes differ.

Increasing Plot Length Reduces Experimental Error of On-Farm Tests

Stewart B. Wuest,* Baird C. Miller, J. Richard Alldredge, Stephen O. Guy, Russ S. Karow, Roger J. Veseth, and Donald J. Wysocki

Farmer conducted on-farm research is an effective tool for development of crop management practices. The randomized complete-block experimental design is being used in on-farm tests, with blocks consisting of two or more long, narrow, side-by-side plots. This study examined the relationship between plot length and experimental error, and assesses the probable statistical outcome of on-farm tests performed in the dryland region of the Pacific Northwest, USA. Fourteen trials were conducted in wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) fields to measure yield variance of combine-width plots ranging in length from 250 to 1500 ft. The relationship between plot length and variance for each site followed a logarithmic decay model (average $r^2 = 0.88$). Variance declined rapidly as plot length increased from 250 to 750 ft at most sites. Averaging the ten least variable sites, the LSD (0.05) with three degrees of freedom declined from 6.5 to 2.6 bu/acre as plot length increased from 250 to 1500 ft. At the same 10 sites, power for mean separation ($\alpha = 0.05$) of treatments with 4 bu/acre true difference was >0.80 with six replications and 750 ft plot length, or four replications and 1250 ft plot length. With adequate replication and plot length, on-farm tests can be designed for highly variable dryland regions with good control of experimental error.

FARMER CONDUCTED ON-FARM TESTS are an effective tool for development of improved crop management practices. The on-farm test is also useful as an extension tool for the promotion of new technology. The distinction between an on-farm test and a demonstration is the use of a replicated, scientific experimental design in the on-farm test to greatly increase reliability of measured results and allow the use of statistical analysis (Kittrel, 1974). With recent increased interest in on-farm tests, there is a need to check the effectiveness of the experimental designs being used (Lockeretz, 1987).

The most common experimental design promoted for use in on-farm tests is the randomized complete block

with a limited number of treatments and four or more replications. A common plot size for on-farm tests in the Midwest is 20 to 40 ft wide by 1200 ft long, and the treatments are replicated six to eight times in adjacent blocks. The use of this design for on-farm tests was analyzed under Midwest row crop conditions and was shown to give satisfactory results (Rzewnicki et al., 1988). Again in the Midwest, Schmitt et al. (1992) found the randomized complete-block design to be more efficient than the Tester design, which utilizes periodic control plots to analyze unreplicated treatments.

In contrast to row crop production in the Midwest, dryland farming in the Pacific Northwest takes place in a highly variable topography. At some locations, an individual field may have slopes ranging from 0 to 45%. The region's fields are also characterized by variable soil depths, fertility, and aspect. For these reasons, it is not unusual for grain yields to vary up to 50% across a single field. Studies of optimum plot length for on-farm tests under these conditions have not been reported.

The influence of plot size on control of experimental error has been investigated most intensely for researcher-size plots of <0.1 acre (LeClerg et al., 1962, p. 113). Most of these studies assumed that, due to limited available area, use of smaller plots would make increased replication possible, and the goal was to maximize experimental precision by optimizing the relationship between plot size and number of replications. Available field area is often not a limiting factor in on-farm tests. Our experience with on-farm tests indicates that plots should be a minimum of ≈ 0.1 acre for practical reasons, mostly related to precision of harvest measurements. Increasing the length of plots adds little to labor or cost and usually does not limit the number of replications possible. The most practical design we have found for on-farm tests consists of blocks of long, side-by-side plots that are wide enough to be harvested by a grain combine. It can be argued from a theoretical standpoint that the longer and closer these narrow sample areas are, the smaller the variance between the plots should be, assuming they do not run parallel to a gradient.

We report on a study designed to determine the relationship between experimental error and the length of combine-width plots. We also compare experimental error between on-farm tests and small plot research, and examine the least significant difference and the power for mean separation that an individual, single-location on-farm test will probably produce.

S.B. Wuest and B.C. Miller, Dep. Crop and Soil Sciences, and J.R. Alldredge, Program in Statistics, Washington State Univ., Pullman, WA 99164; S.O. Guy and R.J. Veseth, Dep. Plant, Soil, and Entomological Sciences, Univ. of Idaho, Moscow, ID 83843-4196; R.S. Karow, Dep. of Crop and Soil Science, Oregon State Univ., Corvallis, OR 97331-3002; and D.J. Wysocki, Columbia Basin Ag. Research Center, Pendleton, OR 97801. Washington State Univ. Dep. of Crop and Soil Sciences paper no. 9301-41. Received 5 Aug. 1993. *Corresponding author.

Table 1. Mean annual precipitation, soil taxonomic description, crop, average yield of uniformity trial, log-log regression coefficients, predicted variance of 1500 ft long plots, and high or low variance group assignment for the 14 uniformity trial sites.

Year	Location	Mean annual precipitation in	Soil series	Soil taxonomic description	Crop	Yield bu/acre	Regression			Predicted variance at 1500 ft	Group
							Intercept	Slope	r		
1991	Craigmont, ID	22	Setters	Fine montmorillonitic, frigid Ultic Paleixeroll	Winter wheat	79	5.01	1.96	0.99	0.60	1
	Moscow, ID	23	Southwick	Fine-silty, mixed, mesic Argiaquic Xeric Argialboll	Spring wheat	55	2.90	-0.61	0.92	9.34	2
	Dayton, WA	15	Athena	Fine-silty, mixed, mesic Pachic Haploxeroll	Winter wheat	93	3.23	-0.93	0.96	1.94	1
	Condon, OR	11	Condon	Fine-silty, mixed, mesic Typic Haploxeroll	Winter wheat	34	1.06	-0.29	0.98	1.37	1
	Helix, OR	14	Walla Walla	Coarse-silty, mixed, mesic Typic Haploxeroll	Winter wheat	88	2.66	-0.41	0.96	22.63	2
	Creston, WA	13	Bagdad	Coarse-silty, mixed, mesic Calcic Argixeroll	Spring wheat	40	2.80	-0.90	0.87	0.90	1
	1992	Craigmont, ID	22	Setters	Fine montmorillonitic, frigid Ultic Paleixeroll	Winter wheat	79	3.33	-0.98	0.88	1.69
Moscow, ID		23	Southwick	Fine-silty, mixed, mesic Argiaquic Xeric Argialboll	Winter wheat	71	2.37	-0.51	0.56	5.82	2
Dayton, WA		15	Palouse	Fine-silty, mixed, mesic Pachic Ultic Haploxeroll	Spring wheat	41	4.13	-1.39	0.96	0.53	1
Condon, OR		11	Condon	Fine-silty, mixed, mesic Typic Haploxeroll	Winter wheat	22	2.95	-0.87	0.99	1.55	1
Helix, OR		14	Walla Walla	Coarse-silty, mixed, mesic Typic Haploxeroll	Winter wheat	41	1.80	-0.35	0.79	4.78	1
Cornelius, OR		37	Laurelwood	Fine-silty, mixed, mesic Ultic Haploxeralf	Winter wheat	47	5.17	-1.49	0.80	2.68	1
Pomeroy, WA		20	Athena	Fine-silty, mixed, mesic Pachic Haploxeroll	Spring barley	46	5.21	-1.36	0.84	7.63	2
Davenport, WA		15	Broadax	Fine-silty, mixed, mesic Calcic Argixeroll	Spring barley	38	1.34	-0.34	0.81	1.85	1
Group 1, low variance							3.36	-1.02	0.50	1.34	
Group 2, high variance							3.29	-0.72	0.36	9.91	

MATERIALS AND METHODS

Uniformity trials were harvested in 14 wheat or barley fields farmed by commercial growers in 1991 and 1992. A uniformity trial is used to measure the variability of a potential research site, and consists of marking out plots and measuring yields where no treatments have been applied. Locations were representative of the dryland small grain growing region of the Inland Pacific Northwest, and ranged from Spokane, WA, in the north, Condon, OR, in the south, and as far east as Craigmont, ID (Table 1). Mean annual precipitation ranges from 11 to 37 in. at the locations sampled. Although 1991 and 1992 were below the normal in precipitation, yields were within the typical range for each of the sites. Crop rotations, tillage, fertilizers, and herbicide applications varied among the sites. Soil taxonomic descriptions for each site are included in Table 1.

Fields were selected by farmer cooperators. An area suitable for side-by-side test plots was selected using the farmer's knowledge of the field along with the field's visual appearance. We attempted to minimize between-strip variation by careful placement of the plots. After measuring and marking the appropriate lengths, the farmer harvested full header-width strips, leaving 2 to 5 ft of uncut crop between strips to ensure a full-width cut. Combine header widths ranged from 16 to 25 ft. After each strip was cut, the grain was unloaded into a weigh wagon. Weights were converted to bushel per acre, based on 60 lb/bu for wheat and 48 lb/bu for barley.

The uniformity trials were arranged as eight 1500-ft long side-by-side strips at six locations in 1991 (Fig. 1). The strips were in pairs, the first consisted of six 250 ft segments (end-to-end), and the second a single 1500 ft segment. The full 1500 ft strip was a check for measure-

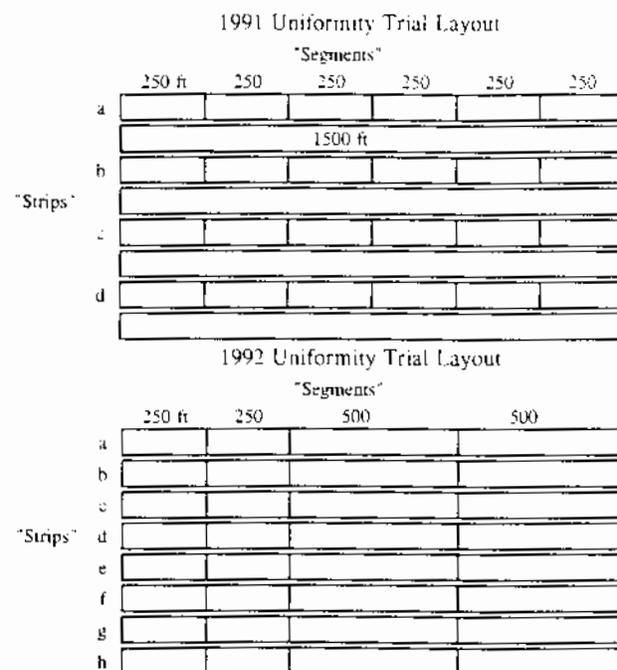


Fig. 1. Uniformity trial layouts for 1991 and 1992.

ment error accumulated when the six 250 ft strips were summed to estimate a long strip. There was no significant difference ($P > 0.05$) in yield between the full strips and the segmented strips. In three of the six fields, the variance between full strips was higher than between segmented strips, but we presently have no explanation as to why. Full strips are not included in the remainder of the analysis.

A small-plot version of the uniformity trial was placed adjacent to each large uniformity trial in 1991. This small uniformity trial had the same plot arrangement, but the segments were 25 ft long and the harvest width was 4 ft. Two feet of unharvested crop was left between harvest strips.

At eight sites in 1992, eight side-by-side strips 1500 ft long were harvested. Each of the strips was harvested in four segments: 250, 250, 500, and 500 ft long (Fig. 1). Full length 1500 ft strips were not harvested as was done in 1991.

In one field in 1991 (Dayton) the uniformity trial was harvested in two separate parts of a field. Four strips were in one contour strip and the other four in another, near-by contour strip. They were separated by a fallow contour strip. Similarly, each half of the 1992 Davenport trial were different barley varieties. At Helix in 1992, the strips were only 1000 ft long, with the two 250 ft and one 500 ft segments being harvested.

To assess variability that would be encountered within one block of a two treatment, randomized complete-block experiment, sample variances were calculated for all pairs of adjacent plots. Since there were four (1991) or eight (1992) strips, there were three or seven pairs of adjacent plots to estimate variance (Fig. 1). The variance estimates are for all possible pairs of plots of all possible lengths. Longer plots were made by combining segments in end-to-end fashion and computing a combined yield. The number of possible pairs of plots for each length in 1991 were: 18 at 250 ft, 15 at 500 ft, 12 at 750 ft, 9 at 1000 ft, 6 at 1250 ft, and 3 at 1500 ft. The 1992 layout provided the following number of possible pairs: 14 at 250 ft, 21 at 500 ft, 7 at 750 ft, 14 at 1000 ft, 7 at 1250 ft, and 7 at 1500 ft. For example, in 1991 there were a total of

four ways to combine the segments in a strip to make 750 ft plots: the first, second, and third segments; the second, third, and fourth segments; third, fourth, and fifth segments; and fourth, fifth, and sixth segments. The four adjacent strips allowed three estimates of variance for a pair of side-by-side plots. Therefore, the variance data for 750 ft plots at each site in 1991 were the average of 12 single degree-of-freedom estimates. These estimates are not independent, and we do not use them for tests of significance or statistical inference. This analysis and the following model are used to describe and summarize the data.

Based on an empirically derived formula relating variance per unit area to plot size (Smith, 1938), the model

$$\log(\text{variance}) = b \times \log(\text{plot length}) + \text{constant}$$

where b is the regression coefficient, was used to relate variance to plot length for each of the 14 sites. Regression coefficients and r^2 are given in Table 1.

Sites were separated into two groups according to variability predicted by the model at 1500 ft plot length (Table 1). Ten of fourteen sites had variance of < 5 at 1500 ft length, and were classified as *low* variability. Variance at the remaining four were above five and were classified as *high* variability. A single curve was fit to the combined data of each group for use in examining least significant difference and power for mean separation.

Least significant differences were calculated using predicted variances for a hypothetical randomized complete-block design with two treatments and four replications (i.e., three degrees of freedom for Student's t). Power for mean separation was estimated as a function of predicted variances, and hence plot length, through the relationship:

$$1 - P[-(t_{\alpha/2} + d/s_d) < t < (t_{\alpha/2} - d/s_d)]$$

where $t_{\alpha/2}$ is the two-tailed critical value of Student's t at the selected significance level, d is the desired detectable difference in mean values, s_d is the estimated stan-

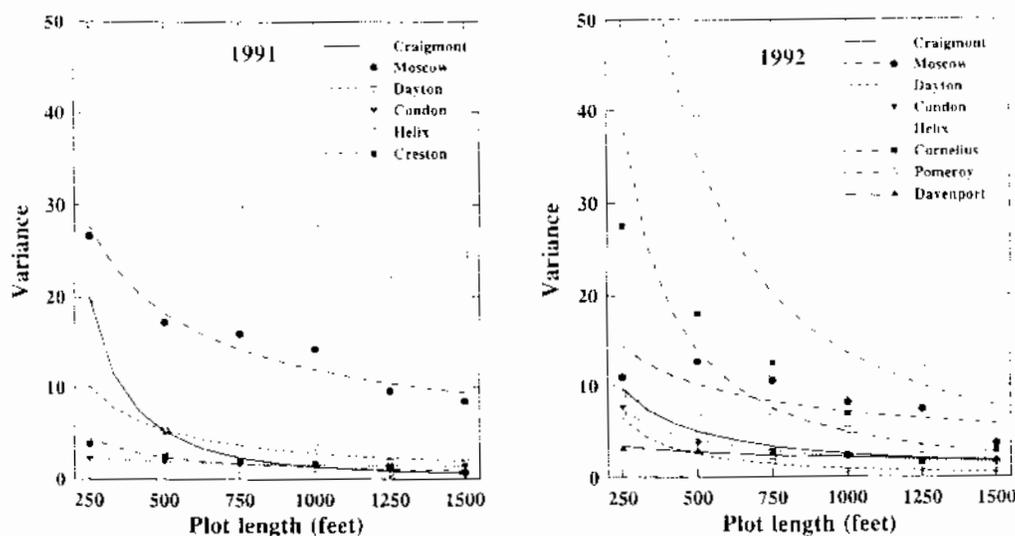


Fig. 2. Variance vs. plot length at six sites in 1991, and eight sites in 1992. Lines are regression of $\log(\text{variance})$ on $\log(\text{length})$.

standard error of the mean difference and equals $(2s^2/n)^{1/2}$, where n is the number of replications for each of the treatments (Carmer and Walker, 1988). Desired detectable difference in mean values of 4 bu/acre for the low variability group and 7 bu/acre for the high variability group were chosen for discussion because they produced acceptable power given sufficient replication and plot length.

RESULTS AND DISCUSSION

Variance decreased with increasing plot length at all uniformity trial sites in a decay function pattern (Fig. 2). This result differs from most previous work on the relationship between plot size and variance that indicated little decrease beyond 1/40 acre (LeClerg et al., 1962, p. 114). The shape and size of our plots compared with common research plots may explain the discrepancy. We used long, narrow, side-by-side plots, and typical field variability was probably localized in patches. As our plot length increased without becoming wider, more and more separate patches of heterogeneity were included in each plot. Since the plots were a constant width and distance from each other, the field areas they sampled did not become more distant with increasing size, as occurs with rectangular plots.

In the small uniformity trials placed near the large trials in 1991, variances were much like those of the large trials (Fig. 3). Despite a 10-fold difference in lengths, the response of variance to increasing plot length follows the same decay pattern. Of the six small uniformity trials, three had lower and three had higher variance than their large counterparts. It is remarkable that a 10-fold reduction in lengths and 5-fold reduction in widths would produce data so similar in magnitude and in the relationship of plot length to variance. It might be expected that the magnitude of variance would continue to decline from the 150 ft plot length of the small trial to the 250 ft plot

length of the large trial, but it does not. A partial explanation for the return to high variance levels at plot lengths of 250 ft compared with very low variance levels for 150 ft plots may involve measurement errors such as incomplete unloading of the farmer's combine. We may further speculate that the different widths of harvest (4 ft vs. 16 to 25 ft) caused the small and large trials to encounter different scales of field heterogeneity, or perhaps the proportion of width to length of a plot and proximity to the adjacent plot has a major influence on variance. Although the data presented here do not provide an explanation for the similarities between the small and large uniformity trial data, they do provide conclusive evidence that on-farm tests can be designed to produce data with experimental errors that compare favorably to those of small research plots.

The small trials covered a total area of only about one-third acre, and this makes selection of a uniform site easier. Uniformity is an important factor when the number of treatments is large, as is often the case under small plot research conditions. One could question how well small plot data represent an entire field, and while not eliminating the problem, large plots decrease this potential bias. This study does not allow us to speculate about the optimum plot size and shape for research plots shorter than 25 ft, but we can conclude that 75 ft plots are less variable than 25 ft plots.

Data from 23 farmer-implemented on-farm tests (Wuest et al., 1992) located in the same region as this study confirm that the variability estimated by the large uniformity trials represent the degree of variability which actual on-farm tests encounter. Most of the on-farm tests involved two tillage treatments with two to four replications in wheat, barley, or canola (*Brassica napus* L.) fields. Plot lengths ranged from 100 to 2500 ft. After subtraction of block and treatment effects, mean square error for yield (in bu/acre) ranged from < 1 to 135, with 19 of the 23 cases below 11, and 13 cases below 5. This agrees with data from the uniformity trials, where variance ranged from < 1 to 112 over all plot lengths. A greater range of estimates of variances would be expected in

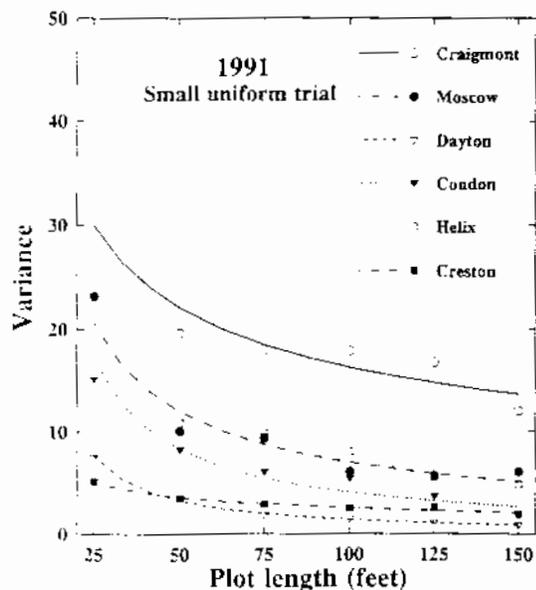


Fig. 3. Variance vs. plot length of small plots (4-ft width) placed near each large uniformity trial in 1991. Lines are regression of $\log(\text{variance})$ on $\log(\text{length})$.

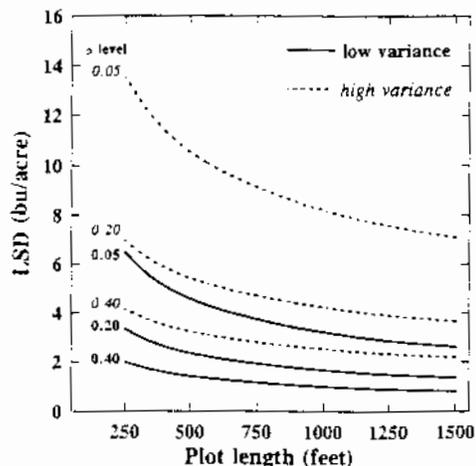


Fig. 4. Least significant difference (LSD) vs. plot length for an experiment with four replications at α levels of 0.05, 0.20 and 0.40. Variances are based upon log-log regression of grouped data: low variance (10 of 14 sites), and high variance (4 of 14 sites).

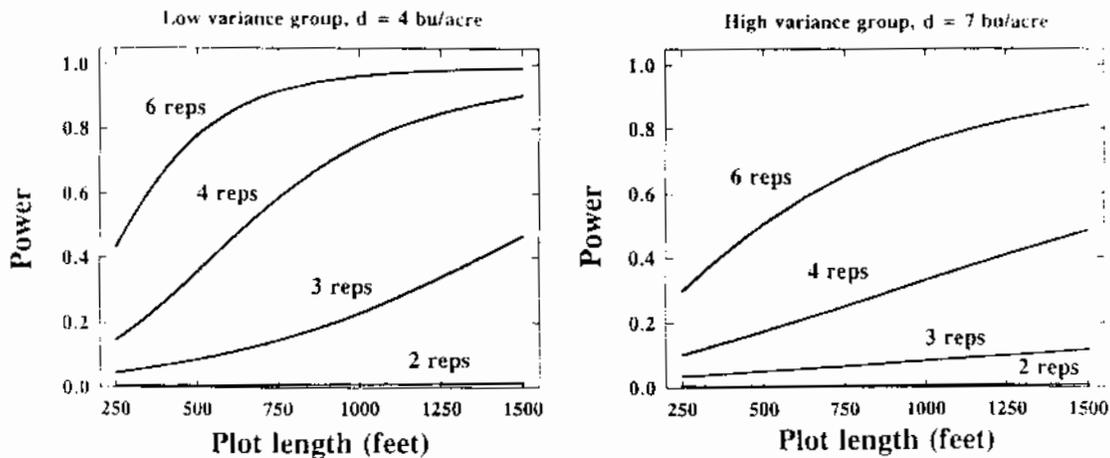


Fig. 5. Power for mean separation of experiments with different plot lengths and number of replications (reps). Power is shown for detecting a true difference ($\alpha = 0.05$) of 4 bu/acre for low variance data, and 7 bu/acre for high variance data.

the on-farm tests due to the lower number of replications and sometimes very short plot lengths. These were the first on-farm tests performed by these particular farmers, and increased experience may reduce experimenter errors and improve designs.

Least significant differences for sites classified in low and high variance groups were calculated at α levels of 0.05, 0.20, and 0.40 (Fig. 4). These calculations are appropriate for an individual two-treatment, four-replication test. The larger α levels are presented because of increasing interest among production management researchers in significance levels that are selected based upon risk assessments instead of scientific hypothesis testing (Carmer & Walker, 1988). The low variability sites should be able to produce LSDs of ≈ 6 bu/acre at a high significance level ($\alpha = 0.05$) even with relatively short plot length. This approaches 4 bu/acre at plot lengths > 1000 ft. For management decisions with less risk involved ($\alpha = 0.20$ or 0.40), LSDs of < 2 bu/acre are possible with long plots. At sites with high variability, short plots and a demand for high significance will result in high LSDs, above 12 bu/acre (Fig. 4). Plots > 1000 ft long or α levels > 0.05 can produce LSDs of 8 bu/acre or less.

Given an estimate of variance, it is possible to calculate the probability that a proposed experiment will correctly detect a true difference between treatments. This probability is called power for mean separation. Power calculations are useful for designing experiments that are likely to accomplish predetermined goals. If it is desired that a test have 80% or better probability of detecting a mean difference of 4 bu/acre (at $\alpha = 0.05$), four replications of 1250 ft plots at a low variability site would be required according to the results of this study (Fig. 5). Six replications would allow a higher probability of detecting a 4 bu/acre difference if plots are over 500 ft long. Two or three replications are not likely to detect a 4 bu difference at any plot length.

Sites with high variability are very unlikely to detect a 4 bu/acre difference at $\alpha = 0.05$; even at 1500 ft the power rises to only 0.38 with six replications (not shown).

At a 7 bu/acre desired detection level, six replications and 1250 ft plots are needed to ensure power > 0.80 .

We conclude that, in most fields of the dryland Pacific Northwest, 750 ft or longer side-by-side on-farm test plots will have much less experimental error than 250 or 500 ft plots. With four or more replications, properly designed on-farm tests can be expected to produce LSDs ($\alpha = 0.05$) of 5 bu/acre or less.

ACKNOWLEDGMENTS

Invaluable assistance in the collection of data was provided by the following county extension agents: Ben Barstow, Dave Bragg, Tim Miller, and Roland Schirman; and the following farmers: Dale Dietrich, Norm Goetze, Gary Houser, Howard Nelson, Jack Osterlund, Dick Southern, Stan Timmerman, Wayne Westburg, and George Wood.

REFERENCES

- Carmer, S.G., and W.M. Walker. 1988. Significance from a statistician's viewpoint. *J. Prod. Agric.* 1:27-33.
- Kittrell, B.U. 1974. Result demonstration technique—history, philosophy, and contemporary nature. *J. Agron. Educ.* 3:90-94.
- LeClerg, E.L., W.H. Leonard, and A.G. Clark. 1962. *Field plot technique*. 2nd ed. Burgess Publ. Co., Minneapolis, MN.
- Lockeretz, W. 1987. Establishing the proper role for on-farm research. *Am. J. Alter. Agric.* 2:132-136.
- Rzewnicki, P.E., R. Thompson, G.W. Lesoing, R.W. Elmore, C.A. Francis, A.M. Parkhurst, and R.S. Moomaw. 1988. On-farm experiment designs and implications for locating research sites. *Am. J. Alter. Agric.* 3:168-173.
- Schmitt, M.A., S.J. Openshaw, and M.W. Davis. 1992. Effects of control-based adjustments on precision of large-plot yield trials. *J. Prod. Agric.* 5:295-298.
- Smith, H.F. 1938. An empirical law describing heterogeneity in the yields of agricultural crops. *J. Agric. Sci.* 28:1-23.
- Wuest, S.B., B.C. Miller, R.J. Veseth, S.O. Guy, D.J. Wysocki, and R.S. Karow. 1992. 1992 Pacific Northwest on-farm test results. *Dep. of Crop and Soil Sciences Technical Rep. 92-4*, Washington State Univ., Pullman.