Increasing Plot Length Reduces Experimental Error of On-Farm Tests

Stewart B. Wuest, Baird C. Miller, J. Richard Aldridge, 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 on highly variable dryland ideal production fields. 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 30 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 100 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 5 to 32 (high variance). Figure 1 shows least significant differences at α = 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 5000 ft.

To ensure the best results, we recommend that plots be as long as is practical.
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.

![Graph](image)

**Fig. 1.** The effect of increasing plot length on LSD values at α = 0.05 is shown for experiments with different numbers of replications (trials). (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. Wiener,* Baird C. Miller, J. Richard Allredge, Stephen O. Guy, Russ S. Knowl, Roger J. Veech, and Donald J. Wysocki

I t is well established that conducting 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 plots, narrow, side-by-side plots. This study examined the relationship between plot length and experimental error, and assessed the potential statistical outcome of on-farm tests performed in the dryland region of the Pacific Northwest, USA. Experiments were conducted in wheat (Triticum aestivum L.) and barley (Hordeum vulgare L.) fields in measure yield, varieties of combine-wide 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 e = -0.58). Variance declined rapidly as plot length increased from 250 to 750 ft or, at mid sites. Averaging the ten least variable sites, the LSSE (0.01) with three degrees of freedom decreased from 6.5 to 2.5 at 250 ft and length increased from 250 to 1500 ft. At this site, 10 sites, power for mean separation was 0.05% of treatments with 4 beds were true effect of with six replications and 750 ft plot length, or if replicating and 1250 ft plot length, with appropriate replication and plot length, on-farm tests can be designed for highly variable dryland region with good control of experimental error.

Experiments conducted on-farm tests are an effective tool for development of improved crop management practices. The on-farm test is also used as an extension tool for the promotion of new technology. The estimation between an on-farm test and a demonstration and in the use of a range of scientific experimental design in the on-farm test to greatly increase reliability of measured results and show the use of statistical analysis (Klotz, 1987; Wysocki, 1984). With an increased interest in on-farm tests, there is need to check the effectiveness of the experimental design being used (Lackenby, 1987).

The most common design promoted for use in on-farm tests in the randomized complete block with a limited number of treatments and four or more replications. A common plot size on on-farm tests of the Midwest is 20 to 40 ft wide by 120 ft long and the treatments are replicated six to eight times in adjacent blocks. The use of this design for on-farm tests was assessed under the Midwest new crop conditions and was shown to give satisfactory results (Klotz, 1987). Again in the Midwest, Schmitz et al. (1982) found the randomized complete block design to be more efficient than the Tesar 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 40%.

The region's fields are also characterized by variable soil depth, 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 research-size plots of <0.1 acre (Lackenby, 1987; Schmitz, 1982). 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 experimen
tal precision by optimizing the relationship between plot size and number of replications. Available field area often was 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 at 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 larger 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-wide plots. We also compare experimental error between on-farm tests and small plot research, and examine the loss of significant differences and the power for mean separation that an individual, single-location on-farm test will probably produce.


Published in J. Prod. Agric. 7:215-218 (1994).

J. Prod. Agric., Vol. 7, no. 2 1994 211
TABLE 1. Mean annual precipitation, soil taxonomy description, crop, average yield of unifactorial trials, long row spacing treatments, predicted variance of 1500 ft long plots, and high or low variance group assignment for the 3 uniformity trial areas.

<table>
<thead>
<tr>
<th>Area</th>
<th>Mean annual precipitation</th>
<th>Soil taxonomy description</th>
<th>Crop</th>
<th>Average yield of unifactorial trials</th>
<th>Predicted variance of 1500 ft long plots</th>
<th>High or low variance group assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern</td>
<td>42.0</td>
<td>Verte, eutric, rendzina, fine loam, high in organic matter</td>
<td>Wheat</td>
<td>79.1</td>
<td>3.40</td>
<td>High</td>
</tr>
<tr>
<td>Western</td>
<td>75.1</td>
<td>Verte, eutric, rendzina, fine loam, high in organic matter</td>
<td>Sugar</td>
<td>68.1</td>
<td>3.60</td>
<td>Low</td>
</tr>
<tr>
<td>Southern</td>
<td>84.0</td>
<td>Verte, eutric, rendzina, fine loam, high in organic matter</td>
<td>Sorghum</td>
<td>60.2</td>
<td>3.80</td>
<td>Low</td>
</tr>
</tbody>
</table>

MATERIALS AND METHODS

Uniformity trials were harvested in 14 wheat or barley fields farmed by commercial growers in 1991 and 1992. A uniformity trial set to measure the variability of soybean growing regions in the mid-Paciic Northwest, and

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 consisting of six 230 ft segments (end-to-end), and the second a single 1500 ft segment. The full 1500 ft strip was a check for measure-
even error accumulated when the six 50 ft strips were
varied from 400 ft; in 1991 strips were included in the analysis.

A small plot version of the uniformity trial was placed parallel to the larger uniformity trial in 1991. The small uniformity trial had the same plot arrangement; the segments were 25 ft long and the harvest width was 4 ft. Two feet of unharvested strip were left between harvest strips.

At eight sites in 1992, eight side-by-side strips 1500 ft long were harvested. Each of the eight strips was harvested in four segments: 250, 220, 250, and 500 ft long (Fig. 1). Full length 1500 ft strips were not harvested as was done in 1991. In one field in 1991 (Byers), the uniformity trial was harvested in two separate parts of a field. Four sites 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 Byers plot trial were different barley varieties. At Holt 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 treatment, randomized complete-block experiment, single 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. 3). The variance estimates are for all possible pairs 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 350 ft, 15 at 500 ft, 12 at 750 ft, 9 at 1000 ft, 6 at 1500 ft, and 2 at 2000 ft. The 1992 layout provided the following number of possible pairs: 11 at 350 ft, 21 at 500 ft, 7 at 750 ft, 14 at 1000 ft, 7 at 1500 ft, and 7 at 2000 ft. For example, in 1991 there were a total of four ways to combine the segments in a strip to make 750 ft; first, second, and third segments; the second, third, and fourth segments; third, fourth, and fifth segments; and fourth, fifth, and sixth segments. The two 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 (Table 1) were the average of 12 single stage of freedom estimates. These estimates are not independent, and we do not use them for tests of significance or by 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, 1963), the model for

\[
\text{log(variance)} = b \times \text{log(plot length) - constant}
\]

where b is the regression coefficient, was used to relate variance to plot length for each of the 14 sites. Regression coefficients and b 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 estimating least significant difference and power for these separations. Least significant differences were calculated using predicted variances for a hypothetical randomized complete block design with two replicates and four replications (i.e., three degrees of freedom for Student's t). Power for mean separation was estimated as a fraction of predicted variances, and hence mean length, through the relationship:

\[
1 - P[ -\bar{c}_{0.025} - d/\bar{s}_0 < \bar{c}_{0.025} - d/\bar{s}_0]
\]

where \(\bar{c}_{0.025}\) is the two-tailed critical value of Student's t at the selected significance level, d is the desired detectable difference in mean values, \(\bar{s}_0\) is the estimated mean.

Fig. 3. Variance vs. plot length as for sites in 1991, and eight sites in 1992. Lines are regression of log(variance) on log(length).

RESULTS AND DISCUSSION

Variability 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 field size decreases beyond 1/4-acre (Koelner et al., 1982, p. 24). The shape and size of our plots compared with commercial research plots might explain the discrepancy. We used long, narrow, side-by-side plots, and typical field variability was probably located in pasture. As our plot length increased without becoming larger, more and more separate patches of heterogeneity were included in each plot. Since the plots were a constant width and distance from one another, 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 less than 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 fewer and three had higher variance than their large counterparts. It is remarkable that a 10-fold reduction in lengths and 3-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 sampling or cull of the farmer's combine. We may further speculate that the different widths of harvest (4 ft vs. 10 to 25 ft) showed the small and large trials to counter 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 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 cover 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 15 ft plots are less variable than 25 ft plots.

Data from 23 farmer-implemented on-farm tests (Wien 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 (Bruguiera napus L.) fields. Plot lengths ranged from 100 to 2,000 ft. After subtraction of block and treatment effects, mean square error for yield (bu/acre) ranged from <1 to >125, with 19 of the 23 cases below 15, and 13 cases below 5. This agrees with data from the uniformity trials, where variance ranged from <1 to 1.15x, with all plot lengths. A greater range of estimates of variances would be expected in
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 a level of 0.05, 0.20, and 0.40 (Fig. 2). These calculations are appropriate for an individual two-treatment, four-replication test. The larger 0.20 levels are presented because of increasing interest among production management researchers in significance levels that are selected based upon risk assessment instead of scientific hypotheses testing (Cornel & Walker, 1980). The low variability sites should be able to produce LSDs of 4 bu/acre at a high significance level (α = 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 (α = 0.20 or 0.40), LSDs of 3 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. 2). Plots > 1000 ft long or a level > 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 design experiments that are likely to accomplish predetermined goals. It is desired that a test have 80% or higher probability of detecting a mean difference of 4 bu/acre at α = 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/acre difference at any plot length.

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

As 7 bu/acre desired detection level, six replications and 1200 ft plots are needed to ensure power > 0.95.

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 (α = 0.05) of 5 bu/acre or less.

ACKNOWLEDGMENTS

Valuable assistance in the collection of data was provided by the following county extension agents: Ben Barglow, Dave Bragg, Tim Miller, and Roland Schranz, and the following farmers: Dale Dietrich, Norm Goette, Gary Horvat, Howard Nelson, Jack Overland, Dick Southern, Stan Timmerman, Wayne Westburg, and George Wood.

REFERENCES

Carraro, S.C., and W.M. Walker. 1980. Significance from a variabil-
Kermit, R.U. 1954. Field determination techniques—soil, 
Lettier, E.L., W.E. Loveless, and A.C. Clark. 1962. Field plot tech-
niques. 2d ed. Burgess Publ. Co., Minneapolis, MN.
Locken, W. 1967. Establishing proper role for on-farm research. 
Mertenski, P.E., R. Thompson, G.W. Lupton, R.W. Elmers, C.A. 
Penninga, A.M. Peartree, and B.S. Mummert. 1980. On-farm treat-
ment design and report forms for testing research sites. Am. J. Agri. 
Sci. 130:1-12.
control treatments applicable to precision of large scale field 
J. Prod. Agric. 5:232-236.
Smith, R.F. 1935. An empirical line measuring yield in the fields 
Univ., Pullman.