Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2002 - 14th Annual Conference Proceedings

UNREPLICATED SPATIAL DESIGNS COMPARED USING OPTIMALITY CRITERIA

Tisha Maas

David Marx

Dr. Jeff Pedersen

UNREPLICATED SPATIAL DESIGNS COMPARED USING OPTIMALITY CRITERIA

Tisha Maas and Dr. David Marx, Department of Biometry
Dr. Jeff Pedersen, USDA-Agricultural Research Service
University of Nebraska—Lincoln
Lincoln, NE 68583-0712

ABSTRACT

In field trials including large numbers of varieties, it is often impossible or impractical to replicate each variety. In these situations, the researcher may choose to use only one replicate of each test variety and to include a "check" variety every so often so that the spatial variability of the field may be determined. Five different check patterns were purposefully designed, each possessing distinct characteristics. The purpose of this study is to determine which spatial patterns for the check variety are better able to identify the spatial structure in a field and to rank the experimental varieties accurately. The problem was approached in two ways. First, the check patterns were compared using optimality criteria. Then, the patterns were applied to an actual field experiment, and the data collected was used to identify the spatial structure of variation in the field and to test for experimental variety differences. It is shown that the results from the optimality criteria were not necessarily comparable to what was actually observed in the field.

1. INTRODUCTION

One problem in the analysis of variety trials in the field concerns the complete absence of replication among the test varieties, either because it is impossible or unfeasible to replicate each of them. This problem is often intensified by significant spatial variation over the experimental area. In this situation, the researcher may choose to include a "check" variety every so often in the midst of the experimental varieties so that the spatial variability of the field may be determined and thus taken into account in the analysis. The question arises as to how these "checks" should be arranged—what pattern yields the most information concerning the spatial structure of the field? After this information is obtained, the variety means can then be adjusted to remove this field variation and thus give a more representative picture of true variety differences.

To investigate this problem, five distinct check patterns were designed, each possessing some intuitive advantage or disadvantage (Figure 1). In each of the designs, an X represents a check plot, and each empty box represents a different experimental variety. Next, each of the check patterns is discussed in detail.

DESIGN A. This pattern consists of 20% checks. Design A was modeled after the "Knight's Move" 5x5 Latin Square, which has historically performed quite well in spatial correlation simulations (Marx, 1992). This involves a systematic arrangement of the check varieties, i.e., one down and two over. This design is usually credited to a Norwegian, Knut Vik, and is sometimes referred to as the Knut Vik Square (Federer, 1955). The underlying philosophy of the Knight's Move is that the treatments should be as well separated as possible. Since this separation principle was applied in the construction of Design A, this pattern is anticipated to be optimal. Also, it should be noted that in this pattern, the density of checks was twice that of the other designs.

DESIGN B. The density of the checks for this design and all that follow is roughly 10%. Though the checks are not as dense as in Design A, Design B embraces the same separation principle. Notice that the check varieties are well spread out over the entire experimental area.

DESIGN C. Again, the checks seem to be spread out over the experimental area. However, there is one distinction between Design C and the others. This pattern gives the potential advantage of having some very closely paired check varieties, which may be useful in detecting spatial structure if the range of spatial correlation is small.

DESIGN D. As in the previous designs, the check variety is spread out over the entire experimental area. In Design D, some pairs of checks are closer together than others, but the pairs of checks that are placed right next to each other are in only one area of the field, unlike Design C where the pairs are placed throughout the entire field .

DESIGN E. This design completely ignores the principle of separation. Moreover, most of the test varieties are not even close in proximity to a check; consequently, this design is expected to be the least optimal.

2. DISCUSSION OF A-OPTIMALITY

One criterion that is widely used in the comparison of various designs is that of A-optimality, which is a particularly simple and intuitively appealing measure of optimality since it is equivalent to minimizing

i. the average variance of the treatment effects,
ii. the average variance of an estimated difference in treatment effects, and
iii. the expected value of the treatment sums of squares when there are no treatment differences (Martin, 1986).

A-optimality was originally developed for independent observations and is calculated as follows:

$$\phi_{a-opt} = \sum \left(\frac{1}{\lambda_i} \right)$$

where λ_i is the i^{th} eigenvalue of X'X and X is the design matrix.
This study, however, assumed the presence of spatial correlation. Martin (1986) proposed an extension to the A-optimality criterion for spatially correlated data. His criterion uses the

$X'V^{-1}X$ matrix as opposed to just the X'X matrix, where V^{-1} is the inverse of the covariance structure of a spatially correlated design. Though several spatial correlation models exist, this study examines only the results of three specific spatial structures: exponential, spherical, and gaussian. The covariance structures are given as follows:

Exponential: $[v_{ij}]=Cov\{exp(-3h/r)\}$
Spherical: $[v_{ij}]=Cov\{1-1.5h/r + .5(h/r)^3\}$
Gaussian: $[v_{ij}]=Cov\{exp(-3h^2/r^2)\}$

Here, $V=[v_{ij}]$ where h is the distance between observation i and observation j and r is the range of the semivariogram. Thus, these spatial structures are incorporated into V, and the calculation of A-optimality follows. Note that the A-optimality is dependent on the range of the semivariogram. Thus, both a "small" and "large" value for the range were used in the computations. For each design in this study, the A-optimality was computed using a nugget of zero and both a range of 5 and a range of 20 for each of these spatial structures.

3. RESULTS OF A-OPTIMALITY

The results are displayed in Table 1. As expected, Design A was the most optimal no matter what spatial structure or range was used in the calculation. Also, Design E consistently performed the worst. Of interest were the intermediate designs B, C, and D. Design B appeared to be the most optimal of the three, aside from the situations where the range was small and the spatial structure was exponential or where the range was large and the spatial structure was gaussian. The cause of this discrepancy is not exactly known.

However, what is happening between Designs C and D is intuitively appealing. Design C was always more optimal than Design D when the range was small. On the contrary, Design D consistently performed better than Design C when the range was large. In Design C, the checks are much closer together; thus, if the range were small, this design would be much more likely to capture the spatial structure of the field. In Design D, the checks are more spread out over the experimental area, which makes this design preferable as long as the range is not too small. If the range is large, the checks can still be used to adequately estimate the spatial variability, which explains why Design D was better than Design C when the range was large.

Finally, notice that the average variance of treatment differences is calculated for each design and is displayed in Table 1. It should be noted that for spatially correlated observations, the best design does not always yield the smallest variance as it does for independent observations.

4. DISCUSSION OF THE ACTUAL FIELD EXPERIMENT

A field experiment was conducted in conjunction with the USDA-Agricultural Research Service in Lincoln, NE. Roughly 2,000 sorghum plant introduction varieties were to be planted in a research field near Mead, NE, in the summer of 2001. Resources allowed for only one plot of each experimental variety to be planted. The sorghum variety B Wheatland was used as a check variety, and the sorghum was planted in the research area according to the five check patterns.

Three response variables were recorded for each individual plot: average height, bloom date, and percent stand. Other variables such as yield are also of interest, but due to laboratory constraints, that data was not available. The variable bloom date was used in the analysis that follows since this variable showed the most consistent results concerning spatial structure.

The data was analyzed in three ways: using just the checks to estimate spatial structure and then analyzing the entire data set with that spatial structure, using the checks and experimental varieties to simultaneously estimate the spatial structure and analyze the data, and using the average height of the neighboring plots as a covariate. The reasoning behind the first two analyses is as follows: the results obtained using just the checks should be identical to the results obtained when the experimental varieties are also included in the analysis since the spatial structure is based on only the checks, regardless. Thus, the first two analyses were performed to confirm this. The third analysis, which included a covariate, was carried out because of what was observed in the field after the plants started to grow and the study took an unforeseen turn. The average height of the check variety B Wheatland is about 85 cm, while the average height of the plant introductions ranged anywhere from 75 cm to 350 cm. As a result, the checks were possibly affected by the shading of the taller experimental varieties, and the bloom dates of the check varieties could have been influenced by the height of the surrounding plants. Since the spatial variation of the field was to be determined from these bloom dates, it was determined that the heights of the nearest neighbors should be incorporated as covariates in the analysis so that the proper spatial structure could in turn be determined.

The following SAS (SAS Institue, 2000) code was used in the spatial analysis. It should be noted that since the nugget effect did not appear to be significant, a no-nugget model was assumed.

```
proc mixed covtest;
class pedigree;
model bloomdate=pedigree neighbor_height/ddmf=kr;
repeated/subject=intercept type=sp(exp)(lat lng);
parms (9)(2.5);
run;
```

The *covtest* option was included to obtain estimates and asymptotic standard errors for the sill and range in the SAS listing. The *model* statement was adjusted for each of the analyses: for checks only, the model was bloomdate= , and when the treatments were included, the model was bloomdate=pedigree. The programming statements shown here are for the third analysis which incorporated the covariate. The *repeated* statement incorporates the spatial structure into the model. This example is for the exponential model—other options include (*sph*) for spherical and (*gau*) for gaussian. The variables lat and lng denote the coordinate system. In this study, the coordinates were given simply as measurements in feet over the experimental area. Finally, the *parms* statement includes reasonable starting values for the sill and range which were obtained from an empirical semivariogram created in GEOEAS using only the checks as a data set. Prior research (Marx and Stroup, 1993) indicates that it is essential to specify initial estimates of the range and sill; PROC MIXED's default initial values frequently lead to unreasonable estimated values.

5. RESULTS OF THE ACTUAL FIELD EXPERIMENT

For each of the five designs, this program was run for all three spatial structures and also without the repeated statement (to simulate the situation where no spatial structure exists). Estimates for the range and sill were calculated along with model fitting criteria for each situation, and the results are shown in Table 2.

First of all, it should be noted that for Designs A, B, and C, the estimates for the range and sill calculated using only the checks in the analysis were identical to the estimates obtained including the experimental varieties, which was anticipated.

Second, contrary to what was expected, the analysis incorporating the heights of neighboring plants as a covariate was not particularly useful. The covariates were only marginally significant, so it was decided to use the analysis that did not include a covariate.

Also, notice that when the experimental varieties were included in the analysis for Designs A and C, the gaussian model failed to converge. This could be attributed to an inherent disadvantage of the Gaussian model; it approaches its origin with a zero gradient which can lead to bizarre effects when used for estimation (Webster and Oliver, 2001). When the experimental varieties were included in the analysis, the covariance matrix became quite large because of the considerable number of data points and unstable because of the parabolic behavior of the Gaussian function near the origin. Notice that this problem did not occur when the data set consisted of only the checks. Thus, in situations such as these, it may be reasonable to use only the checks in the analysis to obtain estimates for the spatial parameters.

Finally, while competing covariance models cannot be tested directly, they can be compared using their model fitting criteria. Here, the models were compared based on their AIC_c. It appears that Designs A and C have a slight exponential spatial structure, and that Design B has no spatial structure. However, the analysis of Design B still gave significant parameter estimates for the range and sill; this could possibly be due to the fact that the estimated asymptotic error of the range may not be accurate, hence generating a test statistic that is incorrect. Finally, Designs D and E detected no spatial structure whatsoever. The estimate for the range in these two designs was zero; i.e., the observations were essentially independent.

6. CONCLUSIONS

Results from the optimality criteria may not necessarily be compatible with what actually occurs in the field. Designs such as A where the check varieties are separated as much as possible yield the best optimality but may not be adequate in detecting underlying spatial structure in practice if the range is too small. Essentially, some feasible estimate of the range is needed prior to a study in order to determine which check pattern is best in an actual field experiment. If one is fortunate enough to have a good estimate of the range before selecting a check pattern, then the concept of separation can be applied if the range is larger than the minimum distance between checks. If not, then another design, such as C or D, should be used in the experiment. Recall that Design C would probably work better in this instance since the range would be "small" as compared to the minimum distance between checks. Finally, Design C has no checks on the edge of the design which will allow experimental varieties to be, on the average, closer to a check.

References

Federer, Walter T. 1955. *Experimental Design.* New York: Macmillan Company.

Martin, R. J. 1986. "On the design of experiments under spatial correlation." *Biometrika* **73**, 2, pp. 247-77.

Marx, David B. 1992. "Designed Experiments in the Presence of Spatial Correlation." *Proceedings of the 1992 Kansas State University Conference on Applied Statistics in Agriculture.* Kansas State University, Manhattan, KS.

Marx, David B. and Walter W. Stroup. 1993. "Analysis of Spatial Variability using PROC MIXED." *Proceedings of the 1993 Kansas State University Conference on Applied Statistics in Agriculture.* Kansas State University, Manhattan, KS.

Webster, Richard and Margaret A. Oliver. 2001. *Geostatistics for Environmental Scientists.* West Sussex, England: Wiley.

SAS Institute. 2000. SAS OnlineDoc, Version 8. Cary, NC: SAS Institue, Inc.

Figure 1.

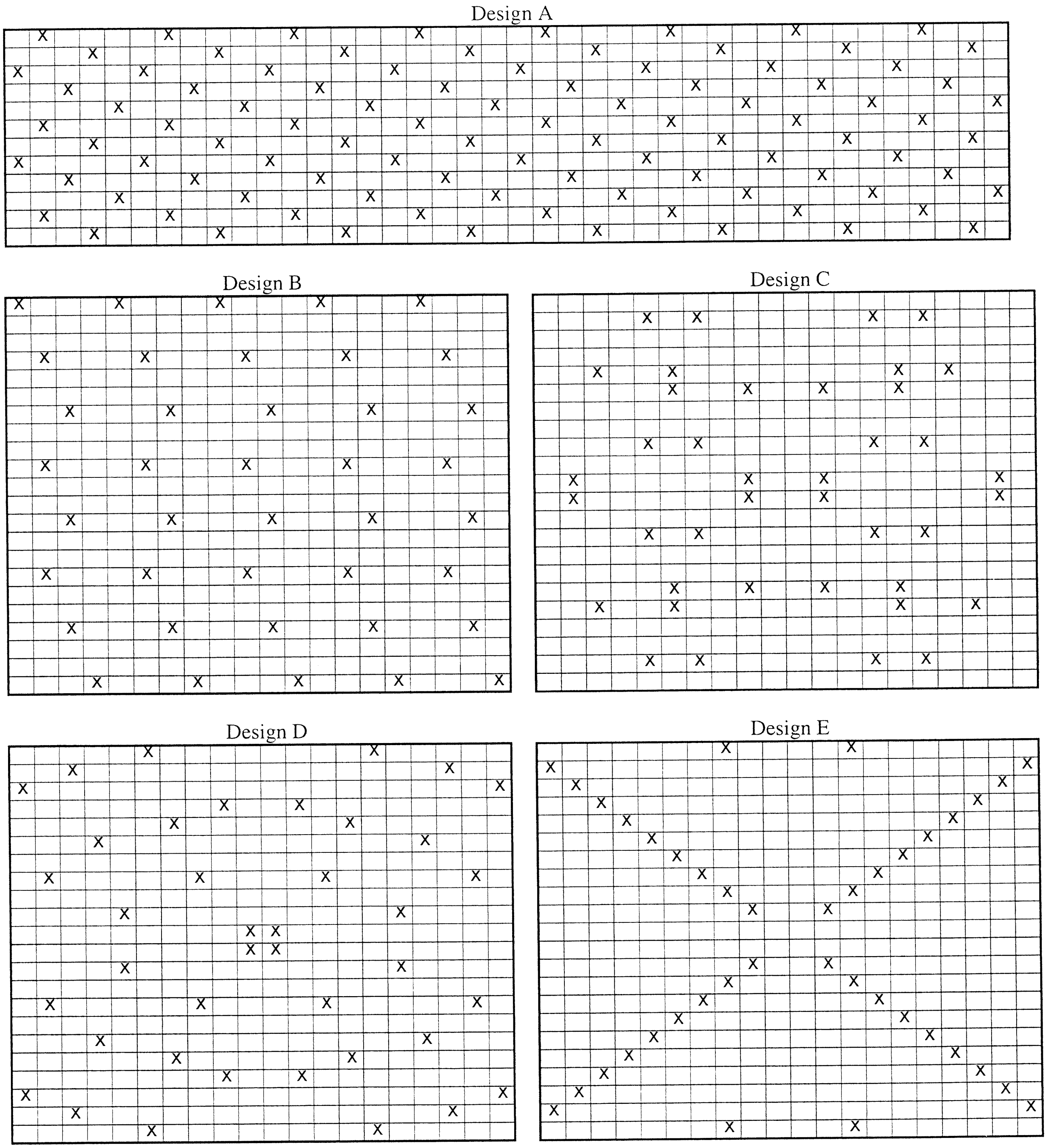

Table 1.

A-OPTIMALITY FOR SMALL RANGE (RANGE=5).

Exponential

Design	A-optimality	Mean Variance	Var(Variance)
A	240.69	1.1993	0.00703
B	302.31	1.5048	0.00222
C	302.14	1.5284	0.03533
D	315.06	1.5125	0.02577
E	340.61	1.64	0.05049

Spherical

Design	A-optimality	Mean Variance	Var(Variance)
A	146.81	0.7117	0.00647
B	228.08	1.1209	0.03695
C	236.93	1.1857	0.07312
D	242.46	1.1511	0.04834
E	301.99	1.4482	0.11177

Gaussian

Design	A-optimality	Mean Variance	Var(Variance)
A	28.89	0.0712	0.00251
B	118.52	0.5158	0.05192
C	126.49	0.5661	0.11998
D	127.75	0.5439	0.07119
E	237.17	1.0911	0.21101

A-OPTIMALITY FOR LARGE RANGE (RANGE=20).

Exponential

Design	A-optimality	Mean Variance	Var(Variance)
A	133.59	0.3467	0.00146
B	184.16	0.5147	0.00984
C	192.9	0.5607	0.02164
D	187.55	0.5178	0.00931
E	225.16	0.7079	0.04367

Spherical

Design	A-optimality	Mean Variance	Var(Variance)
A	105.54	0.1761	0.00045
B	142.7	0.2707	0.00384
C	150.37	0.3016	0.00892
D	143.27	0.2699	0.00307
E	168.15	0.3992	0.02011

Gaussian

Design	A-optimality	Mean Variance	Var(Variance)
A	63.99	0.00003	2.90 E -10
B	84.07	0.00019	1.43 E -07
C	87.34	0.00029	4.00 E -07
D	82.19	0.00009	1.46 E -08
E	91.81	0.02088	0.000719

Kansas State University

Table 2.
Results of Analyses for Each Design.

DESIGN A	Spatial Structure	Range	Sill	AICC
checks only	exponential	29.91	1.53	310.4
	spherical	23.47	1.53	311.4
	gaussian	32.52	1.53	311.5
	none	----	----	315.4
with treatments	exponential	29.91	1.53	310.4
	spherical	23.47	1.53	311.4
	gaussian			
	none	----	----	315.5
with covariate	exponential	22.53	1.45	330.2
	spherical	18.78	1.44	330.3
	gaussian			
	none	----	----	329.9

DESIGN B	Spatial Structure	Range	Sill	AICC
checks only	exponential	18.36	2.74	156.7
	spherical	14.74	2.73	157.1
	gaussian	21.33	2.73	157.1
	none	----	----	155.6
with treatments	exponential	18.36	2.74	156.7
	spherical	14.74	2.73	157.1
	gaussian	21.33	2.73	157.1
	none	----	----	155.6
with covariate	exponential	18.03	2.65	171.2
	spherical	14.73	2.64	171.6
	gaussian	21.33	2.64	171.6
	none	----	----	170.1

DESIGN C	Spatial Structure	Range	Sill	AICC
checks only	exponential	21.39	2.82	151.8
	spherical	14.42	2.78	151.9
	gaussian	19.44	2.8	151.9
	none	----	----	153.2
with treatments	exponential	21.39	2.82	151.8
	spherical	14.42	2.78	151.9
	gaussian			
	none	----	----	153.2
with covariate	exponential	30.96	2.89	171.3
	spherical	14.41	2.78	172.5
	gaussian			
	none	----	----	174.5

DESIGN D	Spatial Structure	Range	Sill	AICC
checks only	exponential	0	1.54	133.3
	spherical	0	1.54	133.3
	gaussian	0	1.54	133.3
	none	----	----	133.3
with treatments	exponential	0	1.54	133.3
	spherical	0	1.54	133.3
	gaussian	0	1.54	133.3
	none	----	----	133.3
with covariate	exponential	0	1.6	151.9
	spherical	0	1.6	151.9
	gaussian	0	1.6	151.9
	none	----	----	151.9

DESIGN E	Spatial Structure	Range	Sill	AICC
checks only	exponential	0	1.1	120.3
	spherical	0	1.1	120.3
	gaussian	0	1.1	120.3
	none	----	----	120.3
with treatments	exponential	0	1.1	120.3
	spherical	0	1.1	120.3
	gaussian	0	1.1	120.3
	none	----	----	120.3
with covariate	exponential	0	1.83	150.5
	spherical	0	1.83	150.5
	gaussian	0	1.83	150.5
	none	----	----	150.5