

USING REMOTE SENSING TO EVALUATE SPATIAL DEPENDENCE IN A WHEAT BREEDING TRIAL

Shane T. Ball*, Bruce E. Frazier, and Calvin F. Konzak
Washington State University
Pullman, WA 99164-6420, USA

ABSTRACT

Significant spatial dependence may mask the proper selection of crop genotypes in yield performance trials, often leading to the release of genotypes less adapted than expected. Our objectives were to estimate the effects of spatial dependence on remotely-sensed vegetation and grain yield data, and evaluate a combined remote sensing-spatial method to improve the analyses of breeder trials. Variograms of the classical residuals demonstrated little spatial structure. Conversely, variograms of deviations from the mean showed large spatial dependence, from 40 to 95% of the sample variance. A combined remote-spatial method was used to successfully remove spatial dependency effects in the breeder trial experiment. The mean square error of the classical ANOVA was reduced from 24 to 65% by applying the combined remote-spatial method, significantly increasing the precision with which differences could be distinguished.

Key Words: *Triticum aestivum*, reflectance, analysis of covariance, crop breeding, spatial heterogeneity, nearest-neighbor analysis

* Corresponding author. Project Nos. 0323, 2670, 5568, and 5570.

INTRODUCTION

Grain yield of spring wheat can be considered as the integrated response of genotype to a specific field environment. Yields are affected by large- and small-scale spatial dependencies, which inflate the error variance in the ANOVA of field-plot experiments (Ball et al., 1992a,b; Bhatti et al., 1991; Mulla et al., 1990). Since spatial dependence contributes to the error term of the ANOVA, this variation reduces the precision with which plant breeders can identify real genetic differences. Recent research indicates that from 10 to 54% of the sample variation in the grain yields of wheat field-plots may be spatially structured (Samra et al., 1990).

In order to identify and remove spatial dependency from breeder trials, researchers need methods to quantify this heterogeneity which do not require elaborate statistical designs. Complicated designs result in practical limitations for plant breeders and consequently they are rarely employed in the analyses of agricultural field experiments (Ripley, 1988). Although applications are yet few, it has been recommended that low-altitude aerial photography be used as a new complementary method in the analysis of field-plot breeder trials (Ball et al., 1992a; Clevers, 1988; Curran, 1988; Dancy et al., 1986).

Recent progress in remote sensing suggests that aerial photographs, combined with spatial statistics, may be modified to provide plant breeders with the tools needed to adjust for the effects of spatial dependence in field experiments (Ball et al., 1992a; Jupp et al., 1989; Rees, 1990). The objectives of this study were to estimate the effects of spatial dependence on remotely-sensed vegetation and measured grain yield data, and evaluate a combined remote sensing-spatial method to improve the analyses of breeder trials.

MATERIALS AND METHODS

The Washington State Uniform Spring Wheat Variety Trial was planted on 0.72 ha sites near Lind, Davenport, Mayview, Pullman, Royal Slope, and Walla Walla, WA during the 1990 growing season. These locations were chosen to represent differences in climatic zones and soil conditions. Cultural practices appropriate for the area were employed for all experimental materials. Plot size for the breeder trial was approximately 1.4 m wide by 4.9 m long. Each plot was eight rows, sown 15 cm apart, with a 0.3 m gap between plots. Grain was combine-harvested from each plot at a time slightly past physiological maturity of the latest maturing genotypes.

Infrared aerial photographs of the crop were taken on June 23 at Davenport and Walla Walla, on June 30 at Pullman and Royal Slope, and on July 7, 1990 at Lind and Mayview. Equipment included a Cessna 172 aircraft, a Canon F1 camera with a 50 mm lens and Wratten #12 gelatin filter, and Kodak 2236 infrared film. Flight altitude was about 300 m and ground speed was 100 km h⁻¹. Each flight was carried out

under sunny atmospheric conditions. Thus, the camera settings were not altered between locations and flights. All of the photographs were of the variety trial and surrounding area.

Near infrared and red wavelengths were selected from the images because the reflectance in each wavelength is related to crop cover. The green wavelength was not used because its values were highly correlated to the red values. Transmittance of the 35 mm transparencies were measured with a Howtek scanner at 1024 x 1024 resolution. Digital images which consisted of brightness values were related to the density of the film emulsion and these values converted to reflectance. The infrared and red bands from the images were selected for statistical analysis because the reflectance in each wavelength is related to crop cover.

The 33 genotypes were arranged in a RCB design with four replicates and the main effects were estimated using (Neter et al., 1990):

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \epsilon_{ij}, \quad i=1, \dots, I; j=1, \dots, J \quad (1)$$

Where Y_{ij} is the response for the j^{th} treatment in i^{th} block; $\mu_{..}$ is the overall mean; ρ_i are the block effects; τ_j are the treatment effects; and ϵ_{ij} are the random errors that are assumed to be NID (0, σ^2). The classical residuals obtained from the RCB model are:

$$Z_{ij} = Y_{ij} - \mu_{..} - \rho_i - \tau_j \quad (2)$$

The classical residuals can be considered as regionalized variables for plot errors and it is assumed that the main effects are correctly estimated.

Because the main effects may not be accurately known when significant spatial correlation between plots exists, we considered an additional regionalized variable (measure of error). These data are the deviations from the mean and they were computed using:

$$Z_{ij} = Y_{ij} - \tau_j \quad (3)$$

where τ_j is the genotype mean, equivalent to the treatment effects in equations (1) and (2). In contrast to the classical residuals, deviations from the mean do not remove the replication or overall mean effects.

Nearest-neighbor analysis (NNA) is an iterative method with which the mean effects are estimated from a comparison of yields in neighboring plots of different genotypes. The NNA model is (Mulla et al., 1990):

$$Y_{ij} = \mu_j + \delta_i + \epsilon_i \quad (4)$$

where Y_{ij} is the yield of the i^{th} plot having the j^{th} genotype; μ_j is the estimate of the treatment mean; δ_i is the estimate of the soil trend; and ϵ_i is the effect of local uncorrelated errors assumed to be NID $(0, \sigma^2)$. The NNA adjusted infrared and red values, as well as the yield data for each plot were applied to the RCB design and analysis (Eq. 1) to obtain the predicted genotype means. Detailed descriptions of detrending and nearest-neighbor analyses are given by Bhatti et al. (1991) and Mulla et al. (1990).

Variograms of classical and non-classical deviations for yield, infrared, and red values were used to evaluate the statistical assumption of independent errors. Slopes > 0 and intercepts < 1 indicate a spatial dependency. The semivariogram was estimated from the formula:

$$\gamma_k(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [Z_{ij}(X_k) - Z_{ij}(X_{k+h})]^2 \quad (5)$$

where $\gamma_k(h)$ is the semivariogram of random variable Z_{ij} ; $N(h)$ is the number of deviation pairs within a given range of distance h ; X_k is the data point location, and X_{k+h} is the data point separated by h from X_k . We used normalized semivariances for each variogram model.

The best fit, according to the least squares criteria, to the semivariance values was given by the linear model:

$$\gamma(h) = C_0 + C_1 h \quad (6)$$

where C_0 (nugget variance) is due to experimental and measurement error and genotypic variation that occurs within a distance shorter than h . The estimated slope function is C_1 . When $C_1 = 0$, the variogram indicates purely random variation which is the condition required by the classical RCB ANOVA.

In an attempt to compensate for the effect of spatial correlation on genotype estimates an analysis of covariance using near infrared and red deviations as covariates was conducted on the NNA-adjusted yield data. This model was:

$$Y_{ij} = \mu_j + \rho_i + \tau_j + \omega_1(Z_{ij1}) + \omega_2(Z_{ij2}) + \epsilon_{ij} \quad (7)$$

Table 1. Spatial structure demonstrated by the classical residuals and the deviations from the mean of grain yield and reflectance parameters at six diverse locations in eastern, Washington.

Parameter	Measure of error	Interpretation of spatial structure		
		<u>Davenport</u>	<u>Lind</u>	<u>Mayview</u>
Grain yield	Residual	Random	Random	Random
	Deviation	Random	Structured	Structured
Near infrared	Residual	Random	Random	Random
	Deviation	Structured	Structured	Structured
Red	Residual	Random	Random	Random
	Deviation	Structured	Structured	Structured
		<u>Pullman</u>	<u>Royal Slope</u>	<u>Walla Walla</u>
Grain yield	Residual	Random	Structured	Random
	Deviation	Structured	Structured	Structured
Near infrared	Residual	Random	Random	Random
	Deviation	Structured	Structured	Structured
Red	Residual	Structured	Random	Random
	Deviation	Structured	Structured	Structured

where μ_j , ρ_i , τ_j , and ϵ_{ij} were previously defined for the RCB model (Eq. 1) and ω_1 and ω_2 are regression coefficients for the relation between Y_{ij} and Z_{ij1} and Z_{ij2} . The covariates were expressed as a deviation from the overall mean. The ϵ_{ij} are random errors assumed to be NID $(0, \sigma^2)$. The SAS procedures, GLM, LSMEANS, MEANS, and UNIVARIATE, were used to perform the analyses of yield, as well as the near infrared and red values (SAS, 1988).

RESULTS AND DISCUSSION

Except for the grain yield data at Royal Slope, variograms of the classical residuals from the RCB model showed no spatial structure (Table 1). Thus, considering only the variograms of classical

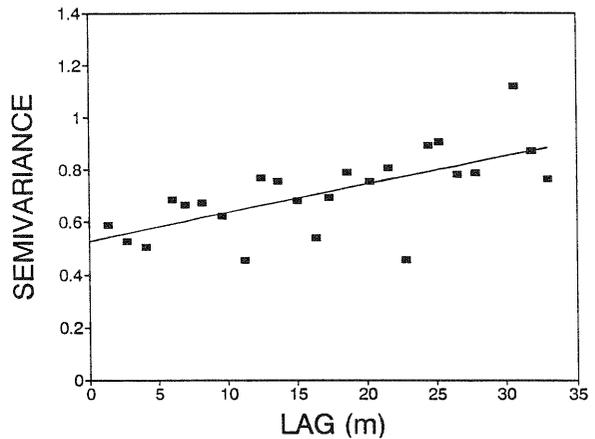


Fig. 1. Plot of the normalized linear variogram using deviations from the mean for grain yield data near Royal Slope, WA (see Table 1 for interpretation).

residuals, we might conclude that the errors are spatially independent. To further evaluate the assumption of independence and small-scale spatial dependence between plot errors, we conducted spatial analyses using non-classical deviations (deviations from the mean) for grain yield. These variograms generally showed that the errors between plots were not randomly distributed, because the slopes were non-zero and intercepts were less than unity (Figs. 1, 2, 3). An exception was grain yield at Davenport, possibly because of nonuniform cultural and management practices (Table 1). The variograms increased smoothly with increases in separation distance for grain yield, near infrared and red values (Figs. 1, 2, 3). Although the remotely-sensed data had larger slopes and smaller intercepts, the spatial patterns were consistent

Table 2. Classical and the remote-spatial ANOVA of grain yield from a breeder trial experiment conducted at six diverse locations in eastern, Washington.

Source of variation	df	Davenport	Lind	Mayview	Pullman	Royal Slope	Walla Walla
Mean square							
<u>Classical ANOVA</u>							
Replication	3	**	*	**	ns	**	**
Genotype	32	ns	ns	ns	ns	ns	*
Error	96						
R ² value		35	32	51	30	55	63
<u>Remote-spatial ANOVA</u>							
Replication	3	ns	ns	ns	ns	ns	ns
Genotype	32	**	**	**	**	**	**
NIR covariate	1	**	ns	ns	**	ns	**
Red covariate	1	**	ns	ns	**	*	**
Error	94						
R ² value		77	73	81	81	70	95
MSE reduction		24	46	46	60	42	65

ns,*,** Denote nonsignificance, significance at the .05, and .01 levels of probability. MSE reduction = % reduction in mean square error compared to the classical ANOVA.

with those for grain yield. These results demonstrated that the assumed independence of errors required for the breeder trial RCB ANOVA was not satisfied. This was because the local controls i.e., replication and randomization, did not successfully neutralize the small-scale spatial dependence effects between plots.

Genotypic effects at all locations, except Walla Walla, were not statistically significant when using the classical ANOVA (Table 2). These

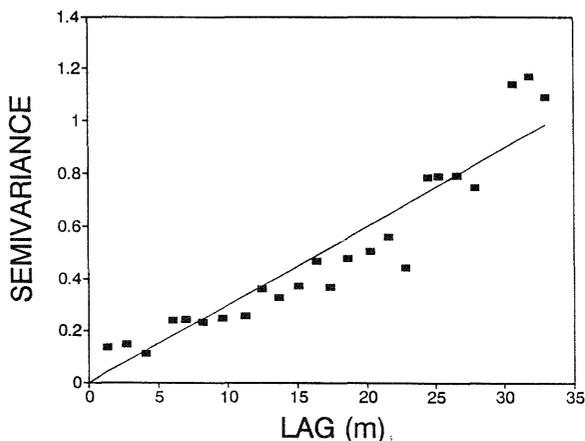


Fig. 2. Plot of the normalized linear variogram using deviations from the mean for near infrared data near Walla Walla, WA (see Table 1 interpretation).

results suggest that classical ANOVA failed to compensate adequately for local-scale spatial effects, showing that spatial heterogeneity decreases precision. Grain yield at Davenport and Pullman showed nonsignificant replication effects, probably due to large mean-square-error (MSE) values, implying that small-scale dependence exists at these field-sites.

The remote-spatial method was employed to remove the spatial trends and correlated errors between adjacent plots (Table 2). At four of six field-sites (Davenport, Pullman, Royal Slope, and Walla Walla), the F-values for covariate effects were

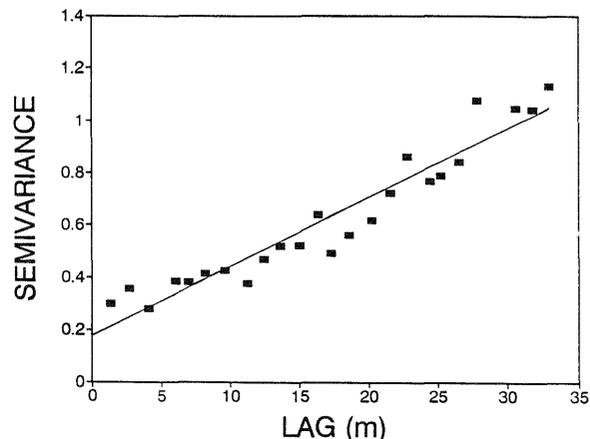


Fig. 3. Plot of the normalized linear variogram using deviations from the mean for red data near Pullman, WA (see Table 1 for interpretation).

significant ($P \leq 0.05$). In contrast to the classical ANOVA, genotype effects were highly significant ($P \leq 0.01$) for grain yield at each field-site. Replication effects were always nonsignificant with the remote-spatial method. The reason for the significant genotype effect was the large reduction in MSE values. Dramatically reduced MSE values (from 24 to 65%) increases the precision with which differences among the spring wheat genotypes can be identified. These results indicated that the remotely-sensed data explained a substantial proportion of the sample variance and illustrates why plant breeders should model spatial heterogeneity in yield trial experiments.

CONCLUSIONS

These results suggested that inexpensive, digitized aerial photographs can be combined with spatial statistical methods to improve the accuracy in estimating grain yields of spring wheat. This approach increased our understanding of the diverse problems posed by field spatial dependence. Complicated spatial patterns also may be elucidated

by this approach, resulting in a greater level of confidence with which lower performers are eliminated at different locations. A combined remote-spatial tool such as presented here may be an effective means to account for the spatial heterogeneity in grain yield patterns observed in breeder experiments and other field studies. The results suggest that other practical applications of this method could include: (i) estimating optimal plot and block configurations, and (ii) mapping of commercial fields to guide management practices. For plant breeders, the bottom line is that substantial spatial dependence may distort the true ranking of genotypes in breeder trial experiments.

ACKNOWLEDGEMENTS

The authors are grateful to Dan Thompson for taking the infrared photographs of each field-site. We acknowledge Mike Davis and Gary Shelton for their technical support in managing the breeder trials. Also, we thank Chana Dyer for her help in typing this manuscript.

REFERENCES

- Ball, S. T., B. E. Frazier, and C. F. Konzak. 1992a. The use of aerial photographs in a variety trial. *Cereal Res. Commun.* 00:000-000 (in review).
- Ball, S. T., D. J. Mulla, and C. F. Konzak. 1992b. Evaluating the independence of errors in wheat breeding experiments using variograms. *Crop Sci.* 00:000-000 (in review).
- Ball, S. T., D. J. Mulla, and C. F. Konzak. 1992c. Removing spatial heterogeneity in wheat breeding experiments using cross-validation kriging. *Crop Sci.* 00:000-000 (in review).
- Bhatti, A. U., D. J. Mulla, F. E. Koehler, and A. H. Gurmani. 1991. Identifying and removing spatial correlation from yield experiments. *Soil Sci. Soc. Am. J.* 55:1523-1528.
- Clevers, J. G. P. W. 1988. Multispectral aerial photography as a new method in agricultural field trial analysis. *Int. J. Remote Sens.* 9:319-322.
- Curran, P. J. 1988. The semivariogram in remote sensing: An introduction. *Remote Sens. of Environ.* 12:491-499.
- Dancy, K. J., R. Webster, and N. O. J. Abel. 1986. Estimating and mapping grass cover and biomass from low-level photographic sampling. *Int. J. Remote Sens.* 7:1679-1704.
- Jupp, D. L. B., A. H. Strahler, and C. E. Woodcock. 1989. Autocorrelation and regularization in digital images. *IEEE Trans. on Geosci. and Remote Sens.* 27:247-258.
- Mulla, D. J., A. U. Bhatti, and R. Kunkel. 1990. Methods for removing spatial variability from field research trials. *Adv. Soil Sci.* 13:201-213.
- Neter, J., W. Wasserman, and M. H. Kutner. 1990. *Applied Linear Statistical Models. Regression, Analysis of Variance and Experimental Designs.* 3rd ed., R. D. Irwin, Inc., IL.
- Rees, W. G. 1990. *Physical principles of remote sensing.* Cambridge University Press, New York, N.Y.
- Ripley, B. D. 1988. *Statistical inference for spatial processes.* Cambridge University Press, New York, NY.
- Samra, J. S., R. Anlauf, and W. E. Weber. 1990. Spatial dependence of growth attributes and local control in wheat and oat breeding experiments. *Crop Sci.* 30:1200-1205.
- SAS Institute. 1988. *SAS/STAT User's guide*, release 6.03 edition. SAS Institute, Cary, NC.