STATISTICAL ANALYSIS OF NON-REPLICATED EXPERIMENTS IN FARMERS' FIELDS. A case of balanced fertilization trials for bean in Burundi

Joaquin Sanabria[1] and John Wendt[2]

[1] International Fertilizer Development Center (IFDC) Head Quarters, Muscle Shoals, AL USA. [2] International Fertilizer Development Center (IFDC) East and Southern Africa Division, Nairobi, Kenya

Abbreviations:

AIC, Akaike Information Criterion
ANOVA, analysis of variance
CCC, Cubic Clustering Criterion
CEC, cation exchange capacity
Chi-Sq., value from the Chi Square distribution
DF, degrees of freedom
GLMM, Generalized Linear Mixed Model
NPK, fertilizer containing nitrogen, phosphorus, and potassium
NPKSZnB, fertilizer containing nitrogen, phosphorus, potassium, sulfur, zinc, and boron
OM, organic matter
RCBD, randomized complete block design

Abstract

Very often agricultural experiments in smallholder farms are conducted without actual replications due to land restrictions. Using fields as replications and performing the ANOVA with a randomized complete block design (RCBD) model can result in misleading conclusions about treatment's performance due to a biased error variance matrix estimated under the wrong assumption of independence between fields. A better alternative is using the spatial variability modeling features of the GLMM to generate an error term to perform unbiased hypothesis tests. Non-replicated trials to estimate bean response to three fertilization treatments were conducted across 175 farmers' fields in Burundi. Yields were used to compare the performance of an ANOVA using a GLMM where the exponential autocorrelation pattern of the residuals was used to model the error variance-covariance matrix against an ANOVA following a RCBD model where the fields are handled as blocks. The fields were grouped in three clusters prior to the ANOVA. AIC and the Generalized Chi-Sq. / DF ratio values for the model involving spatial modeling were 165.3 and 1, respectively. The same fit statistic values from the RCBD model were 330.1 and 0.1, respectively. The near-half magnitude of AIC in the spatial model relative to the RCBD model indicates higher model goodness of fit for the spatial model. The 0.1 value for the Generalized Chi-Sq. / DF ratio in the RCBD model suggests underdispersion and violation of the independence-between-fields assumption. Detection of a significant CLUSTER*TREATMENT by the spatial model corroborates the superiority of the ANOVA model involving spatial variability.

Introduction

The randomized complete block design (RCBD) is the experimental design of highest use in field agricultural research. The blocks play the role of repetitions; they allow to estimate the error variance from the within-treatment variability and to separate estimation of the random variability (mainly due to soil characteristics) from the variability associated with the treatment's effect. The block is called "complete" because each block in the experiment contains all the treatments, and the block is "randomized" because in each block a different randomization is performed for the assignment of treatment to plots. The randomization of treatments within blocks has the purpose of making the treatment responses independent of location inside the block (Steel et al., 1997).

Often, experiments conducted in farmers' fields in developing countries are unreplicated and consider a farm as a replication. Replicated trials cannot be conducted due to farmers' small land-holdings, particularly in Africa and Asia. For example, in Burundi where this research is conducted, 37% of farmers own a quarter of a hectare or less, and 57% have farms of half a hectare or less (World Bank, 2008). The burden for farmers can be reduced by conducting experiments in which the set of treatments is applied only once at each of numerous farmers' fields selected across entire countries or regions. While the lack of actual replications brings complications for the conventional use of the analysis of variance (ANOVA), the experiments conducted in a large number of farmers' fields across large geographical areas provide the opportunity to observe the treatments' performance and the adaptation of the bean variety in the diverse soil conditions and environments where they will be used by farmers.

Using the farmers' fields as replications to estimate the error variance needed to test hypotheses about treatment differences is problematic because treatment responses from the fields are not independent like they are in a RCBD experiment. Correlation between fields may result in biased estimation of treatment and error variances and misleading conclusions about treatment effects. A better alternative is to use the random spatial variability from field to field to model the error variance needed in the ANOVA to test hypotheses about treatment differences using the theory and applications of the Generalized Linear Mixed Models (GLMM) (Gbur et al., 2012). The error variance-covariance matrix estimated this way is used to test hypotheses about differences between treatment responses without bias. The off-diagonal terms in the matrix will be non-zero values while the same matrix from a RCBD where yield determinations are independent has zero in all off-diagonal positions.

Hu and Spilke (2009) and Gbur et al. (2012) show a diversity of methods to take advantage of spatial variability to model the experimental error variances. Some model the error variance-covariance matrix (R-side matrix) of the ANOVA model, fitting specific models following the spatial variation pattern, and some of the model choices are autoregressive, exponential, and power models. When the spatial variability does not follow a defined pattern, unstructured variance-covariance error matrices can also be estimated through the use of the GLMM features. Modeling the random variability at the left side of the GLMM model (G-side matrix), the random spatial-correlated variability can be extracted from the error variance matrix to leave it free of autocorrelation to perform unbiased hypothesis tests about treatment differences (Gbur et al., 2012). Another technique that works by extracting autocorrelated random variability from the error variance matrix is the use of the geographical coordinates or soil properties associated with the spatial structure pattern as covariables in the GLMM model to adjust treatment responses (Gbur et al., 2012). Nearest neighbor adjustment to extract autocorrelated spatial variability from the error variance can also be used in experimental settings with non-independent experimental units (Stroup, 2002; Kravchenko et al., 2006).

There are also methodologies for the setting and analysis of agricultural experiments that are completely outside the conventional experimental design and the ANOVA contexts and that rely on the theory and methodologies of geostatistics (Isaaks and Srivastava, 1989; Goovaerts, 1997; Nielsen and Wendroth, 2003; Schabenberger and Gotway, 2005). Stroup (2002) warns against misleading inferences that potentially occur when research is conducted ignoring the principles of experimental design and when spatial variability is not controlled effectively.

Estimation of the error variance-covariance matrix can be improved if the farms are grouped based on soil characteristics and geographical coordinates. This grouping of farms produces error variance-covariance matrices with lower variances than the same kind of matrix obtained from the full set of farms. Reduction of the error variances increases the power of the ANOVA to detect treatment differences. Grouping of farms also allows for more specific treatment recommendations to be made based on the type of farms that form the different farm clusters.

The objective of this paper is to demonstrate the use of the field to field spatial variability to model the error variance in a GLMM model to perform unbiased hypotheses about treatment differences in experiments with no replications.

Methodology

Experimental Setting

Three fertilization treatments, Control (no fertilizer), NPK 18-46-30, and NPKSZnB 19-35-24-7-2-1, were tested across 175 fields in Burundi in 2014 (Season A [September-January]). Bean (variety G13607 of climbing type) was used as the test crop. Treatments were not replicated within the fields. Macro, secondary, and micronutrient soil contents were determined together with soil pH, cation exchange capacity (CEC), electric conductivity, organic matter (OM), calcium/magnesium (Ca/Mg) ratio, soil aluminum (Al), soil sodium (Na), and acid saturation. Latitude, longitude, and altitude for the 175 fields were recorded.

In general terms, the 175 farms were classified in three clusters based on soil and geographical characteristics. Then, the analysis of variance was performed in two ways. First, the ANOVA error variance needed for testing hypotheses about treatments, clusters, and the interaction treatment*cluster was estimated using the farm to farm exponential pattern suggested by the residuals semivariogram. Second, an ANOVA analysis was performed assuming independence between farms and using them as blocks using a RCBD ANOVA model. The performance of the two models was compared based on two fit statistics – Akaike Information Criterion (AIC) and the Chi-Sq. / DF ratio (also named Normed Chi-Sq.) – and the significance of the interaction. A detailed description of the five-step methodology follows:

1. **Identification of the most important soil characteristics and geographical coordinates** with multivariate factor analysis (Srivastava, 2002) to be used in the next step to group the farm fields. The full set of soil characteristics and geographical coordinates from the 175 farmers' fields was reduced to a few uncorrelated factors that explain most of the variability

contained in the soil characteristics data measured at each of the farms involved in the project; each factor is made up of linear combinations of the original soil characteristics. The larger linear combination coefficients are associated with the most useful soil/coordinate characteristics from each factor to be used in grouping the 175 farm fields that participated in the experiment. Interpretation and selection of the most important soil characteristics/coordinates were facilitated by a rotation of the factors. The varimax rotation method, which is the most commonly used rotation method, was employed (Srivastava, 2002; Costello and Osborne, 2005). After the rotation, coefficients associated with the most important components of each factor became substantially larger than secondary components, as shown in the highlighted characteristics in Table 1.

2. **Grouping of farm fields**. A cluster analysis (Everitt et al., 2011) was performed using the set of important soil characteristics and coordinates identified with the factor analysis. The Ward's agglomerative process to produce a finite number of disjoint clusters of farmers' fields was used (Murtagh and Legendre, 2014). Field members of a cluster are expected to be more similar in soil characteristics than fields that are members of different clusters. The cubic clustering criterion (CCC) (Figure 1) was used to determine the best number of farm field clusters (SAS Institute Inc., 1983).
3. **Identification of spatial pattern**. The empirical semivariogram of the residuals along altitude and latitude was estimated (Figure 2). Based on the pattern suggested by the semivariogram, an exponential model was used to estimate the error variance-covariance matrix of the ANOVA model.

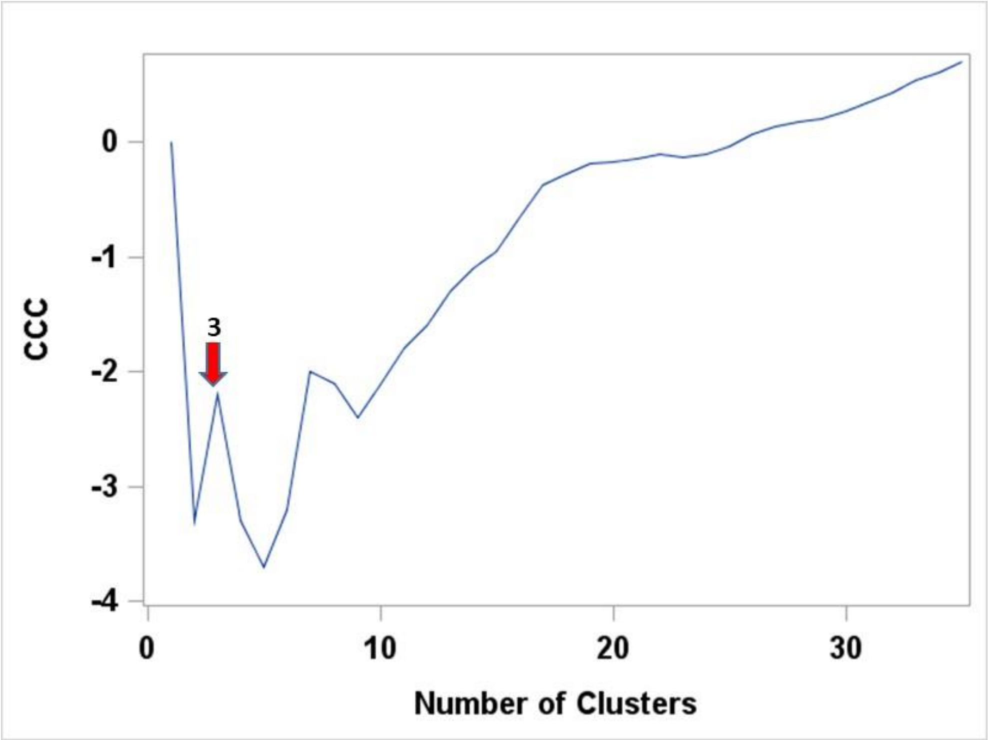

Figure 1. Cubic clustering criterion for deciding the number of farmers' field clusters.

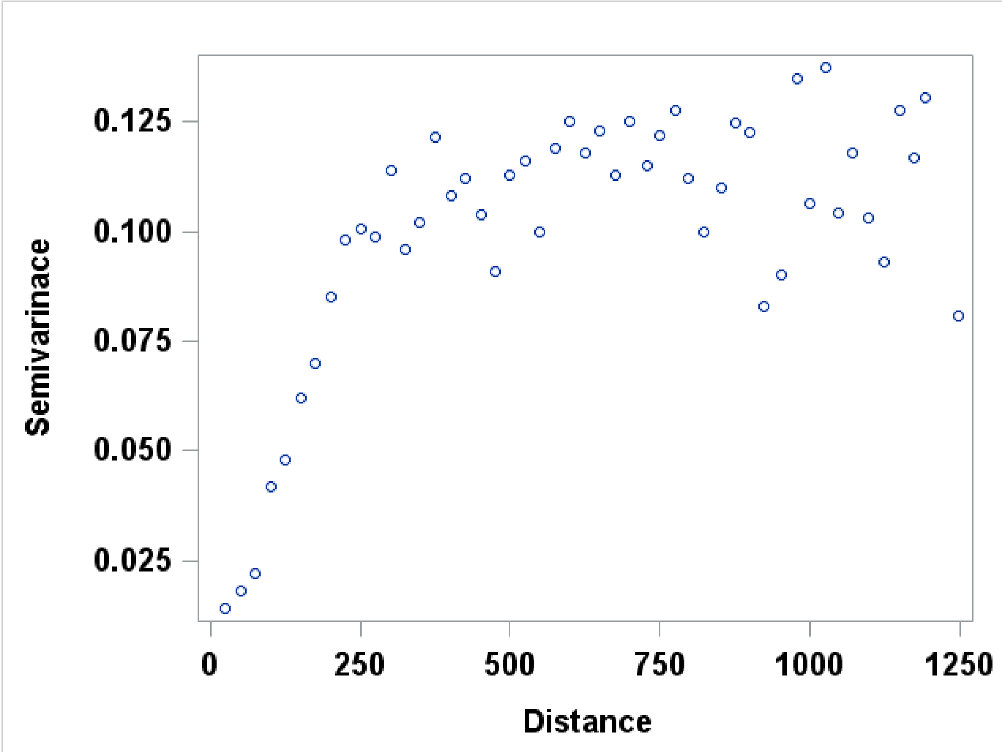

Figure 2. Semivariogram showing the spatial variability pattern of bean yield residuals along the altitude and latitude coordinates.

4. **Analysis of Variance with a Generalized Linear Mixed Model (GLMM)** (Schabenberger, 2005; Gbur et al., 2012).
   A GLMM can be expressed as: $\mathbf{Y = X\beta + Z\gamma + e}$
   Where X and Z are known matrices, β is an unobservable vector of fixed effects, and γ is an unobservable vector of random effects. The random effects γ have a normal distribution with mean 0 and variance matrix G. The distribution of the errors e is normal with mean 0 and variance R. Modeling the G matrix , the columns of the Z matrix and the structure of G are specified. Modeling the R matrix, the covariance structure of the R matrix is directly specified. The SAS code used for the ANOVA involving modeling of the error variance (R matrix) is shown in Figure 3.

```
PROC GLIMMIX DATA=BEANS_MODEL;
CLASS TREATMENT CLUSTER;
MODEL YIELD=TREATMENT|CLUSTER /DDFM=KR;
RANDOM _RESIDUAL_/TYPE(EXP)(ALTITUDE LATITUDE) SUBJECT=FARM GROUP=TREATMENT;
COVTEST HOMOGENEITY;
TITLE "Analysis of Variance Modeling the R-side Matrix";
RUN;
```

Figure 3. SAS code for analysis of variance modeling the error variance-covariance matrix

The SAS program in Figure 3 has several important features; the most important is associated with the RANDOM statement, which fits an exponential model to the farm fields to estimate the error variance-covariance matrix as a function of altitude and latitude, across all farms as indicated by the SUBJECT=FARM statement. The option GROUP=TREATMENT results in the variance-covariance matrix being grouped by treatments. The COVTEST statement together with the HOMOGENEITY option develops a variance homogeneity test for the TREATMENT effect. If the error TREATMENT variances are not homogeneous, the variance-covariance matrix is grouped by treatment. Establishing whether the error variances are homogeneous or not and grouping the error variance-covariance matrix by the effects that do not have homogeneous variances have great influence on the capacity for detection of differences between TREATMENTS, CLUSTERS, and the interaction of TREATMENT*CLUSTER.

The error variance in the RCBD ANOVA was also grouped by treatment to set the two types of ANOVA analysis in the same conditions relative to the non-homogeneity of error variance.

5. **Comparison of the ANOVA performed through modeling of spatial variability against the ANOVA with the fields playing the role of replications** and assuming independence between fields. The two types of ANOVA models are compared using two fit statistics: AIC and the Generalized Chi-Sq. / DF ratio. AIC is used for the selection of the best model in situations when more than one model can be fit to the data. The AIC is calculated as AIC = - 2(Residual Log Likelihood) + 2p. Where p is the model parameters plus any estimated parameters (Akaike, 1974). From a set of models considered, the model with the lowest AIC is said to be the one with the lowest information loss from the information provided by the finite sample available to fit the model. For a low AIC value, the model’s residual goodness of fit estimated by the likelihood function must be low and the number of model parameters must be also low. AIC does not tell about the absolute quality of the model of interest, and it should be complemented by another fitness statistic, such as the Generalized Chi-Sq. / DF ratio used to assess the error variability or dispersion of the ANOVA model (Hooper et al., 2008). When the Generalized Chi-sq. / DF ratio is equal to 1, it is said that the random variability is equidispersed, meaning that there is not discrepancy between observed and expected variability; this situation is an indicator of the appropriate model’s goodness of fit. When the ratio is larger than one, the random variability is overdispersed because the observed variability exceeds the expected variability, and when the ratio is lower than one, the random variability is underdispersed because the observed variability is lower than the expected variability (Kokonendji, 2014). Another comparison criterion between the two models is the capability of the model to reject the null hypothesis associated with the TREATMENT*CLUSTER interaction given the radical importance of the interaction to interpret yield responses to treatments under the particular geographical and soil conditions of the clusters.

Software

The multivariate factor, cluster, spatial, and variance analyses were done with SAS 9.4 using the PROC FACTOR, PROC CLUSTER, PROC VARIOGRAM, and PROC GLIMMIX procedures, respectively (SAS Institute Inc., 2008).

Results

The most important characteristics from soil and geographical coordinates identified by the multivariate factor analysis to be used for the grouping of farmers' fields are contained in three factors (Table 1) that explain 95.5% of the total variability in the data matrix of the soil and coordinate characteristics from the 175 farm fields. The three factors explain 77%, 13.9%, and 4.1% of the total variability, respectively. The most useful variables for the grouping of the bean fields are variables of high variability within the original data matrix, which is reflected in the high loads of the highlighted variables in Table 1: soil pH, Ca soil content, boron (B) soil content, CEC, and Mg soil content in Factor 1; altitude and soil Al in Factor 2; and OM soil content, ratio of Ca/Mg in the soil, and latitude in Factor 3. Altitude and latitude are used to identify the spatial structure variability in the data analysis. It is not surprising that among the most important variables for the classification of the 175 fields are soil characteristics that are determinants of soil fertility, such as soil pH, CEC, soil content of the bases Ca and Mg grouped in Factor 1, and organic matter and Ca/Mg in Factor 3. Other characteristics identified as important by the factor analysis are soil B, altitude, soil Al, and latitude, which are important factors in bean production.

Table 1. Rotated factor pattern from multivariate factor analysis applied to soil characteristics and geographical coordinates of 175 farm fields.

Rotated Factor Pattern			
Characteristic	**Factor 1**	**Factor 2**	**Factor 3**
pH	0.74856	-0.2557	0.02948
Ca	0.73438	-0.2921	0.60281
B	0.70632	-0.1101	0.31253
CEC	0.69673	-0.2599	0.57582
Mg	0.6612	-0.3586	0.3498
EC	0.45514	0.02566	0.11544
Zn	0.394	-0.1809	-0.1433
P	0.30059	-0.0875	0.04406
S	-0.4568	0.39276	0.04976
Al + H	-0.4798	0.1707	-0.0366
Acid Saturation	-0.5705	0.18509	-0.0388
H	-0.7554	0.25484	-0.0596
Altitude	0.07642	0.75815	-0.0061
Al	-0.5759	0.74197	0.32541
Fe	0.02702	-0.1809	0.03001
Cu	0.24746	-0.4207	0.04658
Mn	0.26469	-0.5616	-0.0362
OM	0.02383	0.25487	0.54699
Ca/Mg	0.13767	0.0035	0.5182
Latitude	0.1544	-0.1027	0.3202
Farm Size	0.02915	-0.0252	0.10811
Si	0.06341	-0.224	-0.3395
K	0.58193	0.32814	-0.3814
Longitude	0.00606	0.06284	0.04079
Na	0.02843	-0.0487	-0.0186

Guided by the first peak of the CCC versus number of clusters in Figure 1, it was decided to classify the 175 bean fields in three disjoint clusters with size and classification variable means shown in Table 2. Cluster 3 contains 17 bean fields with the largest means for yield, CEC, OM, Ca, Mg, and a soil pH of 6.2, suggesting that the soils in this cluster have, on average, higher fertility than the other two clusters. Cluster 3 also has the lowest altitude and latitude, which together with the highest bean yield indicates a good adaptation of the G13607 of climbing type variety to the low altitude and latitude regions of Burundi.

Table 2. Cluster size and means for geographical and soil characteristics used for the cluster classification.

Cluster	n	Yield (Mg Ha^{-1})	Latitude	Altitude	pH	Ca (ppm)	B (ppm)	CEC	Mg (ppm)	Al (ppm)	OM (%)	Ca/Mg (%)
1	50	2.44	-3.07	1544.3	5.9	1456.4	0.29	13.6	285.0	1200.1	4.4	3.3
2	108	1.41	-3.32	1626.9	5.4	571.2	0.13	7.4	139.5	1476.8	4.3	2.6
3	17	2.79	-2.86	1462.2	6.2	2766.5	0.44	22.9	526.8	1126.3	5.2	3.5

The ANOVA where the random variability mainly associated with altitude and latitude is used to model the error variance-covariance matrix fit the data far better than the RCBD model as indicated by the AIC and the Generalized Chi-Sq. / DF ratio statistics in the top two rows of Table 3. The AIC value equal to 165.3 in the spatial model relative to 330.1 in the RCBD model shows a random variability error (estimated by the residual Log Likelihood) of magnitude near half of the same variability in the RCBD model. The Generalized Chi-Sq. / DF ratio value equal to 1.0 in the spatial model is associated with equidispersion, or no major discrepancy between observed and expected error variability, compared to the value of 0.1 from the RCBD model that can be interpreted as underdispersion, or lower than expected differences between observed and expected error variability. Underdispersion occurs when model assumptions like independence between subjects are violated (Kokonendji, 2014). Underdispersion in the RCBD model is evidence of the autocorrelation between fields. The higher goodness of fit of the spatial model resulted in enough power to detect the significant interaction of TREATMENT*CLUSTER (left side of Table 3).

Table 3. Comparison of analysis of variance performed with error variance modeling against assuming independence of error variances like in a Randomized Complete Block Design.

Modeling Error Variances					**Assuming Error Independence (RCBD)**				
AIC†	165.3				**AIC**†	330.1			
General. Chi-Sq. / DF‡	1.0				**General. Chi-Sq. / DF**‡	0.10			
Effect	**Num DF**	**Den DF**	**F**	**Pr>F**	**Effect**	**Num DF**	**Den DF**	**F**	**Pr>F**
TREATMENT	2	277.7	2208.9	<.0001	TREATMENT	2	9.454	290.19	<.0001
CLUSTER	2	386.4	9.79	<.0001	CLUSTER	2	515.3	3.78	0.0233
TREATMENT*CLUSTER	4	311.7	4.70	0.0011	TREATMENT*CLUSTER	4	138.5	1.20	0.3105

†Akaike Information Criterion. ANOVA Model with smaller value fits the data better.

‡ Indicator of ANOVA model fitness. The closer to 1.0 the better.

The comparison of means from the interaction of TREATMENT*CLUSTER shows significantly higher yield responses to the NPK fertilization and to the balanced nutrition fertilization (NPKSZnB) that includes the secondary nutrient sulfur (S) and the micronutrients zinc (Zn) and B relative to the no application of fertilizer in the three clusters (Table 4). Bean yield from cluster 3 is 345.5 kg ha^{-1}, significantly higher than in cluster 1, and 317.5 kg ha^{-1}, significantly higher than in cluster 2, when the basic NPK fertilization is complemented with S, Zn, and B (Table 5). The significantly higher yield responses to the application of S, Zn, and B of the bean crops in cluster 3 are explained mainly by the 6.2 pH average of the soils in this cluster, which is within the optimal pH range for bean growth and for the availability of S, Zn, and B to be taken up by the bean crop. Higher OM, CEC, Ca, and Mg and lower Al in cluster 3 soils are also contributing factors to the higher yield responses in this cluster. Identification of the differential performance of the balanced nutrition across geographical and soil conditions of Burundi would not have been possible without an ANOVA model able to detect the significant TREATMENT*CLUSTER interaction.

Table 4. Comparison of treatment means for bean yield within clusters using the Holm-Tukey test

CLUSTER	TREATMENT		Difference	SE	t Value	Adj. P
	i	j	TRT(i) – TRT(j)			
1	CONTROL	NPK18:46:30	-1.669	0.040	-41.65	<.0001
1	CONTROL	NPKSZnB19:35:24:7:2:1	-1.969	0.038	-51.48	<.0001
1	NPK18:46:30	NPKSZnB19:35:24:7:2:1	-0.300	0.052	-5.76	<.0001
2	CONTROL	NPK18:46:30	-1.686	0.058	-28.76	<.0001
2	CONTROL	NPKSZnB19:35:24:7:2:1	-2.002	0.055	-35.79	<.0001
2	NPK18:46:30	NPKSZnB19:35:24:7:2:1	-0.316	0.076	-4.15	<.0001
3	CONTROL	NPK18:46:30	-1.970	0.097	-20.17	<.0001
3	CONTROL	NPKSZnB19:35:24:7:2:1	-2.327	0.092	-24.96	<.0001
3	NPK18:46:30	NPKSZnB19:35:24:7:2:1	-0.357	0.126	-2.81	0.0049

Table 5. Comparison of cluster means for bean yield within treatments using the Holm-Tukey test

TREATMENT	CLUSTER		Difference	SE	t Value	Adj. P
	i	j	CLS(i) – CLS(j)			
Control	1	2	0.00569	0.02367	0.24	0.9687
Control	1	3	0.01287	0.03521	0.37	0.929
Control	2	3	0.00718	0.03799	0.19	0.9805
NPK 18:46:30	1	2	-0.0113	0.06695	-0.17	0.9844
NPK 18:46:30	1	3	-0.2885	0.09956	-2.9	0.0112
NPK 18:46:30	2	3	-0.2772	0.1074	-2.58	0.0278
NPKSZnB 19:35:24:7:2:1	1	2	-0.0279	0.06349	-0.44	0.899
NPKSZnB 19:35:24:7:2:1	1	3	-0.3455	0.09441	-3.66	0.0009
NPKSZnB 19:35:24:7:2:1	2	3	-0.3175	0.1019	-3.12	0.0056

Conclusions

The spatial variability modeling of the error variance-covariance matrix resulted in an ANOVA model of high fitness and power to detect the significant TREATMENT*CLUSTER interaction.

Estimating the error variances with the farms playing the role of blocks in a RCBD resulted in an ANOVA model of lower fitness, underdispersion of random variability due to lack of independence between experimental fields, and insufficient power to detect a significant TREATMENT*CLUSTER interaction.

Detecting the significant TREATMENT*CLUSTER interaction was critical to identify the effects of soil fertility on the responses to basic NPK fertilization and especially to balanced bean nutrition, including sulfur, zinc, and boron. The farms that make up part of cluster 3 have soil pH near neutrality and take advantage of the balanced nutrition, including S, Zn, and B, better than farms in clusters 1 and 2. Without the use of farm clusters in the experimental design or analyzing the data using farms as replications would not have allowed the identification of the differential yield responses to balanced fertilizations across the soil and geographical conditions of Burundi.

Modeling the within-fields spatial variability to produce an error variance matrix would also be very useful in demonstration plots, where the treatments are not replicated. Modeling the error variance applying the concepts of the GLMMs would add research value to the demonstration plots usually deployed in farmers' land.

Acknowledgements

Thanks to the United States Agency for International Development, Bureau of Food Security for providing the funds to conduct the research associated with this paper.

References

Akaike, H. 1974. A new look at the statistical model identification. Trans. Autom. Control 19(6):716-723.

Costello, A.B., and J.W. Osborne. 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Pract. Assess., Res. Eval. 10(7).

Everitt, B.S., S. Landau, M. Leese, and D. Stahl. 2011. Cluster analysis. Wiley Series in Probability and Statistics. John Wiley and Sons, New York.

Gbur, E.E., W.W. Stroup, K.S. McCarter, S. Durham, L.J. Young, M. Christman, M. West, and M. Kramer (Eds.). 2012. Analysis of generalized linear mixed models in the agricultural and natural resources sciences. ASA, CSSA, and SSSA, Madison, WI.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Oxford University Press, New York.

Hooper, D., J. Coughlan, and M. Mullen. 2008. Structural equation modeling: Guidelines for determining model fit. J. Bus. Research Methods 6(1):53-60.

Hu, X., and J. Spilke. 2009. Comparison of various spatial models for the analysis of cultivar trials. New Zealand J. Agric. Res. 52:277-287.

Isaaks, E.H., and R.M. Srivastava. 1989. Applied geostatistics. Oxford University Press, New York.

Kravchenko, A.N., G.P. Robertson, S.S. Snap and A.J.M. Smucker. 2006. Using information about spatial variability to improve estimates of total soil carbon. Agronomy J. 98:823-829.

Kokonendji, C.C. 2014. Over- and under-dispersion models. Methods and applications of statistics in clinical trials: Planning, analysis, and inferential methods. Edited by N. Balakrishnan. John Wiley & Sons, Inc.

Murtagh, F., and P. Legendre. 2014. Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? J. Classif. 31:274-295.

Nielsen, D.R., and O. Wendroth. 2003. Spatial and temporal statistics. Sampling field soils and their vegetation. Genecology Textbook. Reiskirchen, Germany.

Schabenberger, O. 2005. Introducing the GLIMMIX procedure for generalized linear mixed models. SUGI 30. Paper 196-30.

Schabenberger, O., and C.A. Gotway. 2005. Statistical methods for spatial data analysis. Chapman and Hall, Boca Raton, FL.

SAS Institute Inc. 2008. SAS/STAT® 9.2 user's guide. The GLIMMIX procedure (Book Excerpt). Cary, NC.

SAS Institute Inc. 1983. Cubic clustering criterion. Technical report A-108. Cary, NC.

Srivastava, M.S. 2002. Methods of multivariate statistics. Wiley Series in Probability and Statistics. John Wiley and Sons, New York.

Steel, R.G.D., J.H. Torrie, and D.A. Dickey. 1997. Principles and procedures of statistics. A biometrical approach. Third Edition. McGraw-Hill, Boston.

Stroup, W.W. 2002. Power analysis based on spatial effects mixed models: A tool for comparing design and analysis strategies in the presence of spatial variability. J. Agric., Biol., and Environ. Statistics 7(4):491-511.

World Bank. 2008. Country assistance strategy for Burundi. Report 44193-BI. World Bank, Washington, D.C.