Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2000 - 12th Annual Conference Proceedings

YIELD PREDICTION IN 60ft^2 GRIDS

S. Aref

D. G. Bullock

Yield prediction in 60ft^2 grids

By S. Aref and D.G. Bullock
Dept. of Statistics and Dept. of Crop Sciences

Large detailed yield databases incorporating GPS makes it possible to predict yield on a small scale. The objective of this study was to determine how closely yield could be predicted in grids of 60-ft^2 units. Corn and soybean yields were averaged to the 60-ft^2 grid. The yields were modeled on previous yields, soil fertility, soil type, and terrain variables. Soil fertility variables were kriged from a 1-acre grid to the 60-ft^2 grid. Terrain data and soil type data were available at the same scale. Multiple regression models and models with spatial correlation determined from yield semivariograms differed some. Previous yields and wetness were the most significant variables. Soil variables alone were not good predictors.

INTRODUCTION

With the amount of data available from yield monitors it is of interest to try to predict yield at small scales such as a 60-ft^2 grid. Practically the only variables directly available on such fine grids are previous years' yields, since variables such as soil fertility measures are usually only determined on a 1 or 2.5 acre grid at best. To extend soil fertility variables to a finer grid the soil fertility variables can be kriged, but the kriged soil fertility variables are simply interpolations according to certain correlation models and will not be of the same precision as the means of the yield in the 60-ft squares. This is assuming that yield monitors are fairly precise. Prediction equations using the available variables, which are soil fertility measures and/or previous years' yields, produce estimates that are only relative levels of yield, rather than actual levels of yield. Supplementary prediction equations involving weather over a number of years – in particular precipitation – are needed to produce the actual yield level.

The present study grew out of a project where the goal was to determine the feasibility of yield prediction in a 60-ft square grid. The 60-ft length was chosen since the width of fertilizer applicators often is 60 ft. The original study involved eight fields. The yield was to be predicted on a 60-ft^2 grid using the soil fertility data and the soil types. It was also of interest to determine what the previous years' yield data added to the precision of the predictions. If previous yield is included in prediction equations, it is necessary to have a consistent rotation scheme for the previous years. Three of the eight fields had a perfect corn/soybean rotation record. One of these three fields was selected for this study, since this field also had a very detailed elevation map. A thorough discussion of the impact of terrain can be found in Kravchenko and Bullock, 2000. Fig. 1 is the layout of the points of interest in the field: the 1-acre grid, the 60-ft^2 grid, and the points where elevation was measured.

In 1996 and 1997 fertilizer at variable rates (SSM: Site-Specific Management treatment) were applied to some parts of the field, while other parts received whole field management rates (WFM: Whole Field Management treatment). In the case of the SSM treatment nitrogen (N), potassium (P), and phosphate (K) was applied at different rates to small subplots within the field. The applications were based on a 1995 soil test (see fig. 2 for the application map for P in the north part of the field) adjusted by estimated removals by the previous year's yield for P and K, and the estimated removals by the previous year's yield for N. The solid dark larger rectangular areas received a uniform broadcast rate according to the WFM treatment. The other areas received rates based upon interpolated and adjusted soil values and at the rate as identified by the differentiated code according to the SSM treatment in fig. 2. In the SSM treatment where the soil fertility level of P was very high the areas are almost white since these received almost no P and were therefore very low in P application (but not in overall P measure). A solid dark larger rectangular area has received a value based on the field median as part of the WFM treatment. The other areas with very varying levels are in the SSM treatment where the P measure was less than the appropriate level of P and therefore adjusted with respect to the level and the estimated removal by the previous year's yield. The soil tests from 1995 and 1997 were performed on the same 1-acre grid. In 1998 the north half of the field was manured and WFM rates were applied uniformly to the north half of the field.

MATERIALS AND METHODS

Data preparation

The data to be considered here were the yield records constructed from combine data from 1995 through 1998, the 1997 soil test measures on a 1-acre grid, soil types, and the elevation data. The chosen field was a double field where there was a corn/soybean rotation in the northern half from 1995 through 1998 ending in corn. The southern half had the opposite rotation from 1996 through 1998, though crops in 1995 were the same as the northern field (see table 1) and thus did not have a perfect rotation scheme. The two halves of the field had the same four soil types. The southern field furthermore had two 60-ft squares with a fifth soil type.

The data set was in Arcview and was projected using Lambert's Conformal Conic with spheroid GRS80. To predict the yield in the 60-ft^2 grid the yield data were averaged in squares that had midpoints 60 ft apart, using a nearest polygon method. Semivariogram models were determined for yield for each crop and year using SAS proc variogram and proc nlin. Parameter estimates from the semivariogram models were used in the prediction models in proc mixed with spatial correlation.

The predictors were the soil fertility variables and soil types, and possibly the previous yield and elevation data. The soil fertility data were first transformed so that the variables looked fairly normally distributed in order to krig the data from the 1-acre grid to the 60-ft grid. The transformations used were natural log for cation exchange capacity (CEC), organic matter (OM), P, K, sulfur (S), boron, (B), and zinc (Zn) and $\sqrt{(\text{max obs.})} - \sqrt{(\text{max obs.} - \text{obs.})}$ for pH and manganese (Mn). The remaining variables, calcium (Ca), iron (Fe), and magnesium (Mg), were not transformed. The 1997 soil data had these 12 variables while the 1995 soil data only had seven of the variables (pH, CEC, OM, P, K, Ca, and Mg). The 1995 variables needed the same transformations as the corresponding 1997 variables. The transformed variables were then kriged to the midpoints of the 60-ft squares using SAS proc krige. The semivariogram models for the different soil variables were the spherical model for Ca, Ce, Mg, Mn, Om, P, and S, the exponential model for B, the gaussian model for pH, the power model for Fe and Zn, and the two-directional power model for K. The semivariogram models were selected based on the best fitting models using proc nlin, and using the smallest error as the decision criteria. Note that proc variogram and proc krige2d are experimental procedures in SAS version 6.12. While proc variogram worked well, proc krige2d produced reasonable results everywhere on the kriging grid except for the midpoint of the grid where values were completely off for some variables and less so for other variables. Only one soil variable was not affected. In fig. 3 are the kriged values of K plotted against x at y=0 and y at x=0. In fig. 4 is a graph of the kriged K with a more reasonable midpoint value. The soil types were determined for a square according to the soil map by using the soil type with the largest area present in the square.

The elevation data was manipulated in Arcview to produce the terrain map from which the following terrain variables were derived: up-slope, wetness, aspect, slope, profile, tangent, and mean curvature. Pertinent to the 60-ft^2 grid these variables were condensed to values for the midpoints of the polygon. Five terrain variables (wetness, slope, profile, tangent, and mean curvature) were just re-scaled, while the aspect variable was rotated 110° to align to the main direction and the up-slope variable was doubly log transformed due to the extreme skewness of this variable.

Elevation was transformed using elevation$^{1/4}$. The semivariogram based on the transformed variable had drift, see fig. 5. The transformed variable was de-trended using powers and interactions (up to the 4'th degree) of the x and y coordinates. The reduced model was

$$\text{elevation}^{1/4} = \mu + \beta_1 x + \beta_2 y + \beta_3 xy + \beta_4 x^3 + \beta_5 y^3 + \beta_6 x^2 y + \beta_7 x^3 y + \beta_8 x^2 y^2 + \varepsilon,$$

with an R^2 of 91.6% and variance inflation factors that were less than 7.5. A model allowing 5'th degree terms increased the R^2 by 2%, while 4 variance inflation factors were between 12 and 13.5.

The power transformation appeared to be making the elevation variable too skew to the right but the residuals from the de-trended transformed variable looked more normally distributed than if a squareroot transformation was substituted for the power transformation. The experimental semivariogram for the residuals (fig. 6) exhibits a maximum at the lag distance of 0.14 and then levels off 20% lower for lag distances ≤ 0.4. Between the lag distances 0.4 and 0.7 the semivariogram dips down and may indicate there were trends not accounted for. Using residuals from the model with up to fifth degree terms made the semivariogram look worse. A decision was made not to include powers higher than four and only fit the semivariogram at the lag distance ≤ 0.4. Three types of semivariogram models were fitted: the exponential (dotted line), the spherical (dashed line) and the gaussian (solid line). The gaussian model fitted the best overall if the 'bump' at the lag distance 0.14 was ignored. The residuals were kriged to the 60-ft^2 grid and the trend added back in.

Models

Two kinds of prediction equations were used. One kind was based on 1996 yield to predict 1998 yield within the same field, while the other kind was based on 1997 yield to predict 1998 yield in the other half of the field. For each of these prediction equations there were models with and without spatial correlation. For comparison, yield from 1998 was also featured as response in prediction equations.

There were several kinds of full models as well. In the original project a primary focus point was how well the soil fertility data and the soil types predict yield. As mentioned earlier there was not enough resolution in the kriged soil fertility measures to adequately address the yield on the 60-ft grid. The soil measures were more or less just interpolations from a coarser 1-acre grid of 106 points to a twelve times finer 60-ft^2 grid of 1293 points. Thus predicted values cannot reflect the detail in the finer grid. The kriged soil measures have added error on top of what the 1-acre grid soil measures already have, but this error was not taken into account in the models. The error in the regressor variables should produce generally smaller coefficients (in absolute value).

Including previous years' yields in the prediction equations does supply a greater resolution. The full 1997 models for the north field included both 1996 and 1995 yields due to the perfect rotation, while models for the south field could only include 1996 yield due to the lack of rotation from 1995 to 1996. Similarly for the full 1996 models, 1995 yield could be included in the north half of the field but not on the south half of the field. To see how well these prediction equations were performing the 1998 yield was used to predict itself in a similar set of models. The only difference was that all yields from previous years were allowed in the full models that contained previous yield to get the best possible prediction, so that 1995 yield was allowed in the models for the south field too.

The elevation data set consisted of 830 points spread fairly evenly across the field, so elevation produces a resolution that was much closer to the 60-ft grid (fig. 1). Both the estimated $\text{elevation}^{1/4}$ and the terrain variables derived from the elevation data were used together with the already mentioned soil fertility, soil type, and yield variables in the most comprehensive models.

RESULTS and DISCUSSION

Correlation of soil tests

Even though the SSM and WFM treatments were applied in both 1996 and 1997, the soil test variables from 1997 showed strong correlations with the corresponding soil fertility variables except for CEC. CEC and (Ca) should be highly correlated. The correlations of Ca in 1995 and 1997 (Ca95 and Ca97) and CEC97 were very high, the coefficients were 0.7047 between Ca95 and Ca97, 0.6867 between CEC97 and Ca95, and 0.8966 between CEC97 and Ca97. In contrast the correlations between CEC95 and each of Ca95, Ca97, or CEC97 were –0.0584, 0.0238, and

0.1425, respectively, and not significant. The correlations between 1995 and 1997 variables are shown in table 2 where 1995 variables are the headings of the columns, while 1997 variable are headings of the rows. The years are postfixes for the variables. Transformations are indicated by a prefix of 'ln' for natural log and 'is' for √(max obs.) - √(max obs.-obs.). The transformations that made variables look normally distributed also ensured a fairly even spread over the range as can be seen in the plots of P and K from 1997 versus 1995, see fig. 7 and 8.

The strong correlation between corresponding soil variables from 1995 and 1997 imply that though the level of the soil measures may change, relative levels from subplot to subplot do not change much over a couple of years for pH, OM, Ca, and Mg. Soil testing at the level of the one-acre grid was not impacted by the different systems of fertilizer application for the two measures P and K that were involved. The use of the 1997 soil test variables was thus justified as reasonable variables to be used in prediction equations for yield in other years than just 1997.

Prediction models

The variables selected in the reduced models were based on Mallow's cp selection in the regression models and backward elimination in the models with spatial structure. All models were weighted by the relative number of measures of yield in each 60 ft square. The full SAS models were as follows for adjusted dry yield of corn in 1998:

```
proc reg data=kansas.analysis;
model AD98C =
      s1-s5 PB_97 PCA97 PFE97 PK_97 PMG97 PMN97 POM97 PP_97 PPH97 PS_97 PZN97
                     pwelev_r lnlnupsl WETNESS SLOPE PROFILE TANGENT MEANCURV
                                                  /selection=cp best=20 vif;
weight CT98Ca;
run;
```

and

```
proc mixed data= kansas.analysis;
model AD98C =
      s1-s5 PB_97 PCA97 PFE97 PK_97 PMG97 PMN97 POM97 PP_97 PPH97 PS_97 PZN97
               pwelev_r lnlnupsl WETNESS SLOPE PROFILE TANGENT MEANCURV /s p;
parms 298.718  0.820  215.971/noiter;
repeated / subject=intercept local type=sp(exp) (X Y);
weight CT98Ca;
run;
```

The parameter estimates for the reduced models for the most comprehensive model that allows all variables (soil fertility, soil type, previous yield, elevation, and terrain) are shown in table 3 for corn yield and in table 4 for soybean yield. The selection of variables from year to year only showed consistency in the previous years' yields, always selecting these variables (except for soybean in 1997) with positive parameter estimates. Soil types were also selected in all models except in the mixed models for 1997. The soil type parameter estimates for corn yield showed the second soil type was lowest in 1998 and 1996, while the fourth soil type was the lowest in 1997, in the middle in 1998, and the highest in 1996. The remaining three soil types were generally similar and in the middle except in 1996, where soil type five was the highest. The soil type parameter estimates for soybean yield showed that in 1998 soil types one and five were higher than the other soil types, in 1997 just soil type five was higher, while in 1996 soil types two and three were higher than the remaining soil types.

The soil fertility variables were not consistently significant nor with the same sign in the different years. Measures such as P and K did not show a strong presence. K was only part of the regression models without spatial correlation for soybean with a negative relationship and P appeared the same way in 1996 soybean models and 1997 corn models. P had one negative parameter estimate in 1997 in the soybean model with spatial correlation. OM had a positive estimate in the 1996 soybean model with or without spatial correlation. The only other presence was in the 1998 corn model without spatial correlation, where the estimate was negative. Since that was the north field where manure had been added this result was probably not very meaningful.

Using both the elevation variable itself and the derived terrain variables in models supplanted the soil measures as predictors to some extent. In the models with spatial correlation without the terrain variables there were generally fewer variables retained with a corresponding larger MSE. Models without the spatial structure tended to keep the same number of variables and incur a slightly larger MSE.

For all the models the year was a strong factor with respect to how was explained by the variables. The R^2 in the simple regression was similar for 1996 and 1998 while the R^2 in 1997 was half the size of those in the two other years. While the 1997 soybean models did not contain previous years' yields the corn models did. Both the most comprehensive corn and soybean models had R^2 of less than 0.26. For the corn model this might be due to that no previous corn yield was available. The 1996 soybean models did not contain previous years' yields due to the lack of rotation and still the most comprehensive model had an R^2 of 0.5164. The 1996 corn model that contained the 1995 yield had an R^2 of almost 0.70. The values of R^2 are presented in table 5.

Prediction of 1998 yield from the 1996 and 1997 models

To assess how good the predictions of the 1998 yield were, the sum of squared differences between actual and predicted values (adjusted for overall means) were compared to those of the 1998 yield model predicting it's own yield. Also the correlation between predicted and actual yield was checked to further assess the goodness of fit (table 6). The regression models without spatial correlation appeared to give better predictions in general in terms of having smaller mean square errors (MSEs). In the 1996 and 1998 models the MSEs and the correlation coefficients between predicted and actual yield were similar for models with and without spatial correlation. The MSEs and the correlation coefficients differed in the 1997 models. Counter intuitively the MSE were larger for the more comprehensive corn models. The only corn model that has a significant correlation between predicted and actual yield was the simplest mixed model. The soybean models with spatial correlation produced lower MSEs while the only significant correlation was for the model with only soil variables without spatial correlation. This coefficient was negative and thus completely irrelevant. The corresponding model with spatial correlation exhibited a correlation between predicted and actual yield close to 0 and very insignificant, which shows why one should use models with appropriate correlation structure. The 1997 models were the least explanatory models of the yield from the same year.

Using the 60-ft averaged yields from 1998 means and standard deviations were obtained for each crop. The means and standard deviations were weighted by the counts of yield measures in each 60-ft square. A comparison between the standard deviations and the squareroot of the MSEs (SE) showed that most 1996 and 1997 models do worse predictions than if the 1998 overall mean was used as the prediction. For the soybean crop the mean was 30.0 Bu/A and the standard deviation was 6.3 Bu/A. For corn the mean was 128.7 Bu/A and the standard deviation was 23.3 Bu/A. Except for the most comprehensive 1996 corn models all 1996 and 1997 models produced larger SE values (table 6) than the standard deviations discussed above. For the most comprehensive 1996 corn models the gain in precision was about 1 Bu/A. The 1998 model predictions (not surprisingly) do better. Including previous yield in models makes a big difference for both corn and soybean yield predictions. Using models with soil variables alone improve predictions only about 1 Bu/A for corn yield and 0.5 Bu/A for soybean yield over what an overall mean for the field would do.

SUMMARY

The correlation between 1995 and 1997 soil test data, show that the measures did not change much in 1-acre plots relatively from plot to plot, even with an uneven application of N, P, and K over 2 years. The scale of resolution of the soil measures did not work through kriging to be able to predict yield very well, explaining 1998 yield in a field from a 1998 yield model only did 0.5 – 1.0 Bu/Acre better than the overall mean for the field when only soil variables were used for prediction. The prediction from one year to another using only soil variables did worse than simply using the overall mean for the field.

Predictions using previous years' yield did better. Predicting 1998 yield in a field from a 1998 yield model decreased the standard deviation from 23.3 Bu/A to about 15 Bu/A for corn and from 6.3 Bu/A to about 4.8 Bu/A. Using other years yield models to predict soybean yield did not work – at best the models did as well as the mean itself. For corn there was a gain of about 1 Bu/A in precision when using the most comprehensive model in the same half of the field i.e. the 1996 model. Modeling soybean in the same field was hampered by the lack of the unavailable previous year's yield. The 1997 corn models were neither good at explaining their own yield nor do predictions in the other half field, but the models did not have previous corn yield due to the changed rotation scheme in 1996.

The prediction models in the study are only about predicting relative levels of yield. The study shows that without previous yield as predictor, the ability to predict is not available. Even with a good yield record, fields possibly vary a lot in predictability. In the best of circumstances a gain in precision of only about 1 Bu/A over the overall mean was obtained in this study. Yield prediction based on just soil measures was not supported by the results reported here.

REFERENCES

Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. SAS System for Mixed Models. Cary, NC: SAS Institute Inc. 633pp

Kravchenko, A.N. and D.G. Bullock. 2000. Correlation of corn and soybean grain yield with topography and soil properties. Submitted to the Agronomy Journal.

SAS Institute, Inc. 1989. SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2. Cary, NC:SAS Institute Inc.

SAS Institute, Inc. 1997. SAS/STAT Software: Changes and Enhancements through Release 6.12. Cary, NC:SAS Institute Inc.

ACKNOWLEDGEMENT

Thanks to two people from the Geographic Modeling Systems Lab, Dept. of Geography. Diane Timlin, for manipulating the data for me and also showing me to do a few things in Arcview, and Helena Mitasova, for creating the terrain variables from the elevation data.

Thanks also to the reviewer for good suggestions.

Table 1. Crop rotation in the north and south halves of the field.

Year	1998	1997	1996	1995
North	corn	soybean	corn	soybean
South	soybean	corn	soybean	soybean

Table 2. Correlation between 1995 and 1997 soil fertility measures.

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 106

	ISPH95[1]	LNCEC95[2]	LNOM95	LNP95	LNK95	CA95[3]	MG95
ISPH97	**0.58264**[4]	-0.07720	-0.49894	-0.32867	-0.36376	0.13958	0.43573
	0.0001	0.4315	0.0001	0.0006	0.0001	0.1536	0.0001
LNCEC9	0.18051	***0.14247***[5]	0.36017	0.06174	0.22130	0.68670	0.55529
	0.0641	***0.1452***	0.0001	0.5295	0.0226	0.0001	0.0001
LNOM97	-0.46145	0.00497	**0.67392**	0.47581	0.56227	0.27031	-0.14756
	0.0001	0.9597	**0.0001**	0.0001	0.0001	0.0051	0.1312
LNP97	-0.48373	-0.18729	0.67503	**0.70015**	0.71139	0.10703	-0.28742
	0.0001	0.0545	0.0001	**0.0001**	0.0001	0.2748	0.0028
LNK97	-0.39527	-0.08319	0.65876	0.50871	**0.78451**	0.31909	-0.05983
	0.0001	0.3965	0.0001	0.0001	**0.0001**	0.0009	0.5424
LNS97	0.07157	-0.54087	-0.02608	-0.02177	0.18682	0.13548	0.05754
	0.4660	0.0001	0.7907	0.8247	0.0552	0.1661	0.5579
LNB97	0.20364	-0.34466	0.10924	-0.08355	0.13013	0.36241	0.25204
	0.0363	0.0003	0.2650	0.3945	0.1837	0.0001	0.0092
ISMN97	0.35325	-0.20286	-0.24593	-0.06950	-0.16625	0.15853	0.23961
	0.0002	0.0370	0.0110	0.4790	0.0885	0.1046	0.0134
FE97	0.04175	-0.14236	-0.10735	0.15877	0.02577	0.02989	0.00876
	0.6709	0.1455	0.2734	0.1040	0.7931	0.7610	0.9290
LNZN97	-0.23388	-0.11006	0.58057	0.46132	0.55045	0.31706	-0.00153
	0.0158	0.2614	0.0001	0.0001	0.0001	0.0009	0.9876
CA97	0.23266	0.02379	0.26699	0.01561	0.11814	**0.70471**	0.54334
	0.0164	0.8087	0.0057	0.8738	0.2278	**0.0001**	0.0001
MG97	0.53305	0.13358	-0.11129	-0.28543	-0.16666	0.56959	**0.73139**
	0.0001	0.1722	0.2560	0.0030	0.0877	0.0001	**0.0001**

[1] pH and Mn were transformed using $\sqrt{(\text{maximum value})} - \sqrt{(\text{maximum value} - \text{observation})}$.
[2] CEC, OM, P, K, S, B, and Zn were transformed using natural log.
[3] Ca, Fe, and Mg were not transformed.
[4] The correlation coefficients for corresponding soil fertility measures are in bold.
[5] The correlation coefficient for corresponding CEC is different in nature than the other corresponding measures.

Table 3. Parameter estimates for models with and without spatial correlation structure for corn yield for the three years 1996 to 1998. Only estimates for the most comprehensive model where the full model included previous yields, terrain variables, soil types, and soil fertility measures are presented.

Year	1998		1997		1996	
R^2	0.5660		0.2236		0.6871	
	regression	mixed	regression	mixed	regression	mixed
INTERCEP	-40.78	6.88	264.42	176.69	159.67	161.46
S1						3.76
S2	-4.66				-3.59	
S3						7.11
S4			-23.12			
S5	6.73				6.42	9.83
AD96C	0.69	0.60	0.28	0.30	-	-
AD95B	0.51	0.53			0.61	0.52
PFE97		-1.90			45.00	60.42
PB_97	-71.84			88.94	-7.61	-7.18
PCA97	0.04					
PCE97			-11.52	-17.68		
PFE97			0.74			
PMG97	-0.10					
PMN97	7.47		6.77			
POM97	-26.05					
PP_97			-14.30			
PPH97				143.01		
PZN97	-91.20	-224.17	-58.80			
PWELEV	-25.09	-25.75		57.03	-7.16	
WETNESS		-1.53				
LNLNUPSL			8.51	6.91		
ASPECTAD		0.044			0.026	0.021
MEANCURV				-17.82	-18.28	

Table 4. Parameter estimates for models with and without spatial correlation structure for soybean yield for the three years 1996 to 1998. Only estimates for the most comprehensive model where the full model included previous yields, terrain variables, soil types, and soil fertility measures are presented.

Year	1998		1997		1996	
R^2	0.5054		0.2562		0.5164	
	regression	mixed	regression	mixed	regression	mixed
INTERCEP	162.65	6.27	193.00	66.12	38.67	578.71
S1	3.63	3.83				
S2					3.60	2.27
S3					5.94	3.85
S5	1.52	1.85	1.58			
AD97C	0.07	0.08				
AD96B	0.16	0.10				
AD95B	0.24	0.21				
PB_97			7.34			
PCA97			0.003			
PCE97			-6.66			
PFE97	-0.55					
PK_97	-27.48		-18.85			-59.29
PMG97					-0.02	
PMN97	1.61				5.22	4.90
POM97					10.32	8.82
PP_97				-1.70	-3.18	
PPH97			-21.19		20.20	-92.95
PS_97					10.20	-9.47
PZN97			16.04		-47.09	
PWELEV			4.67		-34.41	-43.97
PROFILE	2.41					
SLOPE		1.11			2.98	1.46

Table 5. R^2 for the different kind of regression models without spatial correlation for each year.

Year	1998				1997				1996			
Crop	#	corn	#	soy	#	corn	#	soy	#	corn	#	soy
all variables included[1]	11	0.5660	9	0.5054	8	0.2236	7	0.2543[4]	8	0.6871	-	-
previous yield & soil[2]	11	0.5657	8	0.5020	5	0.1983	7	0.2543[4]	9	0.6747	11	0.5164
soil variables[3]	8	0.2548	6	0.2548	3	0.1323	7	0.2543	8	0.4948	10	0.4307

[1] Full model included previous yields, terrain variables, soil types, and soil fertility measures.
[2] Full model included previous yields, soil types, and soil fertility measures.
[3] Full model included soil types and soil fertility measures.
[4] Neither previous yields nor terrain variables were selected for the final model.

Table 6. Mean square error and standard error of predicted yield, and correlation coefficient and corresponding p-value between actual and predicted yield for each of the crops (corn and soybean).

Corn

full model	year	#obs.	#[1]	MSE	SE	corr	p value	#	MSE	SE	corr	p value
			Regression model					Mixed model				
	1998											
all variables[2]		634	11	206.25	14.36	0.76	0.0001	6	240.88	15.52	0.71	0.0001
prev. yield & soil[3]		634	11	206.79	14.38	0.76	0.0001	4	263.29	16.23	0.68	0.0001
soil variables[4]		653	8	405.01	20.12	0.52	0.0001	5	487.38	22.08	0.38	0.0001
	1997											
all variables		634	8	614.11	24.78	-0.01	0.8520	7	827.62	28.77	0.02	0.5981
prev. yield & soil		634	5	610.75	24.71	-0.04	0.3533	4	634.39	25.19	0.01	0.7897
soil variables		653	3	597.65	24.45	0.04	0.2589	3	611.26	24.72	0.12	0.0031
	1996											
all variables		634	8	493.10	22.21	0.24	0.0001	8	507.23	22.52	0.23	0.0001
prev. yield & soil		634	10	535.16	23.13	0.11	0.0068	6	552.97	23.52	0.11	0.0062
soil variables		653	8	631.90	25.14	0.03	0.5302	6	604.70	24.59	0.18	0.0001

Soybean

full model	year	#obs.	#	MSE	SE	corr	p value	#	MSE	SE	corr	p value
	1998											
all variables		562	9	21.16	4.60	0.68	0.0001	6	23.50	4.85	0.66	0.0001
prev. yield & soil		562	8	21.16	4.60	0.68	0.0001	6	23.05	4.80	0.66	0.0001
soil variables		644	6	33.80	5.81	0.41	0.0001	4	43.75	6.61	0.27	0.0001
	1997											
all variables		634	7	52.97	7.28	-0.32	0.0001	1	40.68	6.38	-0.02	0.6880
	1996											
all variables		578	11	65.59	8.10	0.30	0.0001	9	137.1	11.71	0.38	0.0001
soil variables		578	10	59.54	7.72	0.27	0.0001	6	121.3	11.01	0.29	0.0001

[1] Number of variables in the final model.
[2] Full model included previous yields, terrain variables, soil types, and soil fertility measures.
[3] Full model included previous yields, soil types, and soil fertility measures.
[4] Full model included soil types and soil fertility measures.

Conference on Applied Statistics in Agriculture
Kansas State University

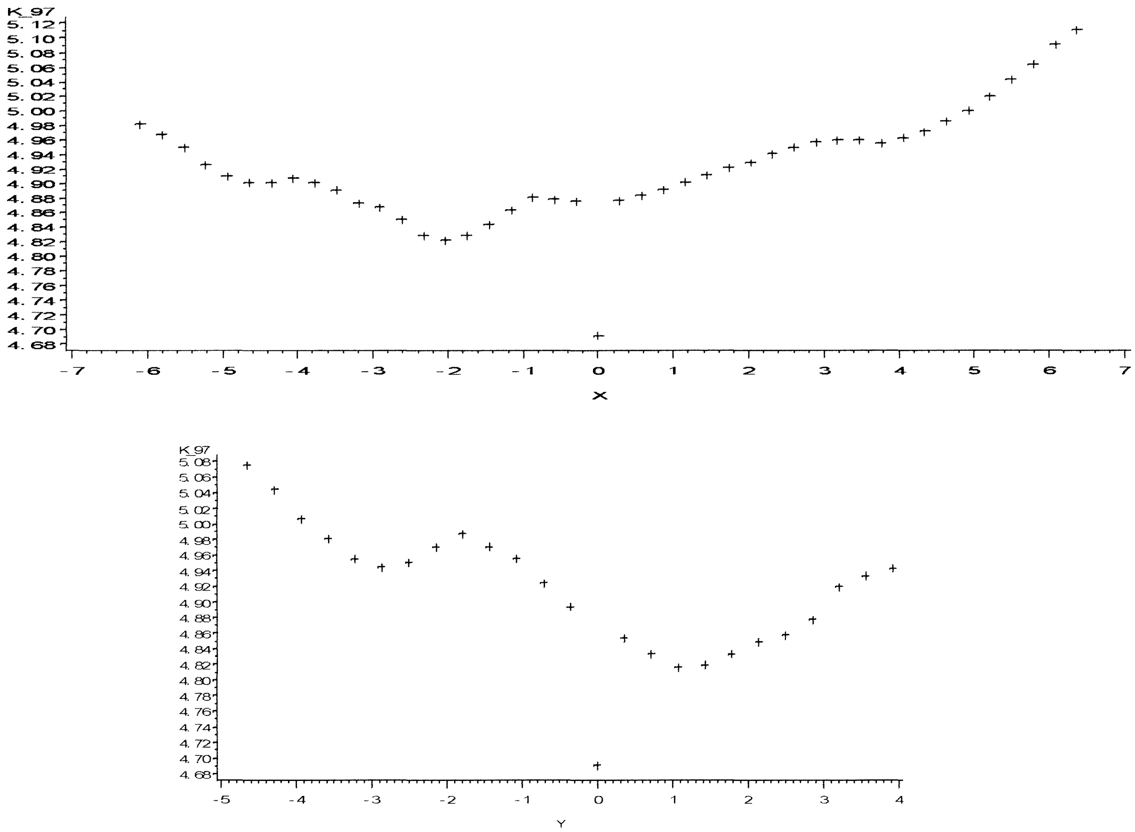

Figure 3. Plots of kriged values of K versus x at y=0 and K versus y at x=0.

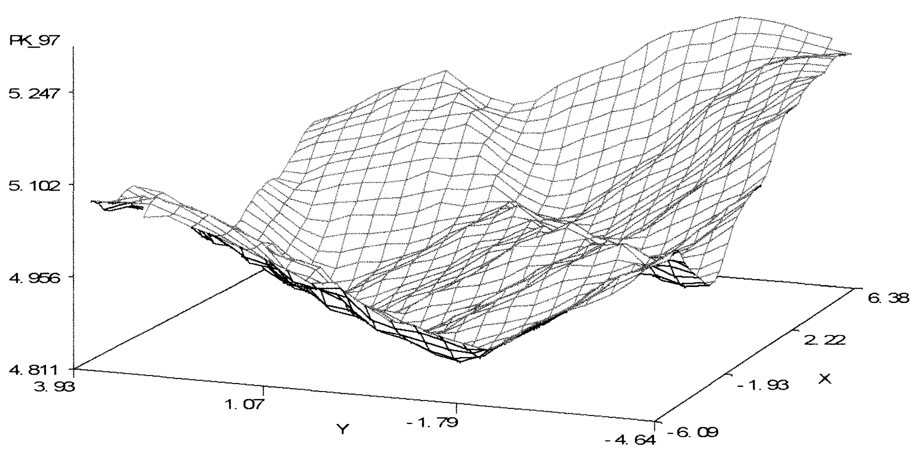

Figure 4. Surface of the kriged K variable.

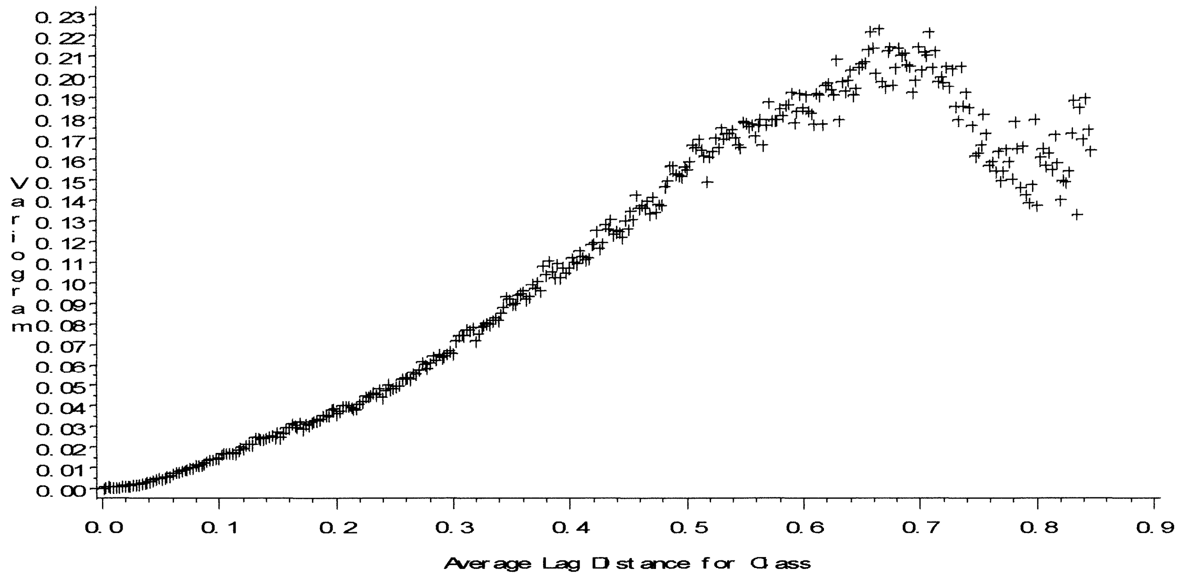

Figure 6. Semivariogram for the residuals of the transformed elevation variable modeled on powers and interactions of x and y.

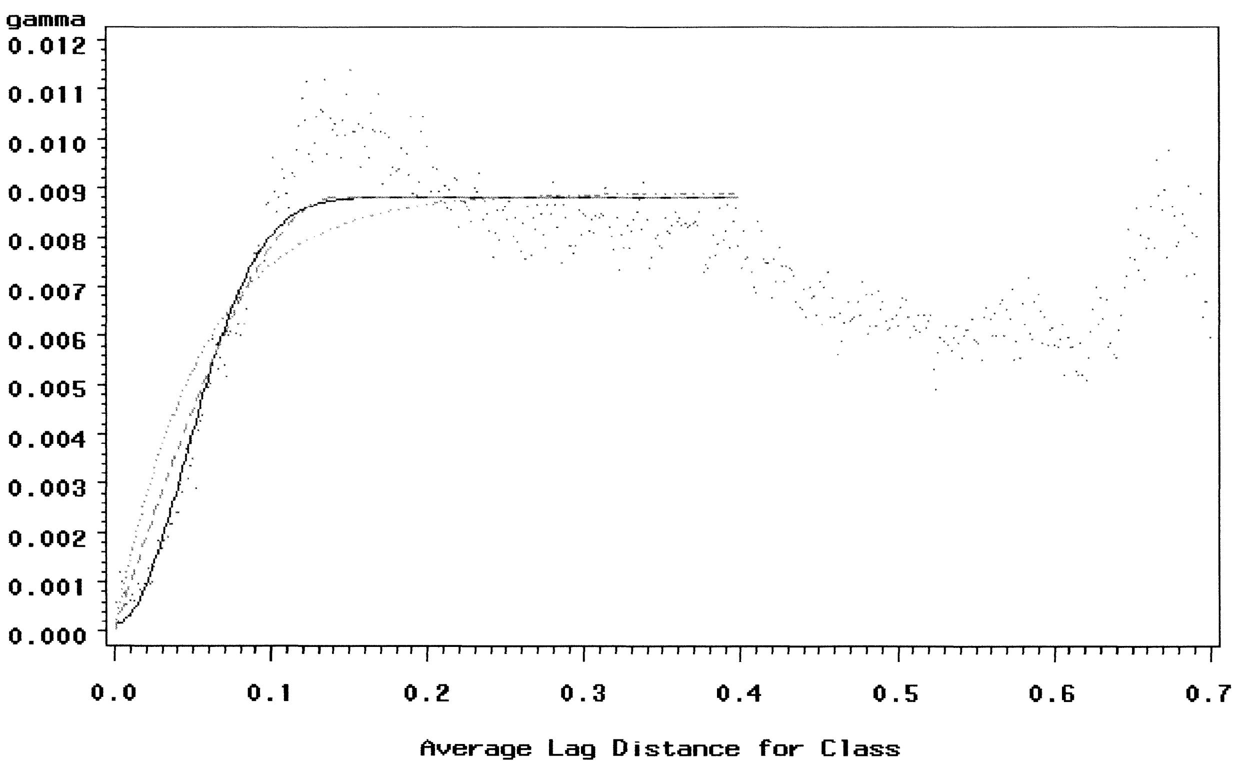

Figure 5. Experimental (large dots), exponential (dotted line), spherical (dashed line), and gaussian (solid line) semivariograms for the transformed elevation variable on less than half the range.

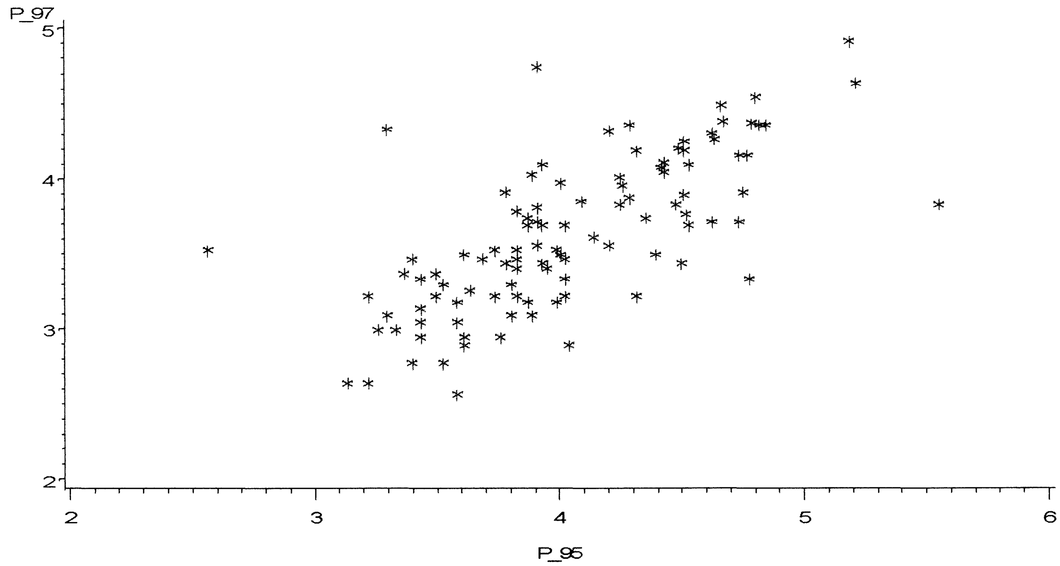

Figure 7. The 1997 P soil test results versus the 1995 P soil test results.

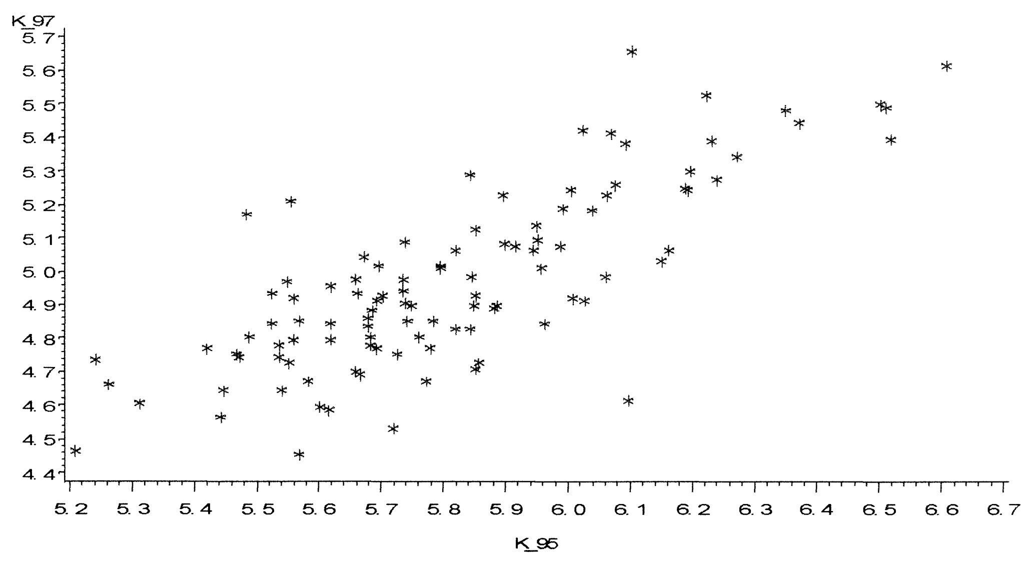

Figure 8. The 1997 K soil test results versus the 1995 K soil test results.