Biometrical Letters
Vol. 43 (2006), No. 2, 67-77

Some considerations on the use of geostatistical methods in agricultural field trials
Part II: Comparison of theoretical considerations with results from real data

Christel Richter*, Bärbel Kroschewski

Department of Biometry and Experimental Design, Humboldt-Universität zu Berlin, Faculty of Agriculture and Horticulture, 10115 Berlin, Invalidenstraße 42, Germany;

Summary

For the description of spatial dependency, the spherical model and the exponential model – both developed for point data – are often utilized. Using both as initial models for different regularizations (transects, plots), we examined the behaviour of the corresponding covariance functions by means of simulation (Richter and Kroschewski, 2006). Now, the theoretical considerations are confronted with results from two uniformity trials. In the first trial, the spatial dependency of single plants was observed. Only by the summarizing of plants to small plots and by the superimposition of environmental effects do there appear covariances which could be modelled by the above-mentioned standard models. In the second trial, larger plots were derived from the small plots and covariance models were fitted. We used the SAS procedure MIXED with all available spatial variance-covariance structures. Using the Akaike criterion, the covariogram functions show mostly anisotropy and a sigmoid shape. Their nugget variance decreases with increasing plot size and/or plant density. So far, the theoretical considerations seem to be confirmed. However, the sensitive estimation of the range does not reflect the theoretical results. Further investigations, taking into account crop, soil and weather-dependent results, are necessary.

Key words: geostatistical methods, point data, regularized data, agricultural field trials

*Corresponding author. D-10115 Berlin, Invalidenstraße 42, phone: +49 30 2093 8483; fax: +49 30 2093 8460, e-mail: ch.richter@agrar.hu-berlin.de

1. Introduction

The starting point of our considerations in part I of this paper was that in agricultural applications of geostatistics the isotropic spherical or exponential model is often used to describe spatial dependency (Richter and Kroschewski, 2006). However, these models were originally intended for point data. Therefore, in the first part we dealt theoretically with the relations between point and regularized data using the layout of transects and plots as is typical for agricultural field experiments. As initial point data we chose the above-mentioned models with or without nugget in the isotropic and geometrically anisotropic form. By means of simulation the characteristics of regularized covariograms are elaborated. Because the omnidirectional calculation of the covariogram function is often used, we focused on this calculation and discussed the relation to the directional covariograms. Except for quadratic plots, the initial isotropic point data change to anisotropic data through regularization. For anisotropic point data the anisotropy does not vanish. For which data the isotropic spherical or the exponential model could be appropriate remains an outstanding question for agricultural experiments.

In classic textbooks on mining geostatistics (Journel and Huijbregts, 1978; Rendu, 1981), there is a strong differentiation between data with punctual support and data which come from higher dimensional support. In mining there is a consensus that data can be observed as point samples when a smaller sample cannot be drawn (i.e. approximately hand-sided) and/or the area/volume is small in relation to the range of the covariance function and to the observed region (Akin and Siemes, 1988).

In field experiments, such experiences are rare. Point data could be interpreted as data from single plants or from small plots as are usually used in uniformity trials (e.g. 1x1 m^2 or 2x2 m^2). At this point, it is indeed questionable whether data from normal plot sizes (e.g. in variety trials with 10 to 15 m^2) can be considered as point data.

For individual plants, Matérn (1986) has argued that the pure competition effect between adjacent plants would provoke a negative covariance. For larger distances, according to Matérn, the covariance function would have to oscillate around zero with decreasing amplitudes. In field experiments, however, we have to take into consideration that competition effects are superimposed by soil effects. Continuously varying soil effects cause positive correlations. Both effects are confounded – only the resulting effect can be observed. Moreover, in most field experiments the purpose of investigation is not the reaction of single plants but of plant stands, e.g. the plot yield. The plot value soaks up the individual variability, competition between the contiguous plants within the plot, and small-scale soil heterogeneity.

The second question concerns the characteristics of the regularized covariogram functions observed (Richter and Kroschewski, 2006). Can real data really confirm these characteristics?

Below, the theoretical considerations will be confronted with results from real data from two uniformity trials in Thyrow and Blumberg (experimental stations of the Humboldt-Universität zu Berlin).

2. Material and Methods

The first trial was laid out in three different fields from 1985 to 1987. In this trial, the yield of single potato plants was recorded. Originally, it was not planned to analyse competition effects, so the distances within (0.33 m) and between rows (0.75 m) conform to the usual practice and were not graduated. Twenty plants stood in each row (1985: 15 rows, 1986 and 1987: 25 rows). We calculated the empirical covariance functions within and between rows.

In the second trial, several crops were grown on the same area (40x48 m^2) in five consecutive years (hemp 1954, winter rye 1955, sugar beet 1956, oats 1957, and hemp 1958). The yield data were recorded on a regular grid with 480 plots of 2x2 m^2.
The latter data were analysed in several publications (e.g. Bätz, 1968; Rabe, 1980) using the empirical variance law of Fairfield Smith (1938) and modified methods. Both authors dealt critically with Smith's heterogeneity coefficient, which plays a fundamental role in his approach. Rabe (1980) gave a first idea of how to use the mean covariance between plots.

The fitting of covariance models assumes second-order stationarity. Because of a trend in the x-direction, we detrended the data by subtraction of the corresponding x-column median. Other methods would also have been possible. In designed experiments, this assumption could be realized by an adequate block construction. From the 2x2 m^2 plots, larger plots are derived with orientation in the x-direction (4x2, 8x2, 10x2) and in the y-direction (2x4, 2x6, 2x8) by summarizing the included 2x2-plot values.

All calculations for the second trial were performed with PROC MIXED from SAS release 9.1.3. We fitted all spatial variance-covariance-structures available in this procedure (SAS notation in brackets):

- isotropic models: exponential (`exp` or *with another parameterization*: `pow`), spherical (`sph`), linear (`lin`), log linear (`linl`), matérn (`matern` or *with another parameterization*: `mathsw`), gaussian (`gau`),
- anisotropic models: several exponential approaches (`powa`, `expa`, `expga`), spherical (`sphga`) and gaussian anisotropic (`gauga`),

each of them with and without nugget (for details cf. SAS 9.1.3).

In cases where two different parameterizations exist for the same model, we use the abbreviation underlined. The corresponding covariance functions with nugget are symbolized by an additional 'N', e.g. `sph N` corresponds to equation (1) and `pow N` to equation (2) in Richter and Kroschewski (2006).

As far as relations between the models exist, they will be characterized as follows: A $\subset$ B denotes that model A is a special case of B. The following representation does not take into consideration a possible nugget effect (in brackets: the number of parameters without nugget and special characteristics).

isotropic anisotropic

`sph` (2) $\subset$ `sphga` (4, geometric anisotropy)

`exp` (2) $\subset$ `expga` (4, geometric anisotropy)
‖
`pow` (2) `powa` (3) $\subset$ `expa` (5, both: separable nisotropy)
$\cap$
`matern` (3)
$\cup$
`gau` (2) $\subset$ `gauga` (4, geometric anisotropy)

`Expa` and `expa N` loom large in the following results. The covariance function of `expa N` is

$$\mathrm{Cov}(h_x, h_y) = \begin{cases} C_0 + C & \text{if } h_x = h_y = 0 \\ C \cdot e^{-t_x \cdot h_x^{px}} \cdot e^{-t_y \cdot h_y^{py}} & \text{otherwise} \end{cases} \tag{1}$$

with h_x and h_y being the distances in the x- and y-direction, while px and t_x, resp. py and t_y are the parameters for the two directions. If px > 1 and/or py > 1, then `expa` has a sigmoid shape in the x- and/or y-direction; otherwise it does not.
`Powa` is a special case of `expa` with px = py = 1 in (1) and has therefore no points of inflection. Both models are separable (Cressie, 1993).
`Gauga and gau` are models with a sigmoid shape. `Matern` may or may not have a point of inflection depending on the smoothness parameter.

`Expga, sphga` and `gauga` are models for geometric anisotropy. In these cases, the isolines of the spatial covariogram lie on concentric ellipses with perpendicular main axes, which, in contrast to `powa` and `expa,` do not necessarily correspond to the x- and y-axis. The corresponding isotropic forms are `pow, sph` and `gau.`
As a consequence of part I (Richter and Kroschewski, 2006) we expected anisotropy associated with a sigmoid shape for regularized data and omnidirectional calculations. `Expa` and `gauga` are the only functions which may display these characteristics.

The REML-procedure in PROC MIXED, which is used for parameter estimation, led to numerical difficulties in many cases. In most situations, we could overcome this problem by default of suited initial values.
For all considered plot sizes, we fitted the omnidirectional covariograms using the option of a potential correlation over the whole area (`subject=intercept`). Furthermore, for the plots 2x4, 2x6 and 2x8 we analysed the covariances in the x-direction, while for the plots 4x2, 8x2 and 10x2 we analysed the covariances in the y-direction (`subject=y` resp. `x`). Depending on the plot shape these are the directions where the larger number of points are available for the estimation. Only for 2x2 plots were the models fitted in both directions.

3. Results and Discussion

In the first trial, the spatial dependence was rather weak in all three years. In 1985 and 1987, the growing conditions were better because of higher rainfalls. Better developed plants used the standing area and competed for resources. So, for the potato yield the 1-lag correlation within the row was -0.2^{+} in 1985 and -0.1^{+} in 1987. For larger distances, the correlation oscillated more or less around zero, but not generally with decreasing amplitudes. Between the rows, the 1-lag correlation was nearly zero. 1986 was a dry year with bad developing conditions. In that year there did not seem to be any competition. Within rows the correlation to the adjacent plant was +0.1⁺ and was positive over the whole length of the rows. The correlations between the rows were positive up to the fifth adjacent row. Because of lack of competition, the similar soil conditions seem to account for this result. 1986 was the only year where, by summarizing adjacent plants to plots, the above-described increase of the *PSS* could be observed (from 0.5 for single plants up to 0.9 for plots with 20 plants per row).

This year-dependent result agrees with Hudson (1941) who noticed that competition will only become operative when the spheres of absorption of the roots overlap. He concludes that differences exist between root crops (wide spacing between plants) and cereals (closer spacing). Competition begins for root crops later than for cereals in the history of plant development (Hudson, 1941).

These considerations have made clear that normally the covariogram models discussed above are not appropriate for single plants. Therefore, in the following we will focus on practical results for plots which we found in the second trial. Selected results of the model fit are given in Tables 1, 2, and 3.

Models with the smallest -2REML Log Likelihood show the best fit (Table 1). To control a possible overfit both for the omnidirectional and for the directional calculation, the model with the smallest Akaike criterion (AIC) is additionally provided in Table 1. This criterion penalizes models with higher numbers of parameters, so that models with fewer parameters will be preferred.

Table 1: Uniformity trial Blumberg: Models with smallest -2 REML Log likelihood and with smallest AIC (if differently: *italic*). [x)] or [y)] : no sigmoid shape in x- or y-direction

crop		plot						
		2 x 2	2 x 4	2 x 6	2 x 8	4 x 2	8 x 2	10 x 2
hemp 1954	omnid.	expa N *powa N*	expa N	expa N *powa N*	expa N *powa*	expa N *powa N*	expa[x)] *powa*	expa[y)] *powa*
	x-dir.	mat[x)]	lin N	mat[x)] *pow*	lin N *pow*			
	y-dir.	mat N *gau N*				lin N	mat N *pow*	sph N
winter rye 1955	omnid.	expa N	gauga N *gau N*	expa N	expa N	expa N	expa N	expa N
	x-dir.	gau N	gau N	mat N *gau N*	mat N *gau N*			
	y-dir.	gau N				gau N	gau N	gau N

sugar beet 1956	omnid.	expa N	expa N *sph N*	sphga N	gauga N *expga*	expa N *gau N*	expga N *gau N*	expga N *expga*
	x-dir.	mat N *pow*	mat N *pow*	sph N *pow*	lin N			
	y-dir.	lin N				gau N	gau N	lin N
oats 1957	omnid.	expa N	expa N *gau N*	gauga N *gau N*	sphga *sph*	expa N	expa$^{x)}$	expa N$^{x)}$ *powa*
	x-dir.	mat N *gau N*	mat N *gau N*	gau N	mat N *gau N*			
	y-dir.	sph N *sph*				mat	mat *lin*	sph
hemp 1958	omnid.	expa N	expa N	expa N *powa*	expa$^{x)}$ *powa*	expa N	expa N$^{x)}$ *sphga*	expa$^{x)}$
	x-dir.	mat N *pow N*	mat N *pow*	mat *pow*	sph N *pow*			
	y-dir.	linl N				sph N	sph N *sph*	sph N *sph*

At first the results for the 2x2 plots will be discussed. The omnidirectional calculation yielded anisotropy for both criteria and points of inflections in both directions with the -2REML criterion. For 4 of 5 crops, the sigmoid shape remains with AIC. So already the smallest observed plots show the previously discussed typical behaviour of regularized data. From these results, we expected that the directional calculations would show a sigmoid shape in every case. However, depending on the criterion, only 6 or 4 cases respectively out of 10 functions had such a shape. The omnidirectional calculation showed the existence of a nugget with both criteria, the directional ones in most cases.
Based on the 2x2 results and part I (Richter and Kroschewski, 2006), we assumed that the anisotropy and the sigmoid shape do not vanish for larger plots.

Altogether, for all 35 omnidirectional cases analysed, we found anisotropy with the -2REML criterion. Except in four cases, the functions have a point of inflection in at least one direction (88.6%). This confirms the expectation that anisotropic sigmoid models are best suited for regularized data. It is remarkable that `expa` or `expa N` occurs 28 times out of 35 cases. These models are with 5 or 6 parameters respectively the most flexible models, but they assume that the main axes of anisotropy correspond to the x- and y-axis. This condition is not satisfied in the 7 cases where geometric anisotropic models give the best fit. In these situations, the x- and y-directional covariograms are not enveloping curves for the omnidirectional one. Comparing the results of both criteria, 19 times we found a model with fewer parameters by using AIC. Partially the anisotropy, the sigmoid form or the nugget vanish; sometimes even two characteristics disappear. Nevertheless, `expa` or `expa N` remain the most frequent models (15 times).

For the directional covariograms the model preference is more heterogeneous, however, following the -2REML criterion, `mat` and `mat N` occur most frequently (16 times out of 40 cases) where in 14 cases a point of inflection exists. Together with `gau` and `gau N`, there are 23 models with a sigmoid shape (57.5%). With 3 and 4 parameters, the models `mat` and `mat N` are more flexible than the others.
By using the AIC for the directional calculations, the percentage of sigmoid curves decreases to 40 %. Now, the 16 situations with a sigmoid form result in the majority of cases from `gau` and `gau N` (15 times) and only once from `mat`. It is remarkable that the cereals especially show the sigmoid shape. For the other crops, the best models change more or less for the different plot sizes. The expectation that the omnidirectional calculation leads to the corresponding results for the directional calculation could only be partly fulfilled.

In Table 2 the percentages of selected model characteristics are summarized.

Table 2: Uniformity trial Blumberg: Percentages of selected model characteristics

omnidirectional (35 cases)	-2 REML Log likelihood percentage	AIC percentage
anisotropy	100.0	80
point(s) of inflection	68.6 (two) + 20 (one)	37.1 (two) + 20 (one)
nugget	82.9	65.7
directional (40 cases)		
point of inflection	57.5	40.0
nugget	85.0	60.0

For the omnidirectional calculation, the development of the parameter estimations depending on the plot size will be demonstrated in Table 3. The parameter development is only reasonably comparable for the same basis model. Because in the majority of cases the best fit was achieved with `expa` or `expa N` (using the -2REML criterion), these models are demonstrated. In the remaining 7 cases, `expa N` or `expa` are chosen depending on the smaller -2 REML Log likelihood.

In accordance with the remarks in part I (Richter and Kroschewski, 2006), in most cases C, C_0 and $C+C_0$ decrease with increasing plot sizes. For the 2x2 plots, $C/(C_0+C)$ grows with the plant density of the crop (sugar beet: 9 plants per m^2, hemp: 50–100, cereals: 350–450 stems per m^2).
Schabenberger and Pierce (2002) and Zimmerman and Harville (1991) have stated that the inclusion of a nugget effect seems to be unnecessary in many field experiments by referring to Besag and Kempton (1986). Their experiments were performed with cereals and had a plot dimension larger than 6.5 m^2.

Table 3: Uniformity trial Blumberg: omnidirectional calculations. Parameter estimation of the models expa or expa N. 'range' corresponds to the practical range (*italic:* 'range' < plot dimension). $^{x)}$ or $^{y)}$: no sigmoid shape in x- or y-direction. (...) models with no smallest -2 REML Log likelihood

crop		plot 2 x 2	2 x 4	2 x 6	2 x 8	4 x 2	8 x 2	10 x 2
hemp 1954	model	expa N	expa N	expa N	expa N	expa N	expa$^{x)}$	expa$^{y)}$
	C	1147.3	994.7	952.1	961.9	1055.2	1031.2	741.2
	C_0	384.2	200.0	107.0	39.6	156.2	0	0
	$C+C_0$	1531.5	1194.7	1059.1	1001.5	1211.3	1031.2	741.2
	$C/(C_0+C)$	0.749	0.833	0.899	0.960	0.871	1	1
	range (x)	13.6	12.1	15.6	18.8	11.9	47.2	23.9
	range (y)	18.0	16.4	16.4	18.6	18.5	17.1	22.1
winter rye 1955	model	expa N	(expa N)	expa N	expa N	expa N	expa N	expa N
	C	3.137	2.026	1.639	1.483	3.076	2.021	1.794
	C_0	0.032	0.354	0.272	0.244	0.162	0.398	0.273
	$C+C_0$	3.168	2.380	1.912	1.727	3.238	2.419	2.067
	$C/(C_0+C)$	0.990	0.851	0.858	0.859	0.950	0.835	0.868
	range (x)	9.4	6.7	6.9	6.9	16.5	*3.0*	*3.1*
	range (y)	9.5	*1.7*	*2.0*	*2.1*	8.6	6.2	6.5
sugar beet 1956	model	expa N	expa N	(expa N)	(expa N)	expa N	(expa N)	(expa)
	C	3861.6	3215.8	2393.1	2566.1	2186.2	2077.3	3146.0
	C_0	4816.4	2610.0	1959.5	1635.5	3815.7	1948.5	0
	$C+C_0$	8678.0	5825.7	4352.7	4201.6	6001.9	4025.8	3146.0
	$C/(C_0+C)$	0.445	0.552	0.550	0.611	0.364	0.516	1
	range (x)	6.9	9.4	10.5	9.6	9.2	9.1	11.2
	range (y)	7.1	5.6	8.1	*2.7*	7.1	7.7	10.0
oats 1967	model	expa N	expa N	(expa N)	(expa N)	expa N	expa$^{x)}$	expa N$^{x)}$
	C	14.485	12.319	12.180	12.027	14.462	15.094	11.451
	C_0	3.429	2.532	1.726	1.236	1.033	0	0.384
	$C+C_0$	17.914	14.851	13.906	13.263	15.495	15.094	11.835
	$C/(C_0+C)$	0.809	0.830	0.876	0.907	0.933	1	0.968
	range (x)	12.2	10.8	10.8	12.9	11.5	28.1	44.1
	range (y)	8.0	7.7	9.5	8.2	8.4	11.3	12.1
hemp 1958	model	expa N	expa N	expa N	expa$^{x)}$	expa N	expa N$^{x)}$	expa$^{x)}$
	C	819.0	685.2	684.0	620.1	729.6	604.1	622.5
	C_0	250.1	153.9	58.8	0	133.6	26.2	0
	$C+C_0$	1069.0	839.0	742.8	620.1	863.2	630.3	622.5
	$C/(C_0+C)$	0.766	0.817	0.921	1	0.845	0.958	1
	range (x)	12.6	9.8	9.9	17.2	18.0	4127.4	∞
	range (y)	9.6	10.3	7.9	9.7	10.0	9.6	10.8

Indeed, for larger plot sizes, the nugget effect more or less vanishes.
The practical ranges react very sensitively, so that the effects described in part I (Richter and Kroschewski, 2006) cannot be detected in the parameters of the fitted EXPA models. Especially in those cases where a point of inflection does not exist in the x-direction, the ranges are large. In some other cases, the range is

smaller than the dimension of the plot in this direction. This indicates that these plots soak up all spatial correlation. If one works with models without points of inflection, e.g. POWA or POW with or without nugget, this reaction is far less sensitive.

4. Concluding remarks

The comments about the behaviour of single plants have shown that single plants have for the most part a totally different correlation structure than could be described by a spherical or exponential model. Only by summarizing of plants to plots and by the superimposition of environmental effects do covariances appear which could be modelled by the discussed standard models. However, in our trial discussed above, the regularization effect becomes apparent even for $2x2m^2$ plots. On the other hand, it must be pointed out that with increasing plot size the points may be so sparse that a point of inflection cannot be observed.

The results of the example discussed and previous experiences with designed experiments have shown that the sigmoid shape and anisotropy of the covariogram functions of plots should be frequently considered. The nugget becomes smaller with increasing plant density, and more or less vanishes for larger plots. In accordance with the principle of parsimony with respect to the number of parameters, simpler models are sometimes sufficient. However, in the case of an omnidirectional calculation, models lacking all three observed characteristics – nugget, anisotropy and point of inflection – are rare. We realize that the concrete results depend on the observed crops, soil and weather conditions. To what extent these results can be generalized is an issue that needs further investigation. In part I of this paper (Richter and Kroschewski, 2006) we mentioned that the exponential and spherical models are very often used in the literature about geostatistics in agricultural field experiments. We assume that the fit of the other models discussed in this article has not been considered.
The use of the covariogram function in the planning phase of an experiment and its comparison with other methods will be discussed in a further paper.

References

Akin H., Siemes H. (1988). Praktische Geostatistik, Springer, Berlin Heidelberg New York.

Bätz, G. (1968). Untersuchungen zur Erhöhung der Aussagekraft von Feldversuchen. Habilitation. Friedrich-Schiller-Universität Jena.

Besag J., Kempton R. (1986). Statistical Analysis of Field Experiments Using Neighbouring Plots. Biometrics 42, 231-251.

Cressie N. A. C. (1993). Statistics for Spatial Data, Wiley, New York.

Hudson H.G. (1941). Population Studies with Wheat. Journal of Agricultural Science 31, 116-137.

Journel A. G., Huijbregts Ch. J. (1978). Mining Geostatistics. Academic Press London.

Matérn B. (1986). Spatial Variation. 2nd edition, Springer Berlin Heidelberg New York.

Rabe, C. (1980). Die Bestimmung optimaler Versuchselemente bei der Planung landwirtschaftlicher Feldversuche, speziell im Gemüsebau. Dissertation, Humboldt-Universität zu Berlin.

Rendu J. M. (1981). An introduction to geostatistical methods of mineral evaluation. South African Institute of Mining and Metallurgy.

Richter C., Kroschewski B. (2006). Some Considerations on the Use of Geostatistical Methods in Agricultural Field Trials. Part I: From Point to Regularized Data – Theoretical Considerations. Biometrical Letters 43(1), 49 – 66.

Schabenberger O., Pierce F.J. (2002). Contemporary statistical models for the plant and soil sciences. CRC Press.

Smith, H.F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. Journal of Agricultural Science XXVIII, 1-23.

Zimmerman D. L., Harville D. A. (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. Biometrics 47, 223 – 239.