Biometrical Letters
Vol. 43 (2006), No. 1, 19-37

The influence of spatial covariances on the type I error and the power for different evaluation models

Xiyuan Hu[1], Joachim Spilke[2*], Christel Richter[3]

[1] College of Agriculture, Northwest-Sci-Tech University of Agriculture and Forestry Yangling, China
[2] Biometrics and Informatics in Agriculture Group, Martin-Luther-University Halle-Wittenberg, Germany
[3] Institute of Crop Sciences, Humboldt-Universität zu Berlin, Germany

SUMMARY

This paper deals with the consequences of spatial covariances for the analysis of field experiments. Based on a Monte Carlo simulation, we investigated the Type I error and the power of three analysis models (block design, exponential and spherical model) in the context of two true spatial models (exponential and spherical). The block design leads to considerable bias of the nominal type I error and to a loss of power. In contrast, the application of the spatial models for analyses guarantees the maintenance of the nominal statistical type I error and a significantly higher power. Our results confirm the necessity of model selection in the analysis of field experiments.

Key words: spatial covariances, comparison of analysis models, type I error, power

1. Introduction and Challenges

In order to be able to estimate unbiased differences between treatments with high sufficient precision, field experiments have long made use of the principle of considering environmental variations, for example those caused by the heterogeneity of the soil, through the formation of blocks. If fixed block effects are to be assumed, every block is assigned a fixed effect. When assuming random block effects in the statistical model, a block-wise covariance modification of the observation vector is taken for granted. Thus a discontinuous environmental

*Corresponding author. D-06108 Halle/Saale, Ludwig-Wucherer-Straße 82-85, phone: +49 345 5522690; fax: +49 345 5527128, e-mail: joachim.spilke@landw.uni-halle.de

transformation is assumed both for fixed and random block effects. However, this approach is suboptimal, since the discretization of the environmental differences performed here is an artificial one – from a biological point of view, one has to assume a continuous modification in most cases.

The inclusion of geostatistic approaches in linear models in particular allows for flexible consideration of spatially dependent error sources. A number of authors have pointed out the importance and necessity of such approaches over recent years (Zimmermann and Harville, 1991; Stroup et al., 1994; Littell et al., 1996; Gilmour, 1997; Federer, 1998; Schabenberger and Pierce, 2002; Richter and Kroschewski, 2002; Stroup, 2002).

Still, with regard to practical application, various unsolved problems have to be dealt with. Among other issues, these are concerned with the selection of a suitable model, i.e. with the criteria that form the basis for a user's decision as to which spatial model to prefer and whether or not to use a spatial model at all. Another point in this regard is the fact that – from a practical point of view – both the existence of spatial covariances and a spatial trend have to be taken into consideration. When selecting a suitable model, importance lies in the question of whether or not both components of spatial dependencies have to be included and whether or not the neglecting of one component has any serious consequences. Furthermore, consideration is given to the question of how practically relevant data structures with usually low numbers of replications can employ spatial models with acceptable estimation errors at all in spite of existing spatial dependencies.

In this contribution, we will examine potential consequences for the testing of hypotheses of fixed effects in such cases where a certain spatial model exists but where a different spatial model or a block design is used for the evaluation. We will analyse the kind of observable effect on the statistical Type I error in the case of a valid null hypothesis or on the power in the case of a valid alternative hypothesis.

In order to do this, we make use of a Monte Carlo simulation and use different spatial dependencies and data structures.

2 Material and Method

2.1 Analysed Spatial Models

The spatial models analysed here are the result of the outcomes of blindfold experiments (Hu, 2004) and field experiments with treatment effects (Spilke, 2005). As it turns out, the exponential model and the spherical model constitute particularly well-suited spatial approaches for large numbers of evaluations.

We now contrast these models with a model of randomized complete block design (RCBD) with discrete spatial effects in the form of random block effects.
When describing the model, we start from the well-known matrix notation for mixed linear models:

$$\underline{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\underline{\mathbf{u}} + \underline{\mathbf{e}} \qquad (1)$$

with

$\boldsymbol{\beta}$	= p × 1 vector for unknown fixed effects,
$\underline{\mathbf{u}}$	= q × 1 vector for unknown random effects,
$\underline{\mathbf{e}}$	= n × 1 vector for unknown random rest effects,
$\mathbf{X}, \mathbf{Z}$	= known experiment plan matrices for fixed and random effects respectively.

(Throughout this article, all random variables in model equations are underscored.)

Additionally, the following is assumed:

$\underline{\mathbf{u}} \sim N(0,\mathbf{G})$, $\underline{\mathbf{e}} \sim N(0,\mathbf{R})$,

$E(\underline{\mathbf{y}}) = \mathbf{X}\boldsymbol{\beta}$, $Var(\underline{\mathbf{y}}) = \mathbf{ZGZ}' + \mathbf{R} = \mathbf{V}$,

$$\mathbf{Var}\begin{pmatrix} \underline{\mathbf{y}} \\ \underline{\mathbf{u}} \\ \underline{\mathbf{e}} \end{pmatrix} = \begin{bmatrix} \mathbf{V} & \mathbf{ZG} & \mathbf{R} \\ \mathbf{GZ'} & \mathbf{G} & 0 \\ \mathbf{R} & 0 & \mathbf{R} \end{bmatrix}$$

(Henderson 1963, 1975, 1990 p. 1ff).

The differences between the models included here can be represented clearly, if in the case of a block design all random effects are described in the matrix R when assuming random block effects.
For k=1,...,r blocks, the covariance matrix of the error effects R_{RCBD} is transformed into a block-diagonal matrix, again with r blocks of the following shape:

$$R_{RCBD} = \begin{bmatrix} \sigma_R^2+\sigma_{RB}^2 \cdots & \sigma_{RB}^2 & \cdots & 0 & & 0 \\ \vdots & \vdots & & & & \\ \sigma_{RB}^2 \cdots & \sigma_R^2+\sigma_{RB}^2 & & & & 0 \\ & & \ddots & & & \vdots \\ 0 & & & \sigma_R^2+\sigma_{RB}^2 \cdots & & \sigma_{RB}^2 \\ & & & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{RB}^2 \cdots & & \sigma_R^2+\sigma_{RB}^2 \end{bmatrix}.$$

Random variables located in the same block have the same covariance regardless of the concrete spatial order, random variables not located in the same block have a covariance of zero. By contrast, covariances in the spatial model are flexibly formulated with dependence on the actual spatial distance. The covariance matrix for the error effects takes the following form:

$$R_{spatial} = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1j} & \cdots & \sigma_{1\,n-1} & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & & & & & \sigma_{2n} \\ \vdots & & \ddots & & & & \vdots \\ \sigma_{i1} & & & \sigma_{ij} & & & \sigma_{in} \\ \vdots & & & & \ddots & & \vdots \\ \sigma_{n-1\,1} & & & & & \sigma^2 & \sigma_{n-1\,n} \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nj} & \cdots & \sigma_{n\,n-1} & \sigma^2 \end{bmatrix}.$$

For the diagonal called *sill*, it holds true that $\sigma^2 = \sigma_R^2 + \sigma_0^2$, with σ_R^2 being called the *nugget* and σ_0^2 going by the name of *partial sill*. For the non-diagonal elements, it holds that $\sigma_{ij} = \sigma_0^2 f(d_{ij})$. Here, d_{ij} is the Euclidian distance between two spatial points *i* and *j*:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\,.$$

For example, the coordinates i (x_i, y_i) and j (x_j, y_j) describe the median points of 2 plots. In this example, *x* denotes the horizontal direction, and *y* denotes the vertical direction.

The function $f(d_{ij})$ is dependent on the spatial model. If ρ denotes the range parameter, the following applies to the spatial models considered in this contribution (Schabenberger and Pierce, 2002 581ff):

(1) Exponential Model

$$f(d_{ij}) = \exp(-d_{ij}/\rho),$$

(2) Spherical Model

$$f(d_{ij}) = [1 - \frac{3d_{ij}}{2\rho} + \frac{d_{ij}^3}{2\rho^3}](d_{ij} \le \rho, \text{else } 0).$$

While the covariance is 0 for $d_{ij} = \rho$ in the spherical model, it moves asymptotically against 0 in the exponential model and lies at $0.05\sigma_0^2$ for $d_{ij} \approx 3\rho$. For this reason, 3ρ is often called the *practical range* or *effective range*.
The employed models are isotropic, i.e., the covariance is independent of the direction. However, an extension to anisotropic models is possible and can be implemented at the software level as well (SAS, 2003).

2.2 Simulated Data Structures and Parameter Variants

The simulated data structures and analysed parameter variants have been derived from the analysis of Hu (2004) and Spilke (2005). This will help cover a broad range of potential application scenarios with regard to the generalisation of our results.
We analyse two data structures with a=10 and a=20 treatments respectively, and with r=3 and r=4 replications respectively. When doing so, we assume a plot size of 3m x 4m. The assignment of the treatments to the plot was performed during a single use of the SAS procedure PLAN (SAS, 2003).
The analysed parameter variants refer to partial sill=1 and range parameter (5, 20, 50) in case of no nugget effects (nugget=0); and the combinations (partial sill, nugget): (1,10), (1,1), (10,1) for range parameters (5,20,50). This results in 12 simulation variants. The proportion of the spatial structure $PSS = \sigma_0^2/(\sigma_0^2 + \sigma_R^2)$ takes the values 1 (no nugget), 10/11, 1/2 and 1/11. Each of these variants is simulated according to the exponential and the spherical model with 10,000 runs under the null and alternative hypothesis (with delta=1 for all contrasts described in more detail below). This simulation sample size will guarantee that the width of a 95% confidence interval for the empirical Type I error rate α is only 0.0086 when α=0.05. All simulations are performed using the DATA step, and the IML and MIXED procedure respectively of the SAS System (SAS, 2003).

The simulated data are then evaluated according to a block design, an exponential model, and a spherical model. This contribution thus initially assumes the existence of spatial co-variances only. Accordingly, a spatial trend does not exist. Of the 45 and 190 possible contrasts between analysis elements for a=10 and a=20 respectively, we analysed 3 contrasts each, all of which were selected by the criterion that on average, a comparatively low, medium and high spatial distance should exist over all replications of the treatments for the respective plan. Thus, the contrasts differ with regard to their median Euclidian distance $\bar{d}_{ij}$.
In particular, the following applies:

for a=10:	small (contrast 1-10)	$\bar{d}_{ij}$ =7.4m (Min 4m, Max 12.6m),
	medium (contrast 5-7)	$\bar{d}_{ij}$ =11.9m (Min 6m, Max 24.3m),
	large (contrast 3-6)	$\bar{d}_{ij}$ =16 m (Min 6m, Max 28.2m),
for a=20:	small (contrast 4-17)	$\bar{d}_{ij}$ =11.3m (Min 3m, Max 26m),
	medium (contrast 2-15)	$\bar{d}_{ij}$ =16.4m (Min 3m, Max 33.6m),
	large (contrast 9-20)	$\bar{d}_{ij}$ =21.4 m (Min 9.8m, Max 29.4m).

This approach allows for a targeted analysis of the influence of the spatial distance within an experimental design.

2.3 Evaluation of the Simulated Data

For the simulated variants and evaluation models, the observed and estimated standard errors as well as the maintenance of the nominal error for a valid null hypothesis and the power for a valid alternative hypothesis respectively are represented.
The mixed model analysis for model (1) uses the case of block design ANOVA, otherwise REML for estimating variance components and different methods for approximating the degrees of freedom. Estimable functions $\mathbf{h}'\hat{\boldsymbol{\beta}}$ of fixed effects are estimated based on $\mathbf{h}'\hat{\boldsymbol{\beta}} = \mathbf{h}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ with V being replaced by a plug-in REML estimate. Null hypotheses of the form H_0: $h'\beta = 0$ are tested by

$$t = \frac{\mathbf{h}'\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{h}'\left(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-}\mathbf{h}}} \sim t(\nu)\,. \tag{2}$$

In general, the test statistic in (2) only has approximate t-distributions and its degrees of freedom must be estimated. The approximate degrees of freedom were determined using different methods:

(1) Satterthwaite method (Satterthwaite (1941): Extended Satterthwaite method of Giesbrecht and Burns (1985) and Fai and Cornelius (1996);
(2) Kenward-Roger method (Kenward and Roger, 1997): This approximation also uses the basic idea of Satterthwaite (1941). Its extension relative to the Satterthwaite method of Giesbrecht and Burns (1985) and Fai and Cornelius (1996) is an asymptotic correction of the estimated covariance matrix of the fixed effects $(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}$ due to Kackar and Harville (1984) in unbalanced data structures.
All evaluations were executed without any parameters given. If the true model did not contain any nugget effect, no such effect was included in the evaluation model either. If starting values were given or if possible nuggets were included in general, the results might differ from those shown here.

3 Results

General remarks
We examined all data structures, model variants and approximations of the degrees of freedom both for the case of a valid null hypothesis and for the case of a valid alternative hypothesis. However, due to space restrictions, we will refrain from including all results in the subsequent section. The complete results are available in a relational database (MS Access) and may be requested from the second author.

3.1 Estimated and Observed Standard Errors of the Contrasts for Different Approximations of the Degrees of Freedom

First, this section will present the influence of the applied evaluation models on the estimated standard error of the analysed contrasts. In order to do this, the estimated and observed standard errors are calculated for each evaluation model and contrast. The estimated standard error is the root of the average of a contrast's error variances estimated per simulation run. The observed standard error results from calculated standard deviation, based on the contrast's estimation values per simulation run. The difference between the former and the latter standard error is represented percentage-wise in relation to the observed standard error. This approach allows for a good evaluation of how the respective evaluation model can correctly reflect the variability of the estimated values as part of the test statistics in use for different data structures, parameter combinations and true models. It should be noted that the differences between the estimated and observed standard errors do not depend on the parameter differences, i.e. if the alternative hypothesis is valid, the results will be the same.

The results in all tables are sorted by the values for the PSS; within identical PSS, they are sorted by range parameter in descending order. This sorting also represents a decrease in the spatial correlation.
The results for the second data structure (a=20, r=4) without correction are given in Table 1. When the exponential and the spherical model are used as simulation and evaluation model respectively, the representation does show the expected underestimation (Kackar and Harville, 1984); however, it is below 5% in most cases. Additionally, the deviations are consistently smaller for the larger amount of data (a=20, r=4), compared with the first data structure (a=10, r=4) (results not reported here). An explicit dependency on the degree of the spatial variability and on the spatial distance of the considered contrasts has not been observed when these evaluation models are applied.
When using the block design as evaluation model, this does not apply. In this case, the most explicit misestimations appear for a comparatively high spatial correlation (nugget=0, partial sill=1, range=50). Furthermore, a clear dependency of the deviation on the spatial distance of the considered treatments can be observed. Treatments with a comparatively small or high distance respectively exhibit the highest bias. Since the block design model provides the same estimation error for all considered contrasts – regardless of the spatial distance – a small spatial distance can result in a partially immense overestimation, and a large spatial distance might result accordingly in a considerable underestimation of the true standard error. For low spatial dependencies (nugget=10 and partial sill=1), however, the standard error can be estimated with only small deviations even when using the block design. In this case, differences are even lower compared to the use of the exponential and spherical model respectively. The reason becomes obvious when we consider the small sample size, since the less complex model of the block design is more advantageous here due to the smaller number of parameters to be estimated.
Aside from the approximation of the degrees of freedom based on Satterthwaite (1941), we also analysed the use of the Kenward-Roger method (Kenward and Roger, 1997) (Table 2 for a=20, r=4). This method corrects the underestimation of the standard error of the fixed effects for unbalanced data. The exploitation of the unbalancedness for the correction of the error variances of the fixed effect comes along with the fact that in our case with the block design as evaluation model, there are no differences from the Satterthwaite method. When using the exponential and the spherical model, a partially still significant reduction of the underestimation of the fixed effects standard errors can be achieved for all cases where nugget > 0. It is remarkable, though, that for the variants with nugget=0, by contrast, the correction leads to a significant overestimation for the average of all simulation runs. The reason lies in the rather large values for the standard error for a small number of simulation runs (about 3%). On closer examination, it

Table 1: Bias of the estimated standard errors for the investigated contrasts, expressed as percentage of the observed standard errors (design a=20, r=4; no correction according to Kackar and Harville (1984))

True Model			Exponential Model									Spherical Model								
Evaluation Model			Block Design			Exponential Model			Spherical Model			Block Design			Exponential Model			Spherical Model		
Spatial Distance*			s	m	l	s	m	l	s	m	l	s	m	l	s	m	l	s	m	l
Nugget	Partial Sill	Range Parameter																		
0	1	50	75.6	39.5	-44.2	0.8	1.3	0	-1.8	-1.1	-2.2	82.9	43.7	-45.8	1.1	1.2	0.1	-1.0	-0.3	-1.3
0	1	20	62.0	32.9	-39.1	0.9	1.0	-.1	-3.2	-2.4	-3.2	60.7	29.3	-32.6	3.0	4.1	2.0	-1.5	-0.2	-1.5
0	1	5	27.6	15.6	-17.8	-0.1	0.4	-1.0	-6.6	-6.1	-7.3	5.7	5.2	-0.4	-0.1	0.5	-1.0	-8.2	-7.1	-7.1
1	10	50	33.8	21.2	-36.1	-2.9	-3.5	-2.5	-3.0	-2.6	-3.3	46.2	28.3	-40.5	-3.3	-3.0	-2.6	-2.9	-2.7	-2.8
1	10	20	40.3	23.7	-34.7	-2.0	-1.8	-1.7	-2.0	-2.2	-3.0	45.7	24.0	-30.0	-0.7	0	-0.5	-2.0	-1.4	-2.1
1	10	5	23.3	13.5	-16.2	0	-0.1	-1.4	-0.4	-0.5	-2.9	5.2	4.8	-0.3	-0.5	0.1	-1.3	0.5	0.1	-2.1
1	1	50	5.6	4.3	-13.8	-1.9	-2.8	-5.5	-1.1	-1.9	-3.7	9.4	7.0	-20.2	-2.3	-2.7	-5.0	-1.7	-2.2	-3.0
1	1	20	9.8	6.9	-17.5	-2.4	-2.8	-4.2	-1.0	-1.7	-3.5	14.4	9.2	-17.3	-2.6	-2.7	-2.7	-2.0	-2.0	-2.8
1	1	5	9.6	6.3	-9.0	-1.4	-1.9	-2.7	0.1	-1.0	-2.4	2.6	2.6	0.1	-1.5	-1.2	-1.6	0.1	-0.2	-1.8
10	1	50	0.4	0.7	-1.6	-2.3	-2.1	-2.9	-2.4	-2.3	-2.9	0.8	1.0	-3.1	-2.1	-2.3	-3.8	-2.1	-2.5	-3.1
10	1	20	0.9	1.0	-2.6	-2.3	-2.3	-3.4	-2.2	-2.4	-3.3	1.6	1.5	-3.0	-2.1	-2.3	-3.4	-2.4	-1.7	-3.2
10	1	5	1.2	1.2	-1.4	-2.3	-2.1	-2.6	-1.8	-2.4	-3.2	0.2	0.7	0.3	-2.4	-1.9	-1.4	-1.7	-1.9	-2.0

* s=small, m=medium, l=large

Table 2: Bias of the estimated standard errors for the investigated contrasts, expressed as percentage of the observed standard errors (design a=20, r=4; correction according to Kackar and Harville (1984))

True Model			Exponential Model									Spherical Model								
Evaluation Model			Block Design			Exponential Model			Spherical Model			Block Design			Exponential Model			Spherical Model		
Spatial Distance*			s	m	l	s	m	l	s	m	l	s	m	l	s	m	l	s	m	l
Nugget	Partial Sill	Range Parameter																		
0	1	50	75.6	39.5	-44.2	>100	>100	>100	>100	>100	>100	82.9	43.7	-45.8	>100	>100	>100	>100	>100	>100
0	1	20	62.0	32.9	-39.1	>100	>100	>100	>100	>100	>100	60.7	29.3	-32.6	>100	>100	>100	>100	>100	>100
0	1	5	27.6	15.6	-17.8	>100	>100	>100	>100	>100	>100	5.7	5.2	-0.4	1.3	2.4	2.1	>100	>100	>100
1	10	50	33.8	21.2	-36.1	-1.9	-2.0	-0.7	-1.7	-0.6	-1.0	46.2	28.3	-40.5	-2.4	-1.7	-1.2	-1.5	-0.6	-0.6
1	10	20	40.3	23.7	-34.7	-0.9	-0.3	0	-0.4	0.1	-0.5	45.7	24.0	-30.0	0.2	1.3	1.0	0	1.1	0.9
1	10	5	23.3	13.5	-16.2	1.1	1.5	0.8	1.5	2.0	0.1	5.2	4.8	-0.3	0.7	1.9	1.9	1.7	1.6	0.6
1	1	50	5.6	4.3	-13.8	-0.8	-1.2	-2.4	-0.3	-0.4	-0.8	9.4	7.0	-20.2	-1.3	-1.0	-2.0	-0.8	-0.5	-0.2
1	1	20	9.8	6.9	-17.5	-1.3	-1.1	-1.1	0.1	0.1	-0.6	14.4	9.2	-17.3	-1.3	-0.8	0.2	-0.6	0.1	0.3
1	1	5	9.6	6.3	-9.0	-0.1	0.0	0.7	1.4	0.9	0.8	2.6	2.6	0.1	-0.3	0.5	1.5	1.0	0.9	0.7
10	1	50	0.4	0.7	-1.6	-1.4	-0.9	-0.6	-1.8	-1.3	-0.5	0.8	1.0	-3.1	-1.2	-1.0	-1.3	-1.4	-1.4	-0.5
10	1	20	0.9	1.0	-2.6	-1.3	-1.0	-1.0	-1.4	-1.2	-0.6	1.6	1.5	-3.0	-1.1	-0.9	-0.6	-1.6	-0.3	-0.2
10	1	5	1.2	1.2	-1.4	-1.3	-0.8	0	-1.0	-1.2	-0.7	0.2	0.7	0.3	-1.5	-0.7	0.9	-1.1	-1.1	-0.1

* s=small, m=medium, l=large

becomes obvious that these are cases with particularly high estimates for the range parameter and with large estimation errors for partial sill and range. These have an important impact on the correction and lead to extremely high values for the corrected standard errors of the estimated contrasts.
In cases where nugget > 0, this does not happen, since the variance components for partial sill and nugget can apparently be separated with less difficulty there, despite the fact that an additional parameter has to be estimated compared to cases where nugget = 0.

3.2 Validity of the Null Hypothesis – Maintenance of the Nominal Type I Error

The results of the hypothesis test with the null hypothesis valid are summarized in Tables 3 and 4.
When using the block design as evaluation model, the explicit impact of the spatial distance on the respective contrast becomes obvious. For contrasts whose analysis shows low spatial distances, the empirical Type I error is not exhausted; for contrasts with high spatial distances, it is occasionally exceeded considerably. This is especially true for the plan a=20, since the spatial dependencies have a particularly strong influence here due to the larger number of plots. As a result, an extreme bias can be observed for this plan for all cases with a significant spatial dependence (PSS=1/2...1). Only when the nugget variance is relatively high compared to the partial sill (nugget=10, PSS=1/11) can we assume an acceptable maintenance of the nominal Type I error when using the block design as evaluation model.
Considerable improvements for the preservation of the nominal error can be achieved when using a spatial model as evaluation model. For both plans, however, a tendency towards overestimation cannot be ignored. However, the deviation from the nominal Type I error is much smaller for the plan with a=20 compared to the plan with a=10.
When using a spatial evaluation model, no apparent systematic differences between contrasts with low, medium and high spatial distance can be observed.
Due to the relatively precise match between observed and estimated standard errors (see Section 3.1) and due to the partially extreme impact of the correction based on Kackar and Harville (1984), the use of corrections does not provide any advantage for the experiment plans presented here. Hence we refrain from presenting here the results of the hypothesis test for the use of the approximation of the degrees of freedom based on Kenward and Roger (1997). The authors nonetheless have these results available.

Table 3: Empirical Type I errors for design a=10 and r=3 and all investigated variants for the t-Test under the null hypothesis (nominal Type I error = 0.05)

True Model			Exponential Model									Spherical Model								
Evaluation Model			Block Design			Exponential Model			Spherical Model			Block Design			Exponential Model			Sperical Model		
Spatial Distance*			s	m	l	s	m	l	s	m	l	s	m	l	s	m	l	s	m	l
Nugget	Partial Sill	Range Parameter																		
0	1	50	0.013	0.028	0.080	0.046	0.049	0.049	0.054	0.054	0.056	0.011	0.025	0.081	0.045	0.048	0.051	0.050	0.051	0.053
0	1	20	0.015	0.033	0.073	0.046	0.047	0.048	0.058	0.057	0.060	0.014	0.036	0.056	0.040	0.042	0.042	0.052	0.052	0.052
0	1	5	0.031	0.052	0.062	0.052	0.058	0.056	0.068	0.074	0.076	0.051	0.056	0.053	0.060	0.062	0.060	0.067	0.069	0.069
1	10	50	0.023	0.034	0.073	0.064	0.064	0.067	0.064	0.066	0.069	0.019	0.029	0.074	0.058	0.062	0.062	0.061	0.067	0.062
1	10	20	0.021	0.035	0.069	0.057	0.060	0.058	0.061	0.065	0.064	0.018	0.036	0.055	0.050	0.055	0.050	0.054	0.059	0.057
1	10	5	0.033	0.053	0.060	0.055	0.062	0.061	0.062	0.072	0.068	0.051	0.055	0.053	0.059	0.063	0.059	0.067	0.066	0.066
1	1	50	0.041	0.046	0.059	0.060	0.064	0.070	0.062	0.067	0.072	0.036	0.043	0.062	0.060	0.065	0.071	0.063	0.068	0.075
1	1	20	0.036	0.045	0.060	0.059	0.065	0.071	0.062	0.069	0.069	0.032	0.045	0.055	0.059	0.065	0.068	0.064	0.070	0.069
1	1	5	0.039	0.051	0.055	0.057	0.064	0.066	0.061	0.072	0.070	0.049	0.053	0.051	0.058	0.062	0.060	0.061	0.065	0.061
10	1	50	0.049	0.051	0.050	0.058	0.061	0.061	0.064	0.066	0.064	0.047	0.050	0.051	0.058	0.061	0.062	0.064	0.066	0.065
10	1	20	0.047	0.050	0.050	0.058	0.060	0.061	0.062	0.065	0.064	0.046	0.051	0.049	0.058	0.060	0.060	0.063	0.066	0.065
10	1	5	0.047	0.052	0.049	0.058	0.061	0.060	0.063	0.065	0.064	0.048	0.052	0.049	0.058	0.060	0.059	0.063	0.065	0.065

* s=small, m=medium, l=large

Table 4: Empirical Type I errors for design a=20 and r=4 and all investigated variants for the t-Test under the null hypothesis (nominal Type I error = 0.05)

True Model			Exponential Model									Spherical Model								
Evaluation Model			Block Design			Exponential Model			Spherical Model			Block Design			Exponential Model			Spherical Model		
Spatial Distance*			s	m	l	s	m	l	s	m	l	s	m	l	s	m	l	s	m	l
Nugget	Partial Sill	Range Parameter																		
0	1	50	0.003	0.010	0.261	0.049	0.046	0.046	0.055	0.052	0.052	0.003	0.009	0.280	0.049	0.046	0.046	0.053	0.049	0.050
0	1	20	0.005	0.013	0.212	0.050	0.046	0.046	0.058	0.055	0.055	0.004	0.014	0.171	0.045	0.040	0.042	0.054	0.050	0.051
0	1	5	0.015	0.022	0.099	0.051	0.048	0.051	0.070	0.065	0.065	0.040	0.040	0.049	0.050	0.046	0.051	0.076	0.073	0.068
1	10	50	0.015	0.022	0.189	0.056	0.057	0.053	0.056	0.055	0.057	0.010	0.017	0.221	0.058	0.056	0.054	0.055	0.055	0.055
1	10	20	0.011	0.018	0.177	0.055	0.052	0.052	0.054	0.054	0.056	0.008	0.017	0.156	0.051	0.048	0.048	0.053	0.053	0.054
1	10	5	0.018	0.026	0.092	0.048	0.047	0.052	0.052	0.051	0.054	0.041	0.041	0.048	0.050	0.047	0.051	0.048	0.050	0.052
1	1	50	0.041	0.042	0.084	0.053	0.054	0.062	0.052	0.053	0.055	0.036	0.037	0.108	0.054	0.055	0.060	0.051	0.054	0.056
1	1	20	0.035	0.037	0.099	0.052	0.056	0.058	0.050	0.055	0.056	0.028	0.034	0.096	0.054	0.053	0.054	0.054	0.054	0.055
1	1	5	0.034	0.037	0.071	0.052	0.052	0.054	0.050	0.054	0.053	0.044	0.043	0.046	0.051	0.052	0.050	0.047	0.049	0.052
10	1	50	0.047	0.047	0.051	0.054	0.055	0.054	0.054	0.053	0.053	0.047	0.047	0.054	0.053	0.055	0.055	0.053	0.053	0.054
10	1	20	0.047	0.046	0.053	0.054	0.055	0.055	0.054	0.053	0.054	0.046	0.045	0.055	0.054	0.054	0.054	0.055	0.053	0.051
10	1	5	0.046	0.046	0.050	0.055	0.054	0.053	0.053	0.056	0.053	0.047	0.046	0.048	0.055	0.054	0.050	0.053	0.055	0.052

* s=small, m=medium, l=large

3.3 Validity of the Alternative Hypothesis – Analysis of the Power

In the previous section, a partially extreme bias of the nominal Type I error was shown in particular for the use of the block design. In order to arrive at a comparative evaluation of the power, this effect has to be excluded, since the power will increase accordingly whenever the nominal Type I error is exceeded. Therefore, the power shown in Table 5 is corrected by the bias of the nominal Type I error for the given validity of the null hypothesis in the form: corrected power = power (t-quantile) – [Type I error rate (empirical) - Type I error rate (nominal)]. This approach is not exact but appears to be sufficient for the conclusions to be drawn here.

The effects of this approach shall be illustrated with an example (exponential model: plan a=20, without nugget, range=5, partial sill=1). If the null hypothesis is valid, the empirical Type I error when using the t-quantiles and applying the exponential model is 0.051 for contrast 1 but 0.099 for the block design (Table 4, Figure 1). In other words, when using the exponential model as evaluation model, the assumed t-distribution applies, and the t-quantiles and the empirical quantiles achieved in the simulation (exponential model, empirical quantiles) match. When using the block design, however, this does not apply (block design, t-quantiles). However, the matching quantiles for the nominal error of 0.05 can be determined when using the simulated data (block design, empirical quantiles). These empirical quantiles allow for correction and hence the determination of the actual power.

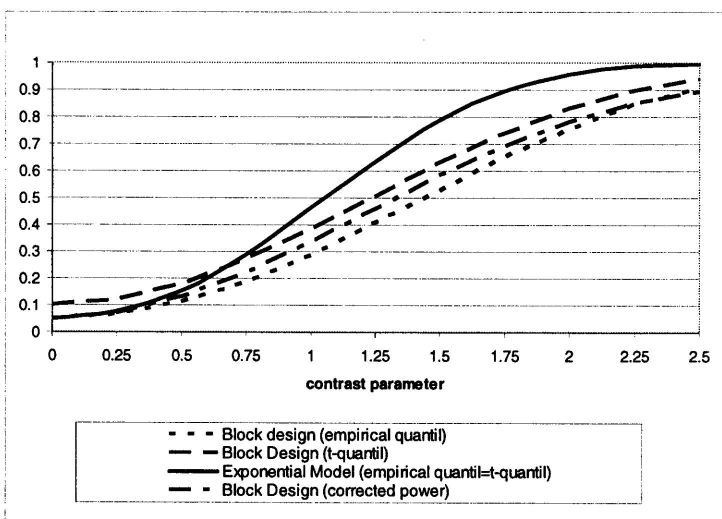

Figure 1: Power for the Exponential Model and Block Design (contrast with large spatial difference, True Model: Exponential Model (Partial Sill=1, Range=5, Nugget=0))

Table 5: Corrected power for design a=20 and r=4 and all investigated variants for the t-Test under the alternative hypothesis (nominal Type I error = 0.05)

True Model			Exponential Model									Spherical Model								
Evaluation Model			Block Design			Exponential Model			Spherical Model			Block Design			Exponential Model			Spherical Model		
Spatial Distance*			s	m	l	s	m	l	s	m	l	s	m	l	s	m	l	s	m	l
Nugget	Partial Sill	Range Parameter																		
0	1	50	0.997	0.967	0.566	1.000	1.000	1.000	0.995	0.998	0.998	0.840	0.815	0.440	1.000	1.000	0.998	0.997	1.000	0.995
0	1	20	0.709	0.694	0.445	0.985	0.984	0.934	0.979	0.978	0.935	0.446	0.461	0.365	0.912	0.906	0.792	0.917	0.910	0.807
0	1	5	0.350	0.365	0.334	0.568	0.563	0.466	0.602	0.592	0.484	0.297	0.297	0.296	0.330	0.332	0.304	0.371	0.359	0.299
1	10	50	0.115	0.128	0.127	0.217	0.207	0.181	0.214	0.213	0.180	0.091	0.101	0.105	0.192	0.191	0.159	0.194	0.186	0.163
1	10	20	0.085	0.095	0.097	0.157	0.155	0.135	0.160	0.158	0.136	0.070	0.078	0.079	0.129	0.130	0.114	0.138	0.132	0.113
1	10	5	0.069	0.076	0.077	0.094	0.098	0.085	0.097	0.094	0.086	0.070	0.073	0.072	0.076	0.078	0.074	0.078	0.073	0.076
1	1	50	0.251	0.254	0.248	0.286	0.280	0.258	0.284	0.280	0.263	0.227	0.234	0.229	0.280	0.270	0.248	0.283	0.271	0.247
1	1	20	0.217	0.223	0.220	0.270	0.263	0.240	0.269	0.254	0.238	0.189	0.194	0.193	0.253	0.253	0.221	0.253	0.251	0.223
1	1	5	0.179	0.184	0.187	0.216	0.212	0.197	0.215	0.209	0.196	0.170	0.171	0.177	0.181	0.179	0.178	0.180	0.175	0.175
10	1	50	0.076	0.075	0.075	0.076	0.075	0.077	0.079	0.076	0.077	0.074	0.075	0.075	0.077	0.075	0.078	0.076	0.075	0.078
10	1	20	0.074	0.075	0.075	0.076	0.075	0.078	0.079	0.075	0.078	0.072	0.076	0.075	0.077	0.076	0.078	0.078	0.075	0.079
10	1	5	0.073	0.074	0.073	0.077	0.075	0.077	0.075	0.071	0.078	0.074	0.074	0.072	0.076	0.073	0.075	0.077	0.072	0.076

* s=small, m=medium, l=large

Figure 1 illustrates the difference between the power based on t-quantiles and on empirical quantiles, both of which have been derived from simulations with various parameter differences. As becomes clear, the corrected power does not match the empirical quantiles but represents the actual power when using the block design much better than would be the case when applying the t-quantiles.
The results show this for both plans. In Table 5 the results are reported for a=20, r=4; results for a=10, r=3 are not reported here. The application of the spatial models realizes a considerably higher power. Again, a comparable power for the block design can be realized only for cases with a high nugget variance (nugget = 10) and with relatively low significance of the spatial covariance (PSS=1/11).

4 Discussion and Conclusions for Practical Evaluations

The great importance of the spatial covariances to be expected from a logical and subject-related perspective is confirmed in a variety of experiments. Accordingly, an adjustment of the evaluation models is necessary.
The analysis of the use of a true model and evaluation models different from it, as presented in this paper, shows a considerable influence on the maintenance of the nominal Type I error rate and on the power. If a block design is used for existing spatial covariances, the nominal Type I error rate is expected not to be reached or to be exceeded considerably, depending on the spatial distance of the observations. In contrast, the spatial distance can be considered in a flexible manner when using a spatial model.
If the bias of the Type I error rate is corrected during the power analysis, a considerable loss becomes apparent for the corrected power compared with the observed power when using the block design.
The cross-wise use of the included spatial models (exponential and spherical) as evaluation models, however, exhibits a rather low influence compared to the respective true model for the validity of both the null hypothesis and the alternative hypothesis.
Additionally, when evaluating with the SAS procedure MIXED, the approximation of the degrees of freedom based both on Satterthwaite (1941) and on Kenward and Roger (1997) was examined. While the Kenward-Roger method is recommended for unbalanced data and achieves a significant improvement in the preservation of the nominal Type I error rate both for alpha-designs (Spilke et al., 2004) and for block, split-plot and strip plot designs with missing values (Spilke et al., 2005), this does not always apply to the spatial models examined in this paper. Apparently, the standard errors of the parameters to be estimated for the spatial functions are estimated without sufficient precision, with the consequence that the subsequent correction of the standard errors of the fixed

effects occasionally delivers unrealistically high values and cannot be recommended for application. Our results agree with Schaalje et al. (2002). They investigated repeated-measures designs with five covariance structures. In their simulation study, the Kenward-Roger method worked as well as or better than the Satterthwaite method in all situations and produced Type I error rates close to the nominal values in case of Compound Symmetry and Toeplitz structures. When the covariance structure became more complex (First-Order-Antedependence), even the Kenward-Roger method had problems and produced inflated error rates. Thus the Kenward-Roger method should be used with particular caution when the random part of the model does not have a simple structure.

Overall, the results demonstrate that the common block design does not represent an adequate analysis model in the presence of important spatial covariances. The use of this analysis model in scenarios with high spatial covariances leads to considerable bias of the Type 1 error or to a loss of power. Taking into account the fact that spatial covariance in practical applications may be quite different, the choice of the analysis model is of fundamental importance. The experiment designer needs to be provided with a decision-support tool in order to arrive at a justified selection of a model. The results at hand are thus the starting point for further research with the main goal being to define the criteria required for a sufficiently justified model selection. Test-based methods are discussed in Stroup (2002) and Cederkvist et al., (2005). Further options arise from the use of analytical criteria such as the Akaike criterion and its modifications (Akaike, 1996; Hurvich and Tsai, 1989) or the ICOMP criterion (Bozdogan, 2000). Methodical experiments on comparison of the effectiveness of these methods are urgently needed and are the subject of current research by the authors.

Acknowledgement

We appreciate the referee's comments, which have led to a considerable number of improvements.

REFERENCES

Akaike, H. (1969): Fitting autoregressive models for prediction, Annals of the Institute of Statistical Mathematics 21: 243-247.

Bozdogan, H. (2000): Akaike's information criteria and resent developments in information criteria. Journal of Mathematical Psychology 44: 62-91.

Cederkvist, H.R.; Aasrveit, A.H.; Næs, T. (2005): A comparison of methods for testing differences in predictive ability. Journal of Chemometrics 19: 500-509.

Fai, A.H.T.; Cornelius, P.L., 1996: Approximate F-Tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. J. Statist. Comput. Simul. 54: 363-378.

Federer, W.T, 1998: Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle designed experiments. Biometrics 54: 471-481.

Giesbrecht, F.G.; Burns, J.C., 1985: Two-stage analysis based on a mixed model: large-sample asymptotic theory and small-sample simulation results. Biometrics 41: 477 - 486.

Gilmour, A.R.; Cullis, B.R.; Verbyla, A.P., 1997: Accounting for natural and extraneous variation in the analysis of field experiments. Journal of Agricultural, Biological, and Environmental Statistics 56: 269-293

Henderson, C.R., 1963: Selection index and expected genetic advance. Statistical Genetics and Plant Breeding, NAS-NRC Publ. No.982, 141 - 163.

Henderson, C.R., 1975: Best linear estimation and prediction under a selection model. Biometrics 31: 423 - 447.

Henderson, C.R., 1990: Statistical method in animal improvement: historical overview. In: Advances in Statistical Methods for Genetic Improvement of Livestock. Springer-Verlag, New York.

Hu, X. (2004): Analysis of two blindfold experiments for spatial covariances. Unpublished paper.

Hurvich, C.M.; Tsai, C.L. (1989): Regression and time series model selection in small samples, Biometrika 76, 297-397.

Kackar, A. N.; Harville, D. A., 1984: Approximation for standard errors of estimators of fixed and random effects in mixed linear models. Journal of American Statistical Association, 79: 853-861.

Kenward, M.G.; Roger, J.H., 1997: Small Sample Inference for fixed effects from restricted maximum likelihood. Biometrics 53: 983-997.

Littell, R. C.; Milliken, G. A.; Stroup, W. W.; Wolfinger, R. D., 1996: SAS system for mixed models. SAS Institute Inc. New York.

McLean, R.A.; Sanders, W.W.; Stroup, W.W., 1991: A unified approach to mixed linear model The American Statistician 45: 54-64

Richter, C.; Kroschewski, B.: Kombination von Block- und räumlicher Struktur bei der Auswertung von randomisierten Blockanlagen. Landwirtschaftliches Versuchswesen. Frühjahrstagung Osnabrück März 2002. Martin-Luther-Universität Halle-Wittenberg S.16-36.

Satterthwaite, F.E., 1941: Synthesis of variance. Psychometrika 6: 309-316.

Schaalje, G. B., McBride, J. B., and Fellingham, G. W., 2002: Adequacy of approximation to distributions of test statistics in complex mixed linear models. Journal of Agricultural, Biological and Ecological Statistics, 7: 512-524.

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

SAS Institute Inc., 2003: SAS Version 9.1.3, Cary, NC: SAS Institute Inc.

Searle, S.R.; Casella, G.; McCulloch, C.E., 1992: Variance Components. John Wiley and Sons, Inc. New York.

Searle, S.R., 1971: Topics in variance component estimation. Biometrics 27: 1 - 76.

Stroup, W.W.; Baenziger, P.S.; Mulitze, D.K., 1994: Removing spatial variation from wheat yield trials: A comparison of methods. Crop Science 86 62-66.

Spilke, J.; Piepho, H.-P.; Meyer, U. (2004): Approximating the degrees of freedom for contrasts of genotype laid out as subplots in an alpha-design in a split-plot experiment. Plant Breeding 123: 193-197.

Spilke, J. (2005): Spatial analysis of data from field experiments. Unpublished paper.

Spilke, J.; Piepho, H.-P. Hu, X., 2005: A simulation study on tests of hypotheses and confidence intervals for fixed effects in mixed models for blocked experiments with missing data. Journal of Agricultural, Biological, and Environmental Statistics 10: 374-389.

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 Journal of Agricultural, Biological and Environmental Statistics 7: 491-511.

Verbeke, G.; Molenberghs, G., 1997: Linear mixed models in practice – a SAS-oriented approach. Springer-Verlag, New York.

Zimmerman, D.L.; Harville, D.A., 1991: A Random Field Approach to the Analysis of Field-Plot Experiments and other Spatial Experiments. Biometrics 47: 223-23.