CBCS
HTTP://AGROBIOL.SGGW.WAW.PL/CBCS

COMMUNICATIONS IN BIOMETRY AND CROP SCIENCE
VOL. 10, NO. 1, 2015, PP. 3–16

INTERNATIONAL JOURNAL OF THE FACULTY OF AGRICULTURE AND BIOLOGY,
WARSAW UNIVERSITY OF LIFE SCIENCES – SGGW, POLAND

REGULAR ARTICLE

Problems in parameter estimation for power and AR(1) models of spatial correlation in designed field experiments

Hans-Peter Piepho [1]*, Jens Möhring [1], Markus Pflugfelder [2], Winfried Hermann [2], Emlyn R. Williams [3]

[1] Institute for Crop Science, Biostatistics Unit, University of Hohenheim, Stuttgart, Germany.
[2] Agricultural Experimental Station, University of Hohenheim, Stuttgart, Germany.
[3] Statistical Consulting Unit, The Australian National University, Canberra ACT 0200, Australia.
*Corresponding author: Hans-Peter Piepho; E-mail: piepho@uni-hohenheim.de

CITATION: Piepho H.P., Möhring J., Pflugfelder M., Hermann W., Williams E.R. (2015). Problems in parameter estimation for power and AR(1) models of spatial correlation in designed field experiments. *Communications in Biometry and Crop Science* 10, 3–16.

Received: 26 September 2014, Accepted: 13 January 2015, Published online: 23 January 2015

ABSTRACT

The AR(1) and power models of spatial correlation are popular in the analysis of field trial data. Numerical difficulties in estimation and interpretation of these models may occur when the autocorrelation parameter ρ tends to either zero or unity. These problems are considered here using three different examples. The first example is based on simulated data for a partially replicated design, where the true underlying variance-covariance structure is known. The other two examples involve real data from a precision farming trial and a plant breeding trial. We suggest four options to deal with the observed numerical problems and illustrate their use with the examples. It is shown in the examples that re-scaling of the spatial coordinates or a re-parameterization of the AR(1) model as an exponential model can be useful to help the model converge. We conclude that individual parameter estimates of the AR(1) model should be interpreted with care, especially when the autocorrelation estimate is close to either zero or unity.

Key Words: *precision farming; convergence problems; autoregressive model; autocorrelation; partially replicated design; linear variance model.*

INTRODUCTION

Spatial variance-covariance structures embedded in a mixed model framework have become popular in the analysis of field trials (Gilmour et al. 1997, Piepho et al. 2008, Gonçalves et al. 2010). Most common spatial models assume a non-linear decay of the spatial correlation with distance. In a randomized experiment with a single observation per plot, a spatial correlation may be assumed between plots in the same block (Williams 1986). Similarly, in on-farm trials with large plots, many spatially referenced measurements (i.e.,

measurements for which spatial coordinates are recorded) may be available per plot, with associated covariate information. A full analysis of such trials requires accounting for the spatial variation among repeated measurements in the same plot (Piepho et al. 2011).

A very popular non-linear spatial model for field trials with equally spaced plots is the first-order autoregressive [AR(1)] model. The model implies an exponential decay of the correlation with spatial distance, measured in terms of the lag distance between plots, equal to the number of intermittent plots plus one. The model can also be extended into two dimensions, the most popular variant being the separable AR(1) ⊗ AR(1) model (Gilmour et al., 1997).

When repeated observations are available per plot, as is often the case in precision farming experiments, observations will usually be irregularly spaced, not forming an exact rectangular grid (Piepho et al. 2011). The AR(1) model is not applicable with such experiments because it requires equal spacing. However, the AR(1) model can be readily extended to unequally spaced measurements, as either the so-called power model or the exponential model (Gotway and Schabenberger 2004). These two models are different parameterizations of essentially the same model (see below).

Frequently, numerical problems are encountered when trying to fit non-linear spatial models to field trial data based on residual maximum likelihood (REML), using various statistical packages offering mixed-model analysis. For example, the REML algorithm may fail to converge, or the final Hessian matrix may fail to be non-positive definite. In this paper we consider such problems with AR(1) and power models and look at possible remedies to achieve convergence to a proper solution. Interpretation of parameter estimates is also considered. Three different examples are used to illustrate typical problems met in practice, and possible solutions are discussed. The first example is based on simulated data for a partially replicated (p-rep) design (Cullis et al. 2006, Williams et al. 2011, 2014), whereas the other two examples involve real data from a precision farming trial and a plant breeding trial.

The paper is organized as follows. First we introduce the three examples. Next, we describe the spatial models. This is followed by a brief description of potential problems when trying to fit these models, possible measures for trouble shooting in case of convergence problems, and an application to the three examples. The paper ends with a brief discussion and our recommendations.

DESCRIPTION OF EXAMPLES

Example 1: We generated a p-rep design for 300 treatments, five locations and a final block size of nine, where three plots per block are devoted to replicated treatments. This was done using the software CycDesigN 5.1 (VSN International; http://www.vsni.co.uk/), which uses a combination of theory, based on upper bounds for the average efficiency factor, and numerical search to generate an optimal or near-optimal design (John and Williams 1995, Williams et al. 2014). Data were simulated according to this design, generating random effects for blocks and plot errors, i.e., there was no spatial correlation within blocks. These data were generated to put spatial models to a hard test in that the true spatial correlation was zero within blocks, which is likely to cause numerical problems of convergence when trying to fit models with a spatial correlation. Plots of each block were assumed to be arranged as a single column of plots. Thus the AR(1) model for correlations among plots in a block can be applied, and its properties can be studied in the limiting case of no real spatial correlation within blocks.

Example 2: An on-farm oilseed rape trial was conducted at the location Ihinger Hof (Universität Hohenheim, Agricultural Experimental Station, Germany). The experiment was laid out in randomized complete blocks with six replicates. On each plot, a large number of yield measurements were recorded using an online monitoring system (e.g., 252

measurements on the first plot), together with the spatial coordinates (degrees WGS84). The treatment factors were tillage method (three levels) and fertilization method (two levels). Five of the six factorial treatment combinations were tested. At each spatial position, a fertility score, called "Ackerzahl", was also recorded. This score assesses the soil quality and subsumes several environmental factors such as soil type, climatic conditions, and landscape characteristics (Schachtschabel et al. 1976). The dataset had a total of 4569 observations.

The analysis aimed to compare the five treatments using the fertility score as a covariate. It was thus necessary to model the covariance among repeated recordings per plot. To illustrate potential problems with spatial models, here we report results of such analyses without this covariate. A complete analysis will be published elsewhere. The measurements within a plot are irregularly spaced so an AR(1) model cannot be used. Instead we will use an isotropic power model for the correlation of measurements within the same plot; the model is described below. Observations from different plots are considered independent because plots are randomization units. Randomization units are experimental units to which levels of a treatment factor are randomly assigned, such as plots in a randomized block experiment or main plots in a split-plot experiment.

Example 3: We consider a wheat trial from South Australia used in Gilmour et al. (1997). The trial comprises 107 varieties sown in three replicates (a replicate was a complete block). Three of these varieties were sown twice per replicate so each replicate had 110 plots arranged as a regular two-dimensional grid of five columns and 22 rows. Gilmour et al. (1997) do not give details of the randomization scheme applied within replicates. From their Table 1 it appears that there was some systematic arrangement in the first replicate, so the allocation of entries to plots within replicates does not seem to have been entirely at random. The example is used here to study the behaviour of various one- and two-dimensional AR(1) models.

THE POWER AND AR(1) MODELS

First, in Example 1 we will develop models for a single column (block) of equally spaced plots. Then, in Example 2 we will discuss extension to isotropic models for two dimensions with possibly unequally spaced data, in which the mode of covariance decay is the same in each spatial direction. Finally, in Example 3 we will consider an anisotropic extension of the AR(1) model to two-dimensional regular grids.

The randomisation-based model: For simplicity of exposition, let us consider a single (say, the *j*-th) block of size *k*. For the *j*-th block the error terms for the plots are denoted as e_{ij} $(i = 1,...,k)$ and may be collected into a vector $\mathbf{e}_j = (e_{1j},...,e_{kj})^T$. Under a randomization-based model, the errors are assumed to be identically independently distributed (i.i.d.), so that

$$\text{var}(\mathbf{e}_j) = \sigma^2 \mathbf{I}_k, \quad (1)$$

where σ^2 is the error variance and $\mathbf{I}_k$ is a *k*-dimensional identity matrix.

In a randomization-based model we also need to account for a block effect b_j. Thus, ignoring treatment effects for simplicity, the observed data y_{ij} can be modelled as

$$y_{ij} = b_j + e_{ij}. \quad (2)$$

Collecting the observed data y_{ij} of the *j*-th block into a random vector $\mathbf{y}_j = (y_{1j},..., y_{kj})^T$ and taking block effects as random with variance σ_b^2, we have the following variance-covariance structure for the data:

$$\text{var}(\mathbf{y}_j) = \sigma_b^2 \mathbf{J}_k + \sigma^2 \mathbf{I}_k, \quad (3)$$

where $\mathbf{J}_k$ is a $k \times k$ matrix of ones.

The power model without nugget: Instead of independent errors, we may assume that errors are serially correlated according to the power model (Gotway and Schabenberger, 2004). This has the variance-covariance structure

$$\text{var}(\mathbf{e}_j) = \sigma_s^2 \{\rho^{d(i,i')}\} = \sigma_s^2 \mathbf{\Omega}, \quad (4)$$

where ρ $(0 < \rho < 1)$ is the autocorrelation parameter, $d(i,i')$ is the Euclidean distance between plots i and i' measured in either one of two dimensions, and σ_s^2 is the spatial variance. The variance of the data is then given by

$$\text{var}(\mathbf{y}_j) = \sigma_b^2 \mathbf{J}_k + \sigma_s^2 \mathbf{\Omega}. \quad (5)$$

When the k plots are equally spaced down a single column (Example 1), the distance can be assessed based on plot indices i, and the distance takes the form $d(i,i') = |i - i'|$. The correlation matrix is then given by

$$\mathbf{\Omega} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{k-1} \\ \rho & 1 & \rho & \cdots & \rho^{k-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{k-1} & \rho^{k-2} & \rho^{k-3} & \cdots & 1 \end{pmatrix}. \quad (6)$$

This model is known as the autoregressive first-order [AR(1)] model, which is a special case of the power model.

When plots are unequally spaced and the coordinates are possibly two-dimensional, the more general definition $\mathbf{\Omega} = \{\rho^{d(i,i')}\}$ can still be applied if isotropy is assumed. This also applies when repeated measurements are taken on the same plot using online monitoring systems, as is common in precision-farming experiments (Piepho et al. 2011). In this case, we can identify the random effect b_j in the basic model (2) with the j-th plot, assuming that plots are the randomization units of the experiment whereas errors e_{ij} represent the spatially referenced repeated measurements taken within the same plot (Example 2).

The power model with nugget: In addition to the spatially correlated term, we may add an independent term (nugget), leading to this model:

$$\text{var}(\mathbf{e}_j) = \sigma_s^2 \mathbf{\Omega} + \sigma^2 \mathbf{I}_k. \quad (7)$$

The nugget term may capture true measurement error as well as random small-scale heterogeneity of soil fertility. The variance of the data is equal to

$$\text{var}(\mathbf{y}_j) = \sigma_b^2 \mathbf{J}_k + \sigma_s^2 \mathbf{\Omega} + \sigma^2 \mathbf{I}_k. \quad (8)$$

PROBLEMS WITH FITTING THE POWER MODEL

If the serial correlation ρ converges to unity, the spatial component of power model converges to $\sigma_s^2 \mathbf{\Omega} = \sigma_s^2 \mathbf{J}_k$, in which case the spatial component and the random block effect are confounded. Conversely, if the correlation ρ converges to zero, the spatial component converges to $\sigma_s^2 \mathbf{\Omega} = \sigma_s^2 \mathbf{I}_k$ and hence is confounded with the nugget term $\sigma^2 \mathbf{I}_k$. In both these situations, convergence problems may arise due to confounding of the correlation with other parameters of the variance-covariance structure, as has been reported repeatedly (e.g., Piepho and Williams 2010, Piepho et al. 2013, Richter et al. 2015). Illustrations of these problems will be given in the examples.

It is worth pointing out that problems similar to those described here for the AR(1) and power models can also arise when the spatial correlation matrix is replaced by some other nonlinear model such as the Gaussian or spherical. In the next section, we list options to tackle the problems with the AR(1) and power models.

TROUBLE SHOOTING IN CASE OF CONVERGENCE PROBLEMS

Apart from trying several starting values or a systematic grid of starting values for the variance parameters, the following options may be considered for achieving convergence with a power model.

Option 1: The spatial coordinates may be rescaled to change the value of the autocorrelation parameter away from the problematic boundary values 0 and 1. For simplicity we drop the plot indices and consider the correlation term ρ^d for a given distance d. If the distance metric is rescaled so that the distance d is replaced by $\tilde{d} = cd$ for some constant $c > 0$, then

$$\rho^d = \tilde{\rho}^{\tilde{d}}, \tag{9}$$

where $\tilde{\rho} = \rho^{1/c}$. If the autocorrelation $\rho \approx 1$, we may choose $c << 1$ (but still $c > 0$) so that the new correlation parameter $\tilde{\rho}$ is well below the upper bound of unity. When $\rho \approx 0$, we may choose $c >> 1$, thus ensuring that $\tilde{\rho}$ is well above the lower bound of zero.

Option 2: The autocorrelation may be parameterized by an exponential function $\rho = \exp(-\theta)$ with $\theta > 0$, such that

$$\rho^d = [\exp(-\theta)]^d = \exp(-\theta d). \tag{10}$$

Similarly, we may re-parameterize the autocorrelation as $\rho = \exp(-1/\theta)$ with $\theta > 0$, such that

$$\rho^d = [\exp(-1/\theta)]^d = \exp(-d/\theta). \tag{11}$$

The exponential model given in equation (11) is the one implemented in the SAS® mixed model procedures MIXED and GLIMMIX. In addition to the exponential re-parameterization, it may also be necessary to re-scale the distance as in Option 1 to circumvent convergence problems when either $\theta \approx 0$ or $\theta \to \infty$. In the exponential model (11), we can replace the distance d by $\tilde{d} = cd$, so that $\exp(-d/\theta) = \exp(-\tilde{d}/\tilde{\theta})$ with $\tilde{\theta} = c\theta$. Thus, for instance when $\theta \approx 0$ ($\rho \approx 0$), we can choose $c >> 1$ so that $\tilde{\theta} >> 0$. Conversely, when θ is very large ($\rho \approx 1$), we may set $c << 1$ (but still $c > 0$).

Option 3: We may set bounds on estimates of the correlation parameter so that critical boundary values cannot be reached during REML iterations; for example, we may choose to require $0.0001 < \hat{\rho} < 0.9999$ (e.g., Piepho et al. 2013). This option is implemented in some packages, e.g., ASReml and GenStat (VSN International; http://www.vsni.co.uk/). It is less suitable when the true value of the autocorrelation is so close to one of the theoretical boundaries 0 and 1 that the numerical bounds imposed for iterations cause non-negligible bias.

Option 4: We can approximate the AR(1) model (or any other non-linear spatial model) by using a linear variance (LV) model (Williams 1986). Over short distances, such as among plots in small blocks of a field trial, this approximation often works very well, and convergence problems are much less likely because the variance-covariance structure is linear in the parameters. In particular, as opposed to the AR(1) model, there can be no confounding of the spatial component with the nugget or block variance (Piepho et al. 2008, Piepho and Williams 2010). The approximation is particularly well suited when the autocorrelation under an AR(1) model is close to unity. In this case, from a first-order Taylor series approximation it follows that $\rho^{|i-i'|} \approx 1 + \log(\rho)|i-i'|$, which is linear in the spatial distance $|i-i'|$ (Piepho and Williams 2010). Note that $\log(\rho) \le 0$ because $\rho \le 1$, so the correlation decays with distance also under the linear approximation.

Under the LV model, using a parameterization that ensures non-negative definiteness, the variance-covariance matrix of the data can be expressed as

$$\mathrm{var}(\mathbf{y}_j) = \sigma_b^2 \mathbf{J}_k + \phi \mathbf{M}_k + \sigma^2 \mathbf{I}_k , \quad (12)$$

where the (i,i')-th element of $\mathbf{M}_k$ is equal to $(k-1)-|i-i'|$ (Piepho and Williams 2010). The spatial component only has a single parameter (ϕ), as opposed to two for the AR(1) model.

ANALYSIS OF EXAMPLES

In Example 1 we deal with one-dimensional AR(1) models applied to blocks of plots equally spaced down single columns. In Example 2 we study the behaviour of the power model used to model spatial covariance of repeated measurements within independent plots. Then, in Example 3, we consider anisotropic two-dimensional AR(1) models. For all analyses, we used the MIXED procedure of SAS® Version 9.3 (TS1M0).

Example 1: Data for the 200 blocks of size nine in the p-rep design were simulated assuming the variance-covariance model (3) with $\sigma_b^2 = \sigma^2 = 1$. Simulated responses were rounded to four decimal places. This simulation model puts the AR(1) model with nugget to a hard test because it is expected to produce cases with confounding of the autocorrelation with the nugget of the block variance from either $\rho \to 0$ or $\rho \to 1$, respectively. The following starting values were provided for the parameters: $\sigma_b^2 = 1$, $\sigma_s^2 = \sigma^2 = 0.5$ and $\rho = 0.1, 0.2, ..., 0.9$. By trial and error we identified seeds for the random number generator in order that this problem occurred. The datasets generated and the program code with the specific seed values for the random number generator are provided in the Supplemental Material. The model used for analysis had fixed effects for treatments and replicates within locations and random effects for blocks within replicates. Furthermore, in all analyses the variance-covariance structure for plot errors in the same block was assumed to have a nugget component as well as a spatial component following a power model.

In the analysis of the first simulated dataset (C2274; dataset names just follow an internal labelling of datasets we generated and have no particular meaning), the correlation moved towards zero in iterations. The REML algorithm converges after 34 iterations with a warning message that the final Hessian is not positive definite. The estimated asymptotic correlation matrix of the parameter vector $\boldsymbol{\vartheta} = (\sigma_b^2, \sigma_s^2, \rho, \sigma^2)$ based on the observed information is shown in Table 1, together with the parameter estimates. The asymptotic correlation of –0.9905 between the estimates of σ_s^2 and σ^2 indicates that the two parameters are confounded. When fitting the exponential parameterization, the numerical difficulties vanish, and the algorithm converges without a warning message pointing to problems with the final Hessian matrix, but the asymptotic correlation among the estimates of σ_s^2 and σ^2 remains about the same (Table 2). Also, the power and exponential spatial models yield about the same likelihood (–2 log L = 5181.7) (log L = logarithm of the maximized residual likelihood) as the model with only block and error variance (–2 log L = 5182.4). Regardless of parameterization of the spatial model, the estimated nugget variance is tiny, although in the simulating model the nugget variance equals unity. This apparent negligibility of the nugget variance results from the confounding with the spatial variance due to the small correlation estimate, which implies a very rapid decay of correlation with spatial distance.

Table 1. Estimates of variance parameters for dataset C2274 (Example 1) and AR(1) model with nugget and corresponding asymptotic variance-covariance matrix based on the observed information (–2 log L = 5181.7).

Parameter	Estimate	Asymptotic correlation matrix of parameter estimates			
		σ_b^2	σ_s^2	ρ	σ^2
σ_b^2	1.1526	1	–0.00784	–0.05937	0.000594
σ_s^2	0.9598		1	–0.1939	–0.9905
ρ	0.02953			1	0.2281
σ^2	0.001948				1

Table 2. Estimates of variance parameters for dataset C2274 (Example 1) and exponential model with nugget and corresponding asymptotic variance-covariance matrix based on the observed information (–2 log L = 5181.7).

Parameter	Estimate	Asymptotic correlation matrix of parameter estimates			
		σ_b^2	σ_s^2	θ	σ^2
σ_b^2	1.1526	1	–0.03128	–0.05455	0.02393
σ_s^2	0.9600		1	–0.1783	–0.9895
θ	0.2838			1	0.2143
σ^2	0.001804				1

For the dataset C2274, the autocorrelation did not converge to zero. But in a second dataset C2288 the autocorrelation did converge to zero (Table 3), again with a non-positive definite final Hessian matrix. Also, in this case the problem of the non-positive-definite Hessian does not vanish with the exponential parameterization (Table 4). Probably the best solution in this case is to drop the spatial component from the model. This is supported by comparison of the reported values of –2 log L, which equal 5237.5 for the models with and without the spatial component.

Table 3. Estimates of variance parameters for dataset C2288 (Example 1) and AR(1) model with nugget and corresponding asymptotic variance-covariance matrix based on the observed information (–2 log L = 5237.5).

Parameter	Estimate	Asymptotic correlation matrix of parameter estimates			
		σ_b^2	σ_s^2	ρ	σ^2
σ_b^2	0.9064	1	–0.05088	-	-
σ_s^2	0.5686		1	-	-
ρ	0			-	-
σ^2	0.4589				-

Table 4. Estimates of variance parameters for dataset C2288 (Example 1) and exponential model with nugget and corresponding asymptotic variance-covariance matrix based on the observed information (–2 log *L* = 5237.5).

Parameter	Estimate	Asymptotic correlation matrix of parameter estimates			
		σ_b^2	σ_s^2	θ	σ^2
σ_b^2	0.9064	1	–0.05088	–0.00013	-
σ_s^2	0.4873		1	0.000314	-
θ	0.06677			1	-
σ^2	0.5402				-

The third simulated dataset (C2373) produces a spatial correlation estimate close to unity, with the block variance converging to zero (Table 5). There are no convergence problems. This result seemingly suggests that the AR(1) component of the model is very important, whereas there is no block effect. By contrast, the randomization-based analysis, in which the AR(1) component is not used, yields a block variance estimate of $\hat{\sigma}_b^2 = 0.9344$ (–2 log *L* = 5180.5), which is not far-off the spatial variance under the AR(1) model ($\hat{\sigma}_s^2 = 0.9538$). This example, with very contrasting block variance estimates depending on whether or not the AR(1) component is added on, illustrates the possible confounding of spatial component and block variance.

Table 5. Estimates of variance parameters for dataset C2373 (Example 1) and AR(1) model with nugget and corresponding asymptotic variance-covariance matrix based on the observed information (–2 log *L* = 5180.0).

Parameter	Estimate	Asymptotic correlation matrix of parameter estimates			
		σ_b^2	σ_s^2	ρ	σ^2
σ_b^2	0	-	-	-	-
σ_s^2	0.9538		1	–0.1989	–0.1942
ρ	0.9945			1	0.5646
σ^2	0.9628				1

Table 6. Estimates of variance parameters for dataset C2295 (Example 1) and exponential model with nugget and corresponding asymptotic variance-covariance matrix based on the observed information (-2 log *L* = 5155.5).

Parameter	Estimate	Asymptotic correlation matrix of parameter estimates			
		σ_b^2	σ_s^2	θ	σ^2
σ_b^2	0	-	-	-	-
σ_s^2	0.8049		1	–0.2156	–0.2366
θ	83.6057			1	0.5672
σ^2	0.9546				1

Finally, when trying to fit the power model to the fourth simulated dataset C2295, the procedure stops after 49 iterations with a "Did not converge" warning message. Because of the lack of convergence, no final Hessian matrix is produced. At the last iteration, the parameter estimates were $\hat{\sigma}_b^2 = 0$, $\hat{\sigma}_s^2 = 0.8049$, $\hat{\rho} = 0.9881$ and $\hat{\sigma}^2 = 0.9546$. Again, the

spatial component moves towards the block effect component, which causes the convergence problem. Switching to the exponential parameterization, convergence is achieved (Table 6) with an equivalent model fit.

For all four datasets, we also fitted the linear variance (LV) model (12). This converged with very few iterations and without problems in all four cases. The model fitting information is shown in Table 7, including the values of the Akaike Information Criterion ($AIC = -2 \log L + 2p$, where L is the restricted likelihood and p is the number of variance parameters) for the LV and AR(1) models. The *AIC* values show that both models give very similar fits in the first case (C2274) and virtually identical fits in the remaining three cases.

Table 7. Fit of linear variance model (LV) for the four artificial datasets of Example 1.

Parameter/	Dataset			
Criterion[#]	C2274	C2288	C2273	C2295
σ_b^2	1.1481	0.9064	0.9120	0.7307
ϕ	0.002551	0	0.005206	0.009210
σ^2	0.9467	1.0275	0.9629	0.9552
AIC(LV)	5182.3	5237.5	5180.0	5155.5
AIC[AR(1)]	5181.7	5237.5	5180.0	5155.5

[#] *AIC* = Akaike Information Criterion

Example 2: We use this example to illustrate Options 1 and 2 for tackling convergence problems. Our baseline model had fixed effects for treatments and blocks, and a simple random effect (b_j) for plots reflecting the fact that plots were the randomization units. It is because the plots were randomization units that the data from different plots were modelled to be uncorrelated in all our analyses (Piepho et al. 2011). The residual error term e_{ij} modelled the variation among repeated measurements on the same plot and was initially modelled as independent. This model had an *AIC* value of 4083.4. We then tried to fit a power model for the correlation among observations in the same plot without a nugget variance. In all analyses, three different starting values were provided for the autocorrelation (0.1, 0.5 and 0.9). Using these starting values with the MIXED procedure of SAS®, this model did not converge. After multiplying both spatial coordinates by a factor of 10^4, we were able to achieve convergence, the spatial correlation estimate ($\hat{\tilde{\rho}} = 0.01245$) taking on a value well away from the problematic boundary value of 0 (Table 8). The fitted correlation for the transformed coordinates corresponds to the correlation estimate $\hat{\rho} = \hat{\tilde{\rho}}^{10000} = 0.01245^{10000} \approx 0$ for the original coordinates, which explains the convergence problems on that scale. The fitted model implied a strong spatial correlation only for short distances. For instance, the first plot had 252 observations. The largest pairwise correlation for adjacent measurements was 0.322, the 95% quantile of all correlations was 0.00964, and the median of all correlations was effectively zero. Hence, for the vast majority of pairs of observations the correlation is virtually zero, and it is therefore not surprising that up to the first decimal place this spatial model had the same *AIC* as the model with independent errors (Table 8).

Similar numerical problems were observed when a nugget variance was added to the model. With the original coordinates, the algorithm stopped after 41 iterations with no proper convergence. After multiplying the coordinates by 10^4, we found a solution, again with the correlation well away from the boundaries ($\hat{\tilde{\rho}} = 0.08587$). According to *AIC*, this spatial model with nugget fitted better than the independent model (Table 8). From this solution the correlation on the original scale can be estimated

as $\hat{\rho} = \hat{\tilde{\rho}}^{10000} = 0.08587^{10000} \approx 0$, which is very close to the boundary of the parameter space, again explaining the convergence problems when the original spatial coordinates were used.

We also tried the exponential model (11). Starting values for the exponential parameter θ were 0.1, 1, 10 and 100. Without nugget and the original coordinates, the REML algorithm converged to the baseline model. Multiplying the coordinates by a factor of 10^4, a different solution was found for the exponential model, this time equivalent to that obtained for the power model. When a nugget was added, the model converged right away with the original coordinates, so a re-scaling was not necessary (but for the re-scaled coordinates we obtained the same fit).

Table 8. Parameter estimates, likelihood criteria for different models fitted to on-farm trial data (Example 2).

Model[$]	Spatial coordinates	σ_b^2	σ^2	σ_s^2	ρ or $\tilde{\rho}$	θ or $\tilde{\theta}$	$-2 \log L$[#]	*AIC*[§]
Baseline	-	0.01660	0.1321	-	-	-	4079.4	4083.4
POW without nugget	Original	0.01660	-	0.1321	0	-	4079.4	4083.4
POW without nugget	Original $\times 10^4$	0.01502	-	0.1322	0.01245	-	3486.2	3492.2
POW with nugget	Original	0.01321	0.0859	0.5949	0	-	Did not converge	-
POW with nugget	Original $\times 10^4$	0	0.0664	0.0954	0.08587	-	2011.7	2017.7
EXP without nugget	Original	0.01660	-	0.1321	-	0	4079.4	4083.4
EXP without nugget	Original $\times 10^4$	0.01502	-	0.1322	-	0.2280	3486.2	3492.2
EXP with nugget	Original	0	0.0664	0.0954	-	0.000656	2011.7	2017.7
EXP with nugget	Original $\times 10^4$	0	0.0664	0.0954	-	6.5645	2011.7	2017.7

[#] $\log L$ = residual log-likelihood
[§] *AIC* = Akaike Information Criterion
[$] POW = power model, EXP = exponential model

Example 3: We here consider several two-dimensional models that impose an anisotropic variance-covariance structure on the plot errors within replicates arranged as a regular grid of *r* rows and *c* columns (Gilmour et al. 1997). These two-dimensional models are extensions of the basic AR(1) model (5). To keep the notation simple, we consider a single replicate. Let e_{ij} represent the error in the *i*-th row and *j*-th column of a replicate and assume that these are collected into a vector $\mathbf{e} = (e_{11},...e_{1c}, e_{21},...., e_{rc})^T$.

A separable two-dimensional AR(1) model for var(**e**) may be defined as $\sigma_s^2(\boldsymbol{\Omega}_r \otimes \boldsymbol{\Omega}_c)$, where $\boldsymbol{\Omega}_r = \{\rho_r^{d_r(i,i')}\}$, $\boldsymbol{\Omega}_c = \{\rho_c^{d_c(j,j')}\}$, $\otimes$ denotes the Kronecker (direct) product operator, and $d_r(i,i') = |i - i'|$ $[d_c(j,j') = |j - j'|]$ is the distance between the *i*-th and i'-th row (*j*-th and j'-th column). The nugget can be represented as $\sigma_s^2(\mathbf{I}_r \otimes \mathbf{I}_c)$, where $\mathbf{I}_r$ and $\mathbf{I}_c$ are *r*- and *c*-dimensional identity matrices. Random row and column effects can be modelled by introducing the components $\sigma_r^2(\mathbf{I}_r \otimes \mathbf{J}_c)$ and $\sigma_c^2(\mathbf{J}_r \otimes \mathbf{I}_c)$, respectively, in the structure for var(**y**), where **y** denotes the observed data vector for a replicate. For further details on this model see Gilmour et al. (1997) and Piepho and Williams (2010).

We fitted different models involving some or all of the four components just described (Table 9). In addition to the two-dimensional AR(1) model, we also fitted one-dimensional AR(1) models, either among plots along the same row, or among plots down the same column. Each model had a fixed effect for varieties and for replicates.

The best-fitting model according to *AIC* was the one containing all random components, i.e., row and column effects, the AR(1) ⊗ AR(1) spatial component, and a nugget (Table 9). The autocorrelation down the columns was very high ($\hat{\rho}_r = 0.9863$), while that along the rows was relatively weak ($\hat{\rho}_c = 0.4117$). At the same time, the column variance σ_c^2 converged to zero, which shows that there is a confounding between the column component of the spatial model (ρ_r) and the column effect. The nugget variance took up a considerable part of the plot-to-plot variation ($\hat{\sigma}^2 = 1490$), compared to the spatial variance ($\hat{\sigma}_s^2 = 13918$). When the nugget was dropped from the model, the autocorrelations were markedly reduced ($\hat{\rho}_r = 0.4332$, $\hat{\rho}_c = 0.1963$), likely because the model was now forced to take up the nugget component into the spatial part of the model. At the same time, because of the much reduced correlation in the spatial components, the column variance was not zero, but took on a very large value ($\hat{\sigma}_c^2 = 11707$), comparable in magnitude to the estimate when the spatial model component was dropped altogether.

The same behaviour as with the two-dimensional AR(1) model was observed when the correlation along the rows was dropped from the model. Again, the column variance was estimated to be zero. There were convergence problems with this model, likely due to the confounding of spatial model and column effects. These could only be resolved by fixing the column variance to zero (Table 9). When this was not done, the column variance came close to zero before the REML algorithm aborted without having reached the maximum of the residual likelihood.

We also fitted a model that had a spatial correlation only along the rows, but not along the columns. Similar to the case of the two-dimensional model, the correlation was quite low ($\hat{\rho}_c = 0.1226$). In fact, the correlation was so low that the fitted covariance was virtually zero at lag 2 or larger along the rows, corresponding to a confounding with the nugget effect. Note that the estimated nugget variance is very tiny here ($\hat{\sigma}^2 = 2$), whereas the spatial component takes up most of the plot variance ($\hat{\sigma}_s^2 = 2507$). This model also had convergence problems, again likely due to the confounding problem, with the final Hessian at convergence being non-positive definite.

Piepho and Williams (2010) fitted the one-dimensional linear variance (LV) model (12) and several of its two-dimensional extensions to the same data as used in this example. The LV models yielded similar fits as the corresponding one- and two-dimensional AR(1) models; convergence was swift with all LV models.

Table 9. Parameter estimates and likelihood criteria for different models fitted to wheat data of Gilmour et al. (1997) (Example 3).

Model				Parameters						Likelihood criteria	
Nugget	Row effect	Column effect	Spatial component	σ^2	σ_r^2	σ_c^2	σ_s^2	ρ_r	ρ_c	−2 log *L*#	*AIC*&
√	√	√		2391	843	12225	-	-		- 2574.2	2580.2
√				13359	-	-	-	-		- 2854.6	2856.6
√			AR(1) ⊗ AR(1)	-	-	-	14721	0.8771	0.3021	2595.0	2601.0
	√	√	AR(1) ⊗ AR(1)	-	382	11707	3052	0.4332	0.1963	2555.0	2565.0
√	√	√	AR(1) ⊗ AR(1)	1490	460	0	13918	0.9863	0.4117	2538.4	2548.4
√	√	√	AR(1) down columns only	2	739	12177	2507	-	0.1226	2573.3§	2583.3
√	√	√	AR(1) down rows only	1436	564	$0	11870	0.9854	-	2542.4	2550.4

§ Final Hessian not positive definite

$ This variance component had to be fixed to zero in order to achieve convergence. When it was not fixed, the value of this variance was also zero when the REML-algorithm aborted

Residual likelihood (REML)

& *AIC* = Akaike Information Criterion

DISCUSSION AND CONCLUSION

We investigated some of the numerical problems that may arise when fitting power and AR(1) models in randomized trials. We also considered some options for achieving convergence to a proper solution of REML-based procedures when default settings fail. The key problem is that the spatial component may be confounded with either the nugget variance or with random effects for experimental units such as blocks (in the case of randomized experiments with a single observation per plot) and plots (in the case of randomized experiments with repeated and spatially referenced measurements taken on the same plot). We showed that care is needed in interpreting the individual parameter estimates. Often, they can create the impression that spatial correlation is substantial whereas randomization-based effects for blocks and independent residual error are small. But the confounding problems elucidated in this paper mean that there is interplay between estimates of different parameters of the model and that a large autocorrelation estimate may in reality be due to a large variance of block effects, and vice versa. Similarly, a small autocorrelation estimate associated with a large spatial variance can, in fact, be confounded with the nugget variance, and the decay of correlation with spatial distance may be so fast that the fitted pairwise residual correlation within blocks is negligible for most pairs of plots. Thus, looking at the spatial variance estimate alone can be misleading. It is perhaps best to avoid strong interpretations of individual parameter estimates in isolation. Instead, the focus should be laid on the whole fitted variance-covariance structure for the data.

In randomized field trials, it is advisable to employ randomization-based models (Nelder 1965) with simple random effects. This can be taken as a point of departure for more refined modelling involving spatial add-on components. Such models can also serve as a fall-back option when spatial models turn out to be difficult to fit or to provide no improvement (Piepho and Williams 2010). We recommend against the common practice of fitting a purely spatial model with no effects for blocks when the design involved blocking. There is some danger that spatial correlation is erroneously reported as high when, in fact, it is low if blocks are fitted to reflect the randomization layout. Such problems can be avoided by starting the modelling process with a randomization-based model and considering spatial covariance merely as an optional add-on component.

In this paper, we have used one particular REML package (the MIXED procedure of SAS). To some extent numerical problems with statistical procedures are always software-specific. Because of differences in implementation of iterative procedures for the REML algorithm, other software can yield different solutions, especially when convergence is difficult to achieve. It was not our intention here to compare different mixed model packages. In our experience, however, the problems with AR(1) and power models are similar with most packages, despite some differences in convergence behaviour for the same dataset. This largely results from the confounding problem when $\rho \to 0$ or $\rho \to 1$, which is an inherent property of the model itself rather than of the particular package used. We have considered options to deal with these problems and illustrated them using three examples. Re-scaling of the spatial coordinates or a re-parameterization of the AR(1) model as an exponential model can be recommended to help the model converge. Which of these options may work in a given situation much depends on the dataset. The LV model is a viable alternative to non-linear spatial models such as AR(1), and in our experience it usually has good convergence behaviour. So we suggest to routinely consider this model when exploring competing spatial add-on components to randomization-based models.

ACKNOWLEDGEMENTS

Hans-Peter Piepho and Jens Möhring acknowledge support by DFG grant PI 377/13-1.

REFERENCES

Cullis B.R., Smith A.B., Coombes N.E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural Biological and Environmental Statistics* 11, 381–393

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* 2, 269–293.

Gonçalves E., St. Aubyn A., Martins A. (2010). Experimental designs for evaluation of genetic variability and selection of ancient grapevine varieties: a simulation study. *Heredity* 104, 552–562.

Gotway C.A., Schabenberger O. (2004). *Statistical methods for spatial data analysis*. Chapman & Hall, London.

John J.A., Williams E.R. (1995). *Cyclic and computer generated designs*. Second edition. Chapman and Hall, London.

Nelder J.A. (1965). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. II. Treatment structure and the general analysis of variance. *Proceedings of the Royal Society of London* A 283, 147–178.

Piepho H.P., Möhring J., Williams E.R. (2013). Why randomize agricultural experiments? *Journal of Agronomy and Crop Science* 199, 374–383.

Piepho H.P., Richter C., Spilke J., Hartung K., Kunick A., Thöle H. (2011). Statistical aspects of on-farm experimentation. *Crop and Pasture Science* 62, 721–735.

Piepho H.P., Richter C., Williams E.R. (2008). Nearest neighbour adjustment and linear variance models in plant breeding trials. *Biometrical Journal* 50, 164–189.

Piepho H.P., Williams E.R. (2010). Linear variance models for plant breeding trials. *Plant Breeding* 129, 1–8.

Richter C., Kroschewski B., Piepho H.P., Spilke J. (2015). Treatment comparisons in agricultural field trials accounting for spatial correlation. *Journal of Agricultural Science* (http://dx.doi.org/10.1017/S0021859614000823)

Schachtschabel P., Blume H.P., Hartge K.H., Schwertmann U. (1976). *Lehrbuch der Bodenkunde*. Enke, Stuttgart.

Williams E.R. (1986). A neighbour model for field experiments. *Biometrika* 73, 279–287.

Williams E.R., John J.A., Whitaker, D. (2014). Construction of more flexible and efficient p-rep designs. *Australian and New Zealand Journal of Statistics* 56, 89–96.

Williams E.R., Piepho H.P., Whitaker, D. (2011). Augmented p-rep designs. *Biometrical Journal* 53, 19–27.