Linear mixed models and geostatistics for designed experiments in soil science: Two entirely different methods or two sides of the same coin?

Johanna I. F. Slaets | Runa S. Boedinghaus | Hans-Peter Piepho

Abstract
Soil scientists are accustomed to geostatistical methods and tools such as semivariograms and kriging for analysis of observational data. Such methods assume and exploit that observations are spatially correlated. Conversely, analysis of variance (ANOVA) of designed experiments assumes that observations from different experimental units are independent, an assumption that is justified based on randomization. It may be beneficial, however, to perform an ANOVA assuming a geostatistical covariance model. Also, it is increasingly common to have multiple observations per experimental unit. Simple ANOVA assuming independence of observations is not appropriate for such data. Instead, a linear mixed model accounting for correlation among observations made on the same plot is required for proper analysis. The purpose of this paper is to demonstrate the benefits of integrating geostatistical covariance structures and ANOVA procedures into a linear mixed modelling framework. Two examples from designed experiments are considered in detail, making a link between terminologies and jargon used in geostatistical analysis on the one hand and linear mixed modelling on the other hand. We provide code in R and SAS for both examples in two supporting companion documents.

Highlights
- Analysis of variance and geostatistical analysis can be joined in a mixed model.
- Randomization justifies the independence assumption in analysis of variance.
- Geostatistical models imply a correlation of errors and can improve efficiency.
- Lacking randomization, spatial correlation can be accounted for in a mixed model.

KEYWORDS
analysis of variance, designed experiment, kriging, mean comparison, nugget, polynomial regression, random effect, randomization, range, semivariogram, sill, spatial analysis
1 | INTRODUCTION

Data from observational studies (surveys) in soil science are often complex. In many cases the main challenge in statistical analysis is to account for and exploit spatial correlation among observations taken in close vicinity, which are typically more similar than observations farther apart. Soil scientists are accustomed to dealing with spatially correlated data, typically using geostatistical methods for analysis, such as the semivariograms and kriging for map interpolation (Diggle & Ribeiro, 2007; Isaacs & Srivastava, 1989; Oliver & Webster, 2014; Webster & Oliver, 2007). Conversely, designed experiments are routinely subjected to analysis of variance (ANOVA) procedures and regression techniques based on linear models, assuming that observations from different experimental units (plots, chambers, pots, etc.) are independently distributed, an assumption that is well justified if the experiment was properly randomized (Bailey, 2009).

Geostatistics on the one side and ANOVA on the other side may appear like two entirely different frameworks. For example, a soil scientist wanting to do an analysis of a designed field experiment to compare several tillage methods may wonder how to account for spatial correlation among several soil samples taken on the same plot. Remembering her geostatistics class, she turns to semivariogram modelling (Oliver & Webster, 2014) and draws up a map of soil properties for the trial, finding that there is in fact substantial spatial correlation both within and between plots. Having reached this far, she then turns back to the research question, which called for a comparison among several group means, and finds that a semivariogram or map does not provide such a comparison. The researcher then remembers how to perform a basic one-way ANOVA using her favourite statistical package but soon realizes that this method does not account for spatial correlation and in fact assumes independence between observations, which leaves her puzzled as to the best way forward.

Even though, on the face of it, geostatistics and ANOVA indeed appear like entirely different methods, they can, in fact, be integrated into a single powerful framework (Lark, Cullis, & Welham, 2006; Robinson, 1991; Stroup, 2002): linear mixed models. It is the purpose of this paper to illustrate this framework and its benefits to soil scientists and to align the jargon used in geostatistics with the terminology used for linear mixed modelling. Two examples from designed experiments serve to introduce different features of the framework.

2 | A FIRST EXAMPLE: A RANDOMIZED FIELD TRIAL TO STUDY THE SPATIAL AND TEMPORAL VARIATION OF ARBUSCULAR MYCORRHIZAL FUNGAL ABUNDANCE IN SOIL

Example 1 A randomized experiment was conducted in the year 2011 to study the small-scale spatio-temporal variation of arbuscular mycorrhizal fungal (AMF) communities under grassland (Buscot, Wubet, Goldmann, & Klemmer, 2018; Goldmann et al., 2020; Regan et al., 2014, 2015, 2017; Regan, Berner, Boeddinghaus, Marhan, & Kandeler, 2019). The experimental area (10 × 10 m) was divided up into 30 blocks (2 × 1.67 m each), arranged in five rows and six columns as shown in Figure 1. Each block had six plots. Soil samples were taken at six dates throughout the year (5 April, 17 May, 27 June, 16 August, 5 October, 21 November). Within each block, the six sampling dates were randomly allocated to plots, so the sampling design was a randomized complete block design with dates as treatments. Each plot was split into two subplots (20 × 20 cm) to assess short-range spatial variation within dates and plots. The subplot samples are thus pseudo-replicates (Hurlbert, 1984; Box S1 in Supporting Information). The main purpose of these subplot samples was to help in estimating the semivariogram at short distances. In total, there were 360 subplots. One soil core was collected per subplot and spatially referenced using coordinates in metres from an origin situated in one corner of the trial area. AMF alpha-diversity was determined for each subplot using 18S rDNA pyrosequencing analysis (Goldmann et al., 2020). Here, we will consider the total abundance over all taxa (total observed taxonomical unit (OTU) richness). This variable was log-transformed (and then multiplied by 10) because all counts were greater than zero (minimum: 11; median: 24; maximum: 45) and diagnostic residual plots indicated that this transformation simultaneously achieved approximate normality and homogeneity of variance. One objective of the study was to assess the variation of AMF communities over time and in space.

Figure 2 shows a kriged map for total AMF OTU richness at each of the six dates. Kriging was performed based on the exponential semivariance model fitted to the subplot data by weighted least squares using the number of observations per distance class (bin) of the
empirical semivariances as weights (http://www.gstat.org/gstat.pdf, Table 4.2). The maps display substantial spatial heterogeneity at each date, and there is a notable evolution of AMF OTU richness over time. An ANOVA was performed for plot means (Table 1), showing that there are significant differences among dates ($F = 30.81, p < .0001$), with the largest total AMF OTU richness occurring in October. The means for the 6 month in Table 2 agree well with the impression from the kriged maps in Figure 2.

The two basic analyses presented so far raise some questions. The ANOVA in Table 1 assumes that the data representing each plot are independent, justified by random allocation of treatments to plots, whereas the kriging done to generate maps in Figure 2 explicitly assumes that the sampling points are spatially correlated across the whole trial area and this correlation is actually the basis for the interpolation between sampling points that is needed to draw kriged maps. So, which of these two analyses is based on the correct assumption? Perhaps surprisingly, the answer is: both are! It will be shown in this paper that the two approaches, kriging and ANOVA, seeming so entirely different at first sight, can be reconciled and integrated into a unifying framework. Before considering that framework, we will review a few basics about spatial statistics and ANOVA.

3 | SOME BASICS ABOUT SPATIAL STATISTICS

In spatial datasets, if we want to estimate the overall variance in the data, we need to account for spatial correlation (i.e., points closer together will be more similar due to their proximity), unless the data are collected based on some kind of random sampling design (de Gruijter, Brus, Bierkens, & Knotters, 2006). Also, interpolation between sampling points is possible by exploiting spatial correlation. In order to model the spatial correlation between observations on a grid of points, one may estimate the semivariance. For random variables $Y_1$ and $Y_2$ the semivariance is defined as half the variance of the difference,
\[ \frac{1}{2} \text{var}(Y_1 - Y_2), \]
and it is estimated by \[ \frac{1}{2}(y_1 - y_2)^2, \]
where \( y_1 \) and \( y_2 \) are the two observations (“realizations”) corresponding to \( Y_1 \) and \( Y_2 \) (Webster & Oliver, 2007). A plot of these semivariances against spatial distance of the two points, \( d_{12} \), is called the empirical semivariogram. A nonlinear function, the semivariance function, may be fitted to describe how the semivariance changes with distance (Figure 3). The semivariance at zero distance, that is, the intercept of the semivariance function, is called the nugget. The semivariance then increases with distance, typically approaching an upper bound (i.e., an asymptote), which is called the sill and tells us the overall variation in the data after adjusting for possible correlations due to proximity. The semivariance is directly related to the covariance between observations. Thus, the covariance function, describing how the covariance between observations depends on spatial distance, can be analytically derived from the semivariance function as:

\[
\text{cov}(Y_1, Y_2) = \frac{1}{2} [\text{var}(Y_1) + \text{var}(Y_2)] - \frac{1}{2} \text{var}(Y_1 - Y_2),
\]

where \( \text{cov}(Y_1, Y_2) \) is the covariance between random variables \( Y_1 \) and \( Y_2 \) and \( \text{var}(Y_1) \) and \( \text{var}(Y_2) \) are their variances (Figure 4). This covariance function can be used to
compute the covariance of the response at an unobserved point, $Y_0$, with the responses at the observed points, $Y_1, ..., Y_n$. Exploiting this covariance, one can obtain an interpolated estimate of $y_0$ as a weighted mean of the observed values $y_1, ..., y_n$. The weights minimizing the squared prediction error can be computed from the variances and covariances (Isaacs & Srivastava, 1989). This method of estimation is called kriging, named after the geologist Krige (1951) who invented it in order to estimate (interpolate) the likely amount of metal ore (e.g., gold nuggets) at an unexplored point based on the amounts found in a few surrounding boreholes of a mining operation. If several boreholes are drilled near the same point, right next to each other, the amount of nuggets varies locally, and the variance between these immediately neighbouring boreholes, whose distance is essentially zero, is therefore called the nugget variance. Generally, in other applications, the observed response will vary when measuring at the same spot due to microscale variation of the soil property and measurement error, and the nugget variance can therefore be associated with both of these sources of variation. Such errors should be accounted for also in applications outside of mining.

Most soil scientists will be familiar with these geostatistical methods. The main point to be reiterated here is that if kriging is applied, then usually a spatial covariance is assumed among all measurements and an explicit model is fitted to describe that covariance via the semivariogram (for exceptions see Webster & Oliver, 2007).

### 4 SOME BASICS ABOUT ANALYSIS OF VARIANCE AND RANDOM EFFECTS MIXED MODELS

The assumption of a covariance among observations made in kriging is in contrast to ANOVA procedures for randomized experiments, which assume independence among experimental units (plots). The justification for this independence assumption is based on randomization, whereby treatments are allocated to experimental units at random (Bailey, 2009; Calinski & Kageyama, 2000). A randomized experiment can be analysed by a linear model with independent errors for the experimental units. With a single observation per experimental unit (plot), this leads to a simple ANOVA (Piepho, Möhring, & Williams, 2013). In the case of a completely randomized design (CRD), the underlying linear model just has an intercept, a treatment effect and a residual error term. When a randomized complete block design (RCBD) is used, the model also includes a block effect.

In Example 1, however, there are two observations per plot, and hence some care is needed in order to furnish a valid analysis. The experimental units are still the plots, but there are two subplots per plot, which constitute subsamples. One way to correctly analyse subsampled data is to compute an average per experimental unit and use this as the response (Piepho, 1997), as was done for Table 1. The linear model for a RCBD is:

$$\bar{y}_{ij} = \mu + \tau_i + b_j + e_{ij},$$

where $\bar{y}_{ij}$ is the plot mean for the $i$-th date (treatment) in the $j$-th block, $\mu$ is a general intercept, $\tau_i$ is the effect of the $i$-th date, $b_j$ is the effect of the $j$-th block, and $e_{ij}$ is the error associated with plot mean $\bar{y}_{ij}$. The plot errors are assumed to be independent, which is justified from randomization. It should be stressed that this independence

### TABLE 1 Analysis of variance for total AMF OTU richness

| Source   | Degrees of freedom | Sum of squares | Mean square | F-value | p-value |
|----------|--------------------|----------------|-------------|---------|---------|
| Dates    | 5                  | 284.6          | 56.92       | 30.81   | <.0001  |
| Blocks   | 29                 | 175.4          | 6.05        | 3.27    | <.0001  |
| Residual | 145                | 267.9          | 1.85        |         |         |

Note: Data were log-transformed and then multiplied by 10 (Example 1). Analysis is based on plot means and assumes independence between plots.

### TABLE 2 Means for six dates for total AMF OTU richness

| Date          | Time (in days) | Mean abundance$^a$ |
|---------------|----------------|--------------------|
| 5 April       | 0              | 30.37$^d$          |
| 17 May        | 42             | 31.35$^c$          |
| 27 June       | 83             | 32.11$^b$          |
| 16 August     | 123            | 30.86$^{cd}$       |
| 5 October     | 173            | 34.25$^a$          |
| 21 November   | 220            | 31.23$^c$          |

Least significant difference ($\alpha = 5\%$) 0.69

Note: Data were log-transformed and then multiplied by 10 (Example 1). Analysis is based on plot means and assumes independence between plots. $^a$Means followed by a common letter are not significantly different according to a $t$-test at a significance level of $\alpha = 5\%$, based on model (4) (see Section “Some basics about analysis of variance and random effects mixed model”). AMF, arbuscular mycorrhizal fungal; OTU, observed taxonomical unit.
assumption is valid no matter what a semivariogram for the same data would tell us. All that is required really is proper randomization! This fact can hardly be over-emphasized. Even if the semivariogram suggests that two adjacent plots are highly correlated, because they tend to be quite similar, we may still proceed with the classical ANOVA based on model (2), which assumes independence among the observations (Piepho et al., 2013).

Randomization ensures that classical ANOVA will be perfectly valid. We will see in the next section that these two different assumptions, that is, spatial covariance for kriging and independence for classical ANOVA, can be put in a common framework that can help to decide which of the two assumptions is preferable. The key observation here is that both assumptions can be justified, and it’s not the case that the one is correct and the other is wrong.

We can also model the observed data at the level of subplots. It is tempting to use the model:

$$y_{ijk} = \mu + \tau_i + b_j + e_{ijk}, \quad (3)$$

where $y_{ijk}$ is the observation of the $i$-th date on the $k$-th subplot on the $ij$-plot within the $j$-th block, $e_{ijk}$ is the subplot error associated with $y_{ijk}$, and the other effects are as defined for model (2). In this model, the subplot errors $e_{ijk}$ are assumed to be independent with zero mean and constant variance $\sigma^2_e$. But is this the right model? No, it isn’t! This model implies that subplots provide true replication, but subplots are just subsamples or pseudo-replications (Hurlbert, 1984; Piepho, 1997; Box S1). The number of true replications in this experiment is given by the number of plots per treatment (overall, there are 30 plots in this experiment), not the number of subplots (there are 60 subplots). A true replication is given by a plot because plots (not subplots) are the experimental units to which treatments (sampling dates in this case) are independently allocated by the randomization procedure, whereas subplots are the observational units. Both subplots in the same plot receive the same treatment, so subplots are not independent replicates. They just represent subsamples or pseudo-replications.

To account for experimental units (plots), the model for observed data needs to be extended as:

$$y_{ijk} = \mu + \tau_i + b_j + p_{ij} + e_{ijk}, \quad (4)$$

where $p_{ij}$ is the random error of the $ij$-th plot within the $j$-th block, assumed to be independently distributed with zero mean and variance $\sigma^2_p$. This linear model has two sources of random variability, in contrast with classical linear models, which have only one source. The two random effects, that is, the plot error $p_{ij}$ and the subplot error $e_{ijk}$, make this a linear mixed model. This model properly represents the randomization structure of the experiment as well as the subsampling per plot (Piepho, 1997). The ANOVA presented in Table 3 is based on this model, which can be used to work out the expected mean squares shown in Table 4, and these indicate that the mean square for dates needs to be tested against the mean square for plots, not the residual mean square for subplots. This is because the dates

---

**FIGURE 3** Explanation of terms sill, partial sill, nugget and range using fit for November from Figure 2 based on exponential semivariogram model. The number next to a plotted point of the empirical semivariogram denotes the number of pairs (semivariances) on which the point is based [Color figure can be viewed at wileyonlinelibrary.com]

**FIGURE 4** Covariance and semivariance function for isotropic power model with nugget using fit for Example 1: $\rho = 0.36, \sigma^2_s = 1.57, \sigma^2_e = 2.31$. The functions are drawn only for distances >0. At a distance of zero, there is a discontinuity (not shown in the figure), with the covariance jumping to $\sigma^2_s + \sigma^2_e = 1.57 + 2.31 = 3.88$ and the semivariance dropping to zero (see Box 2 and Pinheiro & Bates, 2000, p. 231) [Color figure can be viewed at wileyonlinelibrary.com]
mean square has the same expectation as the plots mean square under the null hypothesis of no date effects. Good ANOVA packages will compute these expected mean squares and automatically determine the proper error terms for each effect in the model (Milliken & Johnson, 1992).

The expected mean squares in Table 4 assume balanced data, that is, the blocks are complete and each plot has the same number of subplots. It is important to note that the F-tests for treatments and blocks in Table 3, which are based on an analysis of subplot data using model (4), are identical to the ones obtained for an analysis of plot means based on model (2) as shown in Table 1. The results are identical because the data are balanced and a simple variance components model with independent random effects is used. If a linear mixed model package is used and the model is correctly specified as in (4), the appropriate F-tests for fixed effects will be computed automatically, also with unbalanced data (Piepho & Edmondson, 2018).

It should be stressed that the linear mixed model (4) implies a covariance among the two subplots on the same plot. That covariance is given by the variance $\sigma_p^2$ of the plot effect, which is shared by the two observations made on the same plot. So far, we have not considered the status of the block effect. This effect may also be modelled as random with zero mean and constant variance ($\sigma_b^2$). This effect induces a further covariance among observations that are made in the same block. Thus, according to this model, there are three levels of covariance: (a) observations in different blocks have a covariance of zero, (b) observations in the same block but on different plots have a covariance of $\sigma_p^2$, because these share the same block effect, and (c) observations on the same plot have a covariance of $\sigma_c^2 + \sigma_p^2$, because they share not only the same block effect but also the same plot effect. It may be added that the total variance of an observation is $\sigma_c^2 + \sigma_p^2 + \sigma^2$. These considerations show that the linear mixed model (4) also provides a way to model spatial covariance among observations in the same trial, albeit a somewhat cruder one. Its form is different from the covariance model assumed in kriging, but both represent spatial variances and covariances. The main difference is that kriging assumes that covariance is a smooth function of distance, whereas the randomization-based model implies a discontinuous change of covariance at the block and plot boundaries. For example, at the block boundary the covariance drops to zero, which may seem unnatural, but again this model is justified purely based on randomization and the non-smooth covariance may also be a good approximation to an assumed underlying smooth covariance function.

In the preceding paragraph, we have taken the mixed model (4) as a point of departure and derived implications for spatial covariance. Conversely, starting with a spatial covariance model for residuals, one could infer the between-and-within-block and/or plot variance components using Krig’s relation (Webster & Oliver, 2007, pp. 60–62), which further emphasizes the close connection between the geostatistical and mixed modelling frameworks.

5 | JOINING SPATIAL COVARIANCE MODELLING AND ANOVA

The analysis of variance in Table 1 allowed a comparison among dates in terms of mean total AMF OTU richness.

### Table 4  Expected mean squares in ANOVA for total AMF OTU richness (Example 1) under model (4) based on subplot data

| Source                        | Degrees of freedom | Expected mean square |
|-------------------------------|--------------------|----------------------|
| Dates ($r_t$)                 | $t - 1$            | $\sigma_c^2 + \sigma_p^2 + \sum_{i=1}^{r_t}(t_i - t)^2$ |
| Blocks ($b_i$)                | $r - 1$            | $\sigma_b^2$         |
| Plots nested within blocks ($p_i$) | $(t - 1)(r - 1)$ | $\sigma_c^2 + \sigma_p^2$ |
| Subplots, nested within plots ($e_{ijk}$) | $r(s - 1)$     | $\sigma_c^2$         |

Note: $r$, number of blocks (replicates); $t$, number of treatments (dates); $s$, number of subplots per plot. AMF, arbuscular mycorrhizal fungal; OTU, observed taxonomical unit.

### Table 3  Analysis of variance for total AMF OTU richness

| Source                        | Degrees of freedom | Sum of squares | Mean square | F-value | p-value |
|-------------------------------|--------------------|----------------|-------------|---------|---------|
| Dates                         | 5                  | 569.2          | 113.8       | 30.81\(^a\) | <.0001   |
| Blocks                        | 29                 | 350.8          | 12.1        | 3.27\(^a\) | <.0001   |
| Plots, nested within blocks   | 145                | 535.7          | 3.7         | 1.44\(^b\) | .0107    |
| Subplots, nested within plots | 180                | 463.3          | 2.6         |         |         |

Note: Data were log-transformed and then multiplied by 10 (Example 1). Analysis is based on subplot data. AMF, arbuscular mycorrhizal fungal; OTU, observed taxonomical unit.

\(^a\)Plots mean square was the error term (see Table 4).

\(^b\)Subplots mean square was the error term (see Table 4).
Such a comparison is not possible by the semivariograms computed separately for each date (Figure 2). Conversely, the semivariograms allowed modelling the spatial covariance among observations and generating interpolated maps, an option not available with classical ANOVA, which assumes independence among experimental units. Now wouldn’t it seem more consistent to have a unified modelling approach that allows both map interpolation and makes mean comparisons?

The key to unifying both approaches is to incorporate the modelling of the spatial covariance into the linear mixed model – essentially by modelling the spatial covariance, that is, the spatial dependence structure of the data (Gilmour, Cullis, & Verbyla, 1997; Pinheiro & Bates, 2000). To this end, we can extend model (4) for Example 1 by including a spatially correlated component as:

$$y_{ijk} = \mu + \tau_i + b_j + p_{ij} + s_{ijk} + \epsilon_{ijk},$$

where $s_{ijk}$ is a random subplot effect representing a spatially correlated trend, whereas the independent residual error $\epsilon_{ijk}$ represents the nugget effect and is also associated with subplots. A smooth spatial covariance model is assumed for $s_{ijk}$. For example, one may assume that the covariance for the effects $s_{ijk}$ of two points that are a distance $d$ apart is $\sigma^2 \rho^d$ (Figure 4). According to this model, the decay of covariance decays exponentially with distance, and it does so at the same rate in each spatial direction, which is why the model is called isotropic (Schabenberger & Gotway, 2004). An extension to the isotropic power model is to assume that the decay of covariance differs among the two spatial directions. Such models are called anisotropic. One such model assumes that the covariance among two points is $\sigma^2 \rho_{dx} \rho_{dy}$, where $d_x$ and $d_y$ are the distances in the x-direction and y-direction. A straightforward further extension of this model allows accounting for temporal distance $d_t$ as a third dimension using a time coordinate $t$ (in days), such that the covariance is $\sigma^2 \rho_{dx} \rho_{dy} \rho_{dt}$. Such a model may be more plausible than a purely spatial one in Example 1 because observations are indeed made at different points in time and a decay of covariance is expected with increasing temporal distance among observations. It should be mentioned here that this simple option for spatio-temporal (space–time) modelling has many contenders (see Final Remarks), some of which are rather more complex and not all of which are as easy to implement with a mixed model package.

We may fit model (5) with different spatial covariance models, as well as several sub-models obtained from that model by dropping random effects, in order to see which model best fits the data. To compare models, the Akaike information criterion (AIC) can be used, with smaller values indicating better fit. For details see Box 1. Model fits are reported in Table 5, along with average standard errors of a difference (s.e.d.) to see how this varies with model choice. The fact that the s.e.d. are quite variable shows that model choice matters with respect to statistical inference (significance tests and confidence intervals). It should also be mentioned for clarity that s.e.d. is no guarantee a valid analysis, provided the distributional assumptions of normality and homogeneity of variance hold – these can be inspected visually using the residual plots (Kozak & Piepho, 2018). A great advantage in using this model is that no further model choice is needed; one may just rely on randomization theory and use the one model that is induced by randomization. One can also use a spatial model for the purpose (Table 6), which may entail a gain in precision, but this requires selecting among a larger number of candidate models (using, e.g., the AIC), and one can never be sure to have selected the best model. Here, we only considered the power model for brevity, but there are many more models to choose from (Schabenberger & Gotway, 2004; also see Example 2 for a few other models). Thus, when a randomized design has been used, the randomization-
A SECOND EXAMPLE: A RANDOMIZED FIELD EXPERIMENT WITH MULTIPLE SPATIALLY REFERENCED MEASUREMENTS PER PLOT

We will consider a second example to further illustrate the flexibility of mixed models with spatial modelling of the residual error.

Box 1  Variance–covariance model selection by the Akaike information criterion (AIC)

The Akaike information criterion (AIC) is defined as:

\[ AIC = -2 \log(L_R) + 2q, \]

where \( L_R \) is the residual likelihood and \( q \) is the number of parameters in the variance–covariance model (Wolfinger, 1996). Small values of AIC indicate the better models among a set of considered models, with a difference in AIC of two or less being negligible, and a difference of more than 10 points giving strong preference to the model with the smallest AIC (Burnham & Anderson, 1998). The assumption underlying the AIC is that there is an unknown model that may be assumed to have generated the data but is too complex to be formulated as a statistical model that can be fitted to data. Thus, we can only hope to approximate the true model using simpler statistical models. The AIC is designed to select the model in a candidate set that best approximates the underlying true model.

Likelihoods for different models are comparable only when the exact same data are used. When comparing different variance–covariance models using the AIC computed from the residual likelihood, it is therefore important to use the same specification for the fixed effects throughout. This is because the residual likelihood is based on data that correspond to contrasts (residuals) removing fixed effects and hence the data will change when the fixed effects model is modified. When models with different fixed-effects specifications are to be compared, the estimation method should therefore be changed from residual maximum likelihood (REML) to maximum likelihood (ML) and \( q \) would include both the fixed effects and the covariance parameters. (Note: The “nlme” package defines \( q \) in this way even with REML. Although this does not affect comparison of models having the same fixed-effects structure, it seems more natural to only count the number of covariance parameters when using REML.)

The fixed-effects specification for selecting the variance–covariance model is best chosen to be the most complex model possible or reasonable. In designed experiments with replication, it is preferable to fit the saturated model defining all treatment factors as qualitative. In observational studies, or when covariates are used, it is not generally obvious what the most complex or saturated model is. There are usually competing models, and it is possible that the model used for selecting the variance–covariance structure turns out not to be the final one after the subsequent inspection of the fixed-effects structure. So, there is obviously a chicken-and-egg problem here that is impossible to fully resolve in general, and hence there are different possible variations to this model selection strategy.
Table 5: Fit of several linear mixed models to total AMF OTU richness data (Example 1)

| Model | Design effects (equation 5) | $b_j$ | $p_{ij}$ | $e_{ijk}$ | $s_{ijk}$ | AIC | Mean s.e.d. $^b$ |
|-------|-----------------------------|----|----|----|----|----|---------------|
| CRD$^a$ model ignoring plots$^c$ | $-$ | $-$ | $+$ | $-$ | 1,516.98 | 0.3565 |
| CRD$^a$ model accounting for plots$^c$ | $-$ | $+$ | $+$ | $-$ | 1,486.66 | 0.4121 |
| RCBD$^a$ model ignoring plots$^c$ | $+$ | $-$ | $+$ | $-$ | 1,470.43 | 0.3201 |
| RCBD$^a$ model accounting for plots | $+$ | $+$ | $+$ | $-$ | 1,467.14 | 0.3509 |
| Purely spatial$^a$ model, no nugget | $-$ | $-$ | $-$ | $+$ | 1,469.03 | 0.3455 |
| Purely spatial$^a$ model, with nugget | $-$ | $-$ | $+$ | $+$ | 1,467.36 | 0.3320 |
| RCBD model accounting for plots, spatial$^a$ add-on | $+$ | $+$ | $+$ | $+$ | 1,468.33 | 0.3571 |
| Purely spatial$^a$ model, with nugget, anisotropic | $-$ | $-$ | $+$ | $+$ | 1,458.35 | 0.3236 |
| RCBD model accounting for plots, spatial$^a$ add-on, anisotropic | $+$ | $+$ | $+$ | $+$ | 1,455.72 | 0.3521 |
| Purely spatio-temporal$^a$ model, anisotropic, with nugget | $-$ | $-$ | $+$ | $+$ | 1,453.01 | 0.4732 |

Note: Data were log-transformed and then multiplied by 10. Analysis is based on subplot data. AMF, arbuscular mycorrhizal fungal; OTU, observed taxonomical unit.

$^a$CRD, completely randomized design; RCBD, randomized complete block design.

$^b$s.e.d., standard error of a difference between date means computed using the second-order Kenward-Roger method (Kenward & Roger, 2009; see Box S2).

$^c$These models do not provide valid analyses as they neither have effects for blocks and/or plots, nor do they have a spatial covariance structure to make up for this lack.

$^d$The spatial model considered here is the isotropic power model with covariance $\sigma_x \rho_x^d$. When the model is labelled as anisotropic, we fitted the anisotropic power model with covariance $\sigma_x \rho_x^d \rho_y^d \rho_t^d$. The anisotropic spatio-temporal model has covariance $\sigma_x \rho_x^d \rho_y^d \rho_t^d$, where $x$, $y$ and $t$ are the two spatial and the one temporal coordinates (in metres and days, respectively) and $d_x$, $d_y$ and $d_t$ are the corresponding Euclidean distances.

Example 2: An on-farm trial with oilseed rape was performed at the experiment station Ihinger Hof of the University of Hohenheim using a randomized complete block design with six replicates (Piepho, Möhring, Pfugfelder, Hermann, & Williams, 2015). There were three different tillage methods (mulch seeding, and strip tillage with or without stubble cleaning) and two fertilizer methods (subsoil fertilization and broadcasting). Five of the six possible factorial treatment combinations were tested (Table 7), whereas one combination was omitted because it was not deemed agronomically relevant.

Yield was recorded on all plots using an online monitoring system, providing a total of 4,569 spatially referenced measurements per plot (WGS84). The plot size was 1,500 m$^2$. At each position, a soil fertility score called Ackerzahl was recorded, which accounts for several environmental factors such as soil type, climatic conditions and landscape characteristics (Schachtschabel, Blume, Hartge, & Schwertmann, 1976). The score ranges from 0 to 100, with large values indicating high fertility. The objective of this experiment was to assess the effect of the five treatments as well as of the soil fertility score on yield. The variables available for analysis are shown in Table 8.

Why we need a mixed model: Similar to Example 1, we have two aspects of interest in our dataset: we would like to know the effect of the treatments (tillage and fertilization) on the yield, and we have a spatial heterogeneity (represented as the fertility score) which affects the yield. As a point of departure, to compare treatments only, we may consider an ANOVA-type linear model of the form:

$$y_{ijkh} = \mu + r_k + \tau_j + \phi_i + (\tau \phi)_{ij} + e_{ijkh}, \quad (6a)$$

where $r_k$ is the effect of the $k$-th replicate, $\tau_j$ is the main effect of the $j$-th tillage method, $\phi_i$ is the main effect of the $i$-th fertilization method, $(\tau \phi)_{ij}$ is the $ij$-th interaction effect, and $e_{ijkh}$ is a random residual error for the $h$-th observation on the $ijk$-th field plot. Using the variables in Table 8, the same model can be represented equivalently as:

$$\text{YIELD} = \text{REP} + T + F + T \cdot F, \quad (6b)$$

where REP stands for $r_k$ (REP $\equiv r_k$), $T \equiv \tau_j$, $F \equiv \phi_i$, and $T \cdot F \equiv (\tau \phi)_{ij}$. This symbolic notation is akin to the
Table 6: Means for six dates for total AMF OTU richness

| Date       | Time (in days) | Mean abundance$^\text{§}$ |
|------------|----------------|---------------------------|
| 5 April    | 0              | 30.30$^b$                 |
| 17 May     | 42             | 31.33$^b$                 |
| 27 June    | 83             | 31.94$^b$                 |
| 16 August  | 123            | 30.75$^b$                 |
| 5 October  | 173            | 34.15$^a$                 |
| 21 November| 220            | 31.32$^b$                 |

Average least significant difference ($\alpha = 5\%$)$^\text{§}$ 1.85

Note: Data were log-transformed and then multiplied by 10 (Example 1). Analysis based on RCBD model accounting for plots (see Table 5). AMF, arbuscular mycorrhizal fungal; OTU, observed taxonomical unit.

§ Means followed by a common letter are not significantly different according to a t-test at a significance level of $\alpha = 5\%$, based on subplot data assuming a linear mixed model with covariance $\sigma^2_{ij} \rho^2_{xy} \rho^2_{yt}$, where $x$, $y$, and $t$ are the two spatial and the one temporal coordinates (in metres and days, respectively) and $d_x$, $d_y$, and $d_t$ are the corresponding Euclidean distances.

§ Average least significant difference was computed for every pairwise comparison separately using the second-order Kenward-Roger method (Kenward & Roger, 2009; see Box S2) and then averaged across all pairs.

Selecting the variance–covariance structure for the random effects: By including a random effect for plots in the model, a positive covariance among all measurements taken on the same plot is modelled, the covariance being equal to the variance of plot effects. In such a model, all pairs of points in the same plot are equally correlated – no matter their distance from each other, which may not be a tenable assumption. It may be more realistic – for example, if plots are large – to assume that covariance is a decreasing function of spatial distance. For example, the covariance could decrease linearly, exponentially, or according to a power function with distance. As a result, there are many spatial or geostatistical variance–covariance models suitable for this purpose, including the spherical, Gaussian, power and exponential models (Isaacs & Srivastava, 1989; Pinheiro & Bates, 2000, Table 5.2 on p. 232; Diggle & Ribeiro, 2007; Webster & Oliver, 2007). Box 2 gives an exemplary explanation of the Gaussian model (Figure 6), illustrating that a model can be parameterized in different ways, for example, in terms of the covariance itself or in terms of a semivariogram. The latter parameterization involves geostatistical terms such as sill, partial sill, nugget effect, range, semivariance function and correlation function (see Section “Some basics about spatial statistics”). Different software packages use different parameterizations of spatial covariance models, so the purpose of Box 2 is to clarify the relations between these different parameterizations and statistical terms.

The plot effects $p_{ijk}$ in (7) are assumed to have a normal distribution with zero mean and variance $\sigma^2_p$. These effects are assumed to be independent between plots. Again, this independence assumption is justified because of the randomized allocation of treatments to plots (Hinkelmann & Kempthorne, 1994). Invoking the same randomization argument, we can assume errors $e_{ijkh}$ to be independent between plots, but correlated within plots. Alternatively, we could assume that spatial covariance is also present between errors in different plots, which is not pursued here (but see Example 1).

When both the random plot effect and the residual errors are independent between plots, the responses from different plots are independent as well. They then correspond to “subjects” in terminology that is commonly used with repeated-measures designs, where repeated
Box 2  The Gaussian model of spatial covariance and some alternative parameterizations

Different statistical software packages use different parameterizations (or ways of expressing the same models), which often makes it difficult to compare results. Here we show how linear mixed model outputs from SAS PROC MIXED and the R packages “nlme” and “geoR” can be compared. As an illustration, we use a spatial model with a Gaussian covariance structure.

According to the Gaussian model, the covariance between errors $f_1$ and $f_2$ of two measurements is:

$$\text{cov}(f_1, f_2) = \sigma^2_s \exp(-d_{12}^2/\theta), d_{12} > 0,$$  
(B1)

where $d_{12}$ is the Euclidean distance between the two measurements, $\sigma^2_s$ is a spatial variance component, and $\theta$ is a parameter determining how quickly the covariance decays with distance (it is therefore denoted as the range parameter). The variance of an error $f_1$, corresponding to the covariance of two observations when $d_{12} = 0$, equals:

$$\text{var}(f_1) = \sigma^2 = \sigma^2_s + \sigma^2_e,$$  
(B2)

where $\sigma^2_e$ is the nugget variance. Note that (B2) corresponds to a decomposition of the $h$-th error $f_h$ according to:

$$f_h = s_h + e_h,$$  
(B3)

where $s_h$ is a spatially correlated component (with variance $\sigma^2_s$) and $e_h$ is the independent measurement error (with variance $\sigma^2_e$). For the Gaussian model the semivariance for $d_{12} > 0$ can be derived as:

$$\frac{1}{2} \text{var}(f_1 - f_2) = \frac{1}{2} [\text{var}(f_1) + \text{var}(f_2)] - 2 \text{cov}(f_1, f_2) = \sigma^2_s [1 - \exp(-d_{12}^2/\theta)] + \sigma^2_e,$$

$$= \sigma^2_s \gamma(d_{12}, \theta) + \sigma^2_e,$$  
(B4)

where $\gamma(d, \theta) = 1 - \exp(-d^2/\theta)$ is the semivariogram model (Pinheiro & Bates, 2000, p. 230; also see Table 5.2 on p. 232 for typical alternative choices of $\gamma(d, \theta)$). Thus, the semivariance is a nonlinear function of spatial distance $d_{12}$ having intercept equal to the nugget variance $\sigma^2_e$ and increasing asymptotically to the sill $\sigma^2 = \sigma^2_s + \sigma^2_e$. The difference between the sill and the nugget, denoted as partial sill, is equivalent to the spatial variance $\sigma^2_s$.

The semivariogram model is one way to express or parameterize the spatial covariance model, but there are other possibilities. For example, it may be defined in terms of the correlation function $h$ that is continuous in $d$ (Pinheiro & Bates, 2000, p. 231), so that the correlation at $d = 0$ can be included:

$$h_{\text{nugget}}(d, c_0, \theta) = \begin{cases} (1 - c_0)h(d, \theta), & d > 0 \\ 1, & d = 0, \end{cases}$$  
(B5)

where

$$h(d, \theta) = 1 - \gamma(d, \theta),$$  
(B6)

and $c_0 (0 \leq c_0 \leq 1)$ is referred to as a nugget effect (this is not the same as the nugget variance $\sigma^2_e$ in (B4), but note that other authors refer to the variance $\sigma^2_e$ as the nugget effect (e.g., Diggle & Ribeiro, 2007, p. 56)). The covariance is given by $\sigma^2 h_{\text{nugget}}(d, c_0, \theta)$, where $\sigma^2$ is the variance of an observation (it’s also the sill in the semivariogram; see above). The corresponding semivariogram model (with nugget included) can be written as (Pinheiro & Bates, 2000, p. 232):

$$\gamma_{\text{nugget}}(d, c_0, \theta) = \begin{cases} c_0 + (1 - c_0)\gamma(d, \theta), & d > 0 \\ 0, & d = 0. \end{cases}$$  
(B7)

We can make the following one-to-one mapping of the two parameterizations regardless of the chosen semivariance function $\gamma(d, \theta)$ (see Pinheiro & Bates, 2000, p. 232, Table 5.2 for common choices):

$$\sigma^2 = c_0 \sigma^2_s \text{ and}$$  
(B8)

$$\sigma^2_s = (1 - c_0)\sigma^2.$$

The inverse relations are:

$$\sigma^2 = \sigma^2_s + \sigma^2_e \text{ and}$$  
(B10)

$$c_0 = \frac{\sigma^2_e}{\sigma^2_s + \sigma^2_e}.$$  
(B11)
measures are taken on the same experimental unit (subject) at different points in time (Piepho & Edmondson, 2018). In such designs, a linear regression on time is often used to model time trends, in which case there is a random intercept and a random slope for each subject (Verbeke & Molenberghs, 2000). The random plot effect in (7) corresponds to a random intercept term if plots are regarded as independent subjects, but the model has no corresponding slope term as there is no time covariate. We mention these connections to repeated measures because some mixed-model packages (e.g., “nlme” and “lme4” in R) require modelling the plot effect as an “intercept”, while omitting the slope term, and report the corresponding variance estimate accordingly (see R companion, Box R1), which will be confusing to the novice not familiar with this background.

In order to model spatial correlation among observations in the same plot, we insert a spatially correlated effect $s_{ijkh}$ into the model (7):

$$y_{ijkh} = \mu + r_k + r_j + \phi_i + (\phi \phi)_{ij} + p_{ijk} + s_{ijkh} + e_{ijkh},$$

(8)

where the variance $\sigma^2_e$ of the independent error effect $e_{ijkh}$ plays the role of a nugget effect in geostatistical parlance. In order to select a model for spatial covariance, we here use the fixed-effects specification in (8), adding linear and quadratic terms for the covariate AZ (see Box 3 and Table 11), because preliminary scrutiny of the data revealed that a quadratic model might give a good fit. The extended fixed part of the model can be written as:

$$\text{expected YIELD} = \text{REP} + T + F + T \cdot F + AZ + T \cdot AZ$$

$$+ F \cdot AZ + T \cdot F \cdot AZ + AZ^2 + T \cdot AZ^2 + F \cdot AZ^2 + T \cdot F \cdot AZ^2.$$  

(9)

FIGURE 5  Kriged maps for the six time-points of Example 1 using a spatio-temporal mixed model with nugget and anisotropic power model for the trend effects $s_{ijk}$, with covariance given by $\sigma^2_s \rho^x_{x,x} \rho^y_{y,y} \rho^t_{t,t}$, where $x$, $y$ and $t$ are the two spatial and the one temporal coordinates (in metres and days, respectively) and $d_x$, $d_y$ and $d_t$ are the corresponding Euclidean distances. The colour scales are allowed to be different between subplots so that the colour range is about the same in each subplot. BLUP, best linear unbiased prediction [Color figure can be viewed at wileyonlinelibrary.com]
Box 3  Polynomial regression in factorial experiments

Consider a regression on the quantitative variable AZ. A linear regression can be written as:

\[ \eta(AZ) = \beta_0 + \beta_1 AZ, \tag{B12} \]

where \( \eta(AZ) \) is the expected response for given value of AZ, \( \beta_0 \) is the intercept and \( \beta_1 \) is the slope. The model assumes that the relationship between expected response \( \eta(AZ) \) and AZ is linear. If the response is non-linear, higher-order polynomial terms may be added. For example, a quadratic polynomial can be written as:

\[ \eta(AZ) = \beta_0 + \beta_1 AZ + \beta_2 AZ^2, \tag{B13} \]

where \( \beta_2 \) is the regression coefficient for the quadratic term \( AZ^2 \). It is an important principle (the so-called marginality principle; Nelder, 2000; Piepho & Edmondson, 2018) in polynomial regression for a model with polynomial term up to the \( p \)-th degree, all lower-order terms must be retained as well. In the quadratic model this means we cannot drop the linear term. This is because a model of the form \( \eta(AZ) = \beta_0 + \beta_2 AZ^2 \) would have an optimum at \( AZ = 0 \). Such a response is not usually biologically plausible. For the optimum to be located away from the point \( AZ = 0 \), the linear term is required. Similarly, we do not omit an intercept term when the linear term is significant because a regression through the origin is not usually plausible. Note that the intercept can be regarded as the polynomial term of order zero, as \( \beta_0 AZ^0 = \beta_0 \times 1 = \beta_0 \).

Model (B13) represents a single curve. If the experiment involves treatment factors, there may be a separate curve for each treatment level. For example, with factors tillage and fertilization method, there can be a separate curve for each combination of tillage and fertilizer method. Hence, all polynomial terms need to be indexed by both factors. Using subscripts \( i \) and \( j \) for fertilizers and tillage methods, respectively, the extended quadratic model is:

\[ \eta_{ij}(AZ) = \beta_{0ij} + \beta_{1ij} AZ + \beta_{2ij} AZ^2. \tag{B14} \]

For each of the polynomial terms, we may use a two-way factorial model, as in model (6). Thus, we would assume:

\[ \beta_{0ij} = \mu_0 + \tau_{0i} + \phi_{0i} + (\tau \phi)_{0ij} \tag{B15} \]

\[ \beta_{1ij} = \mu_1 + \tau_{1j} + \phi_{1i} + (\tau \phi)_{1ij} \tag{B16} \]

\[ \beta_{2ij} = \mu_2 + \tau_{2j} + \phi_{2i} + (\tau \phi)_{2ij}. \tag{B17} \]

Plugging these sub-models into (B14), we obtain this model:

\[ \eta_{ij}(AZ) = \mu_0 + \tau_{0j} + \phi_{0i} + (\tau \phi)_{0ij} + \mu_1 AZ + \tau_{1j} AZ + \phi_{1i} AZ + (\tau \phi)_{1ij} AZ + \mu_2 AZ^2 + \tau_{2j} AZ^2 + \phi_{2i} AZ^2 + (\tau \phi)_{2ij} AZ^2. \tag{B18} \]

In symbolic form, and adding a replicate effect, this model can be written as given in Equation (9).

Table 9 shows the results for different spatial models with and without nugget variance when errors are assumed to be independent between plots. There is a strong indication that a nugget variance is needed. The Gaussian model with nugget fits best because it has the smallest value of AIC. The spatial coordinates had to be rescaled in some cases in order to achieve convergence. For details regarding trouble-shooting for numerical problems in fitting spatial models see Piepho et al. (2015). The parameter estimates for the Gaussian model with nugget and random plot effect are shown in Table 10, where we compare output from different packages.

Selecting the fixed effects: Having selected a suitable variance–covariance model (i.e., the Gaussian model with nugget and a random plot effect), we now turn to selecting the fixed-effects part of the model. This is done by sequential Wald-type F-tests (Type I tests in SAS; Piepho & Edmondson, 2018). Observing the marginality principle (Nelder, 2000), we fit main effects before the corresponding interactions and lower-order polynomial terms before higher-order terms. Generally in mixed models, the denominator degrees of freedom must be approximated (Halekoh & Højsgaard, 2014; Kenward & Roger, 1997; Kuznetsova, Brockhoff, Christensen, & Podenphant Jensen, 2019). Here, we use the second-order method of Kenward and Roger (2009) (Box S2 in Supporting Information). The resulting Wald-type F-tests for the fixed effects in model (9) are shown in Table 11.

It emerges that there is no interaction between T and F (terms T•F, AZ•T•F and AZ²•T•F are all non-significant). The regression on AZ requires a linear term for
expected \( \text{YIELD} = \text{REP} + \text{T} + \text{F} + \text{T} \cdot \text{AZ} + \text{F} \cdot \text{AZ} + \text{F} \cdot \text{AZ}^2 \).

We note that the main effects \( \text{AZ} \) and \( \text{AZ}^2 \) were omitted here because they are absorbed by the factorial terms \( \text{T} \cdot \text{AZ}, \text{F} \cdot \text{AZ} \) and \( \text{F} \cdot \text{AZ}^2 \).

**Obtaining the parameter estimates:** The parameter estimates for this model are shown in Table 12. Note that we have adhered to the marginality principle here, which requires for a polynomial of given degree to retain all lower-order terms regardless of their significance (Box 3). Thus, for example, having decided to go for a quadratic model including the term \( \text{F} \cdot \text{AZ}^2 \), we retain the lower-order terms \( \text{F} \cdot \text{AZ} \) and \( \text{F} \) even though they are both non-significant. This ensures that the maximum or minimum of the fitted quadratic curve is allowed to be removed from the point \( \text{AZ} = 0 \). Table 12 shows parameter estimates obtained with two different packages (SAS procedure MIXED and R-package “nlme”). The two polynomial regression terms agree, but all other terms have different estimates, due to differences in restrictions imposed on the parameter estimates. These differences are only due to the parameterization, and the fits are entirely equivalent as discussed in Box 4. The fitted model is plotted in Figure 7. Predictions were computed for the average across replicate effects (Box 4). There is a pronounced curvature for broadcast fertilization (F2), whereas the curve is almost linear for subsoil fertilization (F1). Usually, one would expect yield to increase monotonically with the fertility score, but apparently for broadcast fertilization the crop was not able to make full use of the improved fertility. In fact, the weather after sowing was particularly dry and based on visual impression the broadcast fertilizer was not very effective (Wilfried Hermann, Universität Hohenheim, personal communication).

**7 | FINAL REMARKS**

Some authors have warned users of the risk of numerical instability when using the Gaussian model for kriging (Webster & Oliver, 2007, p. 119). We did not meet such difficulties with Example 2. Goovaerts (1998) suggested

### TABLE 8 Variables available for analysis of Example 2

| Variable | Description |
|----------|-------------|
| REP      | Replicate, complete block |
| PLT      | Plot |
| LAT      | Latitude |
| LONG     | Longitude |
| ID       | ID of measurement position |
| AZ       | Ackerzahl; fertility score ranging from 0 to 100 |
| T        | Tillage method |
| F        | Fertilization method |
| YIELD    | Yield (t/ha) |

**FIGURE 6** Covariance and semivariance function (× 10²) for Gaussian model with nugget using the fitted parameter values for Example 2: \( \theta = 1.33, \sigma_t^2 = 0.0465, \sigma_e^2 = 0.0656 \) (for explanation of parameters see Box 2). The functions are drawn only for distances >0. At a distance of zero, there is a discontinuity (not shown in the figure), with the covariance jumping to \( \sigma_t^2 + \sigma_e^2 = 0.0465 + 0.0656 = 0.1121 \) (× 10²) and the semivariance dropping to zero (see Box 2 and Pinheiro & Bates, 2000, p. 231) [Color figure can be viewed at wileyonlinelibrary.com]

factor T (tillage method) and a linear and quadratic term for factor F (fertilization method). Thus, the final fixed-effects part of the model is:
Why do SAS PROC MIXED and R “nlme” and “lme4” print different parameter estimates? Parameterizations of fixed effects for qualitative factors in linear models and computation of marginal models

Linear models for qualitative factors are overparameterized, meaning there is no unique solution for the effects. Also, different packages may give widely different numerical solutions for the same model, which may be a source of confusion. Luckily, however, these differences are nothing to be worried about, as all of these solutions are entirely equivalent as regards quantities of interest such as treatment means, treatment differences, predicted values in regressions, etc.

To illustrate, consider a simplified version of model (9), written in classical form as:

$$y_{ih} = \mu + \phi_i + e_{ih}, \quad (B19)$$

where $y_{ih}$ is the yield of the $h$-th observation of the $i$-th fertilizer treatment, $\mu$ is a general intercept, $\delta_i$ is the $i$-th fertilizer effect, and $e_{ih}$ is an error term associated with $y_{ih}$. The tangible quantities of interest in this model are the two treatment means (i.e., the average yield for treatment 1, $\eta_1 = \mu + \phi_1$, and for treatment 2, $\eta_2 = \mu + \phi_2$). The model is overparameterized because there are only two means ($\eta_1$ and $\eta_2$), but three parameters ($\mu$, $\phi_1$ and $\phi_2$). This is solved by imposing a restriction on the parameter estimates, such as setting a treatment effect equal to zero (R packages set the first of the treatment effects to zero, whereas SAS packages set the last one to zero.). Whichever restriction we impose, the resulting treatment means are the same. For illustration, assume that the means are $\eta_1 = 50$ and $\eta_2 = 60$. When we impose the restriction $\phi_1 = 0$, these means are obtained with $\mu = 50$ and $\phi_2 = 10$. When the restriction is $\phi_2 = 0$, we must have $\mu = 60$ and $\phi_1 = -10$ to obtain the same means.

There are subtle differences in the way these restrictions are implemented in different packages, and we mention these differences because they do have important consequences as regards the printed output. R packages “nlme” and “lme4” effectively drop the effect(s) that are set to zero from the model, thereby rendering the design matrix for fixed effects, usually denoted as $X$, of full column rank, implying that regular matrix inverses can be used to solve the normal equations/mixed model equations for the effects remaining after dropping the effects that have been set to zero. The important practical consequence of this approach is that the effects that are set to zero do not appear in the printed output. By contrast, SAS retains all parameters and uses so-called generalized inverses to solve the normal/mixed model equations. Thus, solutions are printed for all parameters, including the ones that are set to zero. This is seen in Table 12, where the zero effects are represented by dashes for the output of “nlme”, whereas there are printed zeros for such effects on the output of SAS/MIXED.

Now consider the fitted model (9) for Example 2. It is readily verified from Table 10 that, despite the numerically very different solutions, the important quantities derived from these solutions are the same, including effect differences between factor levels. For example, the difference between the effects of replicates 2 and 3 (REP2-REP3) equals 0.042 with both sets of solutions. Similarly, $F1-F2 = 3.0816$ and $AZ+F1-AZ+F2 = -0.1168$ with both solutions.

In classical notation the model (10) has a separate intercept for each replicate and treatment combination, given by $\mu_0 + r_k + t_0j + \phi_0i$. To report a single model per treatment combination, it is useful to average this across replicate effects, yielding the mean intercept $\mu_0 + \bar{r} + \bar{t}_0j + \bar{\phi}_0i$, where $\bar{r} = (r_1 + r_2 + r_3 + r_4 + r_5 + r_6)/6$. This average intercept was used to draw the five curves for the five treatment combinations in Figure 7. Again, it is readily verified that the mean intercepts are the same with both sets of solutions printed in Table 12.

Either fitting a small nugget effect to evade extreme interpolative properties and artefacts or not using the Gaussian model for kriging at all. We would suggest generally fitting a nugget effect in mixed model implementations and have done so in both examples. In our experience, a nugget effect is very often needed, and if it is not, the REML algorithm is likely to converge to zero or some negligible value for the nugget (Piepho et al., 2015). The Matérn model may be considered as a viable and flexible alternative to the Gaussian model. The Matérn correlation function is given by:
where $\phi > 0$ is the range parameter, $\nu > 0$ is the smoothness parameter, $\Gamma(\cdot)$ is the gamma function, and $K_\nu(\cdot)$ is the modified Bessel function of the third kind of order $\nu$ (Diggle & Ribeiro, 2007, p. 51–53). This model has well-known limiting cases, such as the exponential ($\nu = \frac{1}{2}$) and the Gaussian ($\nu \rightarrow \infty$). The model is notoriously difficult to fit because the parameters are far from orthogonal. Diggle & Ribeiro (2007, p. 114) therefore suggest profiling the likelihood over a grid of values for the smoothness parameter $\nu$. We did this for the data in Example 2, using the grid $\nu = 1, 2, ..., 15$ and found a fit of $\text{AIC} = 3,140.8$ ($\nu = 1$) for the model without nugget and $\text{AIC} = 2,058.3$ ($\nu = 5$) for the model with nugget. The fit for the model with nugget is marginally better than the Gaussian model in Table 9 (a difference smaller than two points, indicating equally good fit; see Box 1). Provision of good starting values was essential to get the Matérn models to converge. In this example, the Gaussian model may be preferred because it is simpler to fit and only a little different from the Matérn model.

Randomization in designed experiments justifies the assumption of independence among experimental units in ANOVA. This paper considered two experiments where several observations were made per experimental unit. Such subsampling data cannot be modelled as independent observations but require accounting for correlation among observations made on the same experimental unit. Geostatistical covariance models can be used for this purpose, as illustrated using two examples. It should be stressed, however, that although geostatistical modelling can to some extent make up for lack of randomization, such modelling is no substitute for randomization.

| TABLE 9 | Selection of variance–covariance model for Example 2 assuming the fixed-effects specification (9) (see Table 10) |
|----------|------------------------------------------------------------------------------------------------------------------|
| Spatial variance–covariance model | AIC No nugget | With nugget |
| Independent | – | 3,250.9 |
| Gaussian | 3,186.3$^c$ | 2059.6$^b$ |
| Spherical | 3,174.2$^b$ | 2,104.5 |
| Power | 3,068.2$^b$ | 2085.0$^b$ |
| Exponential$^a$ | 3,068.2$^b$ | 2085.0 |

Note: In all spatial models, covariance is assumed to be present between measurements taken on the same plot, but observations from different plots are assumed to be independent. All models include a random plot effect as in Equation (8). Models fitted with the MIXED procedure of SAS. AIC, Akaike information criterion.

$^a$Covariance model $\sigma^2\exp(-d/\theta)$; equivalent to power model ($\sigma^2/d^q$) with $\theta = \exp(-1/\theta)$.

$^b$The spatial coordinates LAT and LONG had to be multiplied by $10^4$ to achieve convergence.

$^c$The spatial coordinates LAT and LONG had to be multiplied by $10^2$ to achieve convergence.

| TABLE 10 | Relating the output for Example 2 from different packages to the two parameterizations described in Box 2 |
|----------|---------------------------------------------------------------------------------------------------------------|
| Parameter | Designation in output of package | Parameter estimates$^e$ |
| $\sigma^2_f$ | – | Intercept | – | 0.0121 | 0.0121 |
| $\sigma^2_e$ | Residual | – | Nugget | 0.0656 | – |
| $\sigma^2$ | Variance | – | Partial sill | 0.0466 | – |
| $\theta$ | SP (GAU)$^a$ | Range | Range | 1.3351 | 1.3351 |
| $c_0$ | – | Nugget | – | – | 0.5846 |
| $\sigma^2$ | – | Residual$^b$ | – | – | 0.1122 |
| $\log L_R$ | – | – | – | $-1,003.15$ | $-1,003.13$ |
| AIC$^{cd}$ | – | – | – | 2,014.3 | 2,044.3 |

Note: Note that $\sigma^2_f = c_0\sigma^2$ and $\sigma^2_e = (1-c_0)\sigma^2$ (see Box 2). Results are for the Gaussian model of spatial covariance, assuming that errors are independent between plots. Note that “geoR” reports the nugget variance $\sigma^2$ as “nugget” and “nlme” reports the nugget effect $c_0$ as “nugget”.

$^a$This is the label printed with the Gaussian model – it has a different name if a different covariance structure is selected.

$^b$The “nlme” package reports this as a standard deviation, which is the square root of the variance.

$^c$“geoR” and “geoRglm” cannot fit random intercepts and they cannot model errors to be independent between plots (subjects), so no results can be reported here for these R packages.

$^d$In SAS, $q = 4$ is the number of variance parameters. In R package “nlme”, the parameter count includes the fixed effects: $q = 19$. 

\[
\rho(d; \phi, \nu) = \left\{2^{\nu-1}\Gamma(\nu)\right\}^{-1} \left(\frac{d}{\phi}\right)^\nu K_\nu\left(\frac{d}{\phi}\right),
\]
TABLE 12  Parameter estimates for final model (10) using the MIXED procedure of SAS\(^a\) and the lme() function of the "nlme" package in R

| Parameter\(^b\) | MIXED procedure (SAS) | lme() function of nlme package (R) |
|-----------------|-------------------------|-----------------------------------|
|                 | Estimate | Standard error | Estimate | Standard error |
| Intercept       | -0.7638  | 0.6863          | 2.7581   | 1.1026         |
| REP1            | 0.1712   | 0.0812          | -        | -               |
| REP2            | 0.2416   | 0.0813          | 0.0704   | 0.0789          |
| REP3            | 0.3054   | 0.0814          | 0.1341   | 0.0790          |
| REP4            | 0.3686   | 0.0815          | 0.1974   | 0.0791          |
| REP5            | 0.2752   | 0.0816          | 0.1040   | 0.0793          |
| REP6            | 0        | -               | -0.1712  | 0.0812          |
| T1              | 0.2691   | 0.3054          | -        | -               |
| T2              | -0.5637  | 0.2556          | -0.8328  | 0.3074          |
| T3              | 0        | -               | -0.2691  | 0.3030          |
| F1              | 3.0816   | 1.2500          | -        | -               |
| F2              | 0        | -               | -3.0816  | 1.2367          |
| AZ-T1           | 0.1468   | 0.0267          | 0.0300   | 0.0412          |
| AZ-T2           | 0.1647   | 0.0252          | 0.0479   | 0.0406          |
| AZ-T3           | 0.1515   | 0.0255          | 0.0347   | 0.0411          |
| AZ-F1           | -0.1168  | 0.0477          | -        | -               |
| AZ-F2           | 0        | -               | 0.1168   | 0.0471          |
| AZ-AZ-F1        | -1.0001  | 0.0004          | -0.0001  | 0.0004          |
| AZ-AZ-F2        | -0.0012  | 0.0002          | -0.0012  | 0.0002          |

\(^a\)Identical estimates are obtained with the lme() function of the nlme package in R when the SAS parameterization is used. This can be invoked using options (contrasts = r(ordered = "contr.SAS", ordered = "contr.poly"));

\(^b\)Factors described in Table 8.
Even when only a single observation is modelled per experimental unit and the independence assumption is justified based on randomization (Webster & Lark, 2019), it may still be beneficial to fit a spatial model, as this may improve efficiency (Gilmour et al., 1997). The potential for improved efficiency in comparative experiments stems from the fact that inference is based on pairwise differences among observations and that the variance of a pairwise difference among two random variables \( Y_1 \) and \( Y_2 \) decreases linearly with the spatial covariance (i.e., \( \text{var}(Y_1 - Y_2) = \text{var}(Y_1) + \text{var}(Y_2) - 2\text{cov}(Y_1, Y_2); \) also see Equation (1)). If such spatial modelling turns out to not provide an improved fit, or is deemed to be too complex, the randomization-based model can always be used as a fallback option for analysis, providing an unbiased estimation of treatment effects. Thus, in designed experiments the randomization principle should be adhered to whenever possible, even when use of a geostatistical covariance model is envisioned for analysis (de Gruijter et al., 2006; Williams & Piepho, 2013).

When contemplating the use of spatial models for analysis of a designed experiment, it should be kept in mind that switching from a design- (randomization-) to a model-based analysis will change results. Although the former entails assumptions, these are relatively straightforward and well known (Webster & Lark, 2019). The use of a spatial model imports further assumptions — stationarity of the spatial correlation, in particular. Thus, estimates of treatment means may differ slightly between both analyses, especially when spatial correlation is present over a distance similar to or longer than the length of an experimental field. Furthermore, as discussed by de Gruijter et al. (2006), the standard errors of means behave differently. This is exemplified by the different values for the s.e.d. in Table 5 (Example 1). Furthermore, with spatial analysis there is difference between estimates of the “model mean” (e.g., \( \mu + \tau \) in Equation (5)) (this is equivalent to best linear unbiased estimation) and the “spatial mean”, that is, the prediction of realizations of a random variable across a defined area (de Gruijter et al., 2006) (this is equivalent to BLUP). Here, we have solely considered the former, which is relevant for assessing treatment differences in a designed experiment.

This paper has focused on randomized designed experiments, but it should be emphasized that surveys and observational studies have a similar theme, with random sampling replacing the role of randomization. As with randomization in designed experiments, random sampling justifies the independence assumption in analysis (de Gruijter et al., 2006; de Gruijter & ter Braak, 1990). When sampling is not random but purposive or systematic, spatial modelling, integrated within a linear mixed modelling framework, may be used for a valid analysis (Lark & Cullis, 2004; Wiesmeier et al., 2012, 2013). It should be emphasized, however, that as with randomized experiments, surveys should be designed using the principle of random sampling whenever feasible, because this allows a design-based analysis, which has the virtue of simplicity and robustness. Also, adherence to the random sampling principles minimizes the risk of biased estimates of effects of interest. As with randomized experiments, model-based spatial analysis can always be tried, and this has the potential to improve precision over the purely design-based analysis, but should a suitable spatial model be difficult to identify, one can always use the design-based analysis.

Time series data (longitudinal data, repeated measurements) have commonalities with spatial data in that correlation among observations made at different time-points needs to be modelled. In fact, the models used for time series data bear much resemblance to models used for spatial data, and spatial covariance models are well suited to model temporal correlation, especially when time-points are not equally spaced (Pinheiro & Bates, 2000). Moreover, spatial and temporal covariance can be combined into a spatio-temporal model as shown in Example 1 (Figure 5). Here, we have only used the anisotropic power model for spatio-temporal mixed modelling. More sophisticated mixed models of this type are considered, for example, in Schabenberger &
Gotway (2004, Chapter 9) and de Faveri, Verbyla, Cullis, Pitchford, and Thompson (2017).

In Example 2 we have considered data obtained from yield mapping sensors based on flow-rate measurements in a combine harvester (New Holland, header width of 6 m). The analysis was performed making the simplifying assumption that monitored yields are actual yields. In reality, the monitored yield data represent a convolution of yields from different points simultaneously arriving at the recording device. Different methods may be applied to deal with this problem, such as specialized deconvolution methods (Lark & Wheeler, 2003) or preprocessing of the raw data by smoothing methods (Marchant et al., 2019).

ACKNOWLEDGEMENTS
The data in Example 1 were obtained within the DFG Priority Program 1374 “Infrastructure-Biodiversity-Exploratories”. We thank the staff of the three Exploratories, the BE office and the BExIS team for their work in maintaining the plot and project infrastructure, and Markus Fischer, the late Elisabeth Kalko, Eduard Linsenmair, Dominik Hessenmöller, Jens Nieschulze, Daniel Prati, Ingo Schöning, Francois Buscot, Ernst-Detlef Schulze and Wolfgang W. Weisser for their role in setting up the Biodiversity Exploratories project. We further thank the owners of the two datasets, Francois Buscot, Tesfaye Wubet, Kezia Goldmann and Sandra Klemmer, as well as Kathleen Regan, Doreen Berner, Sven Marhan and Ellen Kandel for conducting the SCALEMIC experiment and providing the data publicly.

Example 2 was obtained from an experiment run at the Ihinger Hof experiment station of the University of Hohenheim. We thank the staff at Ihinger Hof, Wilfried Hermann and Markus Pflugfelder in particular, for conducting this experiment and kindly providing the data.

We thank two anonymous reviewers for very helpful and constructive comments that greatly helped us improve the paper.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in the Biodiversity Exploratories Information System at https://www.bexis.uni-jena.de/PublicData/PublicData.aspx?DatasetId=22548 and at https://www.bexis.uni-jena.de/PublicData/PublicData.aspx?DatasetId=17707.
Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983–997.

Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics and Data Analysis, 53*, 2583–2595.

Koza, M., & Piepho, H. P. (2018). What’s normal anyway? Residual plots are more telling than significance tests when checking ANOVA assumptions. *Journal of Agronomy and Crop Science, 204*, 86–98.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems of the Witswatersrand. *Journal of Chemical Metallurgical and Mining Society of South Africa, 52*, 119–139.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Pødenphant Jensen, S. (2019). *lmerTest*: Tests in linear mixed effects models. R Package Version 3.1-1. Retrieved from https://CRAN.R-project.org/package=lmerTest.

Lark, R. M. (2009). Kriging a soil variable with a simple non-stationary variance model. *Journal of Agricultural Biological and Environmental Statistics, 14*, 301–321.

Lark, R. M., & Cullis, B. R. (2004). Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science, 55*, 799–813.

Lark, R. M., Cullis, B. R., & Welham, S. J. (2006). On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science, 57*, 787–799.

Lark, R. M., & Wheeler, H. C. (2003). A method to investigate within-field variation of the response of combinable crops to an input. *Agronomy Journal, 95*, 1093–1104.

Marchant, B. P., Rudolph, S., Roques, S., Kindred, D., Gillingham, G., Welham, S., ... Sylvester-Bradley, R. (2019). Establishing the precision and robustness of farmers’ crop experiments. *Field Crops Research, 230*, 31–45.

Milliken, G. A., & Johnson, D. E. (1992). *Analysis of messy data. (Vol. I) Designed experiments*. Boca Raton, FL: CRC Press.

Nelder, J. A. (2000). Functional marginality and response surface fitting. *Journal of Applied Statistics, 27*, 109–112.

Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena, 113*, 56–69.

Patterson, H. D. (1997). Analysis of series of variety trials. In R. A. Kempton & P. N. Fox (Eds.), *Statistical methods for plant variety evaluation* (pp. 139–161). London: Chapman & Hall.

Piepho, H. P. (1997). Analysis of a randomized complete block design with unequal subclass numbers. *Agronomy Journal, 89*, 718–723.

Piepho, H. P., Büchse, A., & Emrich, K. (2003). A hitchhiker’s guide to the mixed model analysis of randomized experiments. *Journal of Agronomy and Crop Science, 189*, 310–322.

Piepho, H. P., & Edmondson, R. N. (2018). A tutorial on the statistical analysis of factorial experiments with qualitative and quantitative treatment factor levels. *Journal of Agronomy and Crop Science, 204*, 429–455.

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

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

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. New York, NY: Springer.

Regan, K. M., Boeddinghaus, R. S., Baumgartner, V., Berner, D., Boch, S., ... Marhan, S. (2014). Seasonal controls on grassland microbial biogeography: Are they governed by plants, abiotic properties or both? *Soil Biology and Biochemistry, 71*, 21–30.

Regan, K. M., Nunan, N., Boeddinghaus, R. S., Baumgartner, V., Berner, D., Boch, S., ... Marhan, S. (2015). Corrigendum to ‘seasonal controls on grassland microbial biogeography: Are they governed by plants, abiotic properties or both?’. *Soil Biology and Biochemistry, 86*, 212–214.

Regan, K. M., Stempfhuber, B., Schloter, M., Rasche, F., Prati, D., Philippot, L., ... Marhan, S. (2017). Spatial and temporal dynamics of nitrogen fixing, nitrifying and denitrifying microbes in an unfertilized grassland soil. *Soil Biology & Biochemistry, 109*, 214–226.

Robinson, G. K. (1991). That BLUP is a good thing—The estimation of random effects. *Statistical Science, 6*, 15–51.

Schabenberger, O., & Gotway, C. A. (2004). *Statistical methods for spatial data analysis*. Boca Raton, FL: CRC Press.

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

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. (2000). *Linear mixed models for longitudinal data*. Berlin, Germany: Springer.

Webster, R., & Lark, R. M. (2019). Analysis of variance in soil research: Examining the assumptions. *European Journal of Soil Science, 70*, 990–1000.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). New York, NY: Wiley.

Wiesmeier, M., Hübner, R., Barthold, F., Spörlein, P., Geuß, U., Hangen, E., ... Kögel-Knabner, I. (2013). Amount, distribution and driving factors of soil organic carbon and nitrogen in cropland and grassland soils of southeast Germany (Bavaria). *Agriculture, Ecosystems & Environment, 176*, 39–52.

Wiesmeier, M., Spörlein, P., Geuß, U., Hangen, E., Haug, S., Reischl, A., ... Kögel-Knabner, I. (2012). Soil organic carbon stocks in Southeast Germany (Bavaria) as affected by land use, soil type and sampling depth. *Global Change Biology, 18*, 2233–2245.

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics, 22*, 392–399.

Williams, E. R., & Piepho, H. P. (2013). A comparison of spatial designs for field variety trials. *Australian and New Zealand Journal of Statistics, 55*, 253–258.
Wolfinger, R. D. (1996). Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural Biological and Environmental Statistics*, 1, 205–230.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Slaets JIF, Boeddinghaus RS, Piepho H-P. Linear mixed models and geostatistics for designed experiments in soil science: Two entirely different methods or two sides of the same coin? *Eur J Soil Sci.* 2021;72:47–68. https://doi.org/10.1111/ejss.12976