Journal of Agricultural Science; Vol. 14, No. 12; 2022
ISSN 1916-9752 E-ISSN 1916-9760
Published by Canadian Center of Science and Education

Parametric Models of Covariance Matrices for Repeated Measures: A Simulation Study on Fit, Error and Statistical Power in Mixed Linear Models

Lucas V. Vieira[1], Gabriela N. da Piedade[1] & Maria M. P. Sartori[1]

[1] School of Agriculture, São Paulo State University, Botucatu, Brazil

Correspondence: Lucas V. Vieira, School of Agriculture, São Paulo State University (UNESP), Botucatu, SP, Brazil. Tel: 55-14-98164-5696. E-mail: lucas@lvvieira.com

Received: September 7, 2022 Accepted: October 17, 2022 Online Published: November 15, 2022
doi:10.5539/jas.v14n12p59 URL: https://doi.org/10.5539/jas.v14n12p59

The research is financed by the National Council for Scientific and Technological Development (CNPq).

Abstract

The premise in experiments with repeated measures is that observations taken in the same experimental unit are correlated and that correlations decrease proportionally to the increase in the distance between measurements in time or space. Nevertheless, these experiments are often analyzed as if the correlations between the repeated measures were constant or using methods that only consider correlations different, which may impact on the rejection rate of the null hypothesis, and ultimately type I error rate and statistical power. In this context, this study investigated the application of mixed linear models with different assumptions about the covariance matrix in data sets from simulated experiments with repeated measures. 84 scenarios that varied in terms of the covariance matrix pattern (14 structures), number of repeated measurements (4 and 8) and sample size (4, 8 and 12) were evaluated. 10,000 data sets were simulated for each scenario based on a multivariate normal distribution and were subsequently analyzed using mixed linear models. Type I error rate and statistical power for the hypothesis test of the interaction between treatments and repeated measures were estimated as the proportion of p values less than or equal to 0.01 or 0.05 out of a total of 10,000 tests for each scenario. The models were also evaluated for their ability to fit the data using Bayesian Information Criteria (BIC). Thus, the frequency with which the covariance structures were chosen by the selection criteria was computed. Results indicate that the assumption chosen most frequently by the information criteria resulted from the specified covariance structure that corresponded to the empirical covariance structure of the analyzed data sets, particularly for those with larger number of repeated measures and sample sizes. Results also indicate that the use of covariance models that do not recognize heterogeneous correlations between repeated measures can inflate type I error or reduce it to very conservative levels, which may affect the conclusion of agricultural experiments. For a 5% significance level, type I error bias was greater than 2α, while for 1% significance level, bias was over 5α. In addition, the statistical power was reduced when the assumption about the covariance matrix of the data sets did not correspond to the empirical covariance structure, particularly for those data sets with a smaller sample size.

Keywords: covariance models, F test, robustness, correlated data, type I error, statistical power, information criteria, simulation

1. Introduction

Correlated observations are common in experiments involving biological and physical components such as in agricultural trials. Correlated structures can often arise in studies in which observations are structured in groups, in which observations belonging to the same group tend to present a greater degree of correlation than those from different groups. For example, repeated measurements in the same experimental unit inducing temporal or spatial correlations, and random effects common to several experimental units. In the case of repeated measures experiments, which are conducted to evaluate changes in time (days, months, years) or space (soil samples at different depths) and the effect of treatments on these changes, the experimental units need to be observed at two or more-time intervals or spaces during the conduct of the experiment. Thus, observations in the same experimental unit (plot) tend to be not only proportionally correlated, but also observations that are closer in time or space tend to be more correlated than those that are further away (Condorelli et al., 2018; Gbur et al., 2012; Li et al., 2016; Yang & Juskiw, 2011).

The univariate analysis of variance for repeated measures (RANOVA) based on traditional linear models allied to least squares can restrict the analysis of data from experiments with repeated measures (Kowalchuck, et al, 2004). The structure of errors can be more complex than those imposed by the sphericity assumption. The sphericity assumption is evaluated in the form of compound symmetry, a condition in which all variances of the repeated measures are equal and all pairs of correlations between them are also equal. Compound symmetry is a sufficient but not necessary condition for validating sphericity and therefore not always satisfied. Although not necessary, the lack of compound symmetry indicates that sphericity is unlikely. Although relatively robust against violations of normality, particularly when the sample size is large enough as proposed by the central limit theorem, the application of the RANOVA approach to data sets whose observations were collected with some kind of dependency structure more complex than that of homogeneous variances and homogeneous correlations can lead to loss of robustness of these procedures due to biased estimates of standard errors (Bakdash & Marusich, 2017; Haverkamp & Beauducel, 2017; Marcin et al., 2015).

Robust estimators are important to ensure that estimates from repeated measures experiments are reasonably efficient and asymptotically unbiased (bias tending to zero as the sample size approaches infinity). The concept of efficiency refers to the sampling variability of an estimator. The variance of an estimator measures the degree of spread of the estimates and its information is inversely proportional to the precision, because the smaller the variance of the estimates, the greater the precision of the information. The degree of displacement of this information is given by the bias, so that the smaller the value of the bias, the greater the accuracy of the information. An estimator will be efficient if the variance is as small as possible.

Although the RANOVA approach recognizes the existence of correlations between repeated measures, it ignores the existence of heterogeneity in the correlations by imposing a compound symmetry structure, in which pairs of observations have the same correlation regardless of how far apart they are in time or space. Thus, the risk of obtaining incorrect standard deviations for hypothesis tests and for comparisons between means of different times increases. When the sphericity assumption is violated, F-test calculations can become skewed, which can result in inflated tests, and a greater likelihood that the researcher will conclude that a result has statistical significance, when in fact it happened by chance. The consequence of this violation is positive bias or more liberal tests (the probability of a false positive exceeds that established by the researcher through the significance level α). Although adjustments can be made to tests to account for correlations dependent on repeated measures, such corrections are often inadequate (Yang & Juskiw, 2011).

Another method that has been widely used when RANOVA's assumptions are not supported is the multivariate analysis of variance (MANOVA) (Vasey & Thayer, 1987). Despite allowing heterogeneous correlations, the assumption about the structure of errors can lead to reduced statistical power, since the only assumed covariance structure is one in which all correlations are distinct, resulting in a greater number of estimated parameters (Oberfeld & Franke, 2013).

Assumptions about the structure of variance and covariance of errors in repeated measures experiments that are either very simple, as in the RANOVA approach, or very complex, as in the MANOVA approach, can affect the type I error rate and statistical power (Dixon, 2008). Furthermore, due to the estimation method involved (ordinary least squares) for estimating fixed effects, the ability to deal with missing data in both approaches is quite limited. If a repeated measurement is missing for an experimental unit, the remaining observations for that experimental unit must also be excluded, causing more information to be lost from the dataset. Therefore, the RANOVA and MANOVA approaches are not recommended for unbalanced data sets, which are frequent in agricultural experimentation (Casella & Berger, 2002).

Despite the historical importance for the development of the analysis of repeated measures experiments, the previous methods have disadvantages that have driven the development of other methods more suitable for situations in which data are collected with some dependency structure. Thus, the previously mentioned approaches for experiments with repeated measures have been gradually replaced by another one capable of incorporating different structures of covariance between observations directly in the model. Advances in statistical procedures in the last decades, particularly in the computational area, have been of great interest to agricultural research, as it has provided more suitable approaches for experiments with repeated measures, with random effects common to several experimental units as in block designs, with subdivided plots, complex bioinformatics data such as genetic and molecular studies, in addition to other situations with multiple sources of variation, resulting in the grouping of observations with some degree of correlation (Zhang et al., 2019). Currently, the recommendation has been to use mixed linear models combined with generalized least squares and restricted maximum likelihood estimators for the analysis of repeated measures experiments, allowing greater flexibility for modeling different covariance structures (Li, 2017).

Mixed models extend traditional linear models allied to ordinary least squares through more flexible and complex assumptions about errors, allowing the addition of other random effects, properly considering time-dependent correlations for repeated measurements in experimental units. The MLM approach allows the error to be partitioned between observations and group components, which take the form of additional residual terms, each with its own variance to be estimated. Furthermore, due to maximum likelihood estimation processes, mixed models deal more efficiently with unbalanced data. Still, an important attribute of mixed linear models is the existence of two sample sizes: the number of groups (N) and the number of observations per group (n). In the case of experiments involving repeated measurements, the number of measurements in time or space and the number of experimental units taken in each repeated measurement. Previous research shows that the number of groups is more important than the number of observations per group for maintaining the consistency of the estimators (Duarte & Vencovsky, 2001). On the other hand, the determination of an adequate n is fundamental to obtain the desired statistical power. Many studies lack sufficient power to make correct decisions or detect an effect when it exists (Allegue et al., 2017; Gezan & Carvalho, 2018).

Despite the intense research and the increase in popularity due to the availability in many free statistical programs such as R (R Core Team, 2015) and Python (Rossum & Drake, 2009) and companies such as SAS (SAS Institute Inc., 2015), the application of mixed linear models in the agricultural sciences is still quite timid, particularly for non-statistical researchers. The flexibility with which mixed linear models stand out from traditional linear models is the same that makes them more complex, and consequently less used. Furthermore, the related terminology is quite heterogeneous and confusing. Another problem that makes it difficult to use mixed models is the lack of a definitive method to determine the best variance-covariance structure (Demidenko, 2013).

Gbur et al. (2012) further argue that parsimony and meaning also need to be considered when selecting covariance structures. The decision is often based on comparing models with different covariance structures using the likelihood ratio test or information criteria such as AIC and BIC. Previous work suggests that optimal control of the type I error rate while maintaining a satisfactory level of statistical power to detect possible real differences in experiments would result from the specified covariance structure that corresponds to the empirical covariance structure of the dataset (Duarte & Vencovsky, 2001). However, other works suggest this control may not necessarily result from the combination between the specified covariance model and the empirical covariance matrix of the data, requiring investigation for different covariance models (Piepho et al., 2004).

One way to evaluate the effect of different assumptions on the covariance matrix in experiments with repeated measures is through simulations of a data set with certain characteristics that allow highlighting the effects of applying certain mixed models to specific covariance scenarios. For example, research reveals that the effectiveness of mixed linear models in capturing the variance and covariance structure will depend on data characteristics, estimation procedures, patterns of variance and covariance (Guerin & Stroup, 2000). Therefore, simulation studies are necessary to evaluate the quality of the estimators from mixed linear models considering different experimental scenarios to enable researchers in the agricultural and biological areas to obtain more precise and accurate results from the experiments (Gueorguieva & Krystal, 2004; Stroup, 2013).

In this context, it was hypothesized that the use of mixed linear models with inadequate assumptions about the covariance structure in experiments with repeated measures can affect the statistical power of the hypothesis tests test and change the null hypothesis rejection rate to very conservative or very conservative levels or very liberal, affecting the conclusion of the experiments. Thus, the present study had as general objective the application, evaluation, and selection of mixed linear models with different assumptions about the covariance matrix between repeated measures in data sets from simulated experiments whose covariance matrices were adjusted to different patterns. It was hypothesized that the covariance model chosen from the information criteria and the ideal type I error control corresponds to the empirical covariance structure of the data.

Structured correlation patterns reflect special assumptions about correlations between repeated measures in case they tend to decrease over time or space. Assumptions about the actual covariance structure of the data introduce the risk of choosing a pattern that is too simple, which could inflate the type I error rate, or too complex, which could lead to a loss of statistical power. Despite the studies with mixed linear models, the results are still not conclusive about the most suitable model for the different experimental scenarios, and about the consequence of not considering the structure that provides the best fit in the data analysis. Many of the studies available in the literature are specific to a few scenarios outside the agricultural context, making it difficult to apply more adequate linear mixed models by non-statistical researchers, since the application of mixed models requires the specification of parameters beyond those standards of statistical programs.

In this context, the hypotheses of the present study were: (1) Inadequate assumptions about the covariance structure of repeated measures experiments can alter the type I error rate, altering the conclusion about fixed effects. (2) Specification of very simple covariance structures for data sets with more complex structures can lead to an increase in the type I error rate in inferences about fixed effects. (3) Incorrect specification of very complex covariance structures for repeated measures can lead to loss of statistical power. (4) Specification of covariance structures with intermediate complexities, such as the AR (1) or TOEP pattern, can provide a better relationship between the null hypothesis rejection rate and statistical power. (5) An ideal type I error rate control results from the specified covariance structure that corresponds to the empirical covariance structure of the data set. (6) The number of repeated measures influences the rejection rate of the null hypothesis, as a greater number of repeated means a greater number of parameters to be estimated, and, therefore, a reduction in statistical power. (7) A greater number of repeated measurements requires a greater number of repetitions to reach adequate levels of statistical power.

To address the previous hypotheses, the present study had as its general objective the application and evaluation of mixed linear models in data sets from repeated measures experiments simulated with different experimental scenarios that varied in terms of the covariance structure of random effects, the number of repeated measures, and the sample size. The specific objective was to evaluate the effect of specifying parametric structures of variance and covariance matrices (matrix **G**) on the fit of mixed models, on the estimation of type I error through hypothesis tests on fixed effects, and on statistical power in simulated repeated measures experiments.

2. Materials and Methods

2.1 Simulation of Experimental Scenarios

To enable the evaluation of the effect of using mixed linear models with different assumptions about the covariance matrix in experiments with repeated measures, a series of data sets from simulated experiments based on random samples of a multivariate normal distribution were generated. Data sets were generated so that they had different patterns of covariance between repeated measures, different numbers of repeated measures, and different sample sizes.

The simulated experiments consisted of a completely randomized design with 2 treatments, 4 or 8 repeated measures, 4, 8 and 12 replicates per treatment and 14 patterns of covariance between repeated measures. For each scenario evaluated, 10000 data sets were generated that were later analyzed with mixed linear models with different assumptions about the empirical covariance matrix of the data and evaluated for the type I error rate and the statistical power of the hypothesis tests for the interaction between treatment and repeated measures, and the fit of the model to the data through the BIC information criteria.

2.2 Specification of the Linear Mixed Model

The formulation of the mixed linear model is shown below:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}v + \epsilon \quad (1)$$

where, $\mathbf{y}$ is the response vector, $\mathbf{X}$ is the fixed effects design matrix, $\boldsymbol{\beta}$ is the unknown fixed effects vector to be estimated, $\mathbf{Z}$ is the random effects design matrix, $\mathbf{v}$ is the unknown random effects vector, and $\boldsymbol{\epsilon}$ the random error vector.

The expectation of y was expressed by:

$$\mathrm{E}(\mathbf{y}) = \mathrm{E}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \epsilon) = \mathbf{X}\beta \quad (2)$$

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\beta, \mathbf{ZGZ}' + \mathbf{R}) \quad (3)$$

The random effects v and the errors ϵ were assumed to have a multivariate normal distribution with a mean of zero, are independent, with matrices of variances and covariances $\mathbf{G}$ (q × q) and $\mathbf{R}$ (n × n), respectively.

$$\mathrm{Var}(v) = \mathbf{G} \quad (4)$$

$$\mathrm{Var}(\epsilon) = \mathbf{R} \quad (5)$$

For repeated measures experiments, the formulation of the mixed linear model was given by:

$$y_i = \mathbf{X_i}\beta + \mathbf{Z_i}v_i + \epsilon_i \tag{6}$$

where, y is the response vector ($n_i \times 1$) for experimental unit i, **X**i is the incidence matrix for fixed effects ($n_i \times p$) for experimental unit i, and β is the coefficient vector for fixed effects ($p \times 1$), **Z**i is the incidence matrix for random effects ($n_i \times q$) for experimental unit i, v_i is the random effects vector ($q \times 1$), and ϵ_i is the error vector for experimental unit i with mean 0 and covariance $\mathbf{R}_i$. Also, n_i is the number of repeated measurements on experimental unit i and N is the number of experimental units. While β is the same for all experimental units, v_i can vary depending on the subject.

2.3 Covariance Structures

There are two main issues in the analysis of repeated measures data: the construction of an adequate model for the mean and the selection of an adequate and parsimonious model for the covariance structure of the repeated measures. Intraclass correlation, or correlation between observations taken repeatedly in the same experimental unit (plot), generally follows distinct patterns that allow designing a variety of covariance patterns to reflect such relationships (Condorelli et al., 2018; Gbur et al., 2012; Li et al., 2016).

A model with an adequate covariance structure is essential for the researcher to reach accurate conclusions from the agricultural trial with repeated measures. If the correlation between repeated measures for experimental units is ignored using a very simple covariance model, the type I error rate for hypothesis tests on fixed effects increases. Research shows that statistical inference from repeated measures experiments can be seriously impaired by the inappropriate choice of covariance models (Kenward & Roger, 1997; Keselman et al., 2000).

Some of the most common covariance patterns for applications in agricultural experiments include: the diagonal pattern with zero correlations with homogeneous variances, known as identity (I), and with heterogeneous variances, known as variance components (VC), in which the covariances are null.

$$\mathbf{R_I} = \begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \sigma^2 & \\ & & & \sigma^2 \end{bmatrix}$$

$$\mathbf{R_{CV}} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \sigma_3^2 & \\ & & & \sigma_4^2 \end{bmatrix}$$

$$\mathbf{r} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

Another covariance pattern is the compound symmetry model with homogeneous variances (CS) and with heterogeneous variances (CSH). The CS pattern is the simplest, non-trivial pattern, assuming constant correlation between repeated measures. Such correlation is known as intragroup correlation. The variance and covariance matrix for the CS and CSH patterns, and the correlation matrix are presented below:

$$\mathbf{R_{CS}} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ & & \sigma^2 & \rho\sigma^2 \\ & & & \sigma^2 \end{bmatrix}$$

$$\mathbf{R_{CSH}} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 \\ & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 \\ & & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ & & & \sigma_4^2 \end{bmatrix}$$

$$\mathbf{r} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ & 1 & \rho & \rho \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$$

For those data sets from repeated measures experiments that follow a pattern of constant correlations for all points in time or space, the CS pattern can produce statistically efficient and consistent estimates of the parameters. However, the error structure tends to be often more complex than those imposed by previous standards, particularly for repeated measures experiments, where observations in the same experimental unit tend not only to be correlated, but correlations also tend to decrease with age. increasing distance between points of time or space.

Other patterns based on correlation rules have been developed for the context of repeated measures. The unstructured residual covariance (UN) pattern represents the opposite extreme of the simpler earlier patterns. In it, all covariances are arbitrary, allowing not only unequal variances, but also unequal covariances between observations. In the unstructured pattern (UN), the assumption is that all correlations are different for all elements of the variance and covariance matrix. The variance and covariance matrix for the UN model, and the correlation matrix are presented below:

$$\mathbf{R_{UN}} = \begin{bmatrix} \sigma^2 & \rho_{12}\sigma^2 & \rho_{13}\sigma^2 & \rho_{14}\sigma^2 \\ & \sigma^2 & \rho_{23}\sigma^2 & \rho_{24}\sigma^2 \\ & & \sigma^2 & \rho_{34}\sigma^2 \\ & & & \sigma^2 \end{bmatrix}$$

$$\mathbf{R_{UNH}} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \rho_{14}\sigma_1\sigma_4 \\ & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \rho_{24}\sigma_2\sigma_4 \\ & & \sigma_3^2 & \rho_{34}\sigma_3\sigma_4 \\ & & & \sigma_4^2 \end{bmatrix}$$

$$\mathbf{r_{UN}} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ & 1 & \rho_{23} & \rho_{24} \\ & & 1 & \rho_{34} \\ & & & 1 \end{bmatrix}$$

Despite its flexibility, the UN covariance pattern requires the estimation of many parameters. An unstructured correlation pattern requires the estimation of $p\ (p - 1)/2$ distinct correlations, with p being the number of repeated measurements.

For situations where the data are organized in sequence such as repeated measures experiments, several additional residual variance and covariance patterns are available with intermediate complexities between the composite symmetry pattern imposed by ANOVA for repeated measures and the unstructured pattern. In this case, it is important to consider whether the intervals between repeated measurements are equal or not.

2.2.1 Patterns for Equal Time Intervals

For experiments with equal intervals, one can consider autoregressive (AR) patterns. Among the autoregressive models, those of the first order are perhaps the most used to model the pattern of variance and residual covariance in experiments with repeated measures. In them, the correlation between observations from adjacent levels is ρ, between observations separated by two levels ρ^2, and so on. The first-order autoregressive pattern can assume homogeneous AR (1) or heterogeneous ARH (1) variances between all time points, plus a decreasing intragroup correlation effect ρ^j ($j = 1, ..., p - 1$). This pattern is called first-order because it only depends on the residual term at the previous time point. Given this assumption, this pattern of covariances requires the specification of only two parameters, σ^2 and ρ. Given its simplicity, the autoregressive pattern is used when observations are evenly spaced, and the intragroup correlation does not change abruptly in time or space.

The variance and covariance matrices for the AR (1) and ARH (1) pattern, and the correlation matrix are presented below:

$$\mathbf{R_{AR\,(1)}} = \begin{bmatrix} \sigma^2 & \rho^1\ \sigma^2 & \rho^2\sigma^2 & \rho^3\sigma^2 \\ & \sigma^2 & \rho^1\sigma^{22} & \rho^2\sigma^2 \\ & & \sigma^2 & \rho^1\sigma^2 \\ & & & \sigma^2 \end{bmatrix}$$

$$\mathbf{R}_{\mathbf{ARH\,(1)}} = \begin{bmatrix} \sigma_1^2 & \rho^1\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \rho^3\sigma_1\sigma_4 \\ & \sigma_2^2 & \rho^1\sigma_2\sigma_3 & \rho^2\sigma_2\sigma_4 \\ & & \sigma_3^2 & \rho^1\sigma_3\sigma_4 \\ & & & \sigma_4^2 \end{bmatrix}$$

$$\mathbf{r} = \begin{bmatrix} 1 & \rho^1 & \rho^2 & \rho^3 \\ & 1 & \rho^1 & \rho^2 \\ & & 1 & \rho^1 \\ & & & 1 \end{bmatrix}$$

For example, the correlation matrix for a first-order autoregressive pattern with $\rho = 0.6$ and homogeneous variance is given by:

$$\mathbf{r} = \begin{bmatrix} 1 & 0.6 & 0.36 & 0.216 \\ & 1 & 0.6 & 0.36 \\ & & 1 & 0.6 \\ & & & 1 \end{bmatrix}$$

However, this pattern may not be flexible enough to reflect heterogeneity over time in repeated measures experiments. Therefore, researchers have also developed more refined autoregressive patterns to address more complex patterns of residual covariance.

Another important covariance pattern is called Toeplitz (TOEP), in which the correlation between observations at adjacent factor levels is ρ^d, where d represents the unit of distance between two repeated measurements. This model is slightly less restrictive than an autoregressive model.

In this case, n parameters of variance and covariance are specified, one for constant variance and $n - 1$ for correlations. This structure resembles the AR (1) pattern, except the correlation specification is ρ_d instead of ρ^d, where d represents the unit of distance between two repeated observations. When compared to the AR (1) structure, the TOEP model is more flexible given the specification of ρ_d and, therefore, this pattern can be used to capture more complex patterns of variance. Also, TOEP allows specifying n variance parameters, while AR (1) has only 2.

The variance and covariance matrices for the TEOP and TOEPH standard, and the correlation matrix are presented below:

$$\mathbf{R}_{\text{TOEP}} = \begin{bmatrix} \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 & \sigma^2\rho_3 \\ & \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 \\ & & \sigma^2 & \sigma^2\rho_1 \\ & & & \sigma^2 \end{bmatrix}$$

$$\mathbf{R}_{\text{TOEPH}} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ & & & \sigma_4^2 \end{bmatrix}$$

$$\mathbf{r} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ & 1 & \rho_1 & \rho_2 \\ & & 1 & \rho_1 \\ & & & 1 \end{bmatrix}$$

The correlation matrix for a TOEP standard with $\rho_1 = 0.7$, $\rho_2 = 0.5$ and $\rho_3 = 0.3$ is given by:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.7 & 0.5 & 0.3 \\ & 1 & 0.7 & 0.5 \\ & & 1 & 0.7 \\ & & & 1 \end{bmatrix}$$

The variance and covariance matrices for the TEOP (2) and TOEPH (2) standard, and the correlation matrix are presented below:

$$\mathbf{R}_{\text{TOEP (2)}} = \begin{bmatrix} \sigma^2 & \sigma^2\rho_1 & 0 & 0 \\ & \sigma^2 & \sigma^2\rho_1 & 0 \\ & & \sigma^2 & \sigma^2\rho_1 \\ & & & \sigma^2 \end{bmatrix}$$

$$\mathbf{R}_{\text{TOEPH (2)}} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & 0 & 0 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & 0 \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ & & & \sigma_4^2 \end{bmatrix}$$

$$\mathbf{r} = \begin{bmatrix} 1 & \rho_1 & 0 & 0 \\ & 1 & \rho_1 & 0 \\ & & 1 & \rho_1 \\ & & & 1 \end{bmatrix}$$

The correlation matrix for a TOEP (2) standard with $\rho_1 = 0.7$ is given by:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.7 & 0 & 0 \\ & 1 & 0.7 & 0 \\ & & 1 & 0.7 \\ & & & 1 \end{bmatrix}$$

The variance and covariance matrices for the TOEP (3) and TOEPH (3) standard, and the correlation matrix are presented below:

$$\mathbf{R}_{\text{TOEP (3)}} = \begin{bmatrix} \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 & 0 \\ & \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 \\ & & \sigma^2 & \sigma^2\rho_1 \\ & & & \sigma^2 \end{bmatrix}$$

$$\mathbf{R}_{\text{TOEPH (3)}} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & 0 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ & & & \sigma_4^2 \end{bmatrix}$$

$$\mathbf{r} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & 0 \\ & 1 & \rho_1 & \rho_2 \\ & & 1 & \rho_1 \\ & & & 1 \end{bmatrix}$$

The correlation matrix for a TOEP (3) standard with $\rho_1 = 0.7$, $\rho_2 = 0.5$ is given by:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.7 & 0.5 & 0 \\ & 1 & 0.7 & 0.5 \\ & & 1 & 0.7 \\ & & & 1 \end{bmatrix}$$

The covariance structure in repeated measures experiments can be complex, but it is generally assumed that repeated measures within experimental units (subjects) are correlated and between independent experimental units. Thus, generally, the marginal variance structure is a diagonal block matrix, making it easy to specify different structures for each block. In the case where covariance parameters are expected to change with one-factor groups between subjects, different sets of parameters for the R and G matrices can be specified for each group. As a result, a set of covariance parameters are produced that differ for each level of the variable group.

The patterns of the covariance matrix between the repeated measures for the simulations in this study consisted of 14 structures: I, CS, AR (1), TOEP, TOEP (2), TOEP (3) and BANDED (3), and their respective versions with heterogeneous variances: VC, SCH, ARH (1), TOEPH, TOEPH (2), TOEPH (3), BANDH (3). Furthermore, the size of the **Σ** matrices varied between 4 × 4 or 8 × 8, corresponding to the scenario with 4 and 8 repeated measures in time or space. For each simulation scenario, a vector of multivariate random normal deviations with mean 0 and covariance matrix **Σ** was generated. For each simulation, the vectors were of the following sizes: (32, 1), (64, 1), (96, 1) (160, 1) and (192, 1), and were generated with the multivariate_normal function (μ, $\mathbf{\Sigma}$), where μ is the vector of means, which has been adjusted to μ = [0 0 0 0] or μ = [0 0 0 0 0 0 0 0] depending on the number of repeated measures, and $\mathbf{\Sigma}$ is the covariance matrix of the means, which varied depending on the covariance patterns being evaluated.

For matrices CS, CSH, AR (1) and ARH (1), ρ was set to 0.75. For the TOEP and TOEPH matrices, ρ_1, ρ_2 and ρ_3 were 0.8, 0.6 and 0.4, respectively. For the TOEP (2) and TOEPH (2) matrices, ρ_1 and ρ_2 were 0.8 and 0.6, respectively. For the TOEP (3) and TOEPH (3) matrices, ρ_1 was set to 0.8. For the BAND (3) and BANDH (3) matrices, ρ_1 was set to 0.8. For the construction of the previous matrices, two matrices of size (4 × 4) and (8 × 8) with all elements equal to 1 served as a model, from which repetition structures were used to perform iterations on the elements of the matrices in order to obtain the desired structure.

The identity patterns, variance components and compound symmetry were included in the study as a reference for what would be the data set with the ideal pattern for the assumptions required by the traditional ANOVA and ANOVA for repeated measures. The simulations were obtained through Python code written mostly with Numpy, Pandas and Scipy packages. Subsequently, all data sets were submitted to statistical analysis using a code in SAS mostly through the Glimmix procedure for analysis of mixed linear models with different assumptions about the pattern of covariance between repeated measures.

For each simulated scenario, 10,000 data sets were generated and analyzed using mixed linear models with the accommodation of the following covariance patterns: VC, CS, CSH, AR (1), ARH (1), TOEP, TOEPH, TOEP (2), TOEPH (2), TOEP (3) and TOEPH (3).

2.4 Specification of the Effects of Main Factors and Interaction in Simulations

From the simulated random deviation vectors for each scenario considered, models of experiments with repeated measures were created following the model specified in 2.2 with a factor between subjects (treatment) with two levels, an intra-subject factor (time) with 4 or 8 levels and a treatment × time interaction. To determine the size of the effects, the concept of minimum detectable difference (MDD) and the effect pattern (EP) was considered. Minimum detectable difference is the difference between the largest and smallest effect associated with a factor. This difference represents the smallest difference between two levels of a factor, and the real effect used in the simulation was given by the equation:

$$\text{Effect} = \text{MDD} \times \text{EP} \tag{7}$$

In all simulations with 4 repeated measures, the default factor of repeated measures was +0.5, +0.125, -0.125, -0.5, while for simulations with 8 repeated measures, the default was +0.5, +0.125, 0, -0.125, -0.250, -0.35, -0.45 -0.5. The minimum detectable difference was set to 10.

2.5 Determining the Type I Error Rate and Power of Hypothesis Testing

For the study of the type I error rate of different mixed models (assessment of the rejection rate of H0 when it is false), the main effects pattern was +0.5 and +0.5, and the minimum detectable difference was kept at 1. For the interaction effect, the pattern was the Kronecker product of the main effects patterns, and the minimum detectable difference was also kept at 1. In order not to lose the sense of generalization, the residual variance was adjusted to 1, while the intra-subject variance was 2 2 for data sets with homogeneous variance and ranged between 1,252 and 3,002 for those with heterogeneous variance, in order to maintain a ratio greater than or equal to 4:1 between the highest and lowest variances, a criterion in many tests of heteroscedasticity of variances. Thus, as most observations of a normal distribution are within three standard deviations, variances above 1,252 would virtually eliminate the chance of a significant difference for the interaction factor. Therefore, type I error rates for the interaction hypothesis tests were estimated as the proportion of p-values less than or equal to α out of a total of 10,000 tests.

For the study of the statistical power of mixed linear models (assessment of the rejection rate of H0 when this is true), the main effects pattern was adjusted to -0.5 and +0.5, and the minimum detectable difference was 1. For this purpose of the interaction, the pattern was also the Kronecker product of the main effects patterns, and the minimum detectable difference was also adjusted to 2. In this case, the residual variance was adjusted to 1 2, while the within-subject variance was 2 2 for the data sets with homogeneous variance and ranged between 1002 and 2002 for those with heterogeneous variance, in order to also maintain the ratio greater than or equal to 4:1 between the highest and lowest variances and provide a greater possibility of significant differences in the simulated data sets. The statistical power estimates of the hypothesis tests for the interaction were estimated as the proportion of p-values less than or equal to α out of a total of 10,000 tests.

2.6 Selection of the Covariance Model

The models were also evaluated for their ability to fit the data using the Bayesian information criterion (BIC) provided by the Glimmix procedure. The frequency with which covariance structures were chosen by the selection criteria from a total of 10,000 analyzes was computed.

2.7 Glimmix Procedure for Experiments Involving Repeated Measures

In mixed linear models, the response is modeled as a sum of three terms: the design matrix associated with the fixed effects multiplied on the right by an unknown coefficient vector, the random effects multiplied on the left by the associated design matrix, and a residual error (Littell et al., 1996). The mean of the response is parameterized as the first term in the sum above since random effects and residual error are generally assumed to be zero. As long as this design matrix is full rank, there is no other coefficient vector providing the same mean. The covariance of the response is determined by the covariances of the random effects and the residual error. Both random effects and residual error are assumed to have certain distributions with covariance matrix structures to be specified. Unknown associated parameters are generally restricted to certain spaces and must be estimated (Gbur, et al., 2012).

In mixed linear models, the response covariance matrix is modeled as the sum of two matrices: the product of the random effects covariance matrix with the associated design matrix and the residual error covariance matrix. The construction of a mixed linear model usually involves the selection of covariance matrix structures parameterized for random effects and residual error. However, even if the response covariance matrix is not over parameterized, some specifications of covariance structures may result in the parameters being unidentifiable. When tuning such models, the software may or may not indicate a problem with model identifiably. The Glimmix procedure applies a lower bound of 0 to all variance components in the model. If the lower bound is removed, the actual value of the variance component can be negative. This means that there was not enough variation in the response to attribute any variation to the random effect, controlling for everything else in the model. While there is nothing wrong with the results when the variance components are negative, some statisticians recommend removing random effects from the model, despite being an important component of the model. Other statisticians argue that one should keep the random effect term in the model even if it is 0. For example, if the variance component is the block effect, and if the effect variance were 0, it would still be important to keep it in the model, as it is part of the design of the experiment. Another option is to specify a lower bound to prevent estimates from getting too close to 0. All analyzes with mixed linear models had convergence greater than 99.9%, unless otherwise stated in the tables (Guerin & Stroup, 2000).

2.8 Simulation Algorithm

In summary, this study consisted of the following steps:

(1) Generation of a data set using the Python language with the Numpy library based on a random vector with a multivariate normal response to generate an experiment with a completely randomized design with two treatments, 4, 8 or 12 experimental units (subjects) per treatment, and 4 or 8 repeated measurements.

(2) Statistical analysis of the experiment generated by mixed linear models using the Glimmix procedure in the SAS program with the following 56 options of assumptions about the pattern of the covariance structure for the matrix R: VC, CS, CSH, AR (1), ARH (1), TOEP, TOEPH, TOEP (2), TOEPH (2), TOEP (3), TOEPH (3).

(3) Repeat step (1) and (2) 10,000 times.

(4) Registration of the BIC information criteria for evaluating the ability to fit the data for each data set generated, in addition to estimating the type I error rate or the power of the hypothesis test of the interaction between treatment and repeated measures for the significance level of 0.05 or 0.01.

(5) Repetition of steps (1) to (4) using the following covariance matrices for data simulation: I, VC, CS, CSH, AR (1), ARH (1), TOEP, TOEPH, TOEP (2), TOEPH (2), TOEP (3), TOEPH (3), BAND (3), BANDH (3).

3. Results and Discussion

3.1 Selection of Covariance Structures

Tables A1 to A6 (see Appendix A) provide the percentages of choice of covariance patterns by the BIC information criterion for data sets with the following patterns for the simulated covariance structure of the data: I, CS, AR, TOEP, TOEP (2), TOEP (3), BAND (3), VC, CSH, ARH, TOEPH, TOEPH (2), TOEPH (3), BANDH (3). The values correspond to the proportion of choice of covariance models by the BIC information criterion related to the analysis of 10,000 data sets from simulated experiments with repeated measures involving 4 and 8 repeated measures, sample size of 4, 8 and 12, and the previous covariance patterns. All values correspond to convergence equal to or greater than 99.9%, except when specified in the tables.

Overall, the results indicate that the assumption most frequently chosen by the information criteria resulted from the empirical covariance structure of the analyzed data sets, particularly for those sets with larger sample sizes. Furthermore, the frequency with which covariance models were chosen increased as the number of repeated measures and sample size increased.

For simulations with 4 repeated measures and 4 samples sizes, the VC covariance model was chosen in 54% of the cases when the simulated covariance structure of the data was I (identity) and 40% when the covariance structure was also VC (Table A1). For simulations with 8 samples sizes, the VC covariance model was chosen in 78% of the cases when the simulated covariance structure was I and 44% when the covariance structure was also VC (Table A2). For simulations with 12 samples sizes, the VC covariance model was chosen in 84% of the cases when the simulated covariance structure was I and 35% when the covariance structure was also VC (Table A3).

For simulations with 8 repeated measures and 4 samples sizes, the VC covariance model was chosen in 57% of the cases when the simulated covariance structure of the data was I (identity) and 43% when the covariance structure was also VC (Table A1). For simulations with 8 samples sizes, the VC covariance model was chosen in 79% of the cases when the simulated covariance structure was I and 54% when the covariance structure was also VC (Table A2). For simulations with 12 samples sizes, the VC covariance model was chosen in 85% of the cases when the simulated covariance structure was I and 44% when the covariance structure was also VC (Table A3).

For simulations with 4 repeated measures and 4 samples sizes, the CS covariance model was chosen in 50% of the cases when the simulated covariance structure of the data was also CS and 15% and 12% when the covariance structure was CSH and AR, respectively (Table A1). For simulations with 8 samples sizes, the CS covariance model was chosen in 81% of the cases when the simulated covariance structure was also CS and 7% and 8% when the covariance structure was CSH and AR, respectively (Table A2). For simulations with 12 samples sizes, the CS covariance model was chosen in 90% of the cases when the simulated covariance structure was also CS and 2% and 4% when the covariance structure was CSH and AR, respectively (Table A3).

For simulations with 4 repeated measures and 4 samples sizes, the CSH covariance model was chosen in 34% of the cases when the simulated covariance structure of the data was also CSH (Table A1). For simulations with 8 samples sizes, the CSH covariance model was chosen in 74% of the cases when the simulated covariance structure was also CSH and 19% when the covariance structure was VC (Table A2). For simulations with 12 samples sizes, the CSH covariance model was chosen in 89% of the cases when the simulated covariance structure was also CSH and 27% when the covariance structure was VC (Table A3).

The AR covariance model was the most chosen by different simulated covariance patterns such as AR, ARH, TOEP, TOEPH and TOEP (2). For simulations with 4 repeated measures and 4 samples sizes, the AR covariance model was chosen in 41, 21, 43 and 21% of the cases when the simulated covariance structure of the data was also AR, ARH, TOEP and TOEPH, respectively (Table A1). For simulations with 8 samples sizes, the AR covariance model was chosen in 74, 21 and 71% of the cases when the simulated covariance structure was AR, ARH, TOEP (Table A2). For simulations with 12 samples sizes, the CSH covariance model was chosen in 86, 79 and 31% of the cases when the simulated covariance structure was also AR, TOEP and TOEP (2) (Table A3). For simulation with 8 repeated measures and sample size 4, the AR covariance model was chosen 64, 45, 62 and 43% of the cases when the simulated covariance structure was AR, ARH, TOEP and TOEPH, respectively (Table A4). For simulation with 8 repeated measures and sample size 8, the AR covariance model was chosen 95, 54, 89 and 48% of the cases when the simulated covariance structure was AR, ARH, TOEP and TOEPH, respectively (Table A5). For simulation with 8 repeated measures and sample size 12, the AR covariance model was chosen 98, 36, 92 and 31% of the cases when the simulated covariance structure was AR, ARH, TOEP and TOEPH, respectively (Table A6).

For simulated datasets with covariance pattern more complex than CS and CSH such as AR, ARH, TOEP and TOEPH, the AR covariance model often represented a considerable part of the choice for the BIC information criterion, often being the most chosen model, particularly for simulations with a greater number of repeated measurements and sample size.

For data sets with BAND (3) and BANDH (3) patterns for empirical covariance structure, the TOEP (3) and TOEPH (3) model represented most of the choices by the BIC information criterion and increased with the increase in the number of repeated measurements and in the sample size.

The UN covariance model was never the most chosen considering all simulated scenarios, and the frequency decreased as the number of repeated measures and sample sizes increased. Results indicated that in some cases when datasets are smaller UN pattern can still be a valid option although UN covariance model is computationally costly.

3.2 Study on Type I Error in Mixed Linear Models Applied to Experiments Involving Repeated Measures

Tables B1 to B3 (see Appendix B) provide type I error rates for null hypothesis test from the interaction effect between treatments and repeated measures for linear mixed models associated with CS, AR and TOEPH covariance models with α equals to 0.01. Type I error rates for linear mixed models with others covariance models and for those with α equals to 0.05 are found in the Supplementary Table section.

The results for type I error rates showed that the use of covariance models that do not recognize heterogeneous correlations or any correlation at all between repeated measures can either (a) inflate type I error or (b) reduce it to very conservative levels, which can affect the conclusion of agricultural experiments.

Type I error rates from the others covariance model vary depending on the simulated covariance structures. However, variations on the number of repeated measurements and sample size did not show much of a difference, results were vastly similar. Despite not being the most cost-efficient model, the application of UN covariance structure showed the rate of type I error closer to nominal error rates across all scenarios.

Type I error rates from datasets with equal variances and without correlation between repeated measures (I structure) were close to the nominal rate for all assumptions about the pattern of covariance structure between repeated measures, and did not vary with increasing sample size, nor with the increase in the number of repeated measures for both levels of significance (Supplementary Tables 1 and 12). Results indicated that it is not efficient the application of more complex covariance structures to datasets known not to have any covariance issues. Type I error rates from datasets simulated with unequal variances and without correlation between repeated measures (VC structure) were close to the nominal rate for some assumptions about the pattern of covariance structure between repeated measures, but it was larger surprisingly for other assumptions including VC model itself (Supplementary Tables 2 and 13).

However, if there is non-zero correlation, assumptions about the pattern of the covariance structure in the data affect the rejection rates of H_0 either due to (a) or (b). Covariance patterns that sub model the simulated covariance structure showed a greater tendency to inflate the null hypothesis rejection rate to more liberal levels than those established by the level of significance. For example, the use of LM models with CS covariance model, which does not model heterogeneous correlations, increases the rejection of H_0 in favor of alternative hypothesis (type I error) in several cases.

For a significance level of 1%, the bias towards the nominal rate for type I error was 5 times the standard error (α) when CS model was used to model a TOEP covariance pattern (Table B9). For a 5% significance level, this bias was 2α when CS model was used to model several covariance patterns. In addition, the VC model reduced type I error rate to lower levels compared to α or to virtually zero, except when the simulated pattern had also null correlations for all pairs of repeated measures (I and VC models).

3.3 Study on Statistical Power in Mixed Linear Models Applied to Experiments Involving Repeated Measures

Tables C10 to C12 (see Appendix C) provide statistical power for the null hypothesis test from the interaction effect between treatments and repeated measures for linear mixed models associated with CS, AR and TOEPH covariance models with α equals to 0.01. Statistical power for linear mixed models associated with others covariance models and for those with α equals to 0.05 are found in the Supplementary Table section.

While statistical power increased as the sample sizes increased, it decreased as the number of repeated measures increased. Results indicated that increasing sample sizes give a better return if statistical power is a major concern. Moreover, datasets with simulated covariance structures with heterogeneous variances showed a greater statistical power compared to the homogeneous versions. Overall, an optimum covariance model that it is not trivial such as CV model and that considers some degree of correlation, but it is not that complex such as UN model, provided the larger statistical power for all combinations of simulated covariance structures and applied covariance model.

For datasets with the highest number of repeated measures, statistical power was above 90% in most of the combinations evaluated regardless of the covariance model applied. However, for datasets with the lowest number of repeated measures, statistical powers were at lower levels, particularly for 1% statistical significance. In situations with a few sample sizes, choosing the most appropriate covariance model is a key factor for finding differences among treatments.

4. Discussion

Violation of assumptions in univariate and multivariate ANOVA resulted in sampling bias and inflated rates of type I error, resulting in estimates with little robustness and, consequently, affecting the significance of treatments and the results of experiments. As the observations within the groups do not contribute completely independent information, the effective sample size is smaller than the total number of observations from all groups, and therefore increases the probability of rejecting the null hypothesis, affecting the quality of the estimator.

The statistical power and type I error rate of linear mixed models (LMM) based on residual covariance patterns were also investigated by Gueorguieva and Krystal (2004), who concluded that an ideal type I error control results from the covariance structure specified that corresponds to the empirical covariance structure (observed data). The authors investigated AR and UN covariance model and the results revealed more adequate type I error rates when AR covariance model was applied. However, the simulated data already had an autoregressive structure, implying that previous knowledge about the covariance structure in repeated measure experiment is key to provide correct LMM adjustments.

Goedert et al. (2013) conducted a simulation with a data set with violation of the sphericity assumption. They found greater statistical power of LMM-UN compared to RANOVA, particularly for small sample sizes. Gueorguieva and Krystal (2004) also found similar type I error rates for UN covariance model and RANOVA with Greenhouse-Geisser correction in a simulation study with four time points. However, Type I error rates were slightly more correct for LMM-CS.

Harverkamp and Beauducel (2017) evaluated the error rate of simulated type I error rate from repeated-measure experiments with sphericity violation, different numbers of repeated measures and degrees of correlations. The following methods were considered: RANOVA, LMM-CS and LMM-UN. Results showed a progressive bias for LMM associated with UN covariance model, particularly for small sample sizes and larger number of repeated measures, regardless of sphericity violation. Moreover, LMM associated with CS covariance model and RANOVA showed greater bias when sphericity was violated, reaching over 1.5α. Results suggested that both LMM-CS and RANOVA should not be recommended when sphericity is violated. LMM-UN provided more correct type I error when sample size was larger, and sphericity was violated. However, the study did not investigate intermediate models between compound symmetry and unstructured.

Our study investigated several models from the simplest to the most complex. Results indicated that LMM-AR, an intermediate structure between LMM-CS and LMM-UN, should be recommended since it provides a greater control of type I error and a lower statistical cost while maintaining statistical power at adequate levels, particularly for reduced sample size experiments. On the other hand, LMM-CS applied to data with more complex covariance structure increased type I error, particularly for 1% significance level. Although LMM-CS also displayed greater statistical power, that was followed by a greater chance of rejecting H_0 when it is true, affecting results from experiments.

5. Conclusion

In the analysis of experiments that involve repeated measures, it is essential that assumptions are adequate regarding the covariance structure of the data so that the model has adequate levels of error and statistical power. Moreover, appropriate covariance models become even more important depending on the number of repeated measures and sample sizes. The results of this work reinforce the need for the correct adjustment of mixed models and the care with the use of the standard adjustments of the models offered as a default in statistical programs. In many statistical packages capable of dealing with mixed linear models, the default model is VC or CS, which are the ones with the most irregular results regarding type I error and statistical power in this study. Thus, intermediate covariance models such as the AR should be included among candidate models evaluated by the researcher.

References

Allegue, H., Araya-Ajoy, Y. G., Dingemanse, N. J., Dochtermann, N. A., Garamszegi, L. Z., Nakagawa, S., & Westneat, D. F. (2017). Statistical quantification of individual differences (squid): An educational and statistical tool for understanding multilevel phenotypic data in linear mixed models. *Methods in Ecology and Evolution, 8*, 257-267. https://doi.org/10.1111/2041-210X.12659

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology, 8*(456). https://doi.org/10.3389/fpsyg.2017.00456

Casella, G., & Berger, R. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury.

Condorelli, G. E., Maccaferri, M., Newcomb, M., Andrade-Sanchez, P., White, J. W., French, A. N., … Tuberosa, R. (2018). Comparative aerial and ground based high throughput phenotyping for the genetic dissection of NDVI as a proxy for drought adaptive traits in durum wheat. *Frontiers in Plant Science, 9*(893). https://doi.org/10.3389/fpls.2018.00893

Demidenko, E. (2013). *Mixed models: Theory and applications with R* (p. 768, 2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*(4), 447-456. https://doi.org/10.1016/j.jml.2007.11.004

Duarte, J. B., & Vencovsky, R. (2001). Estimação e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. *Scientia Agricola, 58*(1), 109-117. https://doi.org/10.1590/S0103-9016 2001000100017

Gezan, S. A., & Carvalho, M. P. de. (2018). *Chapter 10: Analysis of Repeated Measures for the Biological and Agricultural Sciences*. https://doi.org/10.2134/appliedstatistics.2016.0008

Gbur, E. E., Stroup, W. W., Mccarter, K. S., Durham, S., Young, L. J., Christman, M., … Kramer, M. (2012). *Analysis of generalized linear mixed models in the agricultural and natural resources sciences* (p. 283). Madison, WI, EUA: American Society of Agronomy, Soil Science Society of America, Crop Science Society of America, Inc. https://doi.org/10.2134/2012.generalized-linear-mixed-models

Goedert, K. M., Boston, R., & Barrett, A. M. (2013). Advancing the science of spatial neglect rehabilitation: An improved statistical approach with mixed linear modeling. *Front. Hum. Neurosci., 7*. https://doi.org/10.3389/fnhum.2013.00211

Gueorguieva, R., & Krystal, J. H. (2004). Move over ANOVA progress in analyzing repeated-measures data and its reflection in papers published in the archives of general psychiatry. *Archives of General Psychiatry, 61*(3), 310-317. https://doi.org/10.1001/archpsyc.61.3.310

Guerin, L., &. Stroup, W. W. (2000). A simulation study to evaluate proc mixed analysis of repeated measures data. *Proceedings of the 12th Annual Conference on Applied Statistics in Agriculture, Manhattan, KS*. Apr. 30-May 2, 2000, Kansas State University, Manhattan, KS. https://doi.org/10.4148/2475-7772.1249

Haverkamp, N., & Beauducel, A. (2017). Violation of the sphericity assumption and its effect on type-i error rates in repeated measures nova and multi-level linear models (MLM). *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01841

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997. https://doi.org/10.2307/2533558

Keselman, H. J., Algina, J., Wilcox, R. R., & Kowalchuk, R. K. (2000). Testing repeated measures hypothesis when covariance matrices are heterogeneous: Revisiting the robustness of the welch-james test again. *Educational and Psychological Measurement, 60*, 925-938. https://doi.org/10.1177/00131640021970998

Li, Z. (2017). A review on mixed models. *Journal of Biometric and Biostatistics, 8*, 350. https://doi.org/10.4172/2155-6180.1000350

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. SAS Institute Inc., Cary.

Marcin, S., Wiesław, M., Adriana, D., Kinga N., & Elżbieta, W. (2015). Four-way data analysis within the linear mixed modelling framework. *Scientia Agricola, 72*(5), 411-419. https://doi.org/10.1590/0103-9016-2014-0333

Oberfeld, D., & Franke, T. (2013). Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. *Behavior Research Methods, 45*, 792-812. https://doi.org/10.3758/s13428-012-0281-2

Piepho, H. P., Büchse, A., & Richter, C. (2004). A mixed modelling approach for randomized experiments with repeated measures. *Journal Agronomy and Crop Sciences, 190*(4), 230-247. https://doi.org/10.1111/j.1439-037X.2004.00097.x

R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

SAS Institute Inc. (2015). *SAS/STAT® 14.1 user's guide*. Cary, NC: SAS Institute Inc.

Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. Scotts Valley, CA: Createspace.

Stroup, W. W. (2013). *Generalized linear mixed models*. Boca Raton, FL. CRC Press.

Vasey, M. W., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measures anova in psychophysiology: A multivariate solution. *Psychophysiology, 24*(4), 479-486. https://doi.org/10.1111/j.1469-8986.1987.tb00324.x

Zhang, X., Guan, Z., & Wang, L. (2020). Combined GWAS and QTL analysis for dissecting the genetic architecture of kernel test weight in maize. *Molecular Genetics and Genomics, 295*, 409-420. https://doi.org/10.1007/s00438-019-01631-2

Yang, R., & Juskiw, P. (2011). Analysis of covariance in agronomy and crop research. *Canadian Journal of Plant Science, 91*(4), 621-641. https://doi.org/10.4141/cjps2010-032

Wang, W. (2013). Identifiability of linear mixed effects models. *Electronic Journal of Statistics, 7*, 244-263. https://doi.org/10.1214/13-EJS770

Appendix A

Selection

Table A1. Percentage of covariance models chosen by the BIC selection criterion from a total of 10000 analyzes with mixed linear models from data from simulated experiments involving 4 repeated measures, and sample size 4

Simulated model	Adjusted model											
	VC	CS	CSH	AR	ARH	T	TH	T (2)	TH (2)	T (3)	TH (3)	UN
	%											
I	53.99	7.47	3.11	4.11	1.59	5.23	1.62	1.72	2.41	6.60	2.22	9.93
VC	39.98	4.69	9.58	2.77	5.64	2.53	4.27	1.36	7.49	4.18	4.45	13.06
CS	1.34	50.01	7.23	11.58	2.75	7.61	3.75	0.49	0.33	2.52	0.89	11.50
CSH	0.93	15.13	33.93	7.71	8.65	1.73	9.75	0.38	0.70	1.47	2.15	17.47
AR	1.18	12.41	2.23	40.88	5.55	4.46	1.83	6.05	2.71	9.91	3.00	9.79
ARH	0.86	5.08	7.91	21.20	21.87	1.42	4.00	3.46	6.44	5.16	8.21	14.39
TOEP	0.37	6.21	1.62	42.62	5.44	4.44	1.85	4.25	2.01	15.29	5.92	9.98
TOEPH	0.25	2.66	4.90	20.59	23.78	1.79	3.95	2.53	4.76	5.86	14.18	14.75
TOEP (2)	5.65	4.07	0.83	18.96	2.90	8.43	2.11	16.16	7.47	19.18	3.62	10.62
TOEPH (2)	0.23	0.44	0.52	17.88	10.51	2.68	1.94	22.96	18.16	15.93	4.87	3.88
TOEP (3)	0.09	0.98	0.52	5.54	1.86	10.14	3.40	0.30	0.04	52.15	12.18	12.80
TOEPH (3)	0.09	0.52	0.89	4.24	3.76	3.49	9.21	0.16	0.20	14.76	41.54	21.14
BAND (3)	0.69	1.57	0.68	2.96	1.13	9.80	2.96	0.14	0.06	53.32	14.39	12.30
BANDH (3)	0.89	1.22	1.68	2.59	2.30	1.65	8.55	0.13	0.08	11.32	50.71	18.88

Note. Convergence: 99.9% > a ≥ 90.0%.; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

All values shown have a convergence greater than 99.9%, unless otherwise specified.

T = TOEP; TH = TOEPH; T2 = TOEP (2); T3 = TOEP (3).

Table A2. Percentage of covariance models chosen by the BIC selection criterion from a total of 10000 analyzes with mixed linear models from data from simulated experiments involving 4 repeated measures, and sample size 8

Simulated model	Adjusted model											
	VC	CS	CSH	AR	ARH	T	TH	T (2)	TH (2)	T (3)	TH (3)	UN
	%											
I	78.32	6.17	1.10	3.57	0.61	2.53	0.13	1.31	0.68	4.76	0.38	0.44
VC	44.16	2.69	19.49	1.97	11.10	0.84	1.70	0.63	10.69	2.37	3.08	1.28
CS	0.02	81.40	3.95	7.58	0.74	4.83	0.45	0.05	0.00	0.42	0.01	0.55
CSH	0.02	7.11	75.09	2.92	6.58	0.22	5.43	0.01	0.05	0.20	0.28	2.09
AR	0.02	8.27	0.64	73.65	3.42	2.23	0.31	2.65	0.40	7.31	0.73	0.37
ARH	0.00	1.64	6.78	20.27	54.79	0.32	1.77	0.93	2.76	1.70	7.31	1.73
TOEP	0.00	1.99	0.12	70.94	3.15	3.24	0.27	1.22	0.16	16.89	1.61	0.41
TOEPH	0.00	0.41	1.74	16.85	53.78	0.56	2.36	0.32	1.28	3.18	17.17	2.35
TOEP (2)	0.48	1.09	0.09	29.2	2.02	7.69	0.38	31.67	2.88	22.64	1.22	0.64
TOEPH (2)	0.23	0.44	0.52	17.88	10.51	2.68	1.94	22.96	18.16	15.93	4.87	3.88
TOEP (3)	0.00	0.05	0.02	1.19	0.25	9.96	1.20	0.00	0.00	81.66	4.48	1.19
TOEPH (3)	0.00	0.00	0.00	0.73	0.93	1.68	9.02	0.00	0.00	8.22	73.64	5.78
BAND (3)	0.00	0.08	0.02	0.16	0.06	9.09	1.21	0.00	0.00	84.04	4.36	0.98
BANDH (3)	0.02	0.03	0.12	0.17	0.44	0.44	8.99	0.00	0.00	3.92	82.03	3.84

Note. Convergence: 99.9% > a ≥ 90.0%.; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

All values shown have a convergence greater than 99.9%, unless otherwise specified.

T = TOEP; TH = TOEPH; T2 = TOEP (2); T3 = TOEP (3).

Table A3. Percentage of covariance models chosen by the BIC selection criterion from a total of 10,000 analyzes with mixed linear models from data from simulated experiments involving 4 repeated measures, and sample size 12

Simulated model	Adjusted model											
	VC	CS	CSH	AR	ARH	T	TH	T (2)	TH (2)	T (3)	TH (3)	UN
	%											
I	84.21	5.53	0.52	3.27	0.36	1.42	0.05	0.56	0.28	3.63	0.15	0.02
VC	35.11	1.38	26.75	1.09	14.87	0.24	1.42	0.29	14.23	1.26	2.85	0.51
CS	0.00	89.69	2.40	4.08	0.16	3.33	0.14	0.00	0.00	0.13	0.00	0.07
CSH	0.00	1.82	88.95	0.55	3.99	0.04	3.6	0.00	0.00	0.01	0.07	0.97
AR	0.00	4.09	0.20	85.58	2.07	1.63	0.04	0.89	0.07	5.11	0.22	0.10
ARH	0.00	0.36	4.43	10.97	74.87	0.12	1.57	0.12	0.93	0.55	5.39	0.69
TOEP	0.00	0.56	0.03	78.54	1.79	4.17	0.15	0.13	0.02	13.69	0.80	0.12
TOEPH	0.00	0.01	0.55	7.78	69.86	0.29	2.83	0.02	0.17	0.93	16.7	0.86
TOEP (2)	0.03	0.22	0.01	30.66	1.30	6.71	0.12	37.21	1.72	20.81	0.90	0.31
TOEPH (2)	0.01	0.08	0.20	13.40	14.37	2.13	1.63	23.37	22.30	12.04	6.58	3.89
TOEP (3)	0.00	0.00	0.00	0.10	0.01	9.02	0.48	0.00	0.00	87.56	2.47	0.36
TOEPH (3)	0.00	0.00	0.00	0.1	0.18	0.45	8.36	0.00	0.00	1.71	85.17	4.03
BAND (3)	0.00	0.03	0.01	0.03	0.01	7.93	0.42	0.00	0.00	88.81	2.55	0.21
BANDH (3)	0.00	0.00	0.01	0.01	0.00	0.08	7.94	0.00	0.00	0.56	89.85	1.55

Note. Convergence: 99.9% > a ≥ 90.0%.; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

All values shown have a convergence greater than 99.9%, unless otherwise specified.

T = TOEP; TH = TOEPH; T2 = TOEP (2); T3 = TOEP (3).

Table A4. Percentage of covariance models chosen by the BIC selection criterion from a total of 10000 analyzes with mixed linear models from data from simulated experiments involving 8 repeated measures, and sample size 4

Simulated model	Adjusted model											
	VC	CS	CSH	AR	ARH	T	TH	T (2)	TH (2)	T (3)	TH (3)	UN
	%											
I	56.5	10.2	1.74	4.84	0.95	7.65	4.08[b]	1.42	1.32[a]	6.95	1.10[a]	nc
VC	42.92	7.53	7.82	3.91	4.54	4.69	8.54[b]	1.33	5.21[a]	5.78	4.21[a]	nc
CS	0.17	74.78	4.73	1.72	0.39	10.80	6.44[b]	0.06	0.00[a]	0.36	0.14[a]	nc
CSH	0.10	26.59	44.93	3.00	1.43	5.01	17.95[b]	0.04	0.08[a]	0.18	0.16[a]	nc
AR	0.00	2.05	0.20	63.66	3.98	9.77	6.60[b]	1.93	0.41[a]	6.71	2.38[a]	nc
ARH	0.02	1.20	1.12	44.48	19.14	5.11	12.30[b]	1.83	1.21[a]	5.33	5.99[a]	nc
TOEP	0.01	5.62	0.56	62.37	5.08	13.41	8.60[b]	0.26	0.11[a]	1.75	0.83[a]	nc
TOEPH	0.03	3.08	2.84	42.77	21.43	7.53	16.63[b]	0.23	0.27[a]	1.58	2.07[a]	nc
TOEP (2)	0.49	0.80	0.13	16.86	1.18	18.96	6.68[b]	12.16	1.65[a]	31.09	7.22[a]	nc
TOEPH (2)	0.56	0.64	0.39	13.33	4.06	12.17	12.70[b]	9.43	5.11[a]	23.15	15.08[a]	nc
TOEP (3)	0.09	0.92	0.04	31.71	1.97	23.32	9.72[b]	0.46	0.23[a]	23.86	5.20[a]	nc
TOEPH (3)	0.06	0.57	0.25	25.28	6.02	16.13	16.05[b]	0.49	0.37[a]	22.10	9.88[a]	nc
BAND (3)	0.05	0.48	0.10	3.84	0.39	15.07	7.38[b]	0.03	0.00a	59.20	10.88[a]	nc
BANDH (3)	0.06	0.36	0.21	2.90	1.25	10.42	10.70[b]	0.03	0.05[a]	51.47	19.98[a]	nc

Note. Convergence: 99.9% > a ≥ 90.0%.; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

All values shown have a convergence greater than 99.9%, unless otherwise specified.

T = TOEP; TH = TOEPH; T2 = TOEP (2); T3 = TOEP (3).

Table A5. Percentage of covariance models chosen by the BIC selection criterion from a total of 10000 analyzes with mixed linear models from data from simulated experiments involving 8 repeated measures, and sample size 8

Simulated model	Adjusted model											
	VC	CS	CSH	AR	ARH	T	TH	T (2)	TH (2)	T (3)	TH (3)	UN
	%											
I	79.30	7.80	0.27	4.30	0.13	0.86	0.02[b]	1.19	0.22[a]	5.83	0.08[a]	0.00
VC	54.06	4.81	14.56	3.17	7.47	0.36	0.88[b]	0.71	7.02[a]	4.08	2.79[a]	0.09
CS	0.00	97.21	0.99	0.09	0.00	1.59	0.11[b]	0.00	0.00a	0.00	0.00a	0.01
CSH	0.00	11.29	85.03	0.29	0.09	0.42	2.76[b]	0.00	0.00a	0.00	0.00a	0.12
AR	0.00	0.21	0.00	94.51	0.85	1.59	0.13[b]	0.11	0.00a	2.55	0.05[a]	0.00
ARH	0.00	0.05	0.08	53.95	40.0	0.35	1.54[b]	0.12	0.11[a]	2.00	1.73[a]	0.07
TOEP	0.00	1.21	0.05	89.08	1.15	8.02	0.38[b]	0.00	0.00[a]	0.09	0.01[a]	0.01
TOEPH	0.00	0.39	0.62	48.01	42.66	2.10	5.81[b]	0.00	0.00[a]	0.16	0.05[a]	0.2
TOEP (2)	0.00	0.03	0.00	12.79	0.18	13.68	0.44[b]	7.47	0.06[a]	63.21	2.09[a]	0.05
TOEPH (2)	0.00	0.06	0.02	9.93	3.73	6.60	5.59[b]	5.05	2.49[a]	40.00	26.21[a]	0.32
TOEP (3)	0.00	0.01	0.00	35.27	0.54	27.81	1.16[b]	0.01	0.00a	34.53	0.58[a]	0.09
TOEPH (3)	0.00	0.01	0.01	27.54	7.72	14.79	10.68[b]	0.00	0.01[a]	31.72	6.88[a]	0.64
BAND (3)	0.00	0.02	0.00	0.51	0.01	5.23	0.32[b]	0.00	0.00[a]	92.50	1.38[a]	0.03
BANDH (3)	0.00	0.00	0.01	0.33	0.13	3.17	2.65[b]	0.00	0.00[a]	74.26	19.30[a]	0.15

Note. Convergence: 99.9% > a ≥ 90.0%.; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

All values shown have a convergence greater than 99.9%, unless otherwise specified.

T = TOEP; TH = TOEPH; T2 = TOEP (2); T3 = TOEP (3).

Table A6. Percentage of covariance models chosen by the BIC selection criterion from a total of 10000 analyzes with mixed linear models from data from simulated experiments involving 8 repeated measures, and sample size 12

Simulated model	Adjusted model											
	VC	CS	CSH	AR	ARH	T	TH	T (2)	TH (2)	T (3)	TH (3)	UN
	%											
I	84.84	6.52	0.04	3.19	0.04	0.3	0.03	0.71	0.07	4.26	0.00	0.00
VC	44.1	2.65	22.24	1.94	12.29	0.07	0.29	0.6	11.01	2.12	2.69	0.00
CS	0.00	99.02	0.37	0.00	0.00	0.61	0.00	0.00	0.00	0.00	0.00	0.00
CSH	0.00	2.14	96.93	0.00	0.00	0.02	0.91	0.00	0.00	0.00	0.00	0.00
AR	0.00	0.01	0.00	98.39	0.23	0.5	0.00	0.00	0.00	0.87	0.00	0.00
ARH	0.00	0.00	0.01	36.25	61.8	0.1	0.65	0.00	0.00	0.42	0.77	0.00
TOEP	0.00	0.24	0.00	91.89	0.45	7.33	0.07	0.00	0.00	0.02	0.00	0.00
TOEPH	0.00	0.04	0.13	31.07	61.69	1.33	5.7	0.00	0.00	0.00	0.03	0.01
TOEP (2)	0.00	0.00	0.00	5.59	0.03	12.61	0.15	2.04	0.02	78.34	1.22	0.00
TOEPH (2)	0.00	0.00	0.00	2.95	2.55	4.97	5.65	1.46	1.19	36.46	44.76	0.01
TOEP (3)	0.00	0.00	0.00	28.09	0.2	40.88	0.38	0.00	0.00	30.12	0.33	0.00
TOEPH (3)	0.00	0.00	0.00	18.26	9.89	19.4	17.59	0.00	0.00	27.11	7.74	0.01
BAND (3)	0.00	0.00	0.00	0.03	0.00	3.78	0.06	0.00	0.00	95.86	0.27	0.00
BANDH (3)	0.00	0.00	0.00	0.04	0.04	1.5	1.8	0.00	0.00	70.45	26.15	0.02

Note. Convergence: 99.9% > a ≥ 90.0%.; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

All values shown have a convergence greater than 99.9%, unless otherwise specified.

T = TOEP; TH = TOEPH; T2 = TOEP (2); T3 = TOEP (3).

Appendix B

Type I error (α = 0.01)

Table B1. Empirical values for type I error from the hypothesis test of the interaction between treatments and repeated measures using mixed models with several covariance structures applied to simulated data with CS covariance structure, with different numbers of repeated measures and sample sizes, and with statistical significance equal to 0.01

Covariance model	Repeated measure	Sample size		
		4	8	12
VC	4	0.0000	0.0000	0.0000
	8	0.0000	0.0000	0.0000
CS	4	0.0092	0.0079	0.0106
	8	0.0105	0.0102	0.0120
CSH	4	0.0039	0.0071	0.0089
	8	0.0015	0.0063	0.0094
AR	4	0.0081	0.0110	0.0139
	8	0.0096	0.0109	0.0130
ARH	4	0.0074	0.0112	0.0148
	8	0.0045	0.0112	0.0143
TOEP	4	0.0152	0.0094	0.0118
	8	0.0363	0.0149	0.0121
TOEPH	4	0.0075	0.0082	0.0097
	8	0.0061[b]	0.0094	0.0102
TOEP (2)	4	0.0079	0.0042	0.0047
	8	0.0053	0.0021	0.0020
TOEPH (2)	4	0.0054	0.0043	0.0047
	8	0.0100[a]	0.0027	0.0024
TOEP (3)	4	0.0087	0.0033	0.0040
	8	0.0065	0.003	0.0027
TOEPH (3)	4	0.0060	0.0050	0.0063
	8	0.0098[a]	0.0045	0.0039
UN	4	0.0117	0.0104	0.0105
	8	nc	0.0097	0.0111

Note. Except where specified, all values represent convergence greater than 99.9%.

Convergence: 99.9% > a ≥ 90.0%; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

Table B2. Empirical values for type I error from the hypothesis test of the interaction between treatments and repeated measures using mixed models with several covariance structures applied to simulated data with AR covariance structure, with different numbers of repeated measures and sample sizes, and with statistical significance equal to 0.01

Covariance model	Repeated measure	Sample size		
		4	8	12
VC	4	0.0005	0.0006	0.0004
	8	0.0048	0.0029	0.0031
CS	4	0.0175	0.0191	0.0181
	8	0.0403	0.0415	0.0393
CSH	4	0.0091	0.0139	0.0162
	8	0.0141	0.0285	0.0321
AR	4	0.0054	0.0084	0.0085
	8	0.0046	0.0074	0.0088
ARH	4	0.0047	0.0092	0.0088
	8	0.0033	0.0084	0.0088
TOEP	4	0.0134	0.0111	0.0093
	8	0.0310	0.0112	0.0105
TOEPH	4	0.0070	0.0100	0.0086
	8	0.0073[b]	0.0118	0.0090
TOEP (2)	4	0.0087	0.0048	0.0038
	8	0.0095	0.0058	0.003
TOEPH (2)	4	0.0121	0.0055	0.0036
	8	0.0225[a]	0.0079	0.0046
TOEP (3)	4	0.0086	0.0063	0.0038
	8	0.0092	0.0071	0.0051
TOEPH (3)	4	0.0094	0.0077	0.0054
	8	0.0176[a]	0.0095	0.0056
UN	4	0.0091	0.0091	0.0083
	8	nc	0.0101	0.0109

Note. Except where specified, all values represent convergence greater than 99.9%.

Convergence: 99.9% > a ≥ 90.0%; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

Table B3. Empirical values for type I error from the hypothesis test of the interaction between treatments and repeated measures using mixed models with several covariance structures applied to simulated data with TOEPH covariance structure, with different numbers of repeated measures and sample sizes, and with statistical significance equal to 0.01

Covariance model	Repeated measure	Sample size		
		4	8	12
VC	4	0.0022	0.0008	0.0016
	8	0.0066	0.0037	0.0030
CS	4	0.0289	0.0304	0.0329
	8	0.0522	0.0504	0.0481
CSH	4	0.0099	0.0122	0.0124
	8	0.0138	0.0166	0.016
AR	4	0.0100	0.0125	0.0156
	8	0.0124	0.0148	0.0164
ARH	4	0.0070	0.0080	0.0105
	8	0.0069	0.0084	0.0100
TOEP	4	0.0180	0.0132	0.0146
	8	0.0294	0.0147	0.0155
TOEPH	4	0.0082	0.0088	0.0093
	8	0.0070	0.0108	0.0087
TOEP (2)	4	0.0133	0.0072	0.0080
	8	0.0114	0.0068	0.0057
TOEPH (2)	4	0.0133	0.0066	0.0056
	8	0.0205	0.0076	0.0049
TOEP (3)	4	0.0117	0.0078	0.0084
	8	0.0132	0.0069	0.0067
TOEPH (3)	4	0.0107	0.0081	0.0072
	8	0.0171	0.0067	0.0049
UN	4	0.0090	0.0104	0.0092
	8	nc	0.0113	0.0106

Note. Except where specified, all values represent convergence greater than 99.9%.

Convergence: 99.9% > a ≥ 90.0%; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

Appendix C

Statistical power (α = 0.01)

Table C1. Statistical power for the hypothesis test of the interaction between treatments and repeated measures using mixed models with several covariance structures applied to simulated data with CS covariance structure, with different numbers of repeated measures and sample sizes, and with statistical significance equal to 0.01

Covariance model	Repeated measure	Sample size		
		4	8	12
VC	4	0.0558	0.2565	0.6236
	8	0.0497	0.2097	0.5890
CS	4	0.6705	0.9893	0.9999
	8	0.8070	0.9987	1.0000
CSH	4	0.4698	0.9809	0.9999
	8	0.5171	0.9954	1.0000
AR	4	0.1574	0.6899	0.9628
	8	0.0817	0.3275	0.6807
ARH	4	0.1261	0.6595	0.9535
	8	0.0491	0.2983	0.6372
TOEP	4	0.4556	0.9632	0.9995
	8	0.5082	0.9847	0.9999
TOEPH	4	0.2823	0.9471	0.9993
	8	0.1384	0.9658	0.9999
TOEP (2)	4	0.1104	0.4250	0.7825
	8	0.0573	0.2384	0.6051
TOEPH (2)	4	0.0600	0.3170	0.6724
	8	0.0356	0.1548	0.4643
TOEP (3)	4	0.1263	0.3536	0.7193
	8	0.0548	0.2423	0.6026
TOEPH (3)	4	0.0519	0.2146	0.5748
	8	0.0271	0.1580	0.4529
UN	4	0.1994	0.9234	0.9988
	8	nc	0.7320	0.9964

Note. Except where specified, all values represent convergence greater than 99.9%.

Convergence: 99.9% > a ≥ 90.0%; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

Table C2. Statistical power for the hypothesis test of the interaction between treatments and repeated measures using mixed models with several covariance structures applied to simulated data with AR covariance structure, with different numbers of repeated measures and sample sizes, and with statistical significance equal to 0.01

Covariance model	Repeated measure	Sample size		
		4	8	12
VC	4	0.0839	0.2914	0.5837
	8	0.1140	0.3283	0.5571
CS	4	0.4331	0.8414	0.9710
	8	0.3596	0.6969	0.8813
CSH	4	0.2741	0.7866	0.9563
	8	0.1937	0.6044	0.8318
AR	4	0.1731	0.6358	0.8971
	8	0.0887	0.3640	0.6400
ARH	4	0.1317	0.6095	0.8851
	8	0.0556	0.3285	0.6005
TOEP	4	0.1756	0.5981	0.8715
	8	0.1271	0.3371	0.5998
TOEPH	4	0.0974	0.5359	0.8412
	8	0.0194	0.2626	0.5358
TOEP (2)	4	0.1564	0.5128	0.8035
	8	0.1127	0.3560	0.6077
TOEPH (2)	4	0.0910	0.3898	0.7011
	8	0.0654	0.2656	0.5063
TOEP (3)	4	0.0798	0.3347	0.6632
	8	0.0804	0.3216	0.5905
TOEPH (3)	4	0.0450	0.2439	0.5594
	8	0.0435	0.2401	0.4943
UN	4	0.0705	0.4682	0.8096
	8	nc	0.1131	0.3634

Note. Except where specified, all values represent convergence greater than 99.9%.

Convergence: 99.9% > a ≥ 90.0%; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

Table C3. Statistical power for the hypothesis test of the interaction between treatments and repeated measures using mixed models with several covariance structures applied to simulated data with TOEPH covariance structure, with different numbers of repeated measures and sample sizes, and with statistical significance equal to 0.01

Covariance model	Repeated measure	Sample size		
		4	8	12
VC	4	0.2551	0.6826	0.9166
	8	0.2299	0.6413	0.8681
CS	4	0.6702	0.9670	0.9972
	8	0.6512	0.9132	0.9961
CSH	4	0.4444	0.9516	0.9969
	8	0.4729	0.9301	0.9983
AR	4	0.4427	0.9221	0.9939
	8	0.2430	0.9120	0.9687
ARH	4	0.3442	0.9196	0.9943
	8	0.2918	0.9045	0.9970
TOEP	4	0.3046[a]	0.8505	0.9839
	8	0.1915	0.8254	0.9429
TOEPH	4	0.1697[a]	0.8348	0.9862
	8	0.0628	0.8142	0.9919
TOEP (2)	4	0.4044	0.8580	0.9814
	8	0.2385	0.8123	0.9195
TOEPH (2)	4	0.2321[a]	0.8196	0.9780
	8	0.1948	0.8023	0.9781
TOEP (3)	4	0.2113	0.7092	0.9465
	8	0.1943	0.6934	0.9297
TOEPH (3)	4	0.1188[a]	0.7076	0.9573
	8	0.1496	0.6973	0.9851
UN	4	0.1265	0.7559	0.9762
	8	nc	0.7435	0.9310

Note. Except where specified, all values represent convergence greater than 99.9%.

Convergence: 99.9% > a ≥ 90.0%; 90.0% > b ≥ 80.0%; 80.0 > c ≥ 70.0%; 70.0 > d ≥ 60.0%; nc < 60%.

Supplementary Tables

Please download the Supplementary Tables at https://ccsenet.org/journal/index.php/jas/article/download/0/0/48020/51593

Copyrights