Published online August 23, 2018

CHAPTER 10: ANALYSIS OF REPEATED MEASURES FOR THE BIOLOGICAL AND AGRICULTURAL SCIENCES

Salvador A. Gezan* and Melissa Carvalho

Abstract

Biological and agricultural experiments often evaluate the same experimental unit over time. Here, a repeated measures analysis is required by combining all measurements into a single complex model that specifies the correlated structure of the experimental data. This is needed, as the assumption of independence between observations is no longer valid, and therefore, an appropriate linear mixed model must be fit. Repeated measures analysis is strongly recommended for analyzing repeated observations because it usually results in reduced standard error of the means, which then produce narrower confidence intervals and increased statistical power. In this chapter, an introduction to the topic of repeated measures analysis is presented in the context of biological studies and particularly those that deal with plant science with emphasis on the specification and evaluation of the variance–covariance matrix. A detailed example illustrates this topic and code is provided for SAS, R, and GenStat focusing on testing and comparison of alternative models.

Many biological and agricultural field and laboratory experiments evaluate one or more responses over a given period of time where repeated measurements are performed on the same experimental unit over the length of the study. Often, conclusions and analyses of these studies require statistical evaluations for each of the time points and the complete set of observations. Evaluating data collected at each of the time points is often straightforward and can be completed by fitting linear models, followed by

Abbreviations: AIC, Akaike information criterion; ANOVA, analysis of variance; AR(1), autoregressive of order 1; ARH(1), heterogeneous autoregressive of order 1; BIC, Bayesian information criterion; BLUE, best linear unbiased estimate; BLUP, best linear unbiased prediction; COR, homogeneous correlation; CORGH, heterogeneous general correlation; CS, compound symmetry; CSH, heterocedastic compound symmetry; DIAG, diagonal; EXP, exponential; FA, factor analytic; GLMM, generalized linear mixed model; ID, independent; LMM, linear mixed model; LRT, likelihood ratio test; ML, maximum likelihood; MVN, multivariate normal; NLMM, nonlinear mixed model; RCBD, randomized complete block; REML, restricted/residual maximum likelihood; ReslogL, residual maximum log-likelihood; SEM, standard error of the mean; TOEP, Toeplitz; TOEPH, heterogeneous Toeplitz; UN, unstructured; US, unstructured.

S.A. Gezan and M. Carvalho, School of Forest Resources and Conservation, University of Florida, 363 Newins-Ziegler Hall, P.O. Box 110410, Gainesville, FL, 32611. *Corresponding author (sgezan@ufl.edu)

doi:10.2134/appliedstatistics.2016.0008

Applied Statistics in Agricultural, Biological, and Environmental Sciences
Barry Glaz and Kathleen M. Yeater, editors

© American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America
5585 Guilford Road, Madison, WI 53711-5801, USA.

the use of Analysis of Variance (ANOVA) tables and prediction of treatment means or evaluation of contrasts of interests to make inferences and develop conclusions. Several statistics books (e.g., Kuehl, 2000; Welham et al., 2014), and other chapters in this book, deal with biological experiments and provide recommendations and guidelines for performing proper statistical analyses in single-point analyses. However, an analysis that combines all time-points is more complex. The main complication is that the assumption of independence of the data may no longer be valid as the same experimental unit is measured several times, and therefore, correlation between repeated measurements of the same experimental unit needs to be incorporated into the linear model. Hence, more sophisticated statistical tools and training are required. Other authors have reported agricultural analyses with repeated measures (e.g., Piepho et al., 2004). The objective of this chapter is to introduce repeated measures analysis by presenting a brief overview and by illustrating repeated measures analysis with a few examples. This chapter is intended as an introduction to this topic. Important recommendations and references are provided to facilitate further study.

Repeated measures occur in experiments in which the same experimental unit is observed several times over time as a result of repeated sampling. For example, when strawberry (*Fragaria* × *ananassa* Duchesne ex Rozier) yield (in kg) for a plot of a given variety comprised of six plants is measured every week over the season, then we have repeated measures over time. In this example, we have some form of temporal correlation between observations belonging to the same experimental unit. Alternatively, we could have spatial correlation; for example, in a given point, several records of soil carbon content are obtained at different depths (e.g., 2, 4, 8, and 20 cm). In this chapter, we focus primarily on temporal correlations, as are typically found in agricultural experiments, and we leave the details of spatial correlations to the literature (for example, Cressie, 1993 and Chapter 12 (Burgueño, 2018). Therefore, repeated measures analysis is the use of statistical tools that deal with correlations between observations. Several approaches exist to analyze this type of data including multivariate techniques (see Chapter 14, Yeater and Villamil, 2018). However, in this chapter, we will focus on extending the use of linear mixed models (LMM) with a single response variable by starting from the original experimental design and its structure. The focus here is based on the assumption that the residuals have an approximate normal distribution. An example of repeated measures with a non-normal response is presented in Chapter 16 by Stroup (2018).

Why perform repeated analyses? Often researchers need to test inferences over time or space on the same experimental unit. For these cases, use of repeated measures analysis using LMM has several important benefits including:

- More efficient analyses, because when data are correlated, repeated measures analysis provides more information.
- Greater statistical power (see also Chapter 4 by Casler, 2018) due to using a more efficient analysis and better control of the factors affecting the process.
- Influence of missing data on the analysis is reduced. This is a benefit of using LMM to model correlations.
- Further biological interpretation (and testing) can be performed with the variance component estimates, such as the temporal correlations.

The main challenge in repeated measures analysis is the definition and incorporation of the correlated structure of the data. This is done by fitting extended LMMs that modify the assumptions of independence of the experimental units by modeling complex error structures which consider correlations among units and heterogeneity of variances. This is the topic of the next section.

Linear Mixed Models

Linear mixed models extend the typical linear model by allowing for more complex and flexible specifications of errors and other random effects by incorporating correlation and heterogeneous variances between the observations or experimental units. An important distinction with LMMs is the definition of fixed and random effects. The former corresponds to factors whose levels are specifically selected (nonrandom) and include the complete population of levels of interest (see also Chapter 16 by Stroup, 2018). In contrast, a random effect corresponds to a factor whose levels are a random sample from a population of a large number of levels. The important distinction is that in the case of fixed effects, the statistical inferences are only made on the specific factor levels selected in the experiment, whereas for the random effects, the inferences are about the complete population of levels, not only those included in the study. For example, consider a LMM to describe a randomized complete block design (RCBD) where all plants from each plot were observed. This model has a fixed block and treatment factor and a random plot factor and can be written as:

$$y_{ijk} = \mu + a_i + t_j + p_{ij} + e_{ijk}$$

where y_{ijk} is the observation from the kth plant from the ith block and jth treatment; μ is the overall mean; a_i is a fixed effect of block i; t_j is a fixed effect of treatment j; p_{ij} is a random effect of plot ij, with $p_{ij} \sim N(0, \sigma_p^2)$; and e_{ijk} is a random error, with $e_{ijk} \sim N(0, \sigma_e^2)$.

As indicated above, an important assumption is that random effects follow a normal distribution with a given variance–covariance structure, as shown for the plot factor. In this particular case, the plot factor needs to be considered random as it is modeling the nature of the experiment where several measurement units (e.g., plants within a plot) with the same treatment are grouped together. Ignoring this will result in inflated degrees of freedom and leads to pseudoreplication (further discussion on this topic can be found in Welham et al., 2014). Also, note that the error term in any linear model is also a random effect that is assumed to be normally distributed with its corresponding variance–covariance structure, that is, $e_{ijk} \sim N(0, \sigma_e^2)$. These variance-covariance structures are the key to most LMMs, and are often described by a matrix of variance-covariance parameters that specifies the properties of these random effects. For example, in matrix notation we assume that $\boldsymbol{e} \sim MVN(\mathbf{0}, \boldsymbol{R})$, where the bold letters identify vectors or matrices, and here the vector of errors $\boldsymbol{e}$, of dimension $n \times 1$, is assumed to follow a multivariate normal distribution with a mean vector of zeros, $\mathbf{0}$, of dimension $n \times 1$, and a variance–covariance matrix $\boldsymbol{R}$, of dimension $n \times n$, where n is the total number of observations. In the example presented above, we have $\boldsymbol{R} = \sigma_e^2 \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is an identity matrix of dimension $n \times n$ with ones in the diagonal and zeros in the off-diagonal, therefore, this $\boldsymbol{R}$ matrix specifies that the errors are all independent and have the same variance σ_e^2 (i.e., homocedastic). Similarly, for the plot effects we

have $\boldsymbol{p} \sim \text{MVN}(\mathbf{0}, \boldsymbol{W}_p)$, where $\boldsymbol{p}$ is a vector of plot effects of dimension $p \times 1$, $\mathbf{0}$ is a vector of zeros of dimension $p \times 1$ and $\mathbf{W}_p$ is a variance-covariance matrix of dimension $p \times p$, that in the above example is $\boldsymbol{W}_p = \sigma_p^2 \boldsymbol{I}_p$. Both of these matrices ($\boldsymbol{R}$ and $\boldsymbol{W}_p$) can be defined in many forms or structures (see more below) to specify correlations and heterogeneity of variances between random effects.

The estimation of the variance components is done by a likelihood-based method, where the most common is the restricted and/or residual maximum likelihood (REML) (Patterson and Thompson 1971). These variance components are later used in the normal equations of the LMM to obtain estimates of the fixed effects (best linear unbiased estimates, BLUEs) and random effects (best linear unbiased predictions, BLUPs) as derived in detail by Henderson (1984). As with linear models, further hypothesis testing by ANOVA tables, predictions of means, and evaluation of contrasts is possible with approximated *F*- and *t*-tests.

In this chapter, we do not provide more details, but good textbooks are available with additional details about LMMs and their properties. We recommend Littell et al. (2006), Pinheiro and Bates (2006), and Chapter 16 (Stroup, 2018) of this book.

Several statistical packages can be used to fit LMMs. These include SAS (SAS Institute Inc. 2011), R (R Development Core Team, 2008), and GenStat (Payne et al., 2011). The SAS package has the procedure PROC MIXED for normal data and PROC GLIMMIX for non-normal data. The R package has a few libraries, such as lme4 (Bates et al., 2015) and nlme (Pinheiro et al., 2016) that can fit several types of linear models. All of these packages and their corresponding libraries have different implementations of LMM methodologies and they use different names for the same variance-covariance structures with an array of options and functions, and in some cases they provide different estimates of the variance components. For this reason, we recommend carefully checking their properties to determine if they can provide the required output and flexibility.

Variance–Covariance Structures

For repeated measures analysis, as indicated earlier, the key is the specification of the variance-covariance matrix of error ($\boldsymbol{R}$), as it is here where we model the correlated nature of the data. To facilitate an understanding of the $\boldsymbol{R}$ matrix, we will assume that we have a group of m individuals (or experimental units) each measured t times; hence, the total database has $n = m \times t$ observations. For now, we will consider that all measurement time points were done at regular intervals (for example, every two days) $\boldsymbol{R}_n = \boldsymbol{I}_m \otimes \boldsymbol{G}_t$. As before, here $\boldsymbol{I}_m$ is an identity matrix of dimension m, which indicates that each of the experimental units is independent from each other, and $\boldsymbol{G}_t$ is a complex variance–covariance of dimension $t \times t$ that will model the correlations between repeated measurements. Also, $\otimes$ is the Kronecker product between matrices, which is an operation between two matrices that produces a block matrix, and it is also known as direct product. Here the key element is the specification of the G_t matrix for which we have several structures. Wolfinger (1996) presents a complete description of the relevant structures for repeated measures, and below we will present some of the most common (see Fig. 1). Note here that for our example we

will assume that we have $t = 4$ observations per individual, and we will define the structure in a generic way and not specifically for one or another statistical package.

The most basic reference structure is the independent matrix, or ID, that, as indicated earlier, assumes independence between observations. This can be extended to become heteroscedastic by assuming that each measurement will have a different error variance, and is often identified as DIAG. One of the simplest structures that allows for correlation between observations is compound symmetry, or CS. This structure and its heteroscedastic counterpart (CSH) have a variance component ρ that represents the correlation between any pair of observations. Hence, it assumes that the correlations between observations that are close or far apart in time will all have exactly the same correlation. For this reason, often this structure is not acceptable. However, if only a few observations are available (e.g., $t < 4$) CS or CSH could provide good approximations. The autoregressive error structure of order 1, AR(1), and its heterogeneous counterpart, ARH(1), are two of the most common structures used; they also have a single correlation variance component ρ, but in this case, there is an exponent that indicates the separation between repeated observations. For example, r^2, indicates that there are two time intervals (but one time point) between observations of the same experimental unit. One restriction of AR(1) and ARH(1) is that both assume that the intervals between observations are all identical. One equivalent structure that allows for modeling unequal intervals is the exponential structure (EXP), which depends on the variable d, that represents the distance (in time) between observations of the same experimental unit. If the intervals are all identical, then this model is a parametrization of the autoregressive. Another flexible model is the Toeplitz, or TOEP, also known as a diagonal constant matrix, which specifies a different correlation (or covariance depending on its parametrization) between observations. Hence, this has additional variance components but allows for greater flexibility. Finally, the unstructured (UN also identified as US), matrix is the most flexible. It specifies a different covariance for every pair and a different variance for each measurement point. This structure has the largest number of variance components to estimate, but as expected, offers the greatest flexibility. Sometimes, this matrix is expressed as an extension of correlation structures with homogeneous (COR) or heterogeneous variances (CORGH, also known as UNR in SAS). Note that different statistical packages (and R libraries) will have different names for the same structures, and they will also include a much larger number of alternative structures. Therefore, we recommend that the reader carefully review software manuals and literature. A good start on this topic is Wolfinger (1996).

New difficulties arise as the complexity of the structure increases (for example, from CS to UN). The first major issue is the increased computational complexity for estimating variance components. In all cases, REML is still used, but the minimization functions are more complex with the risk of nonconvergence of the model fit, or convergence to a local minimum instead of the global minimum. For this reason, it is recommended: i) to start with fitting a simpler model and to increase its complexity carefully; ii) if the software allows it, provide the routine with some starting values to aid with the convergence; and iii) always check that the variance component estimates are biologically reasonable.

ID: Identity

$$\sigma_0^2\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$$

DIAG: Diagonal

$$\begin{bmatrix}\sigma_1^2&0&0&0\\0&\sigma_2^2&0&0\\0&0&\sigma_3^2&0\\0&0&0&\sigma_4^2\end{bmatrix}$$

CS: Compound Symmetry

$$\begin{bmatrix}\sigma_1^2+\sigma_0^2&\sigma_0^2&\sigma_0^2&\sigma_0^2\\\sigma_0^2&\sigma_1^2+\sigma_0^2&\sigma_0^2&\sigma_0^2\\\sigma_0^2&\sigma_0^2&\sigma_1^2+\sigma_0^2&\sigma_0^2\\\sigma_0^2&\sigma_0^2&\sigma_0^2&\sigma_1^2+\sigma_0^2\end{bmatrix}$$

CSH: Heterogeneous Compound Symmetry

$$\begin{bmatrix}\sigma_1^2+\sigma_0^2&\sigma_0^2&\sigma_0^2&\sigma_0^2\\\sigma_0^2&\sigma_2^2+\sigma_0^2&\sigma_0^2&\sigma_0^2\\\sigma_0^2&\sigma_0^2&\sigma_3^2+\sigma_0^2&\sigma_0^2\\\sigma_0^2&\sigma_0^2&\sigma_0^2&\sigma_4^2+\sigma_0^2\end{bmatrix}$$

AR(1): Autorregressive of order 1

$$\sigma_0^2\begin{bmatrix}1&\rho^1&\rho^2&\rho^3\\\rho^1&1&\rho^1&\rho^2\\\rho^2&\rho^1&1&\rho^1\\\rho^3&\rho^2&\rho^1&1\end{bmatrix}$$

ARH(1): Heterogeneous Autorreg. of order 1

$$\begin{bmatrix}\sigma_1^2&\rho^1\sigma_1\sigma_2&\rho^2\sigma_1\sigma_3&\rho^3\sigma_1\sigma_4\\\rho^1\sigma_1\sigma_2&\sigma_2^2&\rho^1\sigma_2\sigma_3&\rho^2\sigma_2\sigma_4\\\rho^2\sigma_1\sigma_3&\rho^1\sigma_2\sigma_3&\sigma_3^2&\rho^1\sigma_3\sigma_4\\\rho^3\sigma_1\sigma_4&\rho^2\sigma_2\sigma_4&\rho^1\sigma_3\sigma_4&\sigma_4^2\end{bmatrix}$$

TOEP: Toeplitz

$$\begin{bmatrix}\sigma^2&\sigma_1&\sigma_2&\sigma_3\\\sigma_1&\sigma^2&\sigma_1&\sigma_2\\\sigma_2&\sigma_1&\sigma^2&\sigma_1\\\sigma_3&\sigma_2&\sigma_1&\sigma^2\end{bmatrix}$$

TOEPH: Heterogeneous Toeplitz

$$\begin{bmatrix}\sigma_1^2&\rho_1\sigma_1\sigma_2&\rho_2\sigma_1\sigma_3&\rho_3\sigma_1\sigma_4\\\rho_1\sigma_1\sigma_2&\sigma_2^2&\rho_1\sigma_2\sigma_3&\rho_2\sigma_2\sigma_4\\\rho_2\sigma_1\sigma_3&\rho_1\sigma_2\sigma_3&\sigma_3^2&\rho_1\sigma_3\sigma_4\\\rho_3\sigma_1\sigma_4&\rho_2\sigma_2\sigma_4&\rho_1\sigma_3\sigma_4&\sigma_4^2\end{bmatrix}$$

CORGH: Heterogeneous General Correlation

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

UN: Unstructured

$$\begin{bmatrix}\sigma_1^2&\sigma_{12}&\sigma_{13}&\sigma_{14}\\\sigma_{12}&\sigma_2^2&\sigma_{23}&\sigma_{24}\\\sigma_{13}&\sigma_{23}&\sigma_3^2&\sigma_{34}\\\sigma_{14}&\sigma_{24}&\sigma_{34}&\sigma_4^2\end{bmatrix}$$

FIG. 1. Common variance–covariance structures used to fit linear mixed models for repeated measures analysis.

Since there are many alternative models to evaluate, it is helpful to have a procedure to select the most appropriate variance–covariance structure. Such a procedure should identify the most parsimonious structure that describes the response variable in a reasonable way. This can be done by using the likelihood ratio test (LRT), which is based on asymptotic derivations. It compares nested models by using a Chi-square test that compares the residual log-likelihood (ReslogL) values between a model with a complex structure (ReslogL_2) and a simpler counterpart (ReslogL_1). The statistic used is:

$$\chi^2_d = 2\,\text{ReslogL}_2 - 2\,\text{ReslogL}_1 \sim \chi^2_{k2\text{-}k1}$$

where ReslogL_1 and ReslogL_2 are the residual maximum log-likelihood values for the corresponding models 1 and 2, and $k_2 - k_1$ is the degrees of freedom calculated as the difference between the number of variance components to estimate each of the models. It is important to note that when REML variance components are used, this test requires that the fixed effects between Models 1 and 2 are exactly the same, otherwise, the test is incorrect and a maximum likelihood procedure (ML) should be used (see also Chapter 11 by Payne, 2018).

In addition, and particularly for non-nested models, it is possible to calculate information criteria, such as the Akaike Information Criterion (AIC, Akaike 1974) and the Bayesian Information Criterion (BIC, Schwarz 1978). Both of these criteria are used to compare the fit between two or more alternative models. The AIC is more liberal and the BIC more conservative (i.e., it chooses models with less parameters), where the latter is not sensitive to prior distributions for large sample sizes. The BIC was developed using a Bayesian approach. Studies by Guerin and Stroup (2000) found that larger values of BIC were associated with larger Type 2 errors. The expressions for these criteria are:

$$\text{AIC} = -2\,\text{ReslogL} + 2k$$

$$\text{BIC} = -2\,\text{ReslogL} + k\log(n)$$

where ReslogL is the residual maximum log-likelihood value of the model, k is the number of effective variance parameters to estimate the model; and n is the sample size. Lower values for AIC and BIC indicate a better model fit. In the examples presented later, the selection of the variance structure will be illustrated in detail.

Fitting a Linear Mixed Model for Data with Repeated Measures

The process to fit a LMM for repeated measures data should be done with care. There are several steps to follow to select a reasonable model. Often, the first step is to fit every single measurement point individually. This step allows to clearly define the linear (mixed) model to use, but also facilitates the detection of departures from normality, and the detection of potential outliers. Also, evaluation of the ANOVA tables provides a preliminary idea of the factors that are significant for the response variable under study. It is also recommended to collect all variance components estimated from this step to be used later as starting values for a more complex model.

The second step consists of extending the original model to construct the repeated measurements LMM. This is done by adding time (as a factor or covariate) individually and with all its interactions (or in some cases as nested effects) with the original model terms. At this stage, it is important to define if the interest is to consider time as a factor, for which a different mean estimate is obtained for each level (or measurement point), or to assume it as a continuous variable (or covariate), for which the repeated measures model will now represent lines (or curves) if this is considered as an explanatory variable. These two approaches have different objectives, when time is a factor we are interested in comparing each of the time points and overall differences; however, when time is a continuous variable, we are interested in the describing the patterns over time of each of the treatments, and we might perform interpolations, as done with regression analyses.

For the construction of this model, the next step is the specification of the error structure. Here it is recommended to start with simpler structures, such as ID or DIAG and then evaluate some other more complex models. We recommend the use of AR(1) or AR1(H) as a baseline. It is often useful to fit the UN structure; however, convergence for the UN is usually difficult. Several error structures should be fitted for the current data, and evaluated using LRT (which is only valid when comparing nested models) or AIC and BIC goodness-of-fit statistics to select the error structure. Note here that we recommend to select the most parsimonious model as the main interest of this analysis is not focused on the interpretation of the error structure. For this reason, often simpler, and somehow incomplete error structures sometimes are selected. Finally, given that the error structure is clear, we can focus our attention on the ANOVA table, which we can follow with prediction of means and comparisons of predetermined contrasts of interest. It is possible at this stage, or it may be necessary to achieve convergence, to drop some factors from the model, or increase its complexity; however, changes in the model often require revisiting the selection of the error structure. Also, residuals need to be examined for potential departures from the basic assumptions. We recommend focusing on Studentized or standardized residuals, as these take into consideration the heterogeneity of variances.

Detailed Example

An experiment was established to assess the effects of three site-preparation treatments (*v*-plow, hand screef or removal of vegetation, and untreated control), two seedling species (Douglas-fir [*Pseudotsuga menziesii* {Mirb.} Franco] and lodgepole pine [*Pinus contorta* Douglas ex Loudon]), and two types of stock (bare root and plug) in a trial located in the Cariboo Forest Region (Nemec, 1996). The experiment used a RCBD with four blocks. Each block contained 12 plots, and a plot consisted of a single row of 25 seedlings. For this example, we will focus on the mean plot height (cm) of the plants. The plants in this experiment were observed annually for 6 yr (from 1984 to 1989), and an initial observation (before planting) is also available. The objective is to determine whether site-preparation treatment, species, stock type, or any of their interactions affect growth. The complete dataset is presented in Table 1.

For fitting this model, we will start by defining the complete factorial model for each of the observed measurement points as:

TABLE 1. Raw data used for example originating from a field trial located in the Cariboo Forest Region (Canada). Source: Nemec (1996). The response variable corresponds to the mean plant height (m) of a plot conformed by 25 seedlings. †

Spp	Stk	Prep	Trt	Blk	Initial	1984	1985	1986	1987	1988	1989
FD	B	S	FD-B-S	1	18.78	22.89	22.44	22.89	22.44	27.56	33.56
FD	B	S	FD-B-S	2	15.92	20.08	21.50	24.75	28.42	39.67	53.67
FD	B	S	FD-B-S	3	20.80	26.60	28.20	27.90	36.90	48.30	59.80
FD	B	S	FD-B-S	4	18.60	22.60	21.40	23.00	22.20	30.60	39.80
FD	B	U	FD-B-U	1	14.10	18.40	20.80	24.70	29.00	40.80	57.70
FD	B	U	FD-B-U	2	17.00	21.00	22.00	18.00	20.00	22.00	25.00
FD	B	U	FD-B-U	3	18.00	22.22	25.22	27.56	29.33	38.67	49.67
FD	B	U	FD-B-U	4	18.14	23.29	25.36	28.36	30.07	38.00	48.50
FD	B	V	FD-B-V	1	16.14	19.81	22.10	25.95	33.43	45.76	59.19
FD	B	V	FD-B-V	2	14.89	19.28	21.11	25.78	30.28	39.89	53.83
FD	B	V	FD-B-V	3	15.08	18.08	19.77	23.08	27.08	37.38	52.08
FD	B	V	FD-B-V	4	19.00	23.06	24.24	28.12	34.47	45.35	58.12
FD	P	S	FD-P-S	1	18.94	26.71	31.06	33.24	42.00	54.12	65.94
FD	P	S	FD-P-S	2	22.82	28.91	34.05	39.91	47.32	59.50	75.64
FD	P	S	FD-P-S	3	22.90	28.15	32.95	39.05	49.20	62.60	74.85
FD	P	S	FD-P-S	4	20.59	25.65	30.12	33.35	39.53	49.29	61.65
FD	P	U	FD-P-U	1	21.56	29.22	31.17	38.00	43.83	54.72	66.78
FD	P	U	FD-P-U	2	20.47	27.11	33.16	39.95	46.74	57.37	70.79
FD	P	U	FD-P-U	3	19.05	27.52	31.33	37.95	46.62	56.95	68.52
FD	P	U	FD-P-U	4	16.29	24.71	31.05	35.76	43.48	55.62	70.86
FD	P	V	FD-P-V	1	18.08	23.63	28.54	37.08	47.83	64.75	86.75
FD	P	V	FD-P-V	2	20.88	26.58	30.50	35.88	42.83	53.71	70.58
FD	P	V	FD-P-V	3	20.19	25.94	28.38	32.38	37.06	48.63	67.63
FD	P	V	FD-P-V	4	20.40	26.25	30.00	34.25	38.35	49.05	65.30
PL	B	S	PL-B-S	1	12.44	20.28	34.67	49.83	65.78	90.83	125.17
PL	B	S	PL-B-S	2	16.33	22.00	32.00	47.33	58.33	81.73	113.07
PL	B	S	PL-B-S	3	11.32	20.05	32.45	47.45	66.77	96.32	132.55
PL	B	S	PL-B-S	4	11.11	16.37	28.74	44.11	65.79	94.05	130.74
PL	B	U	PL-B-U	1	12.40	17.40	26.50	36.80	49.60	75.80	109.90
PL	B	U	PL-B-U	2	13.44	18.44	24.75	41.00	54.31	80.00	111.56
PL	B	U	PL-B-U	3	14.19	19.44	29.63	48.38	62.25	86.69	120.31
PL	B	U	PL-B-U	4	15.22	20.61	29.61	41.83	55.06	70.72	97.78
PL	B	V	PL-B-V	1	12.31	17.08	26.38	41.46	64.08	100.38	147.62
PL	B	V	PL-B-V	2	13.94	19.00	28.41	46.41	81.94	118.41	166.18
PL	B	V	PL-B-V	3	11.53	18.58	29.68	45.47	70.89	105.47	150.32
PL	B	V	PL-B-V	4	12.63	16.63	27.06	43.75	67.38	103.06	146.75
PL	P	S	PL-P-S	1	12.43	23.19	36.71	55.29	75.71	109.48	155.76
PL	P	S	PL-P-S	2	10.23	18.59	33.91	53.59	74.09	108.27	150.64
PL	P	S	PL-P-S	3	9.59	17.82	32.05	49.86	69.50	97.59	133.55
PL	P	S	PL-P-S	4	13.48	21.70	34.26	48.22	73.39	103.83	141.48
PL	P	U	PL-P-U	1	12.00	22.86	34.38	49.00	71.10	105.05	148.71
PL	P	U	PL-P-U	2	9.43	17.14	30.10	43.33	60.95	87.24	125.67
PL	P	U	PL-P-U	3	8.15	15.95	28.60	39.65	58.75	89.00	129.40
PL	P	U	PL-P-U	4	8.75	15.70	27.45	42.55	58.45	85.55	123.85
PL	P	V	PL-P-V	1	12.28	19.52	33.12	55.12	89.24	136.16	193.56
PL	P	V	PL-P-V	2	9.57	17.13	28.74	46.65	74.00	114.22	163.13
PL	P	V	PL-P-V	3	10.25	17.83	29.38	48.00	78.88	116.29	161.50
PL	P	V	PL-P-V	4	7.83	13.58	30.00	53.42	84.71	130.38	186.21

† Spp, seedling species (FD: Douglas-fir, PL: lodgepole pine); Stk, stock type (B: bare root, P: plug); Prep, site-preparation treatment (S, hand screw, U, untreated, V, v-plow); Trt, treatment combination of Spp, Stk and Prep; Blk, block; Initial, mean plot height at planting (1983).

TABLE 2. Summary of results from fitting single point measurement for Cariboo Forest Region data. P-values are presented for each of the model terms, together with residual variance (s_e^2), and average standard error of the mean (SEM) for the treatment combinations.

Effect	df	1984	1985	1986	1987	1988	1989
ht0	1	< 0.0001*	< 0.0001*	0.015*	0.089	0.408	0.821
Blk	3	0.082	0.441	0.814	0.751	0.498	0.402
Trt	11	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
s_e^2	–	1.020	2.845	9.724	26.450	55.483	105.980
SEM	–	0.615	1.026	1.898	3.130	4.533	6.265

* Significant at the 0.05 probability level.

$$y = \mu + \text{ht0} + \text{Blk} + \text{Spp} + \text{Stk} + \text{Prep} + \text{Spp}\times\text{Stk} + \text{Spp}\times\text{Prep} + \text{Stk}\times\text{Prep} + \text{Spp}\times\text{Stk}\times\text{Prep} + e$$

where all factors will be considered as fixed effects, with the exception of the error term *e*. Here, Blk represents the blocks, Ssp the species, Stk the stock, Prep the site-preparation treatment, and the other terms are the two- and three-way interactions. The term ht0 is a covariate representing the initial height (cm) of the plants at the beginning of the experiment (before the treatments were applied). For more details about analysis of covariance, see Chapter 9 by McCarter (2018).

For simplicity, at this stage we will combine all treatment factors (SSp, Stk, and Prep) into a single combined factor with 12 levels, identified as Trt (see also Table 1). Hence, the model is represented as:

$$y_{(t)} = \mu_{(t)} + \text{ht0}_{(t)} + \text{Blk}_{(t)} + \text{Trt}_{(t)} + e_{(t)}, \text{ for } t = 1, \ldots, 6$$

Here, the index (*t*) is used to identify the different measurement points, and the error terms for each are assumed to be $\boldsymbol{e}_{(t)} \sim \text{MVN}(\mathbf{0}, \sigma_e^2 \boldsymbol{I}_n)$, with *n* the number of plots in the experiment. The fitting of the above model was done using SAS 9.3 (SAS Institute Inc. 2011) and the summary of each of the measurement points is shown in Table 2 (see Appendix 1 for SAS, R, and GenStat code). Note that the significance of some factors changes from measurement to measurement point (year to year), however, Trt is always significant. In addition, the estimated residual error, s_e^2, and the standard error of the mean (SEM) for Trt increase with time as would be expected as the plants become larger with time. The presence of heterogeneous variances among years, as with heterogeneous variances among time points or spaces in other experiments, is one of the major reasons that repeated measures analysis is useful.

The next step is to extend the above model to all six measurement points. To do this, we incorporate the factor Time that has a total of six levels. This factor is added alone and with all its potential interactions. Hence, we have:

$$y = \mu + \text{Time} + \text{ht0} + \text{Blk(Time)} + \text{Trt} + \text{Time}\times\text{Trt} + e$$

There are important elements in this model that need further clarification. First, *y* now represents a vector of dimension $n\times1$ that includes observations from all *m* experimental units and all *t* measurement times ($n = m\times6$). We will assume this is sorted by individual and then measurement point within individual; this is not

critical here but for some statistical packages specific data sorting is often required. The constant μ now represents an overall mean across all observations, and therefore, often does not have a reasonable interpretation. The factor Time indicates that for each time point, there is a different expected value. Therefore, it is important to determine if there are trends for height over time in this study. The term ht0 constitutes the covariate that is correcting for initial plant height. The role of this covariate is to adjust the data due to different starting conditions of the plants at the beginning of the experiment. The factor Blk(Time) corresponds to blocks nested within time, and incorporates the different effects due to block at each time point. Having only the factor Blk instead of Blk(Time) would assume that the block effects are all the same for each time point, an assumption that might be incorrect. Also, note that if the Blk term from the single point model would have been assumed to be random, then there would have needed to be variance components associated with this factor. This implies that for the complete repeated measures model, the term Blk(Time) should have a different variance component for each measurement point (hence, six variance components), corresponding to a DIAG structure (see Fig. 1), hence, Blk(Time) ~ MVN($\mathbf{0}$, $\boldsymbol{D}_m \otimes \boldsymbol{I}_b$) with $\boldsymbol{D}_m$ a diagonal matrix of dimension 6 × 6. Note that the choice of random or fixed effects for blocks depends on the assumptions of the scientist, objectives of the study, and the characteristics of the experiment. For this reason we do not discuss this topic any further (see Littell et al., 2006 for an excellent explanation). The factor Trt represents the treatment effect across all time points. Thus, we can view Trt as the effect of a given treatment averaged over the t time points, which often is the main hypothesis of interest. The model factor Time×Trt

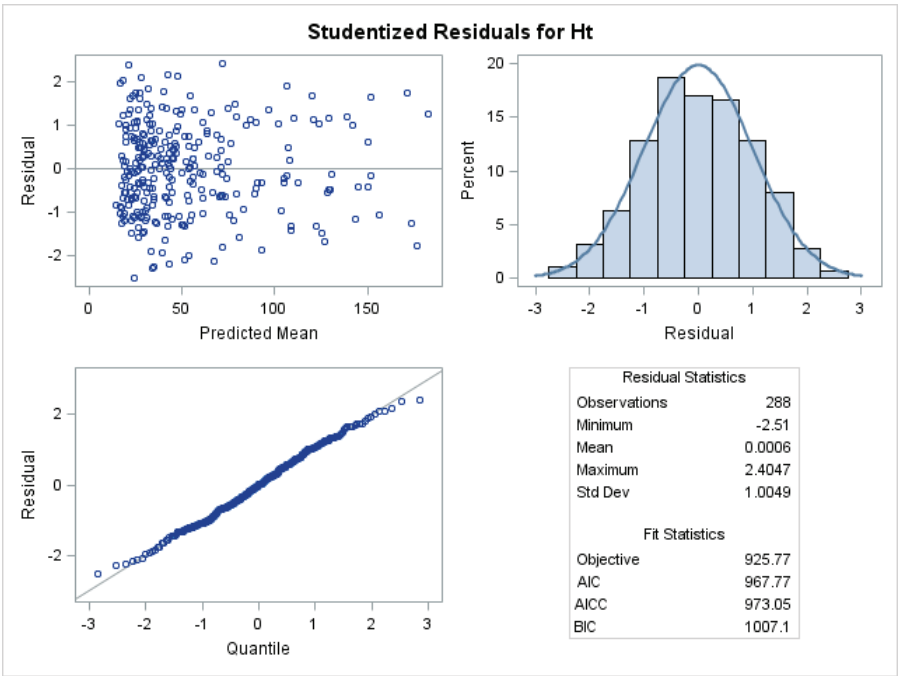

FIG. 2. Panel of residual plots for final model for repeated measures analysis with UN error structure produced with SAS v. 9.3.

TABLE 3. Goodness-of-fit statistics for different error structures evaluated in Cariboo Forest Region data. The statistics presented are: residual maximum log-likelihood (ReslogL), Akaike information criterion (AIC) and Bayesian information criterion (BIC). The parameter *k* indicates the number of variance components estimated by the corresponding error structure.

Structure	*k*	-2 ReslogL	AIC	BIC
ID	1	1390.5	1392.5	1394.3
CS	2	1320.3	1324.3	1328.1
AR(1)	2	1151.5	1155.5	1159.2
TOEP	6	1109.3	1121.3	1132.5
DIAG	6	1204.3	1216.3	1227.5
CSH	7	1092.8	1106.8	1119.9
ARH(1)	7	1008.1	1022.1	1035.2
TOEPH	11	1001.2	1023.2	1043.8
UN	21	925.8	967.8	1007.1

represents the interaction of treatment with time; this is probably one of the most important pieces of information from the repeated measures analysis. Finally, the error term is assumed to be $\boldsymbol{e} \sim \text{MVN}(\mathbf{0}, \boldsymbol{I}_m \otimes \boldsymbol{G}_t)$, where, as indicated earlier, $\boldsymbol{G}_t$ is the matrix to represent repeated measures error structure.

For the above repeated measures model, we have fit several error structures. For all of these models, the number of variance components and the -2 ReslogL, AIC, and BIC are presented in Table 3. According to the AIC and BIC (smaller is better for both), the best error structure is UN, which has a total of 36 variance components. The ARH(1), a simpler model with only seven variance components is the second best model. For this dataset, the UN structure had no difficulties converging, but if this was not the case, ARH(1) would have been selected. To illustrate the use of the LRT, we can compare the models ARH(1) and DIAG. These two structures are nested, and the only difference is the presence of the temporal correlation ρ (see Fig. 1). Hence, the hypothesis that is being tested is H_0: $\rho = 0$ against H_1: $\rho \neq 0$, which is a two-sided hypothesis, with a critical value of χ^2_d = 1204.3– 1008.1 = 196.2, which is compared with a value of 3.84 corresponding to a critical χ^2 with 1 df for a 5% significance level. Therefore, for this example, we have more than enough evidence to conclude that this temporal correlation is highly significant. Comparisons between other nested structures are possible, and we leave this to the interested reader.

Now that we have selected the UN error structure for our model, and after checking for departures form normality and presence of outliers (see Fig. 2, which was part of our SAS output, an excerpt from a panel of Studentized residuals), we can proceed to check the Test of Fixed Effects (Table 4). It is clear form this table that there are several significant interactions, particularly relevant is the Spp×Stk×Prep×Time, interaction which means that we need to report the plant height mean for each combination of treatment and for each time point.

One difficulty that arises from fitting a complex LMM is that the *F*- and *t* tests are no longer valid (Littell et al., 2006). This occurs because these tests are derived assuming there is only one variance component in the linear model, σ_e^2. However, for most LMMs, we have several variance components that first are estimated and then used to construct ANOVA–like tables. For this reason, some packages report

only asymptotic tests (such as χ^2 and z values by using Wald-type statistics), or others perform some corrections on the degrees of freedom (df) to allow for this additional level of approximation and uncertainty on the construction of these tests. One of the most recommended approaches is to use the Kenward–Rogers df correction (Kenward and Rogers, 1997). We have used this correction to generate the results in Table 4 where some denominator df have decimals.

Therefore, for this repeated measures analysis, it is possible to conclude that there are significant effects of most of the treatment factors and that the responses of the treatments (or treatment combinations) depend on the time point under study. We can proceed to predict means for some treatments with confidence intervals and generate different tables and graphs to report these results. In addition, it is also possible to perform some specific comparisons. For example, for this exploratory model we can evaluate, with the slice function in SAS, if there are significant differences between all 12 levels of treatments. These results are presented in Table 5 together with the SEMs for these treatments (these were also presented in Table 1 but for single time point analyses). Note that these SEMs increase with time. This results from having a different error variance for each year (see also Table 6). For the repeated measures analysis based on these data, there are no differences of conclusions compared with the previous model. This is because there were substantial differences among treatments. In addition, there are reductions of the SEM of ~10% in most years, with the exception of 1984. This is a result of a more efficient use of

TABLE 4. Test of fixed effects for final fit model based on an UN error structure for the Cariboo Forest Region data.

Effect	Numerator df	Denominator df	*F*-value	*P*-value
ht0	1	32	107.09	< 0.0001
Time	5	29	582.02	< 0.0001
Blk(Time)	18	49.8	0.84	0.6420
Trt	11	34	71.57	< 0.0001
Spp	1	46.9	653.97	< 0.0001
Stk	1	32.9	72.27	< 0.0001
Spp×Stk	1	35.4	0.38	0.542
Prep	2	33.1	15.90	< 0.0001
Spp×Prep	2	33	12.59	< 0.0001
Stk×Prep	2	33	0.40	0.6730
Spp×Stk×Prep	2	32.9	0.90	0.4170
Trt×Time	55	68.6	19.03	< 0.0001
Spp×Time	5	29	168.22	< 0.0001
Stk×Time	5	29	10.84	< 0.0001
Prep×Time	10	41.3	9.87	< 0.0001
Spp×Stk×Time	5	29	2.14	0.0890
Spp×Prep×Time	10	41.3	5.34	< 0.0001
Stk×Prep×Time	10	41.3	0.29	0.9792
Spp×Stk×Prep×Time	10	41.3	2.62	0.0146

TABLE 5. Summary of results from evaluating hypothesis of differences between treatment combinations (Trt) within a given measurement point from fitting full repeated measure model with UN error structure for the Cariboo Forest Region data. *P*-values are presented together with the average standard error of the mean (SEM) for the treatment combinations.

Effect	df	1984	1985	1986	1987	1988	1989
Trt	11	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
SEM	–	0.635	0.964	1.617	2.574	3.719	5.148

* Significant at the 0.05 probability level.

TABLE 6. Estimated variance-covariance and correlation matrices from fitting full repeated measures model with UN error structure for the Cariboo Forest Region data.

	Variance-covariance matrix						Correlation matrix					
	1984	1985	1986	1987	1988	1989	1984	1985	1986	1987	1988	1989
1984	1.018	1.077	0.834	1.412	1.639	1.984	1	0.606	0.264	0.275	0.220	0.192
1985	1.077	3.103	3.870	5.506	7.220	9.791	0.606	1	0.701	0.615	0.554	0.542
1986	0.834	3.870	9.830	12.668	18.027	25.013	0.264	0.701	1	0.794	0.778	0.777
1987	1.412	5.506	12.668	25.872	35.832	47.031	0.275	0.615	0.794	1	0.953	0.901
1988	1.639	7.220	18.027	35.832	54.682	74.469	0.220	0.554	0.778	0.953	1	0.981
1989	1.984	9.791	25.013	47.031	74.469	105.370	0.192	0.542	0.777	0.901	0.981	1

the information which translated into narrower confidence intervals and increased power to detect differences between treatment levels.

The inspection of the variance components for this analysis can provide some insight into the response variable of interest. For the current data, a matrix of variance-covariance and a matrix of correlations are presented in Table 6. These results, particularly the correlations, indicate that the pair-to-pair correlations increase with time, with high values for years 1988 and 1989. This indicates that the conclusions form these two years will be similar (data not shown). In addition, the first year (1984) has the lowest correlations, indicating that this measurement point is too early in the experiment to provide sufficient information to compare treatments across time.

Finally, it is possible to fit a slightly different, and simplified model, which is common in many repeated measures analyses. This model incorporates a random effect and has the following form:

$$y = \mu + \text{Time} + \text{ht0} + \text{Blk(Time)} + \text{Trt} + \text{Time} \times \text{Trt} + \text{Plot} + e$$

where all the terms were previously defined but here Plot is a random effect factor that identifies each experimental unit (i.e., plots, see Table 1), with Plot ~ MNV($\mathbf{0}$, $\sigma_p^2 \boldsymbol{I}_n$), and the residual errors are all assumed to be independent, that is, $\boldsymbol{e}$ ~ MNV($\mathbf{0}$, $\sigma_e^2 \boldsymbol{I}_n$). This model is equivalent to the previously fit model with CS for the error structure, with the difference that the modeling of the correlation structure is done through the model term Plot. Here, ReslogL values and variance component estimates are identical, and for this reason, this simplification is often preferred. However, it presents the same difficulties as the CS, mainly that they assume a common correlation between any pair of measurement points, which is only recommended whenever the number of time measurements is small ($t < 3$). Nevertheless, the issue of heterogeneity of variances for each of the measurement points still needs to be addressed

properly, and this requires complex models, which are easily implemented by proper specification of the error structure.

Time as a Continuous Variable

In the example presented above we assumed that the time model term was a factor; however, if we were interested in trends, it is possible to change this factor into a continuous variable (or variate) as, for example:

$$y = \mu + \text{Timec} + \text{Timec}^2 + \text{ht0} + \text{Blk(Time)} + \text{Trt} + \text{Timec}\times\text{Trt} + \text{Timec}^2\times\text{Trt} + e$$

In this particular case, we are modeling a quadratic trend, as we are using the square of the continuous variate Timec, which is centered to avoid issues with multicollinearity between continuous variables. A centered variable is one where the mean of the data vector is subtracted to each observation (Welham et al., 2014). Note, that for the above model, we still need to specify the error structure and also, the term Blk(Time) is still included to model for a different block effect nested within each time point; therefore, for this model, Time is still considered as a factor rather than a variate.

In the example presented earlier, we focused on a normally distributed response variable. Often, we have discrete responses that follow a binomial distribution (such as survival with a yes/no or a 0/1 response) or a Poisson distribution (such as count of eggs). For these cases, we recommend the reader to review Chapter 16 by Stroup (2018), which describes in more detail a Generalized Linear Mixed Model (GLMM) approach. In short, this methodology allows the researcher to consider the specific probability distribution of the response, and allows for incorporating random effects

TABLE 7. Data from a soybean experiment that compares two varieties (P, introduction #416937, F, Forrest) measured weekly ten times, starting at 14 d after planting. Only data from year 1988 is considered. Source: Davidian and Giltinan (1995).

Plot	Variety	14	21	28	35	42	49	56	63	70	77
F1	F	0.106	0.261	0.666	2.110	3.560	6.230	8.710	13.350	16.342	17.751
F2	F	0.104	0.269	0.778	2.120	2.930	5.290	9.500	–	16.967	17.747
F3	F	0.108	0.291	0.667	2.050	3.810	6.130	10.280	18.080	20.183	21.811
F4	F	0.105	0.299	0.844	1.320	2.240	4.680	8.820	15.090	14.660	14.005
F5	F	0.101	0.273	0.848	1.950	4.770	6.010	9.910	–	19.262	–
F6	F	0.106	0.337	0.699	1.530	3.870	5.600	9.430	13.730	17.381	19.932
F7	F	0.102	0.275	0.767	1.450	3.950	4.940	9.640	–	17.880	17.724
F8	F	0.103	0.273	0.742	1.410	3.010	5.260	9.810	12.850	18.214	19.680
P1	P	0.131	0.338	0.701	1.660	4.250	9.240	12.150	16.780	15.925	17.272
P2	P	0.128	0.404	0.897	1.780	3.910	7.400	10.070	18.860	17.012	27.370
P3	P	0.131	0.379	1.126	2.440	3.890	6.910	12.490	15.670	23.763	21.491
P4	P	0.154	0.357	1.181	1.830	4.710	10.710	9.910	15.510	14.958	21.800
P5	P	0.139	0.328	0.932	1.990	3.460	7.020	11.790	15.830	15.921	17.442
P6	P	0.139	0.389	1.094	2.130	4.040	7.620	12.480	17.930	14.422	30.272
P7	P	0.145	0.366	0.799	1.610	3.510	6.790	9.950	14.540	19.280	22.573
P8	P	0.130	0.355	1.090	2.280	3.940	4.960	10.920	14.020	17.994	22.371

and specifying complex variance–covariance structures as done with LMMs. Most statistical packages have routines to fit these models. However, fitting a GLMM for repeated measures data, while possible, is often complex due to issues with convergence. We also recommend Gbur et al. (2012) for more information on general implementation of these models and Bolker et al. (2009) for further technical details.

So far, we have presented a few of the most common variance–covariance structures. However, it is possible to use other structures. One popular structure used in agriculture, and particularly in plant breeding, is the factor analytic (FA). This structure has the advantage of providing a good approximation of the UN structure while using a reduced number of variance components. Therefore, the FA structure tends to converge more easily than the UN. For further details and properties of the FA structure, we recommend Smith et al. (2015). Also, of particular relevance are the structures used for spatial statistics that are more flexible and consider irregular time or space measurement intervals, and even different measurement points per experimental unit. Good introductions of this topic are provided in the books from Cressie (1993) and Webster and Oliver (2007).

Finally, it is also possible to model repeated measures with the use of nonlinear models, more specifically, these will be nonlinear mixed models (NLMM). This is the typical case, for example, of a growth curve that models the development of a given experimental unit that is observed several times. The advantage is that the form of the nonlinear model often has good interpretability and better biological background than a linear model. For NLMM, as with LMM, we would need to model the error structure and follow a similar procedure as already described in this chapter. Nevertheless, use of these models is limited due to issues with convergence. Also, statistical tests and confidence intervals associated with NLMM's are based on asymptotic approximations. Additional details of these models, with some interesting examples, can be found in Davidian and Giltinan (1995).

Conclusions

In this chapter we have presented a brief introduction to the topic of repeated measures analysis in the context of biological studies, particularly those that deal with the plant sciences. These analyses rely strongly on the theory and practice of linear mixed models, which are powerful tools for many situations. We have presented only some general topics and illustrated these topics with an example. There are many more aspects that were not presented here. However, there are good references available for these topics. Nevertheless, the analysis of data with repeated measures requires a mixture of solid statistical modeling and practical experience, as each experiment and its corresponding datasets differ. Repeated measures analysis is strongly recommended for analyzing repeated observations of the same experimental unit because it usually results in reduced SEMs which then produce narrower confidence intervals and increased statistical power. However, there are usually concerns, difficulties, and diverse challenges with repeated measures analyses that need careful attention.

Key Learning Points

- If the same experimental unit is observed over time, then the

assumption of independence of the data is no longer valid, and this correlation needs to be incorporated into the statistical analysis.

- Repeated measures analysis has several important benefits, particularly a reduction on the SEMs, which then produce narrower confidence intervals and increased statistical power.
- The specification of the variance-covariance matrix of errors is key for these analyses, and different structures need to be evaluated to properly describe the underlying biological process.
- The analysis of data with repeated measures requires a mixture of solid statistical modeling and practical experience, as each experiment and its corresponding datasets differ.

Review Questions

True or False

1. Spatial correlation is a type of correlation that is present between observations that belong to the same experimental unit.

2. If we have missing data, then repeated measures analysis can't be used.

3. Combining all data from several time points into a single analysis will provide greater statistical power than analyzing every time point separately.

4. For random effects, the statistical inferences are valid only for the levels that are considered in the corresponding factor.

5. The compound symmetry (CS) structure is the simplest structure that can model some form of correlation.

6. The AR(1) and ARH(1) structures do not need identical intervals between measurements.

7. Comparing two models by using the residual log-likelihood (ReslogL) requires that the fixed effects between models are the same.

8. The *F*- and *t*-tests from a repeated measures analysis are no longer valid tests because their degrees of freedom are incorrect.

9. Linear mixed models can only be used on normally distributed response variables.

Exercises

1. Consider the site-preparation experiment presented earlier that was fitted under a RCBD with a factorial model:

   a. Based on Table 3, perform a likelihood ratio test to compare the error structure of heterogeneous autoregressive of order 1 against the unstructured? Use a significance level of 5%. What do you conclude from this result?

b. Repeat the analysis of this data, but this time consider Time as a continuous variable (Timec) in its linear and quadratic form. Do you have similar conclusions to the ones obtained from Table 4? Do you need to consider the quadratic term? Use a significance level of 5%.

2. The data presented in Table 7 corresponds to a soybean experiment that was established to compare growth patterns of an experimental strain against a commercial variety (P: introduction #416937, F: Forrest). This experiment was repeated for several years but only the data from 1988 is considered here, and each plot was measured at weekly intervals eight to ten times starting at 14 d after planting. Average leaf weight from a sample of six plants was calculated. More details are presented in Davidian and Giltinan (1995).

   a. Fit a repeated measures analysis for this data with the fixed effects factors of Time, Variety, and their interaction considering a compound symmetry error structure. Consider a natural logarithm transformation of the data for your analysis. Do you have significant differences between the varieties evaluated? Is there a significant interaction? Use a significance level of 5%.

   b. Evaluate other error structures, such as DIAG and ARH(1), and use LRT, AIC, and BIC to compare your models. Which one do you recommend? Why?

   c. Based on your final selected model, can you indicate if there are significant differences between the varieties at the last measurement time? How about at the first week? Use a significance level of 5%.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19:716–723. doi:10.1109/TAC.1974.1100705

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1):1–48. doi:10.18637/jss.v067.i01

Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens, and J.S.S. White. 2009. Generalized linear mixed models: A practical guide for ecology and evolution. Trends Ecol. Evol. 24(3):127–135. doi:10.1016/j.tree.2008.10.008

Burgueño, J. 2018. Spatial analysis of field experiments. In: B. Glaz and K.M. Yeater, editors, Applied statistics in the agricultural, biological, and environmental sciences. ASA, CSSA, SSSA, Madison, WI.

Casler, M.D. 2018. Blocking principles for biological experiments. In: B. Glaz and K.M. Yeater, editors, Applied statistics in the agricultural, biological, and environmental sciences. ASA, CSSA, SSSA, Madison, WI.

Cressie, N. 1993. Statistics for spatial data: Wiley series in probability and statistics. Wiley & Sons, New York.

Davidian, M., and D.M. Giltinan. 1995. Nonliner models for repeated measurement data. Chapman & Hall/CRC Press, Boca Raton, FL.

Gbur, E.E., W.W. Stroup, K.S. McCarter, S. Durham, L.J. Young, M. Christman, M. West, and M. Kramer. 2012. Generalized linear models. In: E.E. Gbur, W.W. Stroup, K.S. McCarter, S. Durham, L.J. Young, M. Christman, M. West, and M. Kramer, editors, Analysis of generalized linear mixed models in the agricultural and natural resources sciences. ASA, CSSA, SSSA, Madison, WI.

Guerin, L., and W.W. Stroup. 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. 30 Apr.–2 May 2000, Kansas State University, Manhattan, KS.

Kenward, M.G., and J.H. Rogers. 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997. doi:10.2307/2533558

Kuehl, R. 2000. Design of experiments: Statistical principles of research design and analysis. 2nd ed. Duxbury Press, Pacific Grove, CA.

Henderson, C.R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph, ON.

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2006. SAS for mixed models. SAS Institute. Inc., Cary, NC.

McCarter, K.S. 2018. Analysis of covariance. In: B. Glaz and K.M. Yeater, editors, Applied statistics in agricultural, biological, and environmental sciences. ASA, CSSA, SSSA, Madison, WI.

Nemec, A.F.L. 1996. Analysis of repeated measures and time series: An introduction with forestry examples. Working Paper 15. Biometric information handbook No. 6. Province of British Columbia, Victoria, B.C.

Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58(3):545–554. doi:10.1093/biomet/58.3.545

Payne, R.W. 2018. Long-term research. In: B. Glaz and K.M. Yeater, editors, Applied statistics in agricultural, biological, and environmental sciences. ASA, CSSA, SSSA, Madison, WI.

Payne, R.W., D.A. Murray, S.A. Harding, D.B. Baird, and D.M. Soutar. 2011. An introduction to GenStat for Windows. 14th edition. VSN International, Hemel Hempstead, UK.

Piepho, H.P., A. Büchse, and C. Richter. 2004. A mixed modelling approach for randomized experiments with repeated measures. J. Agron. Crop Sci. 190(4):230–247. doi:10.1111/j.1439-037X.2004.00097.x

Pinheiro, J., and D. Bates. 2006. Mixed-effects models in S and S-PLUS. Springer Science & Business Media, Berlin, Germany.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, S. Heisterkamp, and B. Van Willigen. 2016. nlme: Linear and nonlinear mixed effects Models. R package version 3.1-126, http://CRAN.R-project.org/package=nlme (verified 13 Dec. 2017).

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org (verified 13 Dec. 2017).

SAS Institute Inc. 2011. Base SAS 9.3 Procedures Guide. SAS Institute, Inc. Cary, NC.

Schwarz, G.E. 1978. Estimating the dimension of a model. Ann. Stat. 6(2):461–464. doi:10.1214/aos/1176344136

Smith, A.B., A. Ganesalingam, H. Kuchel, and B.R. Cullis. 2015. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. Theor. Appl. Genet. 128:55–72. doi:10.1007/s00122-014-2412-x

Stroup, W. 2018. Non-Gaussian data. In: B. Glaz and K.M. Yeater, Applied statistics in agricultural, biological, and environmental sciences. ASA, CSSA, SSSA, Madison, WI.

Wolfinger, R.D. 1996. Heterogeneous variance-covariance structures for repeated measures. J. Agric. Biol. Environ. Stat. 1:205–230. doi:10.2307/1400366

Welham, S.J., S.A. Gezan, S.J. Clark, and A. Mead. 2014. Statistical methods in biology: Design and analysis of experiments and regression. Chapman and Hall, CRC Press, Boca Raton, FL.

Webster, R., and M.A. Oliver. 2007. Geostatistics for environmental scientists. John Wiley & Sons, Hoboken, NJ. doi:10.1002/9780470517277

Yeater, K.M., and M.B. Villamil. 2018. Multivariate methods for agricultural research. In: B. Glaz and K.M. Yeater, editors, Applied statistics in the agricultural, biological, and environmental sciences. ASA, CSSA, SSSA, Madison, WI.