Pure Error REML for Analyzing Data from Split-Plot and Multi-Stratum Designs

Steven G. Gilmour
King's College London, UK

Peter Goos
KU Leuven, Belgium
Universiteit Antwerpen, Belgium

Heiko Großmann
Otto-von-Guericke-Universität Magdeburg, Germany

July 10, 2017

Abstract

Since the dawn of response surface methodology, it has been recommended that designs include replicate points, so that pure error estimates of variance can be obtained and used to provide unbiased estimated standard errors of the effects of factors. In designs with more than one stratum, such as split-plot and split-split-plot designs, it is less obvious how pure error estimates of the variance components should be obtained and none are given by the popular residual maximum likelihood (REML) method of estimation. We propose a novel method of pure error REML estimation of the variance components, using the full treatment model to obtain the linear combinations of the responses whose likelihood is maximized in the REML estimation approach. This is easily implemented using standard software and improved estimated standard errors of the fixed effects estimates can be obtained by applying the Kenward-Roger correction based on the pure error REML estimates. We illustrate the new method using several data sets and compare the performance of pure error REML with the standard REML method. The results are comparable when the assumed model is exactly correct, but the new method is considerably more robust in the case of model misspecification.

Keywords: Full treatment model; Kenward-Roger correction; linear mixed model; replicates; response surface; split-split-plot design.

1 Introduction

It is increasingly recognized that many industrial and laboratory-based experiments are, or should be, run using split-plot and other multi-stratum structures, especially when some factors have levels which are harder to set than other factors. In the simplest case, we define some easy-to-set factors whose levels can be reset for each experimental unit (often called a subplot or run), and some hard-to-set factors whose levels can only be reset for groups of experimental units (often called whole plots or blocks). A design is then chosen so that the hard-to-set factors' level combinations are randomized to whole plots, while the easy-to-set factors' level combinations are randomized to runs within whole plots. The same principle is easily extended to several levels of experimental units or strata in so-called multi-stratum designs. Provided they are continuous, the response data from such experiments are appropriately analyzed using linear mixed models, which include random effects for each stratum in the design, such as whole plots and subplots, and fixed effects of the treatments on the response, often through a polynomial response surface model. The most common estimation procedure uses residual maximum likelihood (REML), which maximizes the likelihood of a projection of the responses onto a subspace orthogonal to the assumed response surface model, to estimate the variance components, and empirical generalized least squares (GLS), with the REML variance component estimates plugged in, to estimate the fixed parameters. This REML/GLS procedure is available in several statistical computing packages, gives the same results as analysis of variance in orthogonal multi-stratum designs and has good asymptotic properties. This analysis was recommended by Letsinger et al. (1996), the first paper to systematically consider industrial split-plot experiments.

However, it can often be observed in split-plot experiments that the whole-plot or block variance component is estimated to be zero, even when the true variance component being zero is not believable in practice. Goos et al. (2006) showed that this is quite likely to happen even though the true value of the variance component is far from zero. Gilmour

and Goos (2009) discussed this problem and suggested a Bayesian analysis, which uses prior information on the variance components, as a reasonable alternative. This seems to work quite well, but requires specialist software and very careful specification of prior information, which must be substantial to make up for the lack of information on the whole-plot or block variance component in the data.

Rather than a Bayesian analysis, which requires specifying a prior distribution, it might be desirable to perform a robust analysis, based on as few assumptions as possible. In particular, an analysis which allows for possible lack of fit in the assumed response surface model is desirable. Vining et al. (2005) and Vining and Kowalski (2008) recommended a simple analysis based on pure error estimation of each variance component obtained from replicate points, which possesses this robustness property. Vining and Kowalski (2008) recommended that all inference be done using these pure error estimates. However, the method they recommend is only applicable to particular types of designs and only uses replicate points within whole plots and completely replicated whole plots to obtain pure error estimates. Gilmour and Trinca (2000) showed, in the context of blocked response surface designs, that this is a stronger definition of pure error than is used in completely randomized designs, which requires only the use of the full treatment model. Since split-plot designs are also incomplete block designs, with some main effects completely confounded with blocks, the arguments of Gilmour and Trinca (2000) apply to split-plot designs and more general multi-stratum designs as well. Hence, the methods presented in this paper are also applicable to blocked response surface designs, split-plot designs and multi-stratum designs.

The purpose of this paper is to explore in generality the use of pure error estimates of variance components. We do this by using REML with the full treatment model (instead of the polynomial response surface model), which maximizes the likelihood of a projection of the responses onto a subspace orthogonal to the full treatment model, to obtain pure error estimates of the variance components, which do not depend on the assumed quadratic

polynomial (or any other) response surface model form. This allows many designs which do not have explicit replication within whole plots, or repeated whole plots, to be analyzed using pure error estimates. Moreover, even when such replication is available, pure error REML makes more use of the information in the data than the procedure of Vining and Kowalski (2008). We use the variance component estimates from maximizing this pure error REML in empirical GLS to obtain a robust analysis which can be recommended in analyzing data from any multi-stratum design.

The models discussed and the notation used in this paper are clarified in Section 2. A modified REML/GLS method is introduced in Section 3 and applied to some data sets in Section 4. The properties of the new method are studied in Section 5, both when the assumed polynomial response surface model is correct and when it is not. We finish with a discussion of some practical points in Section 6.

2 Models and Notation

In any experimental design, blocking factors arise as restrictions to the randomization, so that some sets of treatments must appear in runs which are in the same block. Unless every block consists of the same set of treatments, some information for comparing treatments is confounded with block effects, although under the assumed model all parameters are estimated orthogonally to block effects in some special designs. If there are nested block factors, e.g. runs within blocks within superblocks, information can be confounded with both the block effects and the superblock effects. By assuming that the block effects are random, the information available in the block totals can be recovered and combined with the usual within-block information. Each level of blocking, e.g. blocks, superblocks, etc., leads to a *stratum* in the analysis. Split-plot designs are block designs with at least one main effect completely confounded with blocks (usually called whole-plots) and split-split-plot designs are nested block designs with at least one main effect confounded

with superblocks (usually called whole-plots) and at least one main effect completely confounded with blocks within superblocks (usually called subplots). More generally, multi-stratum designs are those which have treatments defined by combinations of the levels of several treatment factors with these factors applied in different strata. Therefore, some treatment factors have main effects confounded with the effects of some blocking factors.

The model derived from the randomization, following Hinkelmann and Kempthorne (1994), is

$$\mathbf{Y} = \mathbf{X}_t\boldsymbol{\tau} + \sum_{j=1}^{s} \mathbf{Z}_j\boldsymbol{\delta}_j + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{Y}$ is a random variable of which the response vector $\mathbf{y}$ is assumed to be a realization, $\mathbf{X}_t$ is the $n \times t$ full treatment design matrix, having (i,r)th element equal to 1 if treatment (or design support point) r appears in run i and 0 otherwise, n is the number of runs, t is the number of treatments, $\boldsymbol{\tau}$ is the corresponding vector of treatment means, s is the number of nested blocking factors, implying $s+1$ strata, $\boldsymbol{\delta}_j$ is a vector of random block effects corresponding to the jth stratum, $\mathbf{Z}_j$ is the design matrix for these random effects and $\boldsymbol{\epsilon}$ is the vector of random experimental unit errors. We further assume that $\boldsymbol{\delta}_j \sim N\left(\mathbf{0}, \sigma_j^2\mathbf{I}_{n_j}\right)$, m_j is the number of units in stratum j, $\boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ and $\boldsymbol{\delta}_j$, $j = 1, \ldots, s$, and $\boldsymbol{\epsilon}$ are mutually independent. We refer to model (1) as the full treatment model. The key feature of the full treatment model is that every factor level combination will be viewed as one level, called a *treatment*, of a single categorical factor.

In a typical response surface experiment, we want to further interpret the treatment effects, for example by assuming that

$$\mathbf{X}_t\boldsymbol{\tau} = \mathbf{X}\boldsymbol{\beta}, \tag{2}$$

where $\mathbf{X}$ is the $n \times p$ design matrix for a second order polynomial response surface model, $p = 1+2q+q(q-1)/2$ is the number of parameters in the response surface model, q is the number of treatment factors and $\boldsymbol{\beta}$ is the vector of parameters of this model. Obviously,

adopting the polynomial model is a much stronger assumption than assuming model (1), which allows any pattern of treatment effects. Therefore, we would like our analysis to depend on the assumed polynomial model as little as possible. This is consistent with the presentation of Box and Draper (2007), who emphasize that the popular polynomial models are purely empirical graduating functions. Throughout this paper we use the second order polynomial model, but the ideas apply equally to any other linear response surface model.

3 Estimation

In response surface studies, the main interest is usually in estimating the fixed effects $\boldsymbol{\beta}$ in the polynomial model (2). If the ratios of variance components were known, this would be done optimally using generalized least squares (GLS) to get

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y},$$

where

$$\boldsymbol{\Sigma} = \sigma^2 \left(\sum_{j=1}^{s} \gamma_j \mathbf{Z}_j \mathbf{Z}_j' + \mathbf{I} \right) \tag{3}$$

and $\gamma_j = \sigma_j^2/\sigma^2$. The variance-covariance matrix of the GLS estimator is

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Psi} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}.$$

In practice, of course, the ratios of the variance components are not known, but have to be estimated. The particular method used to estimate the variance components then has an impact on the estimates obtained for the fixed effects.

3.1 Estimating Variance Components

Following the ideas of Letsinger et al. (1996), it has become accepted that the variance components in multi-stratum response surface designs should be estimated using

REML and the fixed effects estimated by empirical GLS. We refer to this approach as the REML/GLS approach. The main idea behind the pure error REML approach introduced here is to avoid the excessive reliance on the polynomial response surface model in (2) made by the usual REML/GLS analysis, because it may be incorrectly specified. We do this by estimating the variance components by REML from the full treatment model (1) and using the resulting variance component estimates in subsequent analyses by GLS based on the polynomial model (2). One advantage of this approach is that it allows the pure error estimation of variance components, based on model (1), as recommended by Vining and Kowalski (2008). We plug the REML variance component estimates from model (1) into the GLS estimator of model (2) and to the corresponding estimated standard errors.

The argument to do so is analogous to that in completely randomized designs as to whether estimated standard errors should be based on the pure error estimate of the error variance or on the estimate for the error variance obtained from the polynomial regression model. As recommended by Box et al. (2005), it is clearly advisable to use the pure error estimates when they are available. The common practice of checking for lack of fit and using the variance component estimates from the regression or response surface model if no lack of fit is found can be thought of as an approximation to this. In that approach, the variance component estimates from the polynomial regression model are only used if they are similar to the pure error estimates, so that one might as well use the pure error estimates. The reason not to is to increase the degrees of freedom for estimating error and so obtain apparently more precise estimates and more powerful tests. This increase in precision is spurious due to the model selection involved.

In any case, in a completely randomized response surface experiment, few would recommend ignoring the separation of pure error from lack of fit (Box and Draper, 2007; Myers et al., 2009). However, this is precisely what is commonly done in multi-stratum response surface designs when the empirical GLS estimates are obtained using the variance component estimates from fitting the polynomial regression model. This is even worse than in

a completely randomized design, since, in multi-stratum designs, not only the estimated standard errors depend on the variance component estimates, but also the estimates of the fixed effects.

The REML method we recommend for estimating variance components from model (1) involves maximizing the likelihood of the residuals, after removing the full treatment model's fixed effects. Specifically, the likelihood of $\mathbf{K}'\mathbf{Y}$ is maximized, where $\mathbf{K}'\mathbf{X}_t = \mathbf{0}$ and $rank(\mathbf{K}) = n - t$, where t is the number of treatments - see, for example, McCulloch et al. (2008). Unlike maximum likelihood estimators, REML estimators of the variance components are unbiased. Moreover, they are identical to the analysis of variance estimators in the case of orthogonal multi-stratum designs.

Following Letsinger et al. (1996), REML is currently usually used with the polynomial regression model (2) to obtain estimators of the variance components, i.e. with $\mathbf{X}_t$ replaced by $\mathbf{X}$ and $\mathbf{K}$ having rank $n - p$ instead of $n - t$. In that case, the REML estimates are based on more degrees of freedom, but they depend on the assumed polynomial regression model, which is not necessarily a good approximation of the correct treatment model.

3.2 Estimating Fixed Effects

Estimating the treatment factors' effects is usually done using the empirical GLS estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Y}, \tag{4}$$

where

$$\hat{\boldsymbol{\Sigma}} = \hat{\sigma}^2 \left(\sum_{j=1}^{s} \hat{\gamma}_j \mathbf{Z}_j \mathbf{Z}_j' + \mathbf{I} \right),$$

$\hat{\gamma}_j = \hat{\sigma}_j^2 / \hat{\sigma}^2$ and $\hat{\sigma}_j^2$ and $\hat{\sigma}^2$ are obtained from REML applied to the polynomial response surface model instead of the full treatment model. The variance matrix of these estimators is usually estimated by

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\Psi}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}. \tag{5}$$

The focus of this paper is on the plug-in estimators of the variance components which are used in the empirical GLS estimator (4). The most crucial property these must have is that the estimated variance components are as close to the true variance components as possible. There is no theoretical reason why the plug-in estimators should be obtained from the polynomial response surface model (2), even though this is the model whose fixed effects we are estimating. We suggest that the unbiased estimators of the variance components obtained from the full treatment model (1) should be used instead. We will refer to this as the pure error REML/GLS, or PE-REML/GLS, method; the standard REML method, based on the polynomial response surface treatment model, will be referred to as RS-REML.

The problem with using the usual estimators of the variance components in (4) is that they are used both to obtain the fixed effects estimates and to assess the quality of these estimates using (5). Consider, for example, a situation in which the polynomial model is inadequate because some higher order terms are missing. Then, the estimated variance components from this model are likely to be overestimated, perhaps considerably so. This leads to overestimated standard errors of the fixed effect estimates, which, in turn, might lead to few effects seeming to be significantly different from zero. We might then make the decision to reduce the order of the model, rather than to increase it. If, on the other hand, the variance components are estimated from the full model, they will be unbiased irrespective of whether the assumed polynomial model is correct or not. The resulting analysis should then correctly show the inadequacy of the assumed polynomial response surface model (Goos and Gilmour, 2017). Of course, since the unbiased estimators are based on less information than the biased estimators, things are not quite so simple. However, this line of reasoning does at least indicate that, in the case of major model inadequacies, the unbiased estimators should be better.

3.3 Estimating Standard Errors of Fixed Effects

The estimated variances and standard errors of the fixed effects' estimates obtained using (5) are known to be negatively biased even if the fixed effects model is correct. A correction, which usually gives less biased estimated variances and standard errors, was suggested by Kenward and Roger (1997) for standard REML estimation. By a direct application of results given by McCulloch et al. (2008) (p. 165-169), the Kenward-Roger correction is trivially adapted to the PE-REML method, i.e. it can be applied regardless of which estimator of variance components is used for estimating the fixed effects.

Simple orthogonal block structures are those made up of crossed and nested blocking factors, in which, for each blocking factor individually, each block contains equal numbers of units (Nelder (1965); see also Gilmour and Trinca (2006)), irrespective of the treatment structure or model. In simple orthogonal block structures, with model (2) for the fixed effects, the approximate variance matrix for fixed effects estimates, with the Kenward-Roger correction, is

$$\widehat{\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})} = \hat{\boldsymbol{\Psi}}_A = \hat{\boldsymbol{\Psi}} + 2\hat{\boldsymbol{\Lambda}},$$

where $\hat{\boldsymbol{\Psi}}$ is from equation (5), $\hat{\boldsymbol{\Lambda}}$ is obtained by plugging the appropriate REML estimates $\hat{\boldsymbol{\sigma}}$ of the variance components into

$$\boldsymbol{\Lambda} = \boldsymbol{\Psi}\left\{\sum_{i=1}^{s+1}\sum_{j=1}^{s+1} u_{ij}\left(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Psi}\mathbf{P}_j\right)\right\}\boldsymbol{\Psi},$$

u_{ij} is the (i,j)th element of the covariance matrix $\mathbf{U} = \mathbf{V}(\hat{\boldsymbol{\sigma}})$, $\hat{\boldsymbol{\sigma}}$ is either the RS-REML or PE-REML estimator of $\boldsymbol{\sigma}$, $\boldsymbol{\sigma} = [\sigma_1^2 \ \cdots \ \sigma_{s+1}^2]'$, $\sigma_{s+1}^2 = \sigma^2$,

$$\mathbf{P}_i = \mathbf{X}_g' \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\sigma}_i}\mathbf{X}_g,$$

$$\mathbf{Q}_{ij} = \mathbf{X}_g' \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\sigma}_i}\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\sigma}_j}\mathbf{X}_g,$$

$\mathbf{X}_g$ is the design matrix corresponding to some fixed effects model (either $\mathbf{X}_t$ or $\mathbf{X}$ in our case) and $\boldsymbol{\Sigma}$ is defined in (3). The correction affects only those standard errors which are influenced by the plug-in estimators of $\sigma_1^2, \ldots, \sigma_s^2$ and not those which depend only on σ^2.

This is still a theoretical approximation, since it depends on the terms σ_i which have to be estimated. In the particular case of a split-plot design with k subplots within each whole plot, $\boldsymbol{\sigma}' = [\sigma_1^2 \ \sigma^2]$,

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I} + \sigma_1^2 \mathbf{Z}\mathbf{Z}'$$

and so

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}\left(\mathbf{I} - \frac{\sigma_1^2}{\sigma^2 + k\sigma_1^2}\mathbf{Z}\mathbf{Z}'\right), \tag{6}$$

using a standard result on matrix inverses, e.g. Searle (1982), p.261. By direct differentiation, we obtain

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_1^2} = -\frac{1}{\left(\sigma^2 + k\sigma_1^2\right)^2}\mathbf{Z}\mathbf{Z}'$$

and

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma^2} = \frac{1}{\sigma^4}\left\{\frac{\sigma_1^2\left(2\sigma^2 + k\sigma_1^2\right)}{\left(\sigma^2 + k\sigma_1^2\right)^2}\mathbf{Z}\mathbf{Z}' - \mathbf{I}\right\}.$$

For the PE-REML estimators, we use the asymptotic sampling variance of REML estimators of variance components, given by McCulloch et al. (2008) for example. As a result,

$$u_{22} = V\left(\hat{\sigma}^2\right) = 2tr(\mathbf{Z}'\mathbf{C}\mathbf{Z}\mathbf{Z}'\mathbf{C}\mathbf{Z})/c,$$

$$u_{11} = V\left(\hat{\sigma}_1^2\right) = 2tr(\mathbf{C}\mathbf{C})/c$$

and

$$u_{12} = u_{21} = Cov\left(\hat{\sigma}^2, \hat{\sigma}_1^2\right) = -2tr(\mathbf{Z}'\mathbf{C}\mathbf{C}\mathbf{Z})/c,$$

where

$$c = tr(\mathbf{C}\mathbf{C})tr(\mathbf{Z}'\mathbf{C}\mathbf{Z}\mathbf{Z}'\mathbf{C}\mathbf{Z}) - \left\{tr(\mathbf{Z}'\mathbf{C}\mathbf{C}\mathbf{Z})\right\}^2$$

and

$$\mathbf{C} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}_t\left(\mathbf{X}_t'\boldsymbol{\Sigma}^{-1}\mathbf{X}_t\right)^{-1}\mathbf{X}_t'\boldsymbol{\Sigma}^{-1}.$$

The corresponding results for the RS-REML method replace $\mathbf{X}_t$ by $\mathbf{X}$ and are used in software packages including SAS (Littell et al., 1996).

One final point to note is that, if the estimated variance component for the blocks, $\hat{\sigma}_1^2$, is zero (or negative if this is allowed), then the Kenward-Roger approximations fail. In this case, we simply let $\hat{\mathbf{\Lambda}} = \mathbf{0}$, in which case the adjusted estimate is equal to the unadjusted estimate. This is also what statistical software packages do. All of the above generalizes to any nested multi-stratum design, including any nested block structure. The results for a split-split-plot structure are given in the appendix.

4 Illustrations

We illustrate the PE-REML method and compare it with the standard response surface RS-REML method using three data sets. These were chosen to show how the new method works in different kinds of designs. The first thing to note is that many multi-stratum designs, especially D-optimal designs, do not allow pure error estimation of variance components, so that PE-REML estimation is infeasible and standard RS-REML estimation is the only possibility to obtain variance component estimates. The designs in our examples include an equivalent estimation split-plot design which has ample degrees of freedom for pure error, a split-plot design constructed by hand which has limited pure error degrees of freedom and an I-optimal split-split-plot design, which illustrates how the method extends to general multi-stratum structures.

4.1 Strength of Ceramic Pipes

The design for the experiment on ceramic pipes reported by Vining et al. (2005) has 12 whole plots, each with four runs, and three of which consist of replicated center points. We name the coded factors in the experiment X_1 (zone 1 temperature), X_2 (zone 2 temperature), X_3 (amount of binder) and X_4 (grinding speed), where X_1 and X_2 are whole-plot factors. Because the design has the equivalent-estimation property (i.e. the

Table 1: Estimates and estimated standard errors for the ceramic pipe data

Parameter	Estimate	Standard Error	
		RS-REML	PE-REML
β_1	4.5579	0.4893	0.3027
β_2	-6.5592	0.4893	0.3027
β_3	-4.9733	0.0648	0.0721
β_4	4.0922	0.0648	0.0721
β_{11}	1.7381	0.8974	0.5551
β_{22}	-0.5407	0.8974	0.5551
β_{33}	-2.3864	0.6059	0.3958
β_{44}	2.5736	0.6059	0.3958
β_{12}	0.8431	0.5993	0.3707
β_{13}	1.4356	0.0688	0.0765
β_{14}	-1.4794	0.0688	0.0765
β_{23}	-1.0019	0.0688	0.0765
β_{24}	1.9856	0.0688	0.0765
β_{34}	-1.0394	0.0688	0.0765

OLS and GLS estimators are equivalent), both the RS-REML and PE-REML methods give the same estimates for the factor effects from empirical GLS. These estimates are shown in Table 1 along with their estimated standard errors. In this case, the Kenward-Roger correction has no effect.

Using the full treatment model to estimate the variance components gives the same variance component estimates ($\hat{\sigma}_1^2 = 0.52626$ and $\hat{\sigma}^2 = 0.09355$), and, hence, the same estimated standard errors of the fixed effects as the sample variance method of Vining et al. (2005) (although their Table 6 actually gives variances, wrongly labeled as standard errors). The estimated standard errors from the polynomial regression model for the linear and quadratic main effects of the whole-plot factors X_1 and X_2 and their interaction effect are larger than those from the PE-REML method, while for the other effects they are smaller. The RS-REML estimates of the variance components are $\hat{\sigma}_1^2 = 1.4176$ and $\hat{\sigma}^2 = 0.07563$.

The fact that the PE-REML method yields the same estimates as the sample variance approach of Vining et al. (2005) shows that, whereas they contrasted their pure error standard error estimates with those obtained from REML, the point is not the method (REML or sample variances) used, but the treatment model used as a starting point for estimating variance components (rather than the RS model).

We note that the standard errors for all effects estimated at least partially in the whole-plot stratum are actually smaller when the pure error estimators of the variance components (based on the full treatment model) are used than when the variance component estimates are based on the polynomial model. The opposite is true for all other effects. This is due to the fact that the pure error estimate for σ_1^2 is smaller than the RS-REML estimate, while the pure error estimate for σ^2 is larger than the corresponding RS-REML estimate.

Table 2: Design with 12 whole plots of five runs for estimating a second order response surface model in two whole-plot factors and two subplot factors, along with simulated responses.

Whole plot	Treatment	X_1	X_2	X_3	X_4	Y
1	1	-1	-1	-1	-1	29.46
1	2	-1	-1	1	-1	31.50
1	3	-1	-1	-1	1	23.41
1	4	-1	-1	1	1	19.12
1	5	-1	-1	0	0	24.38
2	6	1	-1	-1	-1	53.32
2	7	1	-1	1	-1	50.18
2	8	1	-1	-1	1	55.08
2	9	1	-1	1	1	47.97
2	10	1	-1	0	0	49.08
3	11	-1	1	-1	-1	37.10
3	12	-1	1	1	-1	41.39
3	13	-1	1	-1	1	43.22
3	14	-1	1	1	1	38.18
3	15	-1	1	0	0	38.85
4	16	1	1	-1	-1	39.10
4	17	1	1	1	-1	44.05
4	18	1	1	-1	1	58.19
4	19	1	1	1	1	51.32
4	20	1	1	0	0	47.68
5	21	-1	0	-1	0	37.74
5	22	-1	0	1	0	32.18
5	23	-1	0	0	-1	37.27
5	24	-1	0	0	1	34.25
5	25	-1	0	0	0	36.18
6	26	1	0	-1	0	49.91
6	27	1	0	1	0	50.84
6	28	1	0	0	-1	49.24
6	29	1	0	0	1	54.78
6	30	1	0	0	0	50.45
7	31	0	-1	-1	0	40.63
7	32	0	-1	1	0	46.87
7	33	0	-1	0	-1	47.88
7	34	0	-1	0	1	42.95
7	35	0	-1	0	0	47.16
8	36	0	1	-1	0	48.59
8	37	0	1	1	0	49.21
8	38	0	1	0	-1	48.14
8	39	0	1	0	1	53.42
8	40	0	1	0	0	49.59
9	41	0	0	-1	-1	48.61
9	42	0	0	1	-1	51.91
9	43	0	0	-1	1	55.17
9	44	0	0	1	1	50.13
9	45	0	0	0	0	49.47
10	46	0	0	-1	0	49.08
10	47	0	0	1	0	48.77
10	48	0	0	0	-1	51.00
10	49	0	0	0	1	49.52
10	45	0	0	0	0	49.72
11	41	0	0	-1	-1	41.26
11	42	0	0	1	-1	43.83
11	43	0	0	-1	1	57.94
11	44	0	0	1	1	42.02
11	45	0	0	0	0	40.65
12	46	0	0	-1	0	49.43
12	47	0	0	1	0	46.00
12	48	0	0	0	-1	54.96
12	49	0	0	0	1	55.10
12	45	0	0	0	0	44.45

Table 3: Estimates and standard errors for the simulated data from the design in Table 2

Parameter	Estimate		Standard error			
	RS-REML	PE-REML	RS-REML	PE-REML	RS-REML-KR	PE-REML-KR
β_1	8.2320	8.2320	0.8551	1.1169	0.8551	1.1169
β_2	2.6347	2.6347	0.8551	1.1169	0.8551	1.1169
β_3	-0.8825	-0.8825	0.4215	0.5414	0.4215	0.5414
β_4	0.8769	0.8769	0.4215	0.5414	0.4215	0.5414
β_{11}	-6.1579	-6.1591	1.2865	1.6801	1.2867	1.6810
β_{22}	-1.9979	-1.9991	1.2865	1.6801	1.2867	1.6810
β_{33}	-0.3846	-0.3787	0.7137	0.9174	0.7245	0.9578
β_{44}	2.0538	2.0596	0.7137	0.9174	0.7245	0.9578
β_{12}	-4.3080	-4.3080	1.0473	1.3679	1.0473	1.3679
β_{13}	-0.1340	-0.1340	0.5655	0.7264	0.5655	0.7264
β_{14}	2.4995	2.4995	0.5655	0.7264	0.5655	0.7264
β_{23}	0.2105	0.2105	0.5655	0.7264	0.5655	0.7264
β_{24}	2.9180	2.9180	0.5655	0.7264	0.5655	0.7264
β_{34}	-2.4283	-2.4283	0.5162	0.6631	0.5162	0.6631

4.2 Another split-plot example

To illustrate some different points, we simulated data for another design involving four factors, two of which are applied in the whole-plot stratum, and 12 whole plots, each containing five runs. The design was constructed by hand, using standard treatment sets and the ideas of fractional partial confounding (Mead, 1990; Mead et al., 2012) to distribute the 49 treatments between whole plots - see Table 2. Unlike the design for the ceramic pipe experiment, this design has replicated treatments which appear only in different whole plots. All the replicated treatments, labelled 41-49, appear in whole plots 9-12. This example shows that it is possible to get considerable information on pure error variance components using the PE-REML method even from a design which does not have within-whole-plot replicates. This is in contrast to the sample variance method of Vining and Kowalski (2008) which would not allow pure error to be estimated from this design.

When simulating the responses, the true values of the variance components were $\sigma_1^2 = 4$ and $\sigma^2 = 2$. Using PE-REML, the estimate of σ_1^2 is 5.3738 and the estimate of σ^2 is 10.552, while the RS-REML estimates are 3.1085 and 6.3957 respectively. In this case, σ^2 is not very well estimated by either method, even though there are adequate degrees of freedom for its estimation. The true values of the fixed effects were $\boldsymbol{\beta}' = [50\ 8\ 3\ 0\ 0\ -7\ -3\ 0\ 1\ -4\ 0\ 2\ 0\ 3\ -2]$. The fixed effects' estimates and their estimated standard errors, with and without the Kenward-Roger correction, are shown in Table 3. The partially orthogonal structure of the design implies that the fixed effects estimates are the same for each method, except for the quadratic effects of the subplot factors which differ slightly. As a result, the Kenward-Roger correction only affects the standard errors of the estimates of the subplot factors' quadratic effects. In this case, we find that the methods give very similar estimates, but slightly different standard errors. The standard errors obtained using the pure error estimates of the variance components are larger than those based on the polynomial regression model. This is because the pure error estimates of the variance components are larger than their RS-REML counterparts.

4.3 A split-split-plot example

In the same way as for split-plot designs, it is possible to obtain pure error estimates from designs other than split-plot designs, such as split-split-plot designs. To illustrate this, we created artificial data for an I-optimal split-split-plot design produced by the software package JMP. The true fixed effects parameters were the same as in Example 2 and the variance components were $\sigma_1^2 = 4$, $\sigma_2^2 = 2$ and $\sigma^2 = 1$. The split-split-plot design involves six whole plots, twelve subplots and 36 runs. Every whole plot has two subplots of three runs. The I-optimal design, which minimizes the average variance of prediction, includes 30 distinct design points and hence 30 different treatments. The design and the treatment labels are shown in Table 4, along with the simulated data.

The fixed effects' estimates obtained and their estimated standard errors are shown in

Table 4: Design with six whole plots, each containing two suplots of three runs for estimating a second order response surface model in one whole-plot factor, one subplot factor, and two subsubplot factors, along with simulated responses.

Whole plot	Subplot	Treatment	X_1	X_2	X_3	X_4	Y
1	1	1	0	0	-1	1	50.77
1	1	2	0	0	0	0	49.08
1	1	2	0	0	0	0	50.21
1	2	2	0	0	0	0	47.28
1	2	3	0	0	1	-1	48.64
1	2	2	0	0	0	0	49.18
2	3	4	1	-1	1	1	48.68
2	3	5	1	-1	-1	-1	51.67
2	3	6	1	-1	0	0	50.76
2	4	7	1	0	-1	0	52.02
2	4	8	1	0	1	-1	52.36
2	4	9	1	0	0	1	54.37
3	5	10	-1	-1	0	1	22.13
3	5	11	-1	-1	1	-1	37.65
3	5	12	-1	-1	-1	0	28.17
3	6	13	-1	0	-1	-1	37.88
3	6	14	-1	0	1	1	35.33
3	6	15	-1	0	0	0	37.38
4	7	16	0	-1	0	-1	48.69
4	7	17	0	-1	-1	1	44.92
4	7	18	0	-1	1	0	44.14
4	8	19	0	1	-1	-1	46.72
4	8	20	0	1	1	0	50.06
4	8	21	0	1	0	1	54.69
5	9	22	1	0	1	0	49.46
5	9	23	1	0	0	-1	47.67
5	9	24	1	0	-1	1	55.59
5	10	25	1	1	0	-1	43.34
5	10	26	1	1	-1	0	48.88
5	10	27	1	1	1	1	51.32
6	11	28	-1	1	0	0	43.68
6	11	29	-1	1	-1	1	44.96
6	11	30	-1	1	1	-1	44.26
6	12	14	-1	0	1	1	38.21
6	12	13	-1	0	-1	-1	38.72
6	12	15	-1	0	0	0	39.02

Table 5: Estimates and standard errors for simulated data from the design in Table 4

Parameter	Estimate		Standard error			
	RS-REML	PE-REML	RS-REML	PE-REML	RS-REML-KR	PE-REML-KR
β_1	6.6134	6.6134	0.5340	0.5410	0.5340	0.5410
β_2	2.8402	2.8427	0.3856	0.4256	0.5499	0.6250
β_3	0.0218	0.0387	0.2310	0.2014	0.2391	0.2051
β_4	0.1216	0.1046	0.2310	0.2014	0.2391	0.2051
β_{11}	-4.5637	-4.5452	0.9322	0.9430	0.9362	0.9454
β_{22}	-1.9252	-1.8964	0.5460	0.6025	0.7756	0.8812
β_{33}	0.1064	0.0969	0.3995	0.3474	0.4023	0.3495
β_{44}	0.5142	0.5048	0.3932	0.3419	0.3959	0.3440
β_{12}	-3.8645	-3.9355	0.5125	0.5599	0.8285	0.9257
β_{13}	-0.8496	-0.8420	0.2742	0.2386	0.2769	0.2404
β_{14}	2.1437	2.1439	0.2759	0.2397	0.2760	0.2398
β_{23}	-0.0526	-0.0526	0.3107	0.2700	0.3107	0.2700
β_{24}	3.2443	3.2443	0.3107	0.2700	0.3107	0.2700
β_{34}	-1.3678	-1.4290	0.3152	0.2944	0.4401	0.3776

Table 5. The estimates of the whole-plot, subplot and subsubplot variance components, σ_1^2, σ_2^2 and σ^2 were, respectively, 0.799, 0.296 and 1.159 from the RS-REML analysis and 0.743, 0.565 and 0.874 from the PE-REML analysis. This explains the differences we can see in the estimation of the treatment factor effects. It is not clear which analysis is better in this case.

5 Comparison of Methods

The examples in the previous section illustrated several interesting points about the two different REML methods: in Example 1, PE-REML gave considerably smaller standard errors of some fixed effects; in Example 2, the methods gave quite different estimates of the variance components; Example 3 showed quite different estimates of fixed effects being obtained from the two methods. However, no general conclusions can be drawn from these examples about which method is to be preferred. To do this, we conducted a simulation study using the design from the artificial split-plot example in Section 4.2. Initially, we assume that the second order polynomial model is exactly correct. This, of course, is unrealistic, but allows us to compare the new PE-REML method with the usual RS-REML method in the situation that is most favorable for the latter.

5.1 Assumed Model is Correct

We simulated 10,000 data sets from the normal distribution, assuming a second order response surface model with $\boldsymbol{\beta}' = [50\ 8\ 3\ 0\ 0\ -7\ -3\ 0\ 1\ -4\ 0\ 2\ 0\ 3\ -2]$, $\sigma_1^2 = 4$ and $\sigma^2 = 2$ and analyzed all of them using the RS-REML and PE-REML methods. The mean estimated values of σ_1^2 and σ^2 were 4.0215 and 2.0021 respectively from the RS-REML method and 4.1180 and 1.9993 from the PE-REML method. The design has the property that all subplot factor effects, except the quadratic effects of the subplot factors, are

Table 6: Empirical standard errors of quadratic parameter estimates estimated from simulations when $\sigma_1^2 = 4$ and $\sigma^2 = 2$

	PE-REML	RS-REML
β_{11}	1.2939	1.2937
β_{22}	1.2830	1.2832
β_{33}	0.4074	0.4073
β_{44}	0.4117	0.4117

estimated orthogonally to blocks. For this reason, it is only the standard errors of these quadratic effects which differ. The empirical standard errors for these effects, calculated from the sample variances of the parameter estimates in the simulations, are shown in Table 6. In this case, where all necessary assumptions are known to be true, using the variance component estimates from the polynomial model should result in the most precise parameter estimates for the quadratic effects. The difference in precision with the PE-REML method is, however, almost nonexistent. The biases of the fixed effects estimators were also estimated from the simulations; they are similar for both methods and never more than 2.31% of the corresponding standard deviation. For this setup, it is clear that either method is acceptable for estimating the fixed effects parameters.

It is generally more difficult to get good estimators of the standard errors of the fixed effects and these are usually biased. The biases estimated from the simulations, expressed as percentages of the corresponding empirical standard errors, are shown in Table 7. Because $(\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}$ underestimates the true variance, the biases are generally negative. They are small but non-negligible and the biases from the PE-REML method are larger than those from the RS-REML method. This is not surprising, as there are fewer residual degrees of freedom in each stratum when using the full treatment model to obtain variance component estimates, and the estimated standard errors are asymptotically unbiased as these degrees of freedom go to infinity. The Kenward-Roger correction was applied and,

Table 7: Relative biases (%) of uncorrected estimated standard errors when $\sigma_1^2 = 4$ and $\sigma^2 = 2$

	PE-REML	RS-REML
β_1	-7.98	-3.84
β_2	-7.83	-3.68
β_3	-2.38	0.19
β_4	-2.78	-0.22
β_{11}	-8.53	-4.42
β_{22}	-7.75	-3.64
β_{33}	-8.49	-1.42
β_{44}	-9.73	-2.47
β_{12}	-7.32	-3.15
β_{13}	-3.15	-0.60
β_{14}	-3.77	-1.24
β_{23}	-4.01	-1.48
β_{24}	-3.97	-1.44
β_{34}	-3.89	-1.35

Table 8: Relative biases (%) of estimated standard errors for quadratic effects, corrected using the Kenward-Roger method, when $\sigma_1^2 = 4$ and $\sigma^2 = 2$

	PE-REML	RS-REML
β_{11}	-8.51	-4.42
β_{22}	-7.74	-3.64
β_{33}	-6.37	-0.70
β_{44}	-7.64	-1.75

Table 9: Relative biases (%) from estimated standard errors, corrected using the Kenward-Roger method, when the true model has a third order term with a large effect estimated in the whole-plot stratum.

	PE-REML	RS-REML
β_1	-8.99	46.35
β_2	-9.42	45.66
β_3	-3.75	-1.11
β_4	-2.55	0.13
β_{11}	-8.94	46.26
β_{22}	-9.23	45.91
β_{33}	-6.60	0.38
β_{44}	-8.10	-1.12
β_{12}	-8.70	46.81
β_{13}	-4.28	-1.65
β_{14}	-3.17	-0.51
β_{23}	-3.34	-0.68
β_{24}	-1.86	0.83
β_{34}	-3.32	-0.66

for the quadratic effects where it makes some difference, the results are shown in Table 8. For both methods, the correction works well for the quadratic effects of the subplot factors (β_{33} and β_{44}), but less well for the quadratic effects of the whole-plot factors. For the method based on the polynomial regression model, the bias in standard errors is small, and the Kenward-Roger correction reduces that small bias even further.

5.2 Assumed Model is Incorrect

In Section 5.1, the results from using the polynomial treatment model to estimate the random effects showed smaller biases than the results from using the full treatment model. This is not surprising since the data were simulated from this polynomial model. However, if the polynomial model is wrong, the results can change drastically. In Table 9, the relative biases are shown from simulations in which the quadratic by linear interaction effect β_{112} was given the value 5, i.e. similar in size to the other active effects. Also, the model misspecification involves whole-plot factors only, so that we should expect that the whole-plot aspects of the analysis are affected, rather than the subplot parts of the analysis.

It can be seen that the estimated standard errors obtained when using the pure error variance components are quite robust to this model misspecification (having a relative bias of no more than about 10%), while those based on the polynomial model fail completely for the linear effects β_1 and β_2, the quadratic effects β_{11} and β_{22}, and the interaction effect β_{12} (with relative biases greater than 40%). This is due to the fact that the variance component σ_1^2 is overestimated substantially by the RS-REML method. The RS-REML method produced a mean estimate of 9.6323, compared with 4.0358 from the PE-REML method. This results in inflated standard errors for the effects estimated in the whole-plot or block stratum. The estimates of σ^2 have means 1.9980 and 2.0091 from the PE- and RS-REML methods respectively.

The importance of the overestimated standard errors is not only that we might draw wrong conclusions about specific effects, but that we can easily be led to believe that there are few active effects and potentially miss factors which could be very important for process or product improvement. Note that a negative bias of 10% in the estimated standard error will lead to a 5% significance test for that parameter having a true size of 6.69%, while a positive bias of 40% leads to a true size of 1.77%.

Table 10: Relative biases (%) from estimated standard errors, corrected using the Kenward-Roger method, when the true model has a third order term with a large effect estimated in the subplot stratum.

	PE-REML	RS-REML
β_1	-9.02	-4.50
β_2	-9.33	-4.83
β_3	-3.47	87.66
β_4	-2.65	89.25
β_{11}	-8.59	-3.85
β_{22}	-8.81	-4.05
β_{33}	-5.97	76.20
β_{44}	-6.00	75.99
β_{12}	-9.12	-4.61
β_{13}	-2.67	89.22
β_{14}	-4.12	86.40
β_{23}	-3.32	87.95
β_{24}	-3.31	87.97
β_{34}	-4.73	85.21

Table 11: Relative biases (%) from estimated standard errors, corrected using the Kenward-Roger method, when third order terms have small but non-zero effects.

	PE-REML	RS-REML
β_1	-7.36	1.59
β_2	-8.97	-0.17
β_3	-2.87	19.77
β_4	-3.03	19.58
β_{11}	-8.80	0.03
β_{22}	-8.52	0.32
β_{33}	-5.79	19.83
β_{44}	-7.06	18.18
β_{12}	-8.39	0.47
β_{13}	-4.15	18.19
β_{14}	-2.57	20.14
β_{23}	-2.79	19.87
β_{24}	-3.13	19.46
β_{34}	-3.26	19.29

Simulations were also run with the effect β_{334}, which is estimated in the subplot stratum, having value 5. This is a scenario in which the model misspecification is in the subplot stratum rather than in the whole-plot stratum. On average, the RS-REML method estimated σ_1^2 and σ^2 respectively to be 2.8964 and 7.0989, whereas PE-REML again gave reasonable estimates of 4.0006 and 1.9868. The biases of the estimated standard errors, shown in Table 10, are unacceptable for all parameters estimated in the subplots stratum if the RS-REML method is used.

Some simulations were also run in which all third order terms except the pure cubic terms had small, but non-zero, effects of size 0.5 for linear by quadratic interactions and 0.25 for

linear by linear by linear interactions. The result was that σ_1^2 and σ^2 were estimated to be 4.0483 and 1.9888 respectively by PE-REML and 4.2139 and 2.8793 by RS-REML. The biases of the estimated standard errors of the fixed effects, shown in Table 11, indicate that, in general, it is clearly better to use the pure error estimates of the variance components. Using that approach, the relative biases are less than 10%, compared with up to 20% when using the RS-REML method. This time, it is mainly the estimate for σ^2 which is inflated, resulting in substantial upward biases for the estimated standard errors of effects estimated in the subplot stratum when using the RS-REML method. A positive bias of 20% for the standard errors corresponds to a size of 2.91% for a significance test at the 5% level.

6 Discussion

On the basis of our results, we strongly recommend that the variance components in multi-stratum response surface designs should be routinely estimated using PE-REML, and thus based on the full treatment model. REML is implemented in many statistical packages and generally has good properties. Implementing PE-REML simply involves running the available REML procedure with the full set of treatment indicators as the fixed effects, to obtain the estimates of the variance components. These are then plugged in to the generalized least squares formula to obtain the estimates of the fixed effects parameters. The same method could be used for analyzing data from blocked experiments, though in most cases it will make little difference, since most information on treatment effects comes from within the blocks.

The results in Section 5 show that PE-REML gives stable estimates of the fixed effects parameters, along with their standard errors, irrespective of whether or not the assumed model is correct. The standard errors are consistently negatively biased, by up to about 10% in relative terms. This is not a major concern, but it does mean that aspects of

inference, such as p-values and interval estimates at a given level of confidence are not exact. In this case, the Kenward-Roger correction helps only a little and looking for better correction methods might be a fruitful avenue for further research. For now, we recommend the PE-REML method presented here.

Appendix

The derivation of results needed to calculate the Kenward-Roger correction for PE-REML in the split-split-plot design follows the same steps as for the split-plot design, given in Section 3.3, though, of course, there is an additional variance component. In a split-split-plot design with b subplots within each whole plot and k subsubplots within each subplot, $\boldsymbol{\sigma}' = [\sigma_1^2 \ \sigma_2^2 \ \sigma^2]$ and

$$\boldsymbol{\Sigma} = \sigma^2\mathbf{I} + \sigma_1^2\mathbf{Z}_1\mathbf{Z}_1' + \sigma_2^2\mathbf{Z}_2\mathbf{Z}_2'.$$

By twice applying the formula for Schur complements, as in the derivation of (6), we obtain

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}\left\{\mathbf{I} - \frac{\sigma^2\sigma_1^2}{(\sigma^2+\sigma_2^2k)(\sigma^2+\sigma_1^2bk+\sigma_2^2k)}\mathbf{Z}_1\mathbf{Z}_1' - \frac{\sigma_2^2}{\sigma^2+\sigma_2^2k}\mathbf{Z}_2\mathbf{Z}_2'\right\}.$$

Differentiating with respect to each variance component and simplifying, we obtain

$$\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_1^2} = -\frac{1}{(\sigma^2+\sigma_1^2bk+\sigma_2^2k)^2}\mathbf{Z}_1\mathbf{Z}_1',$$

$$\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_2^2} = \frac{1}{(\sigma^2+\sigma_2^2k)^2}\left\{\frac{\sigma_1^2k(2\sigma^2+2k\sigma_2^2kb\sigma_1^2)}{(\sigma^2+\sigma_1^2bk+\sigma_2^2k)^2}\mathbf{Z}_1\mathbf{Z}_1' - \mathbf{Z}_2\mathbf{Z}_2'\right\}$$

and

$$\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma^2} = \frac{1}{\sigma^2}\left\{\frac{\sigma_1^2(2\sigma^4+2\sigma^2\sigma_2^2k+\sigma_1^2\sigma_2^2bk^2)}{(\sigma^2+\sigma_2^2k)^2(\sigma^2+\sigma_1^2bk+\sigma_2^2k)^2}\mathbf{Z}_1\mathbf{Z}_1' + \frac{\sigma_2^2(2\sigma^2+\sigma_2^2k)}{\sigma^2(\sigma^2+\sigma_2^2k)^2}\mathbf{Z}_2\mathbf{Z}_2' - \frac{1}{\sigma^2}\mathbf{I}\right\}.$$

The elements of the matrix $\mathbf{U}$ are obtained by direct numerical inversion.

Acknowledgements

This work was supported by the Royal Society International Joint Project number 2007/R2. Part of it was completed while the authors were participants in the Isaac Newton Institute for Mathematical Sciences program on the Design and Analysis of Experiments.

References

Box, G. E. P. and N. R. Draper (2007). *Response Surfaces, Mixtures and Ridge Analyses, 2nd edition*. New York: Wiley.

Box, G. E. P., W. Hunter, and J. Hunter (2005). *Statistics for Experimenters: Design, Innovation, and Discovery* (2nd ed.). New York: Wiley.

Gilmour, S. G. and P. Goos (2009). Analysis of data from nonorthogonal multi-stratum designs in industrial research. *Applied Statistics 58*, 467–484.

Gilmour, S. G. and L. A. Trinca (2000). Some practical advice on polynomial regression analysis from blocked response surface designs. *Communications in Statistics: Theory and Methods 29*, 2157–2180.

Gilmour, S. G. and L. A. Trinca (2006). Response surface experiments on processes with high variation. In A. I. Khuri (Ed.), *Response Surface Methodology and Related Topics*, pp. 19–46. Singapore: World Scientific.

Goos, P. and S. G. Gilmour (2017). Testing for lack of fit in blocked and split-plot response surface designs. *Journal of Quality Technology 49*, in press.

Goos, P., I. Langhans, and M. Vandebroek (2006). Practical inference from industrial split-plot designs. *Journal of Quality Technology 38*, 162–179.

Hinkelmann, K. and O. Kempthorne (1994). *Design and Analysis of Experiments Volume I: Introduction to Experimental Design*. New York: Wiley.

Kenward, M. G. and J. H. Roger (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics 53*, 983–997.

Letsinger, J. D., R. H. Myers, and M. Lentner (1996). Response surface methods for bi-randomization structures. *Journal of Quality Technology 28*, 381–397.

Littell, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger (1996). *SAS System for Mixed Models*. SAS Institute Inc., Cary, NC.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus (2008). *Generalized, Linear, and Mixed Models* (2nd ed.). New York: Wiley.

Mead, R. (1990). The nonorthogonal design of experiments (with discussion). *Journal of the Royal Statistical Society, Series A 153*, 151–201.

Mead, R., S. G. Gilmour, and A. Mead (2012). *Statistical Principles for the Design of Experiments*. Cambridge: Cambridge University Press.

Myers, R. H., D. C. Montgomery, and C. M. Anderson-Cook (2009). *Response Surface Methodology: Process and Product Optimization using Designed Experiments*. New York: Wiley.

Nelder, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. i. block structure and the null analysis of variance. *Proceedings of the Royal Society of London, Series A 283*, 147–162.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. New York: John Wiley.

Vining, G. G. and S. M. Kowalski (2008). Exact inference for response surface designs within a split-plot structure. *Journal of Quality Technology 40*, 394–406.

Vining, G. G., S. M. Kowalski, and D. C. Montgomery (2005). Response surface designs within a split-plot structure. *Journal of Quality Technology 37*, 115–128.