Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229-5666
Vol. 12, Issue 1 (2019): 117-129

PARAMETER ESTIMATION OF NONLINEAR SPLIT-PLOT DESIGN MODELS: A THEORETICAL FRAMEWORK

[1*]I. J. David, [2]O. E. Asiribo and [3]H. G. Dikko
[1,2,3]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria
E Mail: [1*]davidij@fuwukari.edu.ng; [2]asiribo@yahoo.com;
[3]hgdikko@yahoo.com
*Corresponding Author

Received October 28, 2018
Modified April 11, 2019
Accepted May 07, 2019

Abstract

Split-plot design models are special class of linear models with two error terms, that is the whole plot error and subplot error terms. These models can be remodeled as nonlinearly with variance components. This is a combination of nonlinear model for the mean part of the split-plot design model with additive error terms which describes the covariance configuration of the models. This research work presents estimated generalized least square method for estimating the parameters of the nonlinear split-plot design models. To achieve this, an iterative Gauss-Newton procedure with Taylor Series expansion was implemented. The unknown variance components of the model are estimated via residual maximum likelihood estimation method. The advantage of this technique is that it produces stable numerical values for the parameters mean and variances since it considers the covariance configuration of the model.

Key Words: Nonlinear Split-Plot Design Model, Gauss-Newton, Estimated Generalized Least Square, Restricted Maximum Likelihood Estimation.

1. Introduction

In straightforward terms, split-plot design (SPD) of experiments can be defined as blocked experiments. The blocks represent units in the experimental design for factors division. Also, there are two levels of experimental units namely, whole plot units and subplot units. Randomization of the design is in dual levels. The first randomization aim at assigning block-level treatments to whole plots (*WP*) while the second randomization for the subplot (*SP*) units occurs within each whole plot ([1], [2]). In another view SPD is a combination of two known designed experiments overlaid on each other or as [3] puts it; superimposition of two similar or different form of designs. Investigations have been done in estimating the parameters of the split-plot design linear and response surface models respectively ([1], [2], [4], [6], [5], [7]).

Recently, [8] trellis plots were applied to visualize multivariate data by permitting conditioning throughout the initial data analysis phase of the split plot design data. Also, [9] applied SPD on equipment testing to study three diverse explosive powers inclined by four different intensifiers and four different steel balls. It is cheaper in cost to run a batch of treatments in *SPD* of experiments and statistically efficient and adequate than running the same experiment in a completely randomized fashion. Ju and

Lucas [10] showed that, with one HTC factor or ETC factor, a *SPD* produces a better accuracy in all factor effects estimates involving the subplot and interaction between the WP and SP. However, WP main effect estimate accuracy is low compared to the SP main effect estimate. Design efficiency measure for SPD has been considered over completely randomized designs (CRD) and it has been established that the SPD is more efficient and as well has better optimality determinant than the CRD (see [11] & [12]). This has resulted in many authors and practitioners as well recommending the custom use of SPD even when a CRD is a practicable alternative.

Nonlinear modeling of SPD has attracted little interest mainly in parameters estimation. Although, it follows similar procedure used in estimating the parameters for nonlinear regression. Gumpertz and Rawlings [13] stated that when the objective of fitting a nonlinear function to data from a SP experimental design, a nonlinear model with variance components (whole plot variance, σ^2_γ and split-plot variance, σ^2_ε) is appropriate. The nonlinear SPD model is a distinct type of nonlinear modelling with variance components because the mean part [$f(X, \theta)$] of the model is a nonlinear function with additive random effects (like the *WP* error and *SP* errors). Traditional nonlinear regression follows the assumption that all observations in the data set are uncorrelated and only one source of random error. Biased standard errors will be obtained for the parameter estimates if the traditional nonlinear regression regime is applied for nonlinear SPD model and as well other interesting quantities. This implies that the single variance estimate (i.e. mean square error, *MSE*) obtained will be conciliation between the two random errors (MSE_a and MSE_b) from the nonlinear SPD model analysis ([13], [14], [15]).

Many forms of the split-plot design have been studied in terms of arranging factorial designs (FD), fractional factorial designs (FFD) and response surface designs (RSD) with quadratic and higher order surfaces, optimal designs of first and second orders, sequential and mixture designs. So far the review in this subsection revealed that all works done are based on the estimation of main, interaction, polynomial and higher order effects. Generally, the interest of most researchers is to estimate all the 1^{st}order and 2^{nd}order polynomial terms in a SP response surface model. However, when the order of the term is increased to cubic and higher order terms interpretability issues sets in and in most cases they are of no practical use hence not needed. Therefore, the best option is to model the split-plot experiment nonlinear in parameter, that is, the parameters cannot be linearly transformed. It saves the time of searching at which order a model will produce an estimated optimum yield. In this light, [13] used the Weibull function to model a split-plot fashion of an experimental design for studying an open-top chamber effect of ozone (O_3) and moisture stress regimes on the yield output of two cultivars of soybean. The SPD used was an unbalanced type and the observations within the cultivars are usually correlated with each other. The O_3 and moisture stress regimes are WP effects while the cultivar is the SP effect. Their designed experiment was an unbalanced split-plot design.

Knezevic et al. [15] nonlinearly modelled the effect of three nitrogen rates on critical period for controlling weed (CPCW) in corn. The designed experiment was a SPD with fixed block effect where the three nitrogen rates (*Nr*) were assigned to the *WP* and weed periods [critical timing of weed removal (CTWR) and critical weed free period (CWFP)] as subplot treatment effect. Logistic and Gompertz functions were used to fit the CTWR and CWFP respectively. Blankenship *et al*. [14] stated that the nonlinear SPD model is not different from ANOVA-type model for a standard linear SPD experiment with blocks, but the fixed main and interaction effects are replaced by a nonlinear function relating the connection between treatments and response.

On estimation of the parameters for the nonlinear SPD model developed by [13] was based on EGLS and compared to OLS. They used ANOVA and MIVQUE for estimating the model variance-covariances. They also used maximum likelihood estimation (MLE) technique for estimating the variance components but were dropped from using it for further analyses because the MLE estimates were biased downwards. The MIVQUE did better since it is suitable for unbalance design and generally the EGLS estimates were better than the OLS estimates. However, [15] and [14] added block effect as part of the random effects in their analyses and they used Logistic and Gompertz functions to separately model the WP and SP effects respectively. They estimated their models variance components with residual maximum likelihood estimation (REML) technique.

Estimating variance components (VC) of a nonlinear model, different methods such as MLE by [16], REML introduced by [17], Quasi-MLE, Modified MLE by [18] and [19], ANOVA method by [20], [21] Estimator, Minimum-norm Quadratic Unbiased Estimator (MINQUE), MIVQUE, etc, can be used. The MINQUE and MIVQUE methods were developed by ([22], [23]) and the reason was to find quadratic estimators that are unbiased which are unchanging and minimizes some matrix norm. Rasch and Masata [24] stated that unfortunately, results obtained from these methods depend on the unknown VC. Since they are replaced by estimates from the data, the results are neither unbiased nor quadratic anymore. However, [24] only identified that of MINQUE and MIVQUE in VC estimation but the same is applied to other variance component estimation methods because in almost all cases of modelling with VC, the population VC are unknown. Therefore, since estimated values obtained from the design data which could have outliers, replacement of missing values, etc, the solution might not be unbiased for all other estimation methods.

Weerakkody and Johnson [25] presented a two-step residual-based approach for estimating WP and SP error variances separately. However, as identified by [26] the estimator of the WP error variance in [25]'s approach can be obtained for only the case $a > p$ ($p = 1+p_1+p_2$, where a is the number of runs in each WP unit, p_1 and p_2 are levels of WP and SP effects). This is an impractical strict condition in most situations because it is only suitable for balanced designs. Hasegawa *et al*. [27] suggested a dissimilar estimator for the WP error variance. It has better practical condition than the one introduced by [25]. Yet, both approaches for estimating the two error variances are suitable only for balanced designs. Their approaches were not compared to other methods such as MLE, REML, ANOVA, etc. They are not useful for unbalanced designs, which are often used for reducing experimental runs. Ikeda *et al*. [26] modified the two-step residual-based method proposed by [27] in order to make it readily applicable for balanced and unbalanced designs. Also, they compared their method with REML only. They concluded that their method can be an alternative to REML based on their simulation results. Their alternative method is not a better estimation method because under a different simulation scenario it can perform poorly. Also, their results obtained were not compared with other estimation technique like MLE. However, the methods introduced by [25], [27] and [26] are all implemented only for linear balanced and unbalanced SPD models. In this research we present an iterative Gauss-Newton procedure via Taylor Series expansion using estimated generalized least square (EGLS) method is performed. In this construct, the variance-covariance components which are unknown are estimated by REML method. The purpose for this research is to present a theoretical outline of producing numerically stable parameter estimates.

2. Materials and Method

The nonlinear split-plot model which has whole plot error (WPE) and subplot error (SPE) are special case of a nonlinear model with random effects (also called nonlinear model with variance components, that is, WPE and SPE). The formulated model and assumptions are given as follows.

Let

$$y = f(X,\ \theta) + w + \varepsilon \tag{1}$$

Inserting the levels of the factors to be investigated, (1) is given as follows.

$$y_{ijk} = f(x_{ijk},\ \theta) + w_{ij} + \varepsilon_{ijk} \tag{2}$$

where,

y_{ijk} is the response variable; $i = 1, ..., s$ replicates (***Reps***) or block; $j = 1, ..., a$ levels of the WP factor **A**; $k = 1, ..., b$ levels of the SP factor **B**; w_{ij} is the WPE and ε_{ijk} is the SPE; $f(x_{ijk},\ \theta)$ is the nonlinear function for the mean describing the relationship of fixed main and interaction effects to the response y_{ijk}. The parameters ***Reps***, **A** and **B** are assumed fixed.

Assumption 1: it is presumed that the WPE and SPE are random effects. Also, it is presumed that $w_{ij} \overset{i.i.d.}{\sim} N(0,\sigma^2_{WP})$ and $\varepsilon_{ijk} \overset{i.i.d.}{\sim} N(0,\sigma^2_{SP})$.

Assumption 2: Let $\hat{\theta}$ be the parameter estimate of θ for the model which follows an asymptotic normal distribution with mean θ and variance $\sigma^2(\mathbf{F}'\mathbf{F})^{-1}$, where **F** is the $n \times u$ matrix with elements $\left(\partial f\left(x_{ijk},\theta\right)/\partial\theta'\right)$ which has full column rank, u. This implies that the estimated response $\hat{y}_0$ follows an asymptotic normal distribution with mean y_0 and variance $\mathbf{f}'_x(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}\mathbf{f}_x$ where $\mathbf{f}_x$ is a $u \times 1$ vector with elements $\left(\partial f\left(x_{ijk},\theta\right)/\partial\theta'\right)$ and **V** is the variance-covariance (VC) matrix of the reaction vector.

Assumption 3: if the parameters in the mean function, $f(x_{ijk},\ \theta)$ is p and r is the number of random effects, then n which is the number of observations in the data set must be less than or equal to $p + r + 1$ for all parameters to be estimated. This implies that $n \geq p + r + 1$.

2.1 Estimated generalized Least Square (EGLS) Estimation Method

When the covariance matrix of y is known then the GLS estimator, $\hat{\theta}_{\text{GLS}}$, is found by minimizing the objective function ([8])

$$\left(y - f(X,\ \theta)\right)' \mathbf{V}^{-1}\left(y - f(X,\ \theta)\right) \tag{3}$$

with respect to θ. Where **V** is a known positive definite (non-singular) covariance matrix which arises from the model

$$y_{ijkl} = f(x_{ijkl},\ \theta) + w_{ijk} + \varepsilon_{ijkl} \qquad (4)$$

where, E(w_{ijk}) = 0, Cov(w_{ijk})= $\sigma_w^2\mathbf{I}_N$, E(ε) = 0 and Cov($\sigma_\varepsilon^2\mathbf{I_N}$).

Let the variance-covariance matrix of the observations var(y) be written as

$$\mathbf{V} = \sigma_w^2\mathbf{I_N} + \sigma_\varepsilon^2\mathbf{I_N}$$
$$= \sigma^2\mathbf{I}\ .$$

Using Cholesky decomposition, the inverse of a positive definite matrix **Z** (non-singular matrix) is positive definite with Cholesky factorization if **Z** = L L^t, where L is invertible (its diagonal elements are nonzero) then the right and left inverses of Z are as follows.

- Right inverse of **Z** is T = $L^{-t}L^{-1}$ such that **Z**T = L $L^t L^{-t} L^{-1}$ = L L^{-1} = I
- Left inverse of **Z** is T = $L^{-t}L^{-1}$ such that **Z**T = $L^{-t}L^{-1}$ L L^t= $L^{-t}L^t$ = I

Hence, **Z** is invertible as $\mathbf{Z}^{-1} = L^{-t}L^{-1}$ and $T^{-1} = L L^t$.

Multiplying model (4) by L^{-1} on both sides yield that

$$L^{-1}y_{ijk} = L^{-1}f(x_{ijk},\theta) + L^{-1}(w_{ij}) + L^{-1}(\varepsilon_{ijk}) \qquad (5)$$

Let **I** = T^{-1} = L L^tthen the Cholesky factorization of the error variance is as follows.

$$\begin{aligned} L^{-1}\left[Cov(\varepsilon_{ijkl}) + Cov(w_{ijk})\right]L^{-t} &= L^{-1}Cov(\varepsilon_{ijkl})L^{-t} + L^{-1}Cov(w_{ijk})L^{-t} \\ &= L^{-1}L^{-t}\left[Cov(\varepsilon_{ijk}) + Cov(w_{ij})\right] \\ &= L^{-1}\left(\sigma^2\mathbf{I}\right)L^{-t} \\ &= \sigma^2 L^{-1}LL^tL^{-t} \\ &= \sigma^2\mathbf{I} \end{aligned}$$

Define $\mathbf{T}_{ijk}$= $L^{-1}y_{ijk}$, $\mathbf{M}(x_{ijk},\theta*)$ = $L^{-1}f(x_{ijk},\theta)$ and $\mathbf{\Omega}_{ijk}$ = $L^{-1}(w_{ij}) + L^{-1}(\varepsilon_{ijk})$. Then equation (5) becomes

$$\mathbf{T}_{ijk} = \mathbf{M}(x_{ijkl},\theta^*) + \mathbf{\Omega}_{ijk} \qquad (6)$$

where, $E(\mathbf{\Omega}_{ijk})$ = 0 and $\mathbf{V}(\mathbf{\Omega}_{ijk}) = \sigma^2\mathbf{I}$. Thus the GLS model has been transformed to an OLS model. Hence, model (6) is to be solved using the OLS technique as follows. Taking the summation of both sides of (6) and squaring we have

$$\sum_i^s\sum_j^a\sum_k^b \mathbf{\Omega}_{ijk}^2 = \sum_i^s\sum_j^a\sum_k^b\left[\mathbf{T}_{ijk} - \mathbf{M}(x_{ijk},\theta*)\right]^2 \qquad (7)$$

Let $L(\theta*) = \sum_i^s\sum_j^a\sum_k^b \mathbf{\Omega}_{ijk}^2 = \sum_i^s\sum_j^a\sum_k^b\left[\mathbf{T}_{ijk} - \mathbf{M}(x_{ijk},\theta*)\right]^2$

minimize $L(\theta*)$ w.r.t. $\theta*$, equate to zero and divide both sides by – 2 we have,

$$\frac{\partial L(\theta^*)}{\partial\theta_h^*} = \sum_i^s\sum_j^a\sum_k^b\left[\mathbf{T}_{ijk} - \mathbf{M}(x_{ijk},\ \theta^*)\right]\times\left[\frac{\partial\mathbf{M}(x_{ijk},\ \theta^*)}{\partial\theta_h^*}\right]_{\theta^*=\hat{\theta}^*} = 0 \qquad (8)$$

At this point, equation (8) has no closed form hence will be solved iteratively using the Gauss-Newton method via Taylor series expansion of $\mathbf{M}(x_{ijkl}, \theta^*)$ at first order. Note that the Taylor series expansion is given as

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!} f''(a) + ... + \frac{(x-a)^h}{h!} \times f^{(h)}(a) + R_{h+1} \quad (9)$$

Therefore, we have

$$\mathbf{M}(x_{ijk}, \theta^*) = \mathbf{M}(x_{ijk}, \theta_0^*) + (\theta_1^* - \theta_{10}^*) \left.\frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_1^*}\right|_{\theta^* = \theta_0^*} + (\theta_2^* - \theta_{20}^*) \left.\frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_2^*}\right|_{\theta^* = \theta_0^*} + ... + (\theta_h^* - \theta_{h0}^*) \left.\frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_h^*}\right|_{\theta^* = \theta_0^*} \quad (10)$$

Let $\mathbf{M}(x_{ijk}, \theta^*) = \eta(\theta^*)$ and $d_{ijkl} = \left.\frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_h^*}\right|_{\theta^* = \theta_0^*}$ for all N cases and $\delta = \theta^* - \theta_0^*$ then (10) becomes

$$\eta(\theta^*) = \eta(\theta_0^*) + D_0 \delta \quad (11)$$

where D_0 is the $N \times H$ derivative matrix with elements $\{d_{ijk \times h}\}$ and this is comparable to approximating the residuals for the model, that is, $\mathbf{\Omega}(\theta^*) = \mathbf{T} - \eta(\theta^*)$ by

$$\begin{aligned} \mathbf{\Omega}(\theta^*) &= \mathbf{T} - \left[\eta(\theta_0^*) + D_0 \delta\right] \\ &= \mathbf{T} - \eta(\theta_0^*) - D_0 \delta \\ &= z_0 - D_0 \delta \end{aligned} \quad (12)$$

where $z_0 = \mathbf{T} - \eta(\theta_0^*)$ and $\delta = \theta^* - \theta_0^*$.

The Householder (1958) QR decomposition ([28], [29]) is applied to (28), as a result of its numerical stability characteristics for estimating the model parameters [29]. This is done to decompose D_0 into a product of an orthogonal matrix and an inverted matrix.

Theorem 1: Suppose A is a full column rank matrix of x × y, then A can be written as A = QR where Q is a matrix of x × y whose column vectors create orthonormal basis for the column space of A while R is an y × y invertible upper triangular matrix.

Proof: Let an x × ymatrix have columns $w_1, w_2, ... w_y$ vectors.

Also, let $q_1, q_2, ... q_n, q_{y+1}, ..., q_x$ be orthonormal vectors such that,

$$\|q_i\| = 1, \qquad q_i^t q_j = 0 \text{ if } i \neq j$$

Then Q is $m \times n$ with orthonormal columns such that, $Q^tQ = I$.

If A is a squared matrix ($x = y$), then Q is orthogonal, that is, $Q^tQ = QQ^t = I$, hence, q_i is orthogonal to $w_1, w_2, ... w_y$.

Therefore,

$$w_1 = (w_1 \cdot q_1)q_1$$
$$w_2 = (w_2 \cdot q_1)q_1 + (w_2 \cdot q_2)q_2$$
$$\ldots \tag{13}$$
$$w_y = (w_y \cdot q_1)q_1 + (w_y \cdot q_2)q_2 + \ldots + (w_y \cdot q_y)q_y$$

This implies that A = QR

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_y \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_y \end{bmatrix} \begin{bmatrix} (w_1 \cdot q_1) & (w_1 \cdot q_2) & \cdots & (w_1 \cdot q_y) \\ 0 & (w_2 \cdot q_2) & \cdots & (w_2 \cdot q_y) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (w_k \cdot q_k) \end{bmatrix} \tag{14}$$

Let A = $\begin{bmatrix} w_1 & w_2 & \cdots & w_y \end{bmatrix}$ and $R_{ij} = w_i \cdot q_j$, therefore, equation (14) is written as

$$A = \begin{bmatrix} q_1 & q_2 & \cdots & q_y \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1y} \\ 0 & R_{22} & \cdots & R_{2y} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{yy} \end{bmatrix} \tag{15}$$

Equation (15) shows that R is $y \times y$, upper triangular with nonzero diagonal elements and R is non-singular (since the diagonal elements are nonzero). This means A = QR.

Theorem 2: *If* A *is an* $p \times n$ *matrix with full column rank, and if* A = QR, *a* QR-*decomposition of* A, *then the normal system for* A**x** = **b** *can be expressed as* R**x** = Q^t**b** *and the least squares solution is* $\hat{\mathbf{x}} = R^{-1}Q^t\mathbf{b}$.

Proof: Let $\hat{\mathbf{x}} = (A^tA)^{-1}A^t\mathbf{b}$ be the best approximate solution to A**x** = **b**. Based on the orthonormal and orthogonal property exhibited by QR-decomposition, if

A = QR

then

$A^t = R^tQ^t$.

Therefore,

$$\hat{\mathbf{x}} = (A^tA)^{-1}A^t\mathbf{b} = (R^tQ^t\,QR)^{-1}R^tQ^t\mathbf{b}$$
$$\Rightarrow R^tQ^t\,QR\,\hat{\mathbf{x}} = R^tQ^t\mathbf{b}$$
$$\Rightarrow R^t R\,\hat{\mathbf{x}} = R^tQ^t\mathbf{b}$$

Since $Q^tQ = 1$

$$\hat{\mathbf{x}} = R^{-1}Q^t\mathbf{b}. \tag{16}$$

Based on the two stated and proved theorems on QR-decomposition, the decomposition of D_0 is presented as follows.

Let $D_0 = QR$

where Q is a matrix of $N \times N$ and orthogonal, $Q^tQ = QQ^t = I$ while Ris an $N \times H$ triangular matrix and under the major diagonal Ris zero. Writing Q and R as follows,

$$Q = [Q_1 | Q_2]$$

where Q_1 is the first H columns and Q_2 is the last $N - H$ columns of Q, and

$$R = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}$$

with R_1 a $H \times H$ upper triangular matrix with all elements greater than zero and R_2 is a $(N-H) \times H$ lower matrix of zeros. Also,

$$Q^t = \begin{bmatrix} Q_1^t \\ Q_2^t \end{bmatrix}$$

where Q_1^t and Q_2^t are of dimension $H \times N$ and $(N-H) \times N$ respectively. Therefore,

$$D_0 = QR = Q_1 R_1 \quad (17)$$

Geometrically, the columns of Q define an orthonormal, or orthogonal, basis for the response space based on the property that the H columns cover the expectation plane. Projection onto the expectation plane is simple if the projection is in the coordinate system given by Q [28].

Next is transformation of the response vector, which is

$$g = Q^t z_0 \quad (18)$$

with components

$$g_1 = Q_1^t z_0 \quad (19)$$

and

$$g_2 = Q_2^t z_0 \,. \quad (20)$$

The projection of g onto the expectation plane is simply given as

$$\begin{bmatrix} g_1 \\ 0 \end{bmatrix}$$

in Q coordinates and

$$\hat{\eta}_1 = Q \begin{bmatrix} g_1 \\ 0 \end{bmatrix} = Q_1 g_1 \quad (21)$$

in the original coordinates. So,

$$\delta_0 = R_1^{-1} g_1$$

this implies

$$R_1 \delta_0 = g_1 \quad (22)$$

Equation (22) can now be easily estimated using backward solving ([29]). The point $\hat{\eta}_1 = \eta(\theta_1^*) = \eta(\theta_0^* + \delta_0)$ should now be closer to y than $\eta(\theta_0^*)$, and then move to better parameter value $\theta_1^* = \theta_0^* + \delta_0$ and carryout another iteration by calculating new residuals $z_1 = T - \eta(\theta_1^*)$, a new derivative matrix D_0, and a new increase. Repetition of the process is done until convergence is obtained, that is, until the increment is so small with no useful change in the elements of the parameter vector [28].

It is expected that the new residual sum of squares (RSS) should be less than the initial estimate but if otherwise, a small step in the direction δ_0 is introduced. A step factor λ is introduced and then calculated ([28]):

$$\theta_1^* = \theta_0^* + \lambda \delta_0$$

where λ is chosen to ensure that the new RSS is less than the initial estimate. A common method as suggested by [28] is to begin with $\lambda = 1$ and reduce it by half until it is satisfied that the new RSS is less than the initial estimate.

In actual practice the GLS is impracticable because the VC matrix, **V**, is unknown. Therefore, an estimated **V** is obtained and substituted into equation (3) and the term EGLS is used. There are differenttechniques for estimating the variance components to substitute for **V** in equation (3). In this research work the procedure for REML technique is presented. The technique is presented in section 2.2 below.

2.2 Variance Component Estimation Via REML

It's known that REML procedure does not involve $\hat{\theta}^*$ in the estimation of the variance component. The function of the likelihood is based on vectors in the error space, that is, on linear combinations of y which have expectation to be zero rather than y itself. To obtain these vectors in the error space the linear approximation of the residuals is used $z_0 = D_0\delta + \varepsilon$ as shown in (12).

To estimate the variance components from the nonlinear functions of y that won't involve $\hat{\theta}^*$, vectors of the form $\mathbf{k}^t y$ are formed whereby **k** is selected so that $\mathbf{k}^t D_0 = 0$ which falls in the linear estimate to the error space. $\mathbf{k}^t y$ is called the error contrasts ([30]), that is, the part of the data that is orthogonal to the fixed effects (not dependent on the values of the fixed effect estimates), **k** is a vector from a full rank matrix **K** and maximizing the likelihood on $\mathbf{K}^t y$, the function of the log likelihood on $\mathbf{K}^t y$, is

$$\ln L(\boldsymbol{\Theta}) = -\frac{n}{2}\ln\,(2\pi)\; -\frac{1}{2}\ln\;\left|\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right| - \frac{1}{2}\left(\mathrm{K}^t y - \mathrm{K}^t f\left(X,\,\theta\right)\right)^t \times\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1}\left(\mathrm{K}^t y - \mathrm{K}^t f\left(X,\,\theta\right)\right) \tag{23}$$

where $\boldsymbol{\Theta} = \left(\sigma^{2\,\prime} = \sigma^2_{WP}, \sigma^2_{SP}\right)$, is then approximated by the surface and letting ln$L$ to be $\boldsymbol{\Gamma}$ equation (23) becomes,

$$\begin{aligned}
\boldsymbol{\Gamma}(\boldsymbol{\Theta}) &= -\frac{n}{2}\ln\,(2\pi)\; -\frac{1}{2}\ln\;\left|\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right| - \frac{1}{2}\left(\mathrm{K}^t y - \mathrm{K}^t f(x,\theta)\right)^t\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1}\left(\mathrm{K}^t y - \mathrm{K}^t f(x,\theta)\right)\\
&= C - \frac{1}{2}\ln\;\left|\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right| - \frac{1}{2}\left(\mathrm{K}^t y - \mathrm{K}^t f(x,\theta)\right)^t\left(\mathrm{K}^t y\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1} - \mathrm{K}^t f(x,\theta)\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1}\right)\\
&= C - \frac{1}{2}\ln\;\left|\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right| - \frac{1}{2}\left(\left(\mathrm{K}^t y - \mathrm{K}^t f(x,\theta)\right)^t \mathrm{K}^t y\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1}\right)\\
&\quad + \frac{1}{2}\left(\left(\mathrm{K}^t y - \mathrm{K}^t f(x,\theta)\right)^t \mathrm{K}^t f(x,\theta)\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1}\right)
\end{aligned} \tag{24}$$

The third and fourth terms of equation (24) can be expressed respectively as follows.

$$\begin{aligned}
\frac{1}{2}\left(\left(\mathrm{K}^t y - \mathrm{K}^t f(x,\theta)\right)^t \mathrm{K}^t y\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1}\right) &= \frac{1}{2}\left(\left(\mathrm{K}y^t - \mathrm{K}f(x,\theta)^t\right)\mathrm{K}^t y\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1}\right)\\
&= \frac{1}{2}\left(y^t\mathrm{K}^t\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1} y\mathrm{K} - f(x,\theta)^t\mathrm{K}^t\left(\mathrm{K}^t\,\mathbf{V}\,\mathrm{K}\right)^{-1} y\mathrm{K}\right)
\end{aligned} \tag{25}$$

and

$$\frac{1}{2}\left(\left(\mathrm{K}^{\mathrm{t}}y-\mathrm{K}^{\mathrm{t}}f(x,\theta)\right)^{t}\mathrm{K}^{\mathrm{t}}f(x,\theta)\left(\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right)^{-1}\right)=\frac{1}{2}\left(\left(\mathrm{K}y^{t}-\mathrm{K}f(x,\theta)^{t}\right)\mathrm{K}^{\mathrm{t}}f(x,\theta)\left(\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right)^{-1}\right)$$

$$=\frac{1}{2}\left(y^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right)^{-1}f(x,\theta)\mathrm{K}-f(x,\theta)^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right)^{-1}f(x,\theta)\mathrm{K}\right) \tag{26}$$

respectively. Therefore, equation (24) becomes

$$\begin{aligned}\mathbf{\Gamma}\left(\mathbf{\Theta}\right)=C-\frac{1}{2}\ln\left|\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right|-\frac{1}{2}\left(y^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right)^{-1}y\mathrm{K}-f(x,\theta)^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right)^{-1}y\mathrm{K}\right)\\+\frac{1}{2}\left(y^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right)^{-1}f(x,\theta)\mathrm{K}-f(x,\theta)^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\,\mathbf{V}\,\mathrm{K}\right)^{-1}f(x,\theta)\mathrm{K}\right)\end{aligned} \tag{27}$$

$\mathbf{V}=\sigma^{2}\mathbf{I}=K\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^tK^t$ and $\mathbf{V}\mathbf{V}^{-1}$ can be expressed as given below.

$$\mathbf{V}\mathbf{V}^{-1}=\left(\mathrm{K}\mathbf{V}\mathrm{K}^{t}\right)\left(\mathrm{K}\mathbf{V}\mathrm{K}^{t}\right)^{-1}=\mathbf{V}\left(\mathrm{K}\left(\mathrm{K}\mathbf{V}\mathrm{K}^{t}\right)^{-1}\mathrm{K}^{t}\right)=\sigma^{2}\mathbf{I}\left(\mathrm{Q}_{h}\right)=\sigma^{2}\left(\mathrm{Q}_{h}\mathbf{V}_{j}\right)$$

Inserting $\mathbf{V}$ into equation (27) we have

$$\begin{aligned}\mathbf{\Gamma}\left(\mathbf{\Theta}\right)=C-\frac{1}{2}\ln\left|\mathrm{K}^{\mathrm{t}}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\,\mathrm{K}\right|\\-\frac{1}{2}\left(y^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\,\mathrm{K}\right)^{-1}y\mathrm{K}-f(x,\theta)^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\,\mathrm{K}\right)^{-1}y\mathrm{K}\right)\\+\frac{1}{2}\left(y^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\,\mathrm{K}\right)^{-1}f(x,\theta)\mathrm{K}-f(x,\theta)^{t}\mathrm{K}^{\mathrm{t}}\left(\mathrm{K}^{\mathrm{t}}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\,\mathrm{K}\right)^{-1}f(x,\theta)\mathrm{K}\right)\end{aligned} \tag{28}$$

Differentiate partially equation (28) w.r.t. σ_i^2 and equate to zero. By transformation all other terms in the equation becomes zero since $\mathrm{K}^{\mathrm{t}}D_0=\mathrm{K}^{\mathrm{t}}f(x,\theta)=\mathrm{K}f(x,\theta)^t=0$. Hence we have

$$\frac{\partial\mathbf{\Gamma}\left(\mathbf{\Theta}\right)}{\partial\sigma_i^2}=-\frac{1}{2}\frac{1}{\left|\mathrm{K}^{t}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right|}\left(\mathrm{K}^{t}\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)+\frac{1}{2}y^{t}\mathrm{K}^{t}\left(\mathrm{K}^{t}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)^{-1}\left(\mathrm{K}^{t}\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)\left(\mathrm{K}^{t}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)^{-1}y\mathrm{K}$$

$$\frac{1}{2}\left(\frac{1}{\left|\mathrm{K}^{t}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right|}\left(\mathrm{K}^{t}\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)\right)=\frac{1}{2}y^{t}\mathrm{K}^{t}\left(\mathrm{K}^{t}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)^{-1}\left(\mathrm{K}^{t}\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)\left(\mathrm{K}^{t}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)^{-1}y\mathrm{K} \tag{29}$$

Let $\mathbf{Q}_h=\mathrm{K}^{t}\left(\mathrm{K}^{t}\sum_{i=0}^{2}\sigma_i^2\mathbf{I}_i\mathbf{I}_i^t\mathrm{K}\right)^{-1}\mathrm{K}$ then equation (29) becomes

$$\frac{1}{2}\left(tr\left(\mathrm{Q}_h\mathbf{V}_i\right)\right)=\frac{1}{2}\left(y^{t}\mathrm{Q}_h\mathbf{V}_i\,\mathrm{Q}_h y\right) \tag{30}$$

Multiply the left hand side of equation (30) by $\mathbf{V}\mathbf{V}^{-1}$ we have

$$\frac{1}{2}\left(tr\left(\mathrm{Q}_h\mathbf{V}_i\right)\right)\sigma_{j(h+1)}^2\left(\mathrm{Q}_h\mathbf{V}_j\right)=\frac{1}{2}\left(y^{t}\mathrm{Q}_h\mathbf{V}_i\,\mathrm{Q}_h y\right) \tag{31}$$

$$\left\langle tr\left(\hat{\mathbf{Q}}_{(h)}\hat{\mathbf{V}}_i\;\hat{\mathbf{Q}}_{(h)}\hat{\mathbf{V}}_j\right)\right\rangle\times\left\langle\left(\hat{\sigma}_{j(h+1)}^2\right)\right\rangle=\left\langle\left(y^{\mathrm{t}}\hat{\mathbf{Q}}_{(h)}\hat{\mathbf{V}}_i\;\hat{\mathbf{Q}}_{(h)}y\right)\right\rangle \tag{32}$$

$$\left\langle \left(\hat{\sigma}^2_{j(h+1)} \right) \right\rangle = \left\langle tr\left(\hat{\mathbf{Q}}_{(h)} \hat{\mathbf{V}}_i \ \hat{\mathbf{Q}}_{(h)} \hat{\mathbf{V}}_j \right) \right\rangle^{-1} \times \left\langle \left(y^{\mathbf{t}} \hat{\mathbf{Q}}_{(h)} \hat{\mathbf{V}}_i \ \hat{\mathbf{Q}}_{(h)} y \right) \right\rangle \tag{33}$$

The solutions to the equations might be negative when more iteration does not improve the log likelihood. In such a case, the negative value is retuned to zero afore the next iteration.

3. Conclusion

The estimated generalized least square (EGLS) method presented in this paper is often applied for estimating linear fixed, random and mixed-effect split-plot design models. However, in practical applications, the mean part of the model is often nonlinear due to dynamics involved in the system process. This paper presents the procedure and steps in estimating the parameters for a SP model where the mean part of the model can be any nonlinear function and the variance components $\left(\sigma^{2\,\prime} = \sigma^2_{WP}, \sigma^2_{SP} \right)$ of the model are estimated through REML technique. This is achieved by minimization of the objective function, $\left(y - f(X,\ \theta) \right)^{\mathbf{t}} \mathbf{V}^{-1} \left(y - f(X,\ \theta) \right)$ where the estimates of $\hat{\theta}^*$ and $\sigma^{2\,\prime} = \sigma^2_{WP}, \sigma^2_{SP}$ are iteratively obtained at the $(h + 1)^{st}$ iteration by inserting a prior estimate of $\sigma^{2\,\prime}$ to the estimating equation until it converges. To achieve these iterative procedures for the nonlinear SPD models parameters to be estimated, statistical software such as the %NLINMIX SAS macro can be used to handle all computations.

References

1. Montgomery, D. C. (2008). Design and Analysis of Experiments, 7thed. New York, NY: John Wiley & Sons.
2. Jones, B. and Nachtsheim, C. J. (2009). Split-plot Designs: What, why, and How, Journal of Quality Technology, 41(4), p. 340-361.
3. Hinkelmann, K. and Kempthrone, O. (2008). Design and Analysis of Experiments Vol 1: Introduction to Experimental design, 2nd Edition. New York: Wiley.
4. Letsinger, J. D.; Myers, R. H.; and Lentner, M. (1996). response surface methods for bi-randomization structures, Journal of Quality Technology, 28, p. 381–397.
5. Kowalski, S. M., Cornell, J. A. and Vining, G. G. (2002). Split-plot designs and estimation methods for mixture experiments with process variables, Technometrics, 44, p. 72–79.
6. Draper, N. R. and John, J. A. (1998). Response surface designs where levels of some factors are difficult to change, Australia, New Zealand Journal of Statistics, 40, p. 487–495.
7. Vining, G. G., Kowalski, S. M. and Montgomery, D. C. (2005). Response surface designs within a split-plot structure, Journal of Quality Technology, 37, p. 115–129.
8. Kulachi, M. &Menon, A. (2017). Trellis plots as visual aids for analyzing split plot experiments. Quality Engineering, 29(2), p. 211–225. https://doi.org/ 10.1080/ 08982112.2016.1243248.
9. Huameng, G., Fan, Y. and Lei, S. (2017). Split Plot and Data Analysis in SAS, American Institute of Physics Conference Proceedings 1834, 030024 (2017); doi:10.1063/1.4981589.

10. Ju, H. L. and Lucas, J. M. (2002). L^k factorial experiments with hard-to-change and easy-to-change factors, Journal of Quality Technology, 34, p. 411 – 421.
11. Hasegawa, Y., Ikeda, S., Matsuura, S. and Suzuki, H. (2010). A study on methodology for total design management (the 4th report): A study on the response surface method for split-plot designs using the generalized least squares, In: Proceedings of the 92nd JSQC Technical Conference, Tokyo: The Japanese Society for Quality Control, pp. 235–238 (in Japanese).
12. Anbari, F. T. and Lucas, J. M. (1994). Super-efficient designs: how to run your experiment for higher efficiencies and lower cost, ASQC Technical Conference Transactions, p. 852-863.
13. Gumpertz, M. L. and Rawlings, J. O. (1992). nonlinear regression with variance components: modeling effects of ozone on crop yield, Crop Science, 32, p. 219 – 224.
14. Blankenship, E. E., Stroup, W. W., Evans, S. P. and Knezevic, S. Z. (2003). Statistical inference for calibration points in nonlinear mixed effects models, American Statistical Association and the International Biometric Society Journal of Agricultural, Biological, and Environmental Statistics, 8(4), p. 455 – 468.
15. Knezevic, S. Z., Evans, S. P., Blankenship, E. E., Van Acker, R. C., and Lindquist, J. L. (2002). Critical period for weed control: the concept and data analysis, Agronomy – Faculty Publications, Paper 407.
16. Herbach, L. H. (1959). Properties of model II type analysis of variance tests A: Optimum nature of the F-test for model II in balanced case, Annals of Mathematical Statistics, 30, p. 939–959.
17. Anderson, R. L. and Bancroft T. A. (1952). Statistical Theory in Research, McGraw-Hill, New York.
18. Stein, C. (1969). In admissibility of the usual estimator for the variance of a normal distribution with unknown mean, Annals of the Institute of Statistics and Mathematics (Japan), 16, p. 155–160.
19. Klotz, J. H., Milton, R. C. and Zacks, S. (1996). Mean square efficiency of estimators of variance components, Journal of the American Statistical Association, 64, p. 1383–1402.
20. Klotz, J. H., Milton, R. C. and Zacks, S. (1996). Mean square efficiency of estimators of variance components, Journal of the American Statistical Association, 64, p. 1383–1402.
21. Federer, W. T. (1968). Non-negative estimators for components of variance, Applied Statistics, 17, p. 171–174.
22. Rao C. R. (1971a). Estimation of variance and covariance components: MINQUE theory, Journal of Multivariate Analysis, 1, p. 257–275.
23. Rao C. R. (1972): Estimation of variance and covariance components in linear models, Journal of the American Statistical Association, 67, p. 112–115.
24. Rash, D. and Masata, O. (2006). Methods of variance component estimation, Czech Journal of Animal Science, 51(6), p. 227 – 235.
25. Weerakkody, G. J. and Johnson, D. E. (1992). Estimation of within model parameters in regression models with a nested error structure, Journal of the American Statistical Association 87, p. 708–713.

26. Ikeda, S., Matsuura, S. & Suzuki, H. (2014). Two-step residual-based estimation of error variances for generalized least squares in split-plot experiments, Communications in Statistics: Simulation and Computation, 43(2), p. 342-358, DOI:10.1080/03610918.2012.703280.
27. Hasegawa, Y., Ikeda, S., Matsuura, S. and Suzuki, H. (2010). A study on methodology for total design management (the 4th report): A study on the response surface method for split-plot designs using the generalized least squares, In: Proceedings of the 92nd JSQC Technical Conference, Tokyo: The Japanese Society for Quality Control, pp. 235–238 (in Japanese).
28. Bates, D. M., and Watts, D. G. (1988). Nonlinear Regression Analysis and Its Applications, New York: Wiley.
29. Klotz, J. (2006). A computational approach to statistics, University of Wisconsin, Madison:USA, https://www.mimuw.edu.pl/~pokar/StatystykaI/Literatura/KlotzBook.pdf.
30. Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems, Journal of the American Statistical Association, 72, p. 320–338.