On Parameters Estimation of Nonlinear Split-Plot Design Model with EGLS-MLE

I. J. David*; O. E. Asiribo; H. G. Dikko

Department of Statistics,
Ahmadu Bello University, Zaria, Nigeria.
e-mail: davidij@fuwukari.edu.ng*

***Abstract* — In this paper, a theoretical iterative Gauss-Newton via Taylor Series expansion procedures for estimated generalized least square (EGLS) technique is presented in estimating the parameters of a nonlinear split-plot design (SPD) model where the variance components are unknown and are estimated via restricted maximum likelihood estimation (REML) method.**

Keywords- *Split-plot design model, Nonlinear, Estimated generalized least square, restricted maximum likelihood estimation..*

.

I. Introduction

In simple terms, a split-plot experiment is a blocked experiment, where the blocks themselves serve as experimental units for a subset of the factors. Thus, there are two levels of experimental units. The blocks are referred to as whole plots while the experimental units within blocks are called split plots, split units, or subplots. Corresponding to the two levels of experimental units are two levels of randomization. One randomization is conducted to determine the assignment of block-level treatments to whole plots (*WP*). Then, as always in a blocked experiment, a randomization of treatments to split-plot (*SP*) experimental units occurs within each block or whole plot ([4], [2]). Hence, they are designed experiments that can be viewed as two experiments combined or overlaid on each other or as [3] puts it; superimposition of two similar or different form of designs. Much research has been done in estimating the parameters of the split-plot design linear and response surface models respectively ([5], [7], [6], [10], [2], [1]).

Nonlinear modeling of split-plot design has attracted few researches especially in estimating the parameters of the model. Although, it follows the same procedure used in parameter estimation for nonlinear regression. [8] stated that when the objective of fitting a nonlinear function to data from a split-plot experimental designs, a nonlinear model with variance components (whole plot variance, σ^2_γ and split-plot variance, σ^2_ε) is appropriate.

The nonlinear model with split-plot errors is a special case of the nonlinear model with variance components. This is so because the model contains a nonlinear function for the mean, $\mathbf{g}(\mathbf{X}, \boldsymbol{\theta})$, and the random effects, such as whole plot and subplot errors, are added to the mean function. Standard nonlinear regression programs make the assumption that all the observations in the data set are uncorrelated and that there is only one source of random variation. If they are used to fit models with more than one random error term, they give incorrect standard errors for the parameter estimates and for other quantities of interest. Hence, if an ordinary nonlinear regression program is used to analyze data from a split-plot experiment, the single variance estimate, MSE, will be a compromise between the two error variances, commonly called MSE_a and MSE_b, from the split-plot analysis of variance ([8], [9], [11]).

In this research paper, a theoretical presentation of an iterative Gauss-Newton via Taylor Series expansion procedures for estimated generalized least square (EGLS) technique in estimating the parameters of a nonlinear split-plot design (SPD) model where the variance components are unknown and are estimated via restricted (residual) maximum likelihood estimation (REML) method.

II. Research Methodology

The nonlinear split-plot model which has whole plot error (WPE) and subplot error (SPE) are special case of 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

$$y = f(X, \theta) + w + \varepsilon \qquad (1)$$

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

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

where,

y_{ijk} is the response variable; $i = 1, ..., S$ replicates or block; $j = 1, ..., a$ levels of the *WP* factor **A**; $k = 1, ..., b$ levels of the *SP* factor **B**; w_{ijk} is the WP error and ε_{ijkl} is the *SP* error; $f(x_{ijkl}, \theta)$ is the nonlinear function for the mean.

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

Assumption 2: Let $\hat{\theta}$ be the model parameter estimate of θ which follows an asymptotic normal distribution with mean θ and variance $\sigma^2(\mathbf{F}'\mathbf{F})^{-1}$, where **F** is the $n \times p$ matrix with elements $(\partial f(x_{ijkl}, \theta)/\partial\theta')$ which has full column rank, p. 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 $p \times 1$ vector with elements $(\partial f(x_{0000}, \theta)/\partial\theta')$ and **V** is the covariance matrix of the response vector.

Assumption 3: if the parameters in the mean function, $f(x_{ijkl}, \theta)$ is p and the number of random effects is r, then the number of observations in the data set, n, must be at least $p + r + 1$ in order to estimate all of the parameters. This implies that $n \geq p + r + 1$.

Estimated generalized Least Square (EGLS) Estimation Method

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

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

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, $E(w_{ijk}) = 0$, $Cov(w_{ijk}) = \sigma^2_w \mathbf{I}_N$, $E(\varepsilon_{ijkl}) = 0$ and $Cov(\varepsilon_{ijkl}) = \sigma^2_\varepsilon \mathrm{I}_N$.

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

$$\mathbf{V} = \sigma^2_w \mathbf{I_N} + \sigma^2_\varepsilon \mathbf{I_N}$$

$$= \sigma^2 \mathbf{K}.$$

Using spectral decomposition, it can be established that **V** is positive definite if and only if there exists a non-singular matrix **P** such that

$$\mathbf{V} = \mathbf{P}\mathbf{P}^t.$$

Multiplying model (4) by $\mathbf{P}^{-1}$ on both sides yields

$$\mathbf{P}^{-1}y = \mathbf{P}^{-1}f(x_{ijkl}, \theta) + \mathbf{P}^{-1}(w_{ijk}) + \mathbf{P}^{-1}(\varepsilon_{ijkl}) \qquad (5)$$

where,

$$Cov(\mathbf{P}^{-1}\varepsilon_{ijkl}) + Cov(\mathbf{P}^{-1}w_{ijk})$$
$$= \mathbf{P}^{-1}Cov(\mathbf{P}^{-1}\varepsilon_{ijkl})(\mathbf{P}^{-1})^t + \mathbf{P}^{-1}Cov(\mathbf{P}^{-1}w_{ijk})(\mathbf{P}^{-1})^t$$
$$= \mathbf{P}^{-1}(\mathbf{P}^{-1})^t\left[Cov(\varepsilon_{ijkl}) + Cov(w_{ijk})\right]$$
$$= \mathbf{P}^{-1}\sigma^2\mathbf{K}(\mathbf{P}^{-1})^t$$
$$= \sigma^2\mathbf{P}^{-1}\mathbf{P}\mathbf{P}^t(\mathbf{P}^{-1})^t$$
$$= \sigma^2\mathbf{I}$$

Define $\mathbf{T} = \mathbf{P}^{-1}y$, $\mathbf{M}(x_{ijkl}, \theta*) = \mathbf{P}^{-1}f(x_{ijkl}, \theta)$ and $\mathbf{E} = \mathbf{P}^{-1}(w_{ijk}) + \mathbf{P}^{-1}(\varepsilon_{ijkl})$. Then the model (5) becomes

$$\mathbf{T} = \mathbf{M}(x_{ijkl}, \theta^*) + \mathrm{E} \qquad (6)$$

where, $E(\Phi) = 0$ and $\mathbf{V}(\mathrm{E}) = \sigma^2\mathbf{I}$. 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 square we have

$$\sum_i^S\sum_j^a\sum_k^b\sum_l^c \mathbf{E}^2_{ijkl}$$
$$= \sum_i^S\sum_j^a\sum_k^b\sum_l^c\left[\mathbf{T}_{ijkl} - \mathbf{M}(x_{ijkl}, \theta*)\right]^2 \qquad (7)$$

Let

$$L(\theta*) = \sum_i^S\sum_j^a\sum_k^b\sum_l^c \mathrm{E}^2_{ijkl}$$
$$= \sum_i^S\sum_j^a\sum_k^b\sum_l^c\left[\mathbf{T}_{ijkl} - \mathbf{M}(x_{ijkl}, \theta*)\right]^2$$

minimize $L(\theta^*)$ w.r.t. θ^* and equate to zero we have,

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

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

$$f(x)=f(a)+(x-a)f'(a)+\frac{(x-a)^2}{2!}f''(a)+\ldots+\frac{(x-a)^p}{p!}f^{(p)}(a)+R_{p+1} \qquad (9)$$

where $f'(a)=\dfrac{df(x)}{dx}$ around $x = a$, and R_{p+1} is the remainder term which is reasonably small if p is sufficiently large. Therefore, we have

$$\mathbf{M}(x_{ijkl},\theta^*)=\mathbf{M}(x_{ijkl},\theta_0^*)+(\theta_1^*-\theta_{10}^*)\left.\frac{\partial\mathbf{M}(x_{ijkl},\theta^*)}{\partial\theta_1^*}\right|_{\theta^*=\theta_0^*}+(\theta_2^*-\theta_{20}^*)\left.\frac{\partial\mathbf{M}(x_{ijkl},\theta^*)}{\partial\theta_2^*}\right|_{\theta^*=\theta_0^*}+\ldots+(\theta_p^*-\theta_{p0}^*)\left.\frac{\partial\mathbf{M}(x_{ijkl},\theta^*)}{\partial\theta_p^*}\right|_{\theta^*=\theta_0^*} \qquad (10)$$

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

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

where D_0 is the $N\times P$ derivative matrix with elements $\{d_{ijkl\times p}\}$ and this is equivalent to approximating the residuals for the model, that is, $\mathrm{E}(\theta^*)=\mathbf{T}-\eta(\theta^*)$ by

$$\mathrm{E}(\theta^*)=\mathbf{T}-\left[\eta(\theta_0^*)+D_0\delta\right]$$
$$=\mathbf{T}-\eta(\theta_0^*)-D_0\delta$$
$$=z_0-D_0\delta \qquad (12)$$

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

Applying the Householder (1958) **QR** decomposition ([12], [13]) to (12) and this is due to its numerical stability characteristic for estimating the parameters in the model Klotz (2006). This is done to decompose D_0 into the product of an orthogonal matrix and an inverted matrix.

Theorem 1: *If A is an $m\times n$ matrix with full column rank, then A can be factored as $A = QR$ where Q is an $m\times n$ matrix whose column vectors form an orthonormal basis for the column space of A and R is an $n\times n$ invertible upper triangular matrix.*

Proof: Let $m\times n$ matrix have columns $w_1,w_2,\ldots w_n$ m-vectors.

Also, let $q_1,q_2,\ldots q_n,q_{n+1},\ldots,q_m$ be orthonormal m-vectors such that,

$\|q_i\|=1$, $q_i^T q_j=0$ if $i\neq j$

Then Q is $m\times n$ with orthonormal columns, $Q^TQ=I$. If A is a square matrix ($m = n$), then Q is orthogonal, that is, $Q^TQ = QQ^T = I$, hence, q_i is orthogonal to $w_1,w_2,\ldots w_n$.

Therefore,

$$w_1=(w_1\cdot q_1)q_1$$
$$w_2=(w_2\cdot q_1)q_1+(w_2\cdot q_2)q_2$$
$$\ldots$$
$$w_k=(w_n\cdot q_1)q_1+(w_n\cdot q_2)q_2+\ldots+(w_n\cdot q_n)q_n \qquad (13)$$

This implies that $A = QR$

$$\begin{bmatrix}w_1 & w_2 & \cdots & w_n\end{bmatrix}=\begin{bmatrix}q_1 & q_2 & \cdots & q_n\end{bmatrix}\times\begin{bmatrix}(w_1\cdot q_1) & (w_1\cdot q_2) & \cdots & (w_1\cdot q_n)\\ 0 & (w_2\cdot q_2) & \cdots & (w_2\cdot q_n)\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & (w_k\cdot q_k)\end{bmatrix} \qquad (14)$$

Let A = $\begin{bmatrix}w_1 & w_2 & \cdots & w_n\end{bmatrix}$ and R = $w\cdot q$, therefore, equation (14) is written as

$$A = [q_1 \quad q_2 \quad \cdots \quad q_n] \times \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ 0 & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{nn} \end{bmatrix} \qquad (15)$$

Equation (15) shows that R is $n \times n$, upper triangular with nonzero diagonal elements and R is non-singular (diagonal elements are nonzero).

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

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

$$A = QR$$

then

$$A^T = R^T Q^T.$$

Therefore,

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} = (R^T Q^T QR)^{-1} R^T Q^T \mathbf{b}$$

$$R^T Q^T QR\,\hat{\mathbf{x}} = R^T Q^T \mathbf{b} \;\; (Q^T Q = 1)$$

$$R^T R\,\hat{\mathbf{x}} = R^T Q^T \mathbf{b}$$

$$\hat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}. \qquad (16)$$

Based on the two stated and proved theorems on **QR**-decomposition, the decomposition of M_0 is presented as follows.

Let $M_0 = \mathbf{QR}$

where $\mathbf{Q}$ is an $N \times N$ orthogonal matrix, that is, $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}$ and $\mathbf{R}$ is an $N \times P$ triangular matrix, that is, $\mathbf{R}$ is zero below the main diagonal. Writing $\mathbf{Q}$ and $\mathbf{R}$ as follows,

$$\mathbf{Q} = [\mathbf{Q}_1 \,|\, \mathbf{Q}_2]$$

where $\mathbf{Q}_1$ is the first P columns and $\mathbf{Q}_2$ is the last $N - P$ columns of $\mathbf{Q}$, and

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix}$$

with $\mathbf{R}_1$ a $P \times P$ upper triangular matrix with all elements greater than zero and $\mathbf{R}_2$ is a $(N - P) \times P$ lower matrix of zeros. Also,

$$\mathbf{Q}^{\mathrm{T}} = \begin{bmatrix} \mathbf{Q}_1^{\mathrm{T}} \\ \mathbf{Q}_2^{\mathrm{T}} \end{bmatrix}$$

where $\mathbf{Q}_1^{\mathrm{T}}$ and $\mathbf{Q}_2^{\mathrm{T}}$ are of dimension $P \times N$ and $(N - P) \times N$ respectively. Therefore,

$$M_0 = \mathbf{QR} = \mathbf{Q}_1\mathbf{R}_1 \qquad (17)$$

Geometrically, the columns of $\mathbf{Q}$ define an orthonormal, or orthogonal, basis for the response space with the property that the P columns span the expectation plane. Projection onto the expectation plane is then very easy if the projection is in the coordinate system given by $\mathbf{Q}$ ([12]).

Next is transformation of the response vector, which is

$$g = \mathbf{Q}^{\mathrm{T}} z_0 \qquad (18)$$

with components

$$g_1 = \mathbf{Q}_1^{\mathrm{T}} z_0 \qquad (19)$$

and

$$g_2 = \mathbf{Q}_2^{\mathrm{T}} z_0. \qquad (20)$$

The projection of g onto the expectation plane is then simply

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

in the $\mathbf{Q}$ coordinates and

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

in the original coordinates. So,

$$\delta_0 = \mathbf{R}_1^{-1} g_1$$

this implies

$$\mathbf{R}_1 \delta_0 = g_1 \qquad (22)$$

Equation (22) can now be easily estimated using backward solving ([13]). 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 this better parameter value $\theta_1^* = \theta_0^* + \delta_0$ and perform another iteration by calculating new residuals $z_1 = \mathrm{T} - \eta(\theta_1^*)$, a new derivative matrix M_0, and a new increase. This process is repeated until convergence is obtained, that is, until the increment is so small that there is no useful change in the elements of the parameter vector ([12]).

It is expected that the new residual sum of square 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 such that ([12])

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

where λ is chosen to ensure that the new residual sum of squares is less than the initial estimate. A common method as suggested by [12] is to start with $\lambda = 1$ and halve it until it is satisfied that the new residual sum of squares is less than the initial estimate.

In actual practice the GLS is not possible to be implemented, however, because the variance-covariance matrix, **V**, is unknown. Therefore, an estimated **V** is obtained and substituted into equation (3) and the term Estimated Generalized Least Square (EGLS) is used. There are many methods for estimating the variance components to substitute for **V** in equation (3). In this research work the procedure for residual maximum likelihood estimation (REML) technique is presented. The next subsection presents the technique estimation procedure.

- **Variance Component Estimation Via MLE**

Residual maximum likelihood estimation (REML) procedure does not involve $\hat{\theta}^*$ in the estimation of the variance component. The likelihood function 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 $\mathbf{k}'y$ is formed whereby $\mathbf{k}$ is chosen so that $\mathbf{k}'D_0 = 0$ which falls in the linear approximation to the error space. $\mathbf{k}'y$ is called the error contrasts ([14]), 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 applying maximum likelihood to $\mathbf{K}'y$, the log likelihood function of $\mathbf{K}'y$, is

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

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 Γ equation (23) becomes,

$$\Gamma(\boldsymbol{\Theta}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\mathbf{K}'\mathbf{V}\mathbf{K}\right| - \frac{1}{2}z_0'\mathbf{K}\left(\mathbf{K}'\mathbf{V}\mathbf{K}\right)^{-1}\mathbf{K}'z_0 \quad (24)$$

(24)

By iterative maximization of (24) at $(h + 1)^{st}$ iterations, the system equation yields,

$$\left\langle tr\left(\left[\mathbf{V}_{(h)}^{-1}\left(1 - D_0\left(D_0'\mathbf{V}_{(h)}^{-1}D_0\right)^{-1}D_0'\mathbf{V}_{(h)}^{-1}\right)\right] \times \mathbf{V}_i\left[\mathbf{V}_{(h)}^{-1}\left(1 - D_0\left(D_0'\mathbf{V}_{(h)}^{-1}D_0\right)^{-1}D_0'\mathbf{V}_{(h)}^{-1}\right)\right]\mathbf{V}_j\right)\right\rangle \times \left\langle\left(\sigma^2_{j(h+1)}\right)\right\rangle - \left\langle\left(y'\left[\mathbf{V}_{(h)}^{-1}\left(1 - D_0\left(D_0'\mathbf{V}_{(h)}^{-1}D_0\right)^{-1}D_0'\mathbf{V}_{(h)}^{-1}\right)\right]\right)\mathbf{V}_i \times \left(\left[\mathbf{V}_{(h)}^{-1}\left(1 - D_0\left(D_0'\mathbf{V}_{(h)}^{-1}D_0\right)^{-1}D_0'\mathbf{V}_{(h)}^{-1}\right)\right]y\right)\right\rangle = 0. \quad (25)$$

Let $\mathbf{Q}_{(h)} = \mathbf{V}_{(h)}^{-1}\left(1 - D_0\left(D_0'\mathbf{V}_{(h)}^{-1}D_0\right)^{-1}D_0'\mathbf{V}_{(h)}^{-1}\right)$

then equation (25) becomes

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

$$\left\langle\left(\hat{\sigma}^2_{j(h+1)}\right)\right\rangle = \left\langle tr\left(\hat{\mathbf{Q}}_{(h)}\hat{\mathbf{V}}_i\,\hat{\mathbf{Q}}_{(h)}\hat{\mathbf{V}}_j\right)\right\rangle^{-1} \times \left\langle\left(y'\hat{\mathbf{Q}}_{(h)}\hat{\mathbf{V}}_i\,\hat{\mathbf{Q}}_{(h)}y\right)\right\rangle \quad (26)$$

The solutions to the equations may turn out to be negative when further iteration does not improve the log likelihood. In such a case, the negative value is reset to zero before the next iteration.

III. RESULTS AND DISCUSSION

This paper presents the procedure and steps in estimating the parameters for a split-plot design model where the mean part of the model can be any nonlinear function and the variance components ($\sigma^{2\prime} = \sigma^2_{WP}, \sigma^2_{SP}$) of the model are estimated via residual maximum likelihood estimation (REML) technique. This was achieved by minimizing the objective function,

$\left(y - f(X,\theta)\right)'\mathbf{V}^{-1}\left(y - f(X,\theta)\right)$ where the estimates of $\hat{\theta}^*$ and $\sigma^{2\prime} = \sigma^2_{WP}, \sigma^2_{SP}$ are iteratively obtained at $(h + 1)^{st}$ iteration by substituting a prior estimate of $\sigma^{2\prime}$ to the estimating equation till convergence occurs.

This was done by transforming the generalized least square (GLS) nonlinear split-plot design model into an ordinary least square (OLS) nonlinear split-plot design model using iterative Gauss-Newton via Taylor Series expansion procedure approximated at first order. Householder **QR** decomposition technique was introduced into the estimation system to achieve stability in the estimates. However, in estimating the variance components, $\sigma^{2'} = \sigma_{WP}^2, \sigma_{SP}^2$, $\hat{\theta}^*$ is not involved. The likelihood function was based on vectors in the error space where the linear approximation of the residuals is used $z_0 = D_0\delta + \varepsilon$ and these vectors are not dependent on the values of the fixed effect estimates.

IV. 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 functional form of the mean part of a 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 split-plot model where the mean part of the model can be any nonlinear function and the variance components ($\sigma^{2'} = \sigma_{WP}^2, \sigma_{SP}^2$) of the model are estimated via residual (restricted) maximum likelihood estimation (REML) technique. This is achieved by minimizing the objective function, $(y - f(X,\theta))' \mathbf{V}^{-1}(y - f(X,\theta))$ where the estimates of $\hat{\theta}^*$ and $\sigma^{2'} = \sigma_{WP}^2, \sigma_{SP}^2$ are iteratively obtained at $(h + 1)^{st}$ iteration by substituting a prior estimate of $\sigma^{2'}$ to the estimating equation till convergence is achieved. To achieve these iterative procedures for estimating the parameters of the nonlinear split-plot models, statistical software such as the %NLINMIX SAS macro can be used to handle all computations.

Acknowledgment

The authors are grateful to anonymous reviewers for their valuable comments on the original draft of this manuscript.

References

[1] Jones, B. and Goos, P. (2012). I-optimal versus D-optimal Split-Plot response surface designs. *Research Paper*, D/2012/1169/002. www.ua.ac.be/tew and www.repec.org

[2] Jones, B. and Nachtsheim, C. J. (2009). Split-plot Designs: What, why, and How. *Journal of Quality Technology*, 41(4): 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] Montgomery, D. C. (2008). *Design and Analysis of Experiments*, 7th ed. New York, NY: John Wiley & Sons.

[5] Letsinger, J. D.; Myers, R. H.; and Lentner, M. (1996). Response Surface Methods for Bi-Randomization Structures. *Journal of Quality Technology* 28: 381–397.

[6] 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: 72–79.

[7] 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: 487–495.

[8] Gumpertz, M. L. and Rawlings, J. O. (1992). Nonlinear Regression with Variance Components: Modeling Effects of Ozone on Crop Yield. *Crop Science*, 32: 219 – 224.

[9] 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): 455 – 468.

[10] Vining, G. G., Kowalski, S. M. and Montgomery, D. C. (2005). Response surface designs within a split-plot structure. *Journal Quality Technology*, 37: 115–129.

[11] 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. http://digitalcommons.unl.edu/agronomyfacpub/407

[12] Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

[13] Klotz, J. (2004) *A computational approach to statistics*. University of Wisconsin, Madison:USA. https://www.mimuw.edu.pl/~pokar/StatystykaI/Literatura/KlotzBook.pdf

[14] Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems. *Journal of the American Statistical Association*, 72: 320–338