

Nonlinear Split-Plot Design Model in Parameters Estimation using Estimated Generalized Least Square - Maximum Likelihood Estimation

Ikwoche John David¹, Osebekwin Ebenezer Asiribo²,
and Hussaini Garba Dikko³

^{1,2,3}Department of Statistics, Faculty of Physical Sciences, Ahmadu Bello University
Zaria-Nigeria

¹davidij@fuwukari.edu.ng; ²asiribo@yahoo.com; ³hgdikko@yahoo.com

Received: 24th April 2018/ **Revised:** 24th September 2018/ **Accepted:** 8th October 2018

How to Cite: David, I. J., Asiribo, O. E., & Dikko, H. G. (2018). Nonlinear Split-Plot Design Model in Parameters Estimation using Estimated Generalized Least Square - Maximum Likelihood Estimation. *ComTech: Computer, Mathematics and Engineering Applications*, 9(2), 65-71. <https://doi.org/10.21512/comtech.v9i2.4703>

Abstract - This research aimed to provide a theoretical framework for intrinsically nonlinear models with two additive error terms. To achieve this, an iterative Gauss-Newton via Taylor Series expansion procedures for Estimated Generalized Least Square (EGLS) technique was adopted. This technique was applied in estimating the parameters of an intrinsically nonlinear split-plot design model where the variance components were unknown. The unknown variance components were estimated via Maximum Likelihood Estimation (MLE) method. To achieve the numerical stability in the iterative process of estimating the parameters, Householder QR decomposition was used. The results show that EGLS method presented in this research is available and applicable to estimate linear fixed, random, and mixed-effect models. However, in practical situations, the functional form of the mean in the model is often nonlinear due to the dynamics involved in the system process.

Keywords: Split-plot design, parameters estimation, Estimated Generalized Least Square, Maximum Likelihood Estimation

I. INTRODUCTION

A Split-Plot (SP) experiment is simply a blocked experiment. The experimental units are made of blocks for a compartment of the factors. Thus, there are two levels of experimental units. The blocks are usually called the Whole Plot (WP), and the experimental units within blocks are called split units, subplots, or SP. This implies two levels of randomization on the two levels of experimental units. One randomization determines the assignment of block-level treatments to WP. Then, as always in a blocked experiment, randomization of treatments to SP experimental units occurs within each block or WP (Montgomery, 2008; Jones & Nachtsheim, 2009). Hence, those are designed experiments

that can be viewed as two combined or overlaid experiments on each other. In addition, Hinkelmann and Kempthorne (2008) put it as a superimposition of two similar or different forms of designs. Many researches have been done in estimating the parameters of the linear SP design, response surface, and optimal models respectively (Jones & Nachtsheim, 2009; Jones & Goos, 2012; Lu & Anderson-Cook, 2014; Lu, Robinson, & Anderson-Cook, 2014; Lu, Anderson-Cook, & Robinson, 2011).

SP design in an experiment and its analysis are not new. It is applied in various fields outside agriculture from which it is originated from (Lu & Anderson-Cook, 2012; Lu, Anderson-Cook, & Robinson, 2012; Wang, Kowalski, & Vining, 2009; Myers, Montgomery, & Anderson-Cook, 2009). Jones and Nachtsheim (2009) identified that all experiments in industries were SP experimental design. This is because it contains two types of factors, which are Hard-To-Change (HTC) factor and Easy-To-Change (ETC) factor. The HTC is the WP factor, and the ETC factor is the SP and subplot factor.

Milliken and Johnson (2009) stated that the SP experimental design was a multilevel design because of its hierarchal design structure. It has two units of different experimental sizes that the large experimental unit size is the WP and the small experimental unit size is the subplot or SP. Recently, Kulahci and Menon (2017) applied trellis plots to visualize multivariate data by allowing the condition during the preliminary data analysis stage of the SP experimental design data. Moreover, SP experimental design is applied to the equipment test to study three different explosive powers influenced by four different intensifiers and four different steel balls (Gao, Yang, & Shi, 2017).

Moreover, in the line of cost implication for the design, it has been found that SP experimental design is cheap regarding cost compared to other experimental design types. Despite its cheapness, it retains its statistical efficiency and adequacy compared to other randomized design of experiments. Anderson (2016) compared the

power efficiency of five factors in SP design (one HTC factor at three levels and each four ETC factor at two levels) and randomized design. He found out that the SP design had greater power efficiency for all the subplot factors compared to the randomized design. However, the SP design showed great weakness in power efficiency for the WP factor compared to the completely randomized design. The interaction between the HTC and ETC factors showed great power efficiency compared to the completely randomized design. He identified this interaction power efficiency for the SP experimental design as a bonus and a viable alternative to run an experiment as a completely randomized design when there was the presence of HTC factor like temperature. The same findings were seen in the research of Anderson and Whitcomb (2014). They employed power to the right size design of the experiment.

According to Anderson-Cook, Borror, and Montgomery (2009), there should be additional consideration for different costs associated with making changes at the WP and sub-plot levels. It is observed that for a completely randomized experiment, the total number of observations can be a practical substitute for the cost. Consequently, many of the optimal design criteria use the total number of design points as the penalty to balance the decrease in prediction variance for larger designs. Furthermore, in some SP design cases, it may be prohibitively expensive to do the equipment set-up for each of the WP. Therefore, the total number of WP in the designed experiment can be the only contributor to cost in these situations. In other situations, WP may be slightly more expensive than a sub-plot. However, the total cost of the experiment is more logically to be a weighted average of the number of WP and the number of sub-plots runs.

Estimating the variance component of the nonlinear model also involves the use of different methods. The methods can be the Maximum Likelihood Estimation (MLE), Residual Maximum Likelihood Estimation (RMLE), Quasi-maximum Likelihood Estimation (QMLE), Modified Maximum Likelihood Estimation (MMLE), Analysis of Variance (ANOVA), Minimum-Norm Quadratic Unbiased Estimation (MINQUE), Minimum Variance Quadratic Unbiased Estimation (MIVQUE), and others. The MINQUE and MIVQUE methods are developed to find unbiased quadratic estimators, which are invariant and minimize some matrix norm.

Rasch and Masata (2006) stated that unfortunately, the solution in the most interesting cases depended on the unknown variance components. If they were replaced by estimates from the data, the solution would be neither unbiased nor quadratic any longer. However, they only identified the MINQUE and MIVQUE in variance components estimation. The same thing applied to other variance component estimation methods because in almost all cases of modeling with variance components, the population variance components were unknown. Therefore, since the estimated values obtained from the design data, which could have outliers, replacement of missing values, and others. The solution may not be unbiased for all other estimation methods.

Weerakkody and Johnson (1992) presented a two-step residual-based approach for estimating the WP and subplot error variances separately. However, as identified by Ikeda, Matsuura, and Suzuki (2014), the estimator of the WP error variance in approach by Weerakkody and Johnson (1992) can be obtained only for the case of $a > p$ ($p = 1 + p_1 + p_2$), where a is the number of runs in each WP unit.

Then, p_1 and p_2 are levels of the WP and subplot effects. This condition is strict and not practical in many situations because it is only suitable for balanced designs.

Hasegawa, Ikeda, Matsuura, and Suzuki (2010) proposed a different estimator for the WP error variance having more practical conditions than the one suggested by Weerakkody and Johnson (1992). Nevertheless, both of these approaches for estimating the two error variances can only be used in balanced designs, and their approaches are not compared to other methods such as ML, RMLE, ANOVA, and others. They cannot be used in unbalanced designs, which are frequently employed to reduce the number of experiments.

Ikeda, Matsuura, and Suzuki (2014) did a modification of the two-step residual-based method proposed by Hasegawa *et al.* (2010) to make it available for application in balanced and unbalanced designs. Then, they compared their method with RMLE only. They concluded that their method could be an alternative to RMLE based on their simulation results. Their alternative method was not a better estimation method because it could perform poorly under a different simulation scenario and they did not compare it with other estimation technique like MLE. However, the methods introduced by Weerakkody and Johnson (1992), Hasegawa *et al.* (2010) and Ikeda, Matsuura, and Suzuki (2014) were only implemented for linear balanced and unbalanced SP design models. Nonlinear modeling of SP 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. Gumpertz and Rawlings (1992) stated that when the objective of fitting a nonlinear function to data from SP experimental designs, a nonlinear model with variance components (WP variance, σ^2_{γ} , SP variance, and σ^2_{ϵ}) was appropriate.

Nonlinear modeling of SP data, which has two error terms, is a special modeling case of a nonlinear model with variance components. The reason is that the model contains a nonlinear function for the mean part, $g(X, \theta)$, and the random effects (WP and subplot errors) are added to the mean function. Normal nonlinear regression modeling assumes that all the observations in the data set are uncorrelated and there is only one source of random error. If they are used to fit models with over one random error term, they give standard errors for the incorrect parameter estimates and other important quantities. Therefore, if a standard nonlinear regression program is used to analyze SP design data, the single variance estimate like Mean Squared Error (MSE) will be a compromise between the two error variances, MSE_a and MSE_b, from the SP analysis of variance (Gumpertz & Rawlings, 1992; Knezevic, Evans, Blankenship, Van Acker, & Lindquist, 2002; Blankenship, Stroup, Evans, & Knezevic, 2003).

In this research, a theoretical presentation of an iterative Gauss-Newton via Taylor series expansion procedures for Estimated Generalized Least Square (EGLS) technique for estimating intrinsically nonlinear SPD model parameters will be done. The variance components are unknown, and they are estimated via MLE method. Householder QR decomposition technique is adopted to achieve numerical stability during the iterative process of estimating the parameters in the model.

II. METHODS

The methodological approach for this research is

completely theoretical. No qualitative or quantitative data is used for estimating the intrinsically nonlinear SP design model. In achieving the objective, the use of EGLS method of parameter estimation is adopted. The EGLS is adopted due to the assumption that the SP model variance-covariance matrix is unknown.

The variance-covariance matrix is estimated using MLE. The estimated variance-covariance matrix parameter. Then, V is substituted into the model to estimate the SP model parameters. The model parameters are estimated via an iterative Gauss-Newton via Taylor Series expansion procedure. This iterative system has to be used since the partially derived system of equation does not exist in a closed form. However, to achieve an asymptotically numerically stable parameter estimate, QR decomposition by Householder (1958) is implemented.

III. RESULTS AND DISCUSSIONS

The nonlinear SP model, which has Whole Plot Error (WPE) and Sub Plot Error (SPE) are special cases of the nonlinear model with random effects. It is also called the nonlinear model with variance components. The formulated model and assumptions are given as follows.

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

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

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

Where, y_{ijk} is the response variable and $i = 1, \dots$. Moreover, S is as replicates or blocks and $j = 1, \dots$. Then, a levels the WP factor A and $k = 1, \dots$. Next, b levels the SP factor B, w_{ijk} is the WP error and ε_{ijkl} is the SP error. Meanwhile, $f(x_{ijkl}, \theta)$ is the nonlinear function for the mean.

There are three assumptions in this model. First, it is assumed that the WP and SP errors are random effects. Moreover, it is assumed that $w_{ijk} \stackrel{i.i.d.}{\sim} N(0, \sigma_{wp}^2)$ and

$$\varepsilon_{ijkl} \stackrel{i.i.d.}{\sim} N(0, \sigma_{sp}^2)$$

Second, it lets $\hat{\theta}$ to be the model parameter estimate of θ . It follows an asymptotic normal distribution with mean and variance $\sigma^2(\mathbf{F}'\mathbf{F})^{-1}$, where F is the $n \times p$ matrix with elements $(\partial f(x_{ijkl}, \theta) / \partial \theta')$. It 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 $(\partial f(x_{ijkl}, \theta) / \partial \theta')$ and V_x is the covariance matrix of the response vector.

Third, if the mean function, $f(x_{ijkl}, \theta)$ parameters is p and the number of random effect is r , the number of observations in the data set, n , must be at least $p + r + 1$. It estimates all of the parameters. This implies that it is $n \geq p + r + 1$.

In EGLS estimation method, the covariance matrix of y is known as the Generalized Least Square (GLS) estimator, $\hat{\theta}_{GLS}$. It is to minimize the objective function With respect to θ . (Gumpertz & Rawlings, 1992).

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

Then, V is a known positive definite (non-singular) covariance matrix, which arises from the model as follows.

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

It shows $E(w_{ijk}) = 0$, $Cov(w_{ijk}) = \sigma_w^2 \mathbf{I}_N$, $E(\varepsilon_{ijkl}) = 0$ and $Cov(\varepsilon_{ijkl}) = \sigma_\varepsilon^2 \mathbf{I}_N$.

It lets the variance-covariance matrix of the observations var (y) be written as follows.

$$\begin{aligned} \mathbf{V} &= \sigma_w^2 \mathbf{I}_N + \sigma_\varepsilon^2 \mathbf{I}_N \\ &= \sigma^2 \mathbf{K} \end{aligned} \quad (5)$$

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

$$\mathbf{V} = \mathbf{P}\mathbf{P}' \quad (6)$$

Then, by multiplying model (4) by \mathbf{P}^{-1} on both sides, it yields

$$\begin{aligned} \mathbf{P}^{-1}y &= \mathbf{P}^{-1}f(x_{ijkl}, \theta) + \mathbf{P}^{-1}(w_{ijk}) \\ &\quad + \mathbf{P}^{-1}(\varepsilon_{ijkl}) \end{aligned} \quad (7)$$

$$\begin{aligned} &Cov(\mathbf{P}^{-1}\varepsilon_{ijkl}) + Cov(\mathbf{P}^{-1}w_{ijk}) \\ &= \mathbf{P}^{-1}Cov(\mathbf{P}^{-1}\varepsilon_{ijkl})(\mathbf{P}^{-1})' + \mathbf{P}^{-1}Cov(\mathbf{P}^{-1}w_{ijk})(\mathbf{P}^{-1})' \\ &= \mathbf{P}^{-1}(\mathbf{P}^{-1})' [Cov(\varepsilon_{ijkl}) + Cov(w_{ijk})] \\ &= \mathbf{P}^{-1}\sigma^2\mathbf{K}(\mathbf{P}^{-1})' \\ &= \sigma^2\mathbf{P}^{-1}\mathbf{P}\mathbf{P}'(\mathbf{P}^{-1})' \\ &= \sigma^2\mathbf{I} \end{aligned} \quad (8)$$

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

$$\mathbf{T} = \mathbf{M}(x_{ijkl}, \theta^*) + \mathbf{E} \quad (9)$$

$E(\Phi)$ is 0 and $V(\mathbf{E}) = \sigma^2\mathbf{I}$. Thus, the GLS model has been transformed into an Ordinary Least Square (OLS) model. Hence, equation (9) is solved using the OLS technique. Taking the summation of both sides of equation (9) and square the researches have

$$\begin{aligned} &\sum_i^S \sum_j^a \sum_k^b \sum_l^c \mathbf{E}_{ijkl}^2 \\ &= \sum_i^S \sum_j^a \sum_k^b \sum_l^c [\mathbf{T}_{ijkl} - \mathbf{M}(x_{ijkl}, \theta^*)]^2 \end{aligned} \quad (10)$$

$$\begin{aligned} L(\theta^*) &= \sum_i^S \sum_j^a \sum_k^b \sum_l^c \mathbf{E}_{ijkl}^2 \\ &= \sum_i^S \sum_j^a \sum_k^b \sum_l^c [\mathbf{T}_{ijkl} - \mathbf{M}(x_{ijkl}, \theta^*)]^2 \end{aligned} \quad (11)$$

By minimizing $L(\theta^*)$ w.r.t. θ^* and equating to zero the researches have the equation as follows:

$$\frac{\partial L(\theta^*)}{\partial \theta_i^*} = \sum_i^s \sum_j^a \sum_k^b \sum_l^c [\mathbf{T}_{ijkl} - \mathbf{M}(x_{ijkl}, \theta^*)] \times \left[\frac{\partial \mathbf{M}(x_{ijkl}, \theta^*)}{\partial \theta_i^*} \right]_{\theta^* - \delta^*} = 0 \quad (12)$$

At this point, equation (12) has no closed form. Hence, it will be solved iteratively using the Gauss-Newton method via Taylor series expansion of $\mathbf{M}(x_{ijkl}, \theta^*)$ at first order of

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!} f''(a) + \dots + \frac{(x-a)^p}{p!} f^{(p)}(a) + R_{p+1} \quad (13)$$

It shows that $f'(a) = \frac{df(x)}{dx}$ around $x = a$. Then, is the remainder term which is reasonably small if p is sufficiently large. Therefore, the researches have the equation as follows:

$$\begin{aligned} & \mathbf{M}(x_{ijkl}, \theta^*) \\ &= \mathbf{M}(x_{ijkl}, \theta_0^*) + (\theta_1^* - \theta_{10}^*) \frac{\partial \mathbf{M}(x_{ijkl}, \theta^*)}{\partial \theta_1^*} \Big|_{\theta^* - \theta_0^*} \\ & \quad + (\theta_2^* - \theta_{20}^*) \frac{\partial \mathbf{M}(x_{ijkl}, \theta^*)}{\partial \theta_2^*} \Big|_{\theta^* - \theta_0^*} + \dots \\ & \quad + (\theta_p^* - \theta_{p0}^*) \frac{\partial \mathbf{M}(x_{ijkl}, \theta^*)}{\partial \theta_p^*} \Big|_{\theta^* - \theta_0^*} \end{aligned} \quad (14)$$

Let $\mathbf{M}(x_{ijkl}, \theta^*) = \eta(\theta^*)$

And $d_{ijkl} = \frac{\partial \mathbf{M}(x_{ijkl}, \theta^*)}{\partial \theta_p^*} \Big|_{\theta^* - \theta_0^*}$ for all N cases and

$\delta = \theta^* - \theta_0^*$, then equation (14) becomes

$$\eta(\theta^*) = \eta(\theta_0^*) + D_0 \delta \quad (15)$$

It shows that D_0 is the $N \times P$ derivative matrix with elements of $\{d_{ijkl \times p}\}$. This is equivalent to approximating the residuals for the model that is $E(\theta^*) = \mathbf{T} - \eta(\theta^*)$ by

$$\begin{aligned} E(\theta^*) &= \mathbf{T} - [\eta(\theta_0^*) + D_0 \delta] \\ &= \mathbf{T} - \eta(\theta_0^*) - D_0 \delta \\ &= z_0 - D_0 \delta \end{aligned} \quad (16)$$

It shows that $z_0 = \mathbf{T} - \eta(\theta_0^*)$ and $\delta = \theta^* - \theta_0^*$. Next, the researchers apply the QR decomposition by Householder (1958) to (16). 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, A can be factored as $A = QR$. It shows that Q is an $m \times n$ matrix whose column vectors forms an orthonormal basis for the column space of A and R as an $n \times n$ invertible upper triangular matrix.

Proof: Let $m \times n$ matrix have columns w_1, w_2, \dots, w_n m -vectors. In addition, it lets $q_1, q_2, \dots, q_n, q_{n+1}, \dots, q_m$ be orthonormal m -vectors such as:

$$\|q_i\| = 1, q_i^T q_j = 0 \text{ if } i \neq j \quad (17)$$

Then Q is $m \times n$ with orthonormal columns, $Q^T Q = I$. If A is a square matrix ($m = n$), Q is orthogonal that is $Q^T Q = Q Q^T = I$. Hence, q_i is orthogonal to w_1, w_2, \dots, w_n . Therefore, it shows:

$$\begin{aligned} w_1 &= (w_1 \cdot q_1)q_1 \\ w_2 &= (w_2 \cdot q_1)q_1 + (w_2 \cdot q_2)q_2 \\ &\dots \\ w_k &= (w_n \cdot q_1)q_1 + (w_n \cdot q_2)q_2 + \dots + (w_n \cdot q_n)q_n \end{aligned} \quad (18)$$

This implies that $A = QR$

$$\begin{aligned} [w_1 \ w_2 \ \dots \ w_n] &= [q_1 \ q_2 \ \dots \ q_n] \\ &\times \begin{bmatrix} (w_1 \cdot q_1) & (w_1 \cdot q_2) & \dots & (w_1 \cdot q_n) \\ 0 & (w_2 \cdot q_2) & \dots & (w_2 \cdot q_n) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (w_k \cdot q_k) \end{bmatrix} \end{aligned} \quad (19)$$

It lets $A = [w_1 \ w_2 \ \dots \ w_n]$ and $R = w \cdot q$. Therefore, equation (19) is written as

$$\begin{aligned} A &= [q_1 \ q_2 \ \dots \ q_n] \\ &\times \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n} \\ 0 & R_{22} & \dots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_{nn} \end{bmatrix} \end{aligned} \quad (20)$$

Equation (20) shows that R is $n \times n$. It is upper triangular with nonzero diagonal elements and R is nonsingular. Meanwhile, diagonal elements are nonzero.

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

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

$$\begin{aligned} \hat{x} &= (A^T A)^{-1} A^T b = (R^T Q^T Q R)^{-1} R^T Q^T b \\ R^T Q^T Q R &= R^T Q^T b \quad (Q^T Q = I) \\ R^T R &= R^T Q^T b \\ \hat{x} &= R^{-1} Q^T b. \end{aligned} \quad (21)$$

Based on the two stated and proved theorems on QR-decomposition, the decomposition of M_0 is presented as $M_0 = \mathbf{QR}$.

It shows \mathbf{Q} as an $N \times N$ orthogonal matrix that is $\mathbf{Q}^T \mathbf{Q} = \mathbf{Q} \mathbf{Q}^T = \mathbf{I}$. Then, \mathbf{R} is an $N \times P$ triangular matrix that \mathbf{R} is zero below the main diagonal. The researchers write \mathbf{Q} and \mathbf{R} as follows:

$$\mathbf{Q} = [\mathbf{Q}_1 | \mathbf{Q}_2] \quad (22)$$

It means \mathbf{Q}_1 is the first P columns and \mathbf{Q}_2 is the last $N - P$ columns of \mathbf{Q} . Then, it shows:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix} \quad (23)$$

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. Moreover, it shows:

$$\mathbf{Q}^T = \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \quad (24)$$

It means that \mathbf{Q}_1^T and \mathbf{Q}_2^T are of dimension $P \times N$ and $(N - P) \times N$ respectively. Therefore, it is $M_0 = \mathbf{QR} = \mathbf{Q}_1 \mathbf{R}_1$ geometrically. Then, \mathbf{Q} columns define an orthogonal or orthonormal basis for the response space with the property. Moreover, P columns cover the expectation plane. Projection onto the expectation plane is very easy if the projection is in the coordinate system given by \mathbf{Q} (Klotz, 2006).

Next is the transformation of the response vector, which is

$$\mathbf{g} = \mathbf{Q}^T \mathbf{z}_0 \quad (25)$$

with components of

$$\mathbf{g}_1 = \mathbf{Q}_1^T \mathbf{z}_0 \quad (26)$$

and

$$\mathbf{g}_2 = \mathbf{Q}_2^T \mathbf{z}_0 \quad (27)$$

The projection of \mathbf{g} onto the expectation plane is therefore simply given as $\begin{bmatrix} \mathbf{g}_1 \\ 0 \end{bmatrix}$ in the \mathbf{Q} coordinates and

$\hat{\eta}_1 = \mathbf{Q} \begin{bmatrix} \mathbf{g}_1 \\ 0 \end{bmatrix} = \mathbf{Q}_1 \mathbf{g}_1$ with as n the original coordinates. So, it is $\delta_0 = \mathbf{R}_1^{-1} \mathbf{g}_1$. This implies that $\mathbf{R}_1 \delta_0 = \mathbf{g}_1$ can now be easily estimated using backward solving (Klotz, 2006).

$$\hat{\eta}_1 = \eta(\theta_1^*) = \eta(\theta_0^* + \delta_0) \quad (28)$$

In equation (28), it should be closer to y than $\eta(\theta_0^*)$ and move to a better parameter value of $\theta_1^* = \theta_0^* + \delta_0$. Another iteration is performed by calculating new residuals $\mathbf{z}_1 = \mathbf{T} - \eta(\theta_1^*)$, a new derivative matrix M_0 , and a new increase. This process is continuously repeated until convergence is obtained, and until the increment becomes so small. There is no useful change in the elements of the parameter vector.

It is expected that the new residual sum of square should be less than the initial estimate, but if it is otherwise, a small step in the direction δ_0 is introduced. A step factor

λ is introduced such as $\theta_1^* = \theta_0^* + \lambda \delta_0$ where λ is chosen to ensure that the new residual sum of squares is less than the initial estimate. A common method starts with $\lambda = 1$ and halves it until it is satisfied that the new residual sum of squares is less than the initial estimate. In actual practice the GLS, it is not possible to be implemented. This is because the variance-covariance matrix, \mathbf{V} , is unknown. Therefore, an estimated \mathbf{V} is obtained and substituted into equation (3), and the term of EGLS is used. There are many methods for estimating the variance components to substitute for \mathbf{V} in equation (3). In this research, the procedure for MLE technique is presented.

The MLE procedure for variance components estimation from nonlinear SP design model is presented. The MLE method used for this research is an iterative method for estimating the variance components. The method follows the maximum likelihood algorithm for the linear model with variance components by Hemmerle and Hartley (1973) and the procedure presented by Gumpertz and Rawlings (1992). The mean function of $\mathbf{M}(\mathbf{T}_{ijk}, \theta^*)$ is

the first approximated through a first-order in Taylor series centered at θ_0^* as shown in equation (15). Therefore, the log-likelihood function is given as:

$$\ln L(\Delta) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{V}| - \frac{1}{2} (y - f(X, \theta))^' \times \mathbf{V}^{-1} (y - f(X, \theta)) \quad (29)$$

It means $\Delta = (\theta', \sigma^2)' = (\sigma_{WP}^2, \sigma_{SP}^2)'$ is approximated by

the surface and lets $\ln L$ to Γ become,

$$\Gamma(\Delta) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{V}| - \frac{1}{2} (z_0 - D_0 \delta)' \times \mathbf{V}^{-1} (z_0 - D_0 \delta) \quad (30)$$

It shows that $z_0 = \mathbf{T} - \eta(\theta_0^*)$, $\delta = \theta^* - \theta_0^*$,

$\mathbf{V} = \sigma^2 \mathbf{K} = \sum_{i=0}^1 \sigma_i^2 \mathbf{K}_i \mathbf{K}_i'$, and $\mathbf{V} \mathbf{V}^{-1} = \sum_{i=0}^1 \sigma_i^2 \mathbf{V}^{-1} \mathbf{K}_i \mathbf{K}_i'$. Then,

the gradient is given by:

$$\frac{\partial \Gamma(\Delta)}{\partial (\theta^*)'} = (z_0 - D_0 \delta)' \mathbf{V}^{-1} D_0 \quad (31)$$

$$\frac{\partial \Gamma(\Delta)}{\partial \sigma_i^2} = -\frac{1}{2} \text{tr}(\mathbf{K}_i \mathbf{K}_i' \mathbf{V}^{-1}) + \frac{1}{2} (z_0 - D_0 \delta)' \times \mathbf{V}^{-1} \mathbf{K}_i \mathbf{K}_i' \times (z_0 - D_0 \delta) \quad (32)$$

Multiplying the partial derivative first term by $\mathbf{V} \mathbf{V}^{-1}$ and equating to zero gives the estimating equations as follows:

$$D_0' \mathbf{V}^{-1} D_0 \delta = D_0' \mathbf{V}^{-1} z_0 \quad (33)$$

$$\left\langle \text{tr}(\mathbf{K}_j \mathbf{K}_j' \mathbf{V}^{-1} \mathbf{K}_i \mathbf{K}_i' \mathbf{V}^{-1}) \right\rangle \sigma^2 = \left\langle (z_0 - D_0 \delta)' \mathbf{V}^{-1} \mathbf{K}_j \mathbf{K}_j' (z_0 - D_0 \delta) \right\rangle \quad (34)$$

IV. CONCLUSIONS

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 σ^2 into equation (33) to get an updated estimate of $\hat{\theta}^*$. Then, the updated $\hat{\theta}^*$ and prior estimate of σ^2 are substituted into equation (34) to obtain updated estimates of the variance components. These two steps are iterated till convergence is achieved. Therefore, equations (33) and (34) become

$$\begin{aligned} \hat{D}'_0 \hat{V}'_{(h)} \hat{D}_0 (\hat{\theta}^* - \hat{\theta}_0^*) &= \hat{D}'_0 \hat{V}'_{(h)} z_0 \\ \hat{D}'_0 \hat{V}'_{(h)} \hat{D}_0 \hat{\theta}^* - \hat{D}'_0 \hat{V}'_{(h)} \hat{D}_0 \hat{\theta}_0^* &= \hat{D}'_0 \hat{V}'_{(h)} z_0 \\ \hat{D}'_0 \hat{V}'_{(h)} \hat{D}_0 \hat{\theta}^* &= \hat{D}'_0 \hat{V}'_{(h)} \hat{D}_0 \hat{\theta}_0^* + \hat{D}'_0 \hat{V}'_{(h)} z_0 \\ \hat{\theta}_{(h+1)}^* &= \hat{\theta}_0^* + [\hat{D}'_0 \hat{V}'_{(h)} \hat{D}_0]^{-1} \hat{D}'_0 \hat{V}'_{(h)} z_0 \end{aligned} \quad (35)$$

and

$$\begin{aligned} \hat{\sigma}_{(h+1)}^2 &= \left\{ tr(\mathbf{K}_j \mathbf{K}'_j \hat{V}'_{(h)} \mathbf{K}_i \mathbf{K}'_i \hat{V}'_{(h)}) \right\}^{-1} \\ &\times \left\{ (z_0 - \hat{D}_0 (\hat{\theta}_{(h+1)}^* - \hat{\theta}_0^*)) \hat{V}'_{(h)} \mathbf{K}_j \mathbf{K}'_j \hat{V}'_{(h)} \right. \\ &\left. \times (z_0 - \hat{D}_0 (\hat{\theta}_{(h+1)}^* - \hat{\theta}_0^*)) \right\} \end{aligned} \quad (36)$$

When the further iteration does not improve the log likelihood, the solutions to the equations may turn out to be negative. In such scenario, the negative value resets to zero before the next iteration.

This research presents the procedure and steps in estimating the parameters for a SP design model that the mean part of the model can be any nonlinear function. Then, the variance components ($\sigma^{2'} = \sigma_{WP}^2, \sigma_{SP}^2$) of the model are estimated via MLE technique. This is achieved by minimizing the objective function, $(y - f(X, \theta))' \mathbf{V}^{-1} (y - f(X, \theta))$ that 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 occurs. This is done by transforming the GLS nonlinear SPD model into an OLS nonlinear SPD model using iterative Gauss-Newton via Taylor Series expansion procedure approximated at first order. QR decomposition technique is introduced into the estimation system to achieve stability in the estimates.

The EGLS method presented in this research is available and applicable for estimating linear fixed, random and mixed-effect models. However, in practical situations, the functional form of the mean in the model is often nonlinear due to the dynamics involved in the system process. Since the parameters enter the model nonlinearly in which the model is intrinsically nonlinear, the closed form of the differentiated objective function does not exist. Hence, the use of an iterative Gauss-Newton via Taylor Series expansion procedure approximated at first order is adopted and implemented. These iterative procedures for estimating the parameters of the nonlinear SP models, statistical software such as the %NLINMIX SAS macro or SAS PROC NLMIXED can be used to handle all computations.

This research addresses the variance estimation using MLE for the EGLS. The use of other variance-covariance estimation technique can be used such as RMLE technique, the two-step technique, and others. In addition, the MLE technique can be adopted to estimate intrinsically nonlinear SPD model parameters and its variance components using an iterative scoring method. This will involve a partial derivation at first and second order of the log-likelihood function. The results show that EGLS method presented is available and applicable to estimate linear fixed, random, and mixed-effect models. However, in practical situations, the functional form of the mean in the model is often nonlinear due to the dynamics involved in the system process.

Further research can be done by estimating the variance-covariance parameters using RMLE technique. It has the ability of producing estimates that are not downwardly biased. It is peculiar to MLE technique when the sample size is small. This can as well lead to biased standard error estimates. Moreover, an expectation maximization technique can be used as an alternative iteration technique for estimating the model parameters. This research objective is strictly theoretical. An application of the technique can be further researched on or using other research techniques suggested.

REFERENCES

- Anderson, M. J. (2016). Design of Experiments (DOE): How to handle hard-to-change factors using a split plot. *Chemical Engineering*, 123(9), 12-1-12-5.
- Anderson, M. J., & Whitcomb, P. J. (2014). Employing power to 'right-size' design of experiments. *ITEA Journal*, 35(March), 40-44.
- Anderson-Cook, C. M., Borror, C. M., & Montgomery, D. C. (2009). Response surface design evaluation and comparison. *Journal of Statistical Planning and Inference*, 139(2), 629-641.
- Blankenship, E. E., Stroup, W. W., Evans, S. P., & Knezevic, S. Z. (2003). Statistical inference for calibration points in nonlinear mixed effects models. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(4), 455-468.
- Gao, H., Yang, F., & Shi, L. (2017). Split plot design and data analysis in SAS. In *AIP Conference Proceedings* (Vol. 1834, No. 1, p. 030024). AIP Publishing.
- Gumpertz, M. L., & Rawlings, J. O. (1992). Nonlinear regression with variance components: Modeling effects of ozone on crop yield. *Crop Science*, 32(1), 219-224.
- Hasegawa, Y., Ikeda, S., Matsuura, S., & 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* (pp. 235-238). Tokyo: The Japanese society for quality control.
- Hemmerle, W. J., & Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed AOV model using the W transformation. *Technometrics*, 15(4), 819-831.

- Hinkelmann, K., & Kempthorne, O. (2007). *Design and analysis of experiments, volume 1: Introduction to experimental design* (2nd ed.). John Wiley & Sons.
- 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), 342-358.
- Jones, B., & Goos, P. (2012). I-optimal versus D-optimal split-plot response surface designs. *Journal of Quality Technology*, 44(2), 85-101.
- Jones, B., & Nachtsheim, C. J. (2009). Split plot designs: What, why, and how. *Journal of Quality Technology*, 41(4), 340-361.
- Klotz, J. H. (2006). *A computational approach to statistics*. Madison: University of Wisconsin.
- Knezevic, S. Z., Evans, S. P., Blankenship, E. E., Van Acker, R. C., & Lindquist, J. L. (2002). Critical period for weed control: The concept and data analysis. *Weed Science*, 50(6), 773-786.
- Kulahci, M., & Menon, A. (2017). Trellis plots as visual aids for analyzing split plot experiments. *Quality Engineering*, 29(2), 211-225.
- Lu, L., & Anderson-Cook, C. M. (2012). Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature. *Quality Engineering*, 24(3), 404-422.
- Lu, L., & Anderson-Cook, C. M. (2014). Balancing multiple criteria incorporating cost using Pareto front optimization for split-plot designed experiments. *Quality and Reliability Engineering International*, 30(1), 37-55.
- Lu, L., Anderson-Cook, C. M., & Robinson, T. J. (2011). Optimization of designed experiments based on multiple criteria utilizing a Pareto frontier. *Technometrics*, 53(4), 353-365.
- Lu, L., Anderson-Cook, C. M., & Robinson, T. J. (2012). A case study to demonstrate a Pareto Frontier for selecting a best response surface design while simultaneously optimizing multiple criteria. *Applied Stochastic Models in Business and Industry*, 28(3), 206-221.
- Lu, L., Robinson, T. J., & Anderson-Cook, C. M. (2014). A case study to select an optimal split-plot design for a mixture-process experiment based on multiple objectives. *Quality Engineering*, 26(4), 424-439.
- Milliken, G. A., & Johnson, D. E. (2009). *Analysis of messy data: Designed experiments* (2nd ed., Vol. 1). Chapman and Hall/CRC.
- Montgomery, D. C. (2008). *Design and analysis of experiments* (7th ed.). Wiley.
- Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2009). *Response surface methodology: Process and product optimization using designed experiments* (3rd ed.). USA: John Wiley & Sons.
- Rasch, D., & Masata, O. (2006). Methods of variance component estimation. *Czech Journal of Animal Science*, 51(6), 227-235.
- Wang, L., Kowalski, S. M., & Vining, G. G. (2009). Orthogonal blocking of response surface split-plot designs. *Journal of Applied Statistics*, 36(3), 303-321.
- Weerakkody, G. J., & Johnson, D. E. (1992). Estimation of within model parameters in regression models with a nested error structure. *Journal of the American Statistical Association*, 87(419), 708-713.