

ALTERNATIVE TECHNIQUE FOR PARAMETER ESTIMATION OF NON-LINEAR SPLIT-PLOT MODEL

Eric Monfung Egomo¹; Stephen Sebastian Akpan²; John Effiong Usen¹

¹Department of Statistics, University of Cross River State, Calabar, Nigeria

²Department of Statistics, University of Calabar, Calabar, Nigeria

Correspondence email: eemsat@yahoo.com

ABSTRACT

An alternative procedure to the technique of David, Asiribo and Dikko (2018; 2019) for estimating the parameters of a non-linear split-plot model was proposed in this study. David, Asiribo and Dikko (2018; 2019) – whose study advanced the work of Marcia and John (1990) on the remodeling of split-plot models to non-linear forms, proposed the combined implementation of the EGLS (Estimated Generalized Least Squares) and then MLE (Maximum Likelihood Estimation) techniques for estimating the parameters of the nonlinear split-plot design model. Here, the techniques of EGLS and MLE were used, respectively, to estimate the parameters in the mean function and variance components of the non-linear split-plot model. However, since the latter technique as used by the authors proved inefficient for implementing further analyses especially as the estimates obtained were biased downwards, this study therefore proposed a three-step routine to bypass this pitfall. The first step in the thesis was to transform the non-linear split-plot model to a probability density function via established theorems, following which an MLE routine (which replaces the EGLS technique) is implemented on the obtained density function in order to obtain estimates of parameters in the mean function part of the non-linear split-plot model. Thereafter, and lastly, the MLE was used on a variant of the density function in order to obtain estimates of the variance components of the non-linear split-plot model in contrast to approach of MLE used prior to this study. Findings of the study have shown this proposed routine to be a better choice at bypassing the deficiency of the existing routine to an extent; and, hence, justifying the need for a more detailed comparison to be made as we have recommended.

Key words: Non-linear Split-Plot Model, EGLS, MLE

1.0 INTRODUCTION

An important requirement for making fast and reliable progress in many branches of today's society is the efficient use of experimental design methodology (Bruce, 2018; Montgomery, 2013; Naes, Aastveit & Sahni, 2007). Popular amongst the most frequently used types of experimental designs, are the factorial designs and fractional factorial designs. Experimenters use these two popular design types because they are simple and generally straightforward to generate and analyze, and can be used in a number of different situations. However, in some multifactor experiments, an

experimenter may often encounter cases in which he is unable to completely randomize the order of the experimental runs. In such cases, the basic requirements of randomization and blocking of experimental runs are usually difficult to satisfy due to either economic or practical reasons (Bradley & Christopher, 2009). Circumstances of this nature often prompt a generalization of the factorial design or fractional factorial design (Montgomery, 2013). This generalization is only feasible with the application of various restrictions imposed on the structure of the factorial design or fractional factorial design; hence, resulting in data with more complex

error structure than that of the completely randomized designs (CRD).

One of the most important and frequently used design strategies prompted by such generalization is the “split-plot design (SPD)”, which has recently received much attention in the literature of Bruce (2018), Montgomery (2013), Oehlert (2010), Bradley and Christopher (2009), Naes, Aastveit and Sahni (2007), Cox and Reid (2000), Cochran and Cox (1992), etc.

Recent advancements in the study of the SPDs have focused on the study of split-plot design models, which are defined to be a special class of linear models with two error terms being the whole plot and subplot error terms (Marcia & John, 1990; Johannes, 2010; Toth, 2011; David, Asiribo & Dikko, 2019). In particular, Marcia and John (1990) have shown the possibility of remodelling the split-plot design models in nonlinear forms containing variance components. Such forms consist of a combination of nonlinear model for the mean part of the split-plot design model – being additive error terms which describe the covariance configuration of the models. This study of Marcia and John (1990) was advanced by the independent studies of David, Asiribo and Dikko (2018; 2019) who respectively used the estimated generalized least squares (EGLS) and maximum likelihood estimation (MLE) in estimating parameters of nonlinear split-plot design models.

Sadly, nonlinear modelling of split-plot designs has indeed attracted little interest mainly in estimation of parameters, even though the available procedures for estimating parameters of the nonlinear split-plot model are similar, in some ways, to that used in estimating the parameters for nonlinear regression. In particular, only two recent attempts have been made in succession by David, Asiribo and Dikko (2018, 2019) at estimating parameters of the nonlinear split-plot design model as proposed by Gumpertz and Rawlings (1992) who continued the

study of Marcia and John (1990) on the possibility of remodelling the split-plot design models in nonlinear forms with variance components. In all attempts, the authors used ANOVA and minimum variance quadratic unbiased estimation (MIVQUE) for estimating the model variance-covariances. They also used maximum likelihood estimation (MLE) technique for estimating the variance components, but could not make further analyses with it because the MLE estimates were biased downwards. In lieu of the foregoing pitfall, this study was an attempt at proffering an alternative to the EGLS-MLE technique of parameter estimation for nonlinear SPD models as tendencies abound that certain interesting theoretical and practical results could emerge.

2.0 LITERATURE REVIEW

2.1 Definitions, relevance and applications of the split-plot design

Split-plot designs were originally developed by Sir Ronald Aylmer Fisher in 1925 for use in agricultural experiments. Split-plot designs are blocked experiments in which 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 the blocks are called subplots (or split-plots, or split-units). Corresponding to these two levels of experimental units, are two levels of randomization; the first randomization process is conducted to determine the assignment of block-level treatments to whole plots, while the other randomization which assigns treatments to split-plot experimental units occur within each block or whole plot.

The split-plot design (Montgomery, 2013), is a generalization that results from the inability to completely randomize the order of the runs in some multifactor experiments. Lukas (2014) defined the split-plot design as a special case of a factorial treatment structure

which is used when some factors are harder (or more expensive) to vary than others; hence, it is one which consists of two (2) experiments with different experimental units of different size. This may explain one of the reasons why a split-plot design was described by David, Asiribo and Dikko (2019) as a design which results from a two-stage randomization of a factorial treatment structure. Notwithstanding, Bruce (2018) earlier described it as one in which the levels of one or more experimental factors are held constant for a batch of several consecutive experimental runs called a whole plot.

Naes, Aastveit and Sahni (2007) have emphasized that the use of the factorial or fractional factorial experiments may impose extra challenges on the experimenter, both when designing the experiment, and when analyzing data (in non-replicated cases); but with split-plot designs, these practical limitations can be overcome in relation to cost. David, Asiribo and Dikko (2019) have stressed that a variety of the split-plot design forms can be used for reconstructing factorial designs, fractional factorial designs, response surface designs (having quadratic and higher-order surfaces), optimal designs (of the first and second orders), as well as the sequential and mixture designs.

Chang-Yun (2016) proposed a new method for constructing optimal split-plot designs that are robust for model misspecification. The author provided a general form of the loss function used for the D-optimal minimax criterion and applied it to searching for robust split-plot designs. To more efficiently construct designs, the author developed an algorithm which combines the anneal algorithm and point-exchange algorithm. Hence, the author modified the update formulas for calculating the determinant and inverse of the updated matrix and applied them to increasing the comparing speed for his developed program. Anqi (2016) investigated two different approaches, Neymanian randomization-based method and the Bayesian model-based method,

towards the causal inference for 2^2 split-plot designs, both under the potential outcome framework.

Luis (2015) provided practical strategies to help practitioners in dealing with the challenges presented by second-order block split-plot design, including an end-to-end, innovative approach for the construction of a new form of effective and efficient response surface design referred to as second-order sub-array Cartesian product split-plot design. This new form of design was an alternative to the ineffective split-plot designs that were then in use by the manufacturing/quality control community. The design was economical, the prediction variance of the regression coefficients was low and stable, and the aliasing between the terms in the model and effects that were not in the model, as well as the correlation between similar effects that were not in the model, was low. Based on an assessment using well-accepted key design evaluation criterion, it was demonstrated that second-order sub-array Cartesian product split-plot designs perform as well or better than historical designs that have been considered standards up to this point.

Vahide and Rajael (2015) studied the factors affecting liquidity accepted in stock exchange Agricultural Machinery companies using a Split-Plot design model. In their study they considered all five Agricultural Machinery Companies that were accepted in Tehran Stock Exchange since (1388-1390). Their study adopted a descriptive research based on the target application. In this model $(\tau\beta)_{ij}$ is still the whole plot error but the blocks $\times B$ and blocks $\times AB$ interactions were essentially pooled with ε_{ijk} to form the subplot error. Based on the results, it was observed that ratio type and company type, and interactions ratio type and company type, affected the company's liquidity.

Oluwole, Amahia and Fakorede (2014) studied design effects for the maximum

likelihood estimators of variance components in a split-plot design. Their study used the general linear model with one whole plot factor and one sub-plot factor and assumed that both factor effects are random variables. The main problem studied is how to assign a given number of whole plots with equal sizes to the level of the whole plot factor in a way that will form a balanced one-way design. Their study introduced a method of classifying the 5 variance components to make comparison and presentation meaningful. The resulting optimal designs depended on the true proportional value of the variance components.

The objective of a tutorial paper by Johannes (2010) was to review split-plot designs for full and fractional factorial experiments, and to explain why they often arise in industrial experiments, and provide several illustrative examples.

Beasley and Bruno (2009) reviewed three aligned rank methods for transforming data from multiple group repeated measures (otherwise, split-plot) designs. Univariate and multivariate statistics for testing the interaction in split-plot designs were elaborated. Computational examples were used to provide a context for performing these ranking procedures and statistical tests.

2.2 Linear and non-linear split-plot models

In Montgomery (2013), the linear model for a split-plot design with main treatment A and sub-treatment B is one in which τ_i , β_j , and $(\tau\beta)_{ij}$ represent the whole plot and correspond, respectively, to replicates, main treatments (factor A), and WPE (replicates $\times A$); and γ_k , $(\tau\gamma)_{ik}$, $(\beta\gamma)_{jk}$ and $(\tau\beta\gamma)_{ijk}$ represent the subplot and correspond, respectively, to the sub-plot treatment (factor B), the replicates $\times B$ and AB interactions, and the SPE (replicate $\times AB$). Note that WPE is the replicates $\times A$ interaction and the SPE is the three-factor interaction replicates $\times AB$. The sums of squares for these factors are

computed as in the three-way analysis of variance without replication.

According to David, Asiribo and Dikko (2019), the nonlinear split-plot model which has WPE and SPE is a special case of a nonlinear model with random effects (that is, nonlinear model with variance components WPE and SPE). The assumptions are given as follows.

Assumption 1:

It is presumed that the WPE and SPE are random effects. Also, it is presumed that $w_{ij} \sim N(0, \sigma_{wp}^2)$ and $\varepsilon_{ijk} \sim N(0, \sigma_{sp}^2)$.

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 $\mathbf{F}$ is the $n \times u$ matrix with elements $\partial f(x_{ijk}, \theta) / \partial \theta$ 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-covariance (VC) matrix of the reaction vector.

Assumption 3:

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

3.0 RESEARCH METHODOLOGY

Here, we have presented the procedure of EGLS (for estimating the parameters in the mean function component of the non-linear split-plot model), the variance component estimation by REML, needed concepts in the proposed alternative, and the procedure for implementing the proposed alternative.

3.1 Estimated generalized least squares (EGLS) estimation technique

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

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

with respect to θ ; where V is a known positive definite (non-singular) covariance matrix which arises from the model:

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

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

Let the variance-covariance matrix of the observations $\text{var}(y)$ be written as

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

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

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

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

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

Multiplying model (2) by $\mathbf{L}^{-1}$ on both sides yield that

$$\mathbf{L}^{-1} y_{ijk} = \mathbf{L}^{-1} f(x_{ijk}, \theta) + \mathbf{L}^{-1} w_{ij} + \mathbf{L}^{-1} \varepsilon_{ijk} \quad (3)$$

Let $\mathbf{I} = \mathbf{T}^{-1} = \mathbf{L} \mathbf{L}^t$ then the Cholesky factorization of the error variance is as follows.

$$\begin{aligned} \mathbf{L}^{-1} \{ \text{cov}(\varepsilon_{ijkl}) + \text{cov}(w_{ijk}) \} \mathbf{L}^{-t} &= \mathbf{L}^{-1} \text{cov}(\varepsilon_{ijkl}) \mathbf{L}^{-t} + \mathbf{L}^{-1} \text{cov}(w_{ijk}) \mathbf{L}^{-t} \\ &= \mathbf{L}^{-1} \mathbf{L}^{-t} \{ \text{cov}(\varepsilon_{ijk}) + \text{cov}(w_{ij}) \} \\ &= \mathbf{L}^{-1} (\sigma^2 \mathbf{I}) \mathbf{L}^{-t} \\ &= \sigma^2 \mathbf{L}^{-1} \mathbf{L} \mathbf{L}^t \mathbf{L}^{-t} \\ &= \sigma^2 \mathbf{I} \end{aligned}$$

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

Then equation (3) becomes

$$\mathbf{T}_{ijk} = \mathbf{M}(x_{ijkl}, \theta^*) + \Omega_{ijk} \quad (4)$$

where, $E(\Omega_{ijk}) = 0$ and $\mathbf{V}(\Omega_{ijk}) = \sigma^2 \mathbf{I}$. Thus the GLS model has been transformed to an OLS model. Hence, model (4) is to be solved using the OLS technique as follows. Taking the summation of both sides of (4) and squaring we have

$$\sum_i^s \sum_j^a \sum_k^b \Omega_{ijk}^2 = \sum_i^s \sum_j^a \sum_k^b \{ \mathbf{T}_{ijk} - \mathbf{M}(x_{ijkl}, \theta^*) \}^2 \quad (5)$$

$$\text{Let } L(\theta^*) = \sum_i^s \sum_j^a \sum_k^b \Omega_{ijk}^2 = \sum_i^s \sum_j^a \sum_k^b \{ \mathbf{T}_{ijk} - \mathbf{M}(x_{ijk}, \theta^*) \}^2$$

Minimize $L(\theta^*)$ w.r.t. θ^* , equate to zero and divide both sides by -2 we have,

$$\frac{\partial L(\theta^*)}{\partial \theta_h^*} = \sum_i^s \sum_j^a \sum_k^b \{ \mathbf{T}_{ijk} - \mathbf{M}(x_{ijk}, \theta^*) \} \times \left\{ \frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_h^*} \right\}_{\theta^* = \hat{\theta}^*} \quad (6)$$

At this point, equation (6) has no closed form, and hence will be solved iteratively using the Gauss-Newton method via Taylor series expansion of $\mathbf{M}(x_{ijk}, \theta^*)$ at first order. Note that the Taylor series expansion is given as

$$f(x) = f(a) + (x - a)f'(a) + \frac{(x - a)^2}{2!}f''(a) + \dots + \frac{(x - a)^h}{h!}f^{(h)}(a) + R_{h+1} \quad (7)$$

Therefore, we have

$$\begin{aligned} \mathbf{M}(x_{ijk}, \theta^*) &= \mathbf{M}(x_{ijk}, \theta_0^*) + (\theta_1^* - \theta_{10}^*) \frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_1^*} \Big|_{\theta^* = \theta_0^*} \\ &\quad + (\theta_2^* - \theta_{20}^*) \frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_2^*} \Big|_{\theta^* = \theta_0^*} + \dots \\ &\quad + (\theta_h^* - \theta_{h0}^*) \frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_h^*} \Big|_{\theta^* = \theta_0^*} \end{aligned} \quad (8)$$

$$\text{Let } \mathbf{M}(x_{ijk}, \theta^*) = \eta(\theta^*) \text{ and } d_{ijk} = \frac{\partial \mathbf{M}(x_{ijk}, \theta^*)}{\partial \theta_h^*} \Big|_{\theta^* = \hat{\theta}_0} \text{ for all } N \text{ cases and } \delta = \theta^* - \theta_0^* \text{ then (8)}$$

becomes

$$\eta(\theta^*) = \eta(\theta_0^*) + D_0 \delta \quad (9)$$

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

$$\begin{aligned} \Omega(\theta^*) &= \mathbf{T} - (\eta(\theta_0^*) + D_0 \delta) \\ \Omega(\theta^*) &= \mathbf{T} - \eta(\theta_0^*) - D_0 \delta \end{aligned}$$

$$\Omega(\theta^*) = \mathbf{z}_0 - D_0 \delta \quad (10)$$

where $\mathbf{z}_0 = \mathbf{T} - \eta(\theta_0^*)$ and $\delta = \theta^* - \theta_0^*$.

The Householder (1958) QR decomposition is applied to (26), as a result of its numerical stability characteristics for estimating the model parameters. This is done to decompose D_0 into a product of an orthogonal matrix and an inverted matrix.

Proof:

Let an $x \times y$ matrix have columns $w_1, w_2, \dots, w_y$ vectors.

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

Theorem 1:

Suppose $\mathbf{A}$ is a full column rank matrix of $x \times y$, then $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q}$ is a matrix of $x \times y$ whose column vectors create orthonormal basis for the column space of $\mathbf{A}$ while $\mathbf{R}$ is an $y \times y$ invertible upper triangular matrix.

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

Then $\mathbf{Q}$ is $m \times n$ with orthonormal columns such that, $\mathbf{Q}^t \mathbf{Q} = \mathbf{I}$.

If $\mathbf{A}$ is a squared matrix ($x = y$), then $\mathbf{Q}$ is orthonormal, that is, $\mathbf{Q}^t \mathbf{Q} = \mathbf{Q} \mathbf{Q}^t = \mathbf{I}$, hence, q_i is orthonormal to $w_1, w_2, \dots, w_y$.

Therefore,

$$\begin{aligned} w_1 &= (w_1 \cdot q_1)q_1 \\ w_2 &= (w_2 \cdot q_1)q_1 + (w_2 \cdot q_2)q_2 \\ &\dots \\ w_y &= (w_y \cdot q_1)q_1 + (w_y \cdot q_2)q_2 + \dots + (w_y \cdot q_y)q_y \end{aligned} \quad (11)$$

This implies that $\mathbf{A} = \mathbf{QR}$

$$(w_1 \ w_2 \ \dots \ w_y) = (q_1 \ q_2 \ \dots \ q_y) \begin{pmatrix} (w_1 \cdot q_1) & (w_1 \cdot q_2) & \dots & (w_1 \cdot q_y) \\ 0 & (w_2 \cdot q_2) & \dots & (w_2 \cdot q_y) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (w_k \cdot q_k) \end{pmatrix} \quad (12)$$

Let $\mathbf{A} = (w_1 \ w_2 \ \dots \ w_y)$ and $\mathbf{R}_{ij} = w_j \cdot q_i$ therefore, equation (12) is written as

$$\mathbf{A} = (q_1 \ q_2 \ \dots \ q_y) \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{R}_{1y} \\ 0 & \mathbf{R}_{22} & \dots & \mathbf{R}_{2y} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{R}_{yy} \end{pmatrix} \quad (13)$$

Equation (13) shows that $\mathbf{R}$ is $y \times y$ upper triangular with nonzero diagonal elements and $\mathbf{R}$ is non-singular (since the diagonal elements are nonzero). This means $\mathbf{A} = \mathbf{QR}$.

Theorem 2:

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

Proof:

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

Therefore,

$$\begin{aligned} \hat{\mathbf{x}} &= (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{b} = (\mathbf{R}^t \mathbf{Q}^t \mathbf{QR})^{-1} \mathbf{R}^t \mathbf{Q}^t \mathbf{b} \\ &\Rightarrow \mathbf{R}^t \mathbf{Q}^t \mathbf{QR} \hat{\mathbf{x}} = \mathbf{R}^t \mathbf{Q}^t \mathbf{b} \\ &\Rightarrow \mathbf{R}^t \mathbf{R} \hat{\mathbf{x}} = \mathbf{R}^t \mathbf{Q}^t \mathbf{b} \end{aligned} \quad (14)$$

Based on the two stated and proved theorems on $\mathbf{QR}$ -decomposition, the decomposition of D_0 is presented as follows.

Let $D_0 = \mathbf{QR}$. Where $\mathbf{Q}$ is a matrix of $\mathbf{N} \times \mathbf{N}$ and orthogonal, $\mathbf{Q}^t \mathbf{Q} = \mathbf{Q} \mathbf{Q}^t = \mathbf{I}$ while $\mathbf{R}$ is a $\mathbf{N} \times \mathbf{H}$ triangular matrix and under the major diagonal $\mathbf{R}$ is zero. Writing $\mathbf{Q}$ and $\mathbf{R}$ as follows,

$$\mathbf{Q} = (\mathbf{Q}_1 | \mathbf{Q}_2)$$

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

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

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

$$\mathbf{Q}^t = \begin{pmatrix} \mathbf{Q}_1^t \\ \mathbf{Q}_2^t \end{pmatrix}$$

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

$$D_0 = \mathbf{Q}\mathbf{R} = \mathbf{Q}_1\mathbf{R}_1 \quad (15)$$

Geometrically, the columns of $\mathbf{Q}$ define an orthonormal, or orthogonal, basis for the response space based on the property that the $\mathbf{H}$ columns cover the expectation plane. Projection onto the expectation plane is

$$\mathbf{g} = \mathbf{Q}^t \mathbf{z}_0 \quad (16)$$

with components

$$\mathbf{g}_1 = \mathbf{Q}_1^t \mathbf{z}_0 \quad (17)$$

and

$$\mathbf{g}_2 = \mathbf{Q}_2^t \mathbf{z}_0 \quad (18)$$

The properties of $\mathbf{g}$ onto the expectation plane is simply given as

$$\begin{pmatrix} \mathbf{g}_t \\ \mathbf{0} \end{pmatrix}$$

in $\mathbf{Q}$ coordinates and

$$\hat{\eta}_t = \mathbf{Q} \begin{pmatrix} \mathbf{g}_t \\ \mathbf{0} \end{pmatrix} = \mathbf{Q}_t \mathbf{g}_t \quad (19)$$

in the original coordinates. So,

$$\delta_0 = \mathbf{R}_t^{-1} \mathbf{g}_t$$

this implies

$$\mathbf{R}_t \delta_0 = \mathbf{g}_t \quad (20)$$

Equation (20) can be estimated using backward solving. 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 $\mathbf{z}_t = \mathbf{T} - \eta(\theta_1^*)$, 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

simple if the projection is in the coordinate system given by $\mathbf{Q}$.

Next, is transformation of the response vector, which is

small with no useful change in the elements of the parameter vector.

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 such as $\theta_1^* = \theta_0^* + \lambda \delta_0$ where λ is chosen to ensure that the new RSS is less than the initial estimate. A common method begins with $\lambda = 1$ and reduces 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 in the $\mathbf{VC}$ matrix, $\mathbf{V}$ is unknown. Therefore, an estimated $\mathbf{V}$ is obtained and substituted into equation (1) and the term EGLS is used. Different techniques for estimating the variance components to substitute for $\mathbf{V}$ in equation (1) are available. David, Asiribo and Dikko (2018) used the technique of REML (Residual Maximum Likelihood) presented in section 3.2 below.

3.2 Variance component estimation

It is 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

$$\ln L(\boldsymbol{\theta}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln|\mathbf{K}^t \mathbf{V} \mathbf{K}| - \frac{1}{2} (\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{X}, \boldsymbol{\theta}))^t \times (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} (\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{X}, \boldsymbol{\theta})) \quad (21)$$

where $\boldsymbol{\theta} = \left(\sigma^2 = \sigma_{WP}^2, \sigma_{SP}^2 \right)$ is then approximated by the surface and letting $\ln L$ to be Γ equation (21) becomes,

$$\begin{aligned} \Gamma(\boldsymbol{\theta}) &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln|\mathbf{K}^t \mathbf{V} \mathbf{K}| - \frac{1}{2} (\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}))^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} (\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta})) \\ &= \mathbf{C} - \frac{1}{2} \ln|\mathbf{K}^t \mathbf{V} \mathbf{K}| - \frac{1}{2} (\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}))^t (\mathbf{K}^t \mathbf{y} (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} - \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}) (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1}) \\ &= \mathbf{C} - \frac{1}{2} \ln|\mathbf{K}^t \mathbf{V} \mathbf{K}| - \frac{1}{2} \left((\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}))^t \mathbf{K}^t \mathbf{y} (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \right. \\ &\quad \left. + (\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}))^t \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}) (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \right) \end{aligned} \quad (22)$$

The third and fourth terms of equation (22) can be expressed respectively as follows.

$$\begin{aligned} \frac{1}{2} \left((\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}))^t \mathbf{K}^t \mathbf{y} (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \right) &= \frac{1}{2} \left((\mathbf{K} \mathbf{y}^t - \mathbf{K} f(\mathbf{x}, \boldsymbol{\theta})^t) \mathbf{K}^t \mathbf{y} (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \right) \\ &= \frac{1}{2} (\mathbf{y}^t \mathbf{K}^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \mathbf{y} \mathbf{K} - f(\mathbf{x}, \boldsymbol{\theta})^t \mathbf{K}^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \mathbf{y} \mathbf{K}) \end{aligned} \quad (23)$$

and

$$\begin{aligned} \frac{1}{2} \left((\mathbf{K}^t \mathbf{y} - \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}))^t \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}) (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \right) &= \frac{1}{2} \left((\mathbf{K} \mathbf{y}^t - \mathbf{K} f(\mathbf{x}, \boldsymbol{\theta})^t) \mathbf{K}^t f(\mathbf{x}, \boldsymbol{\theta}) (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \right) \\ &= \frac{1}{2} (\mathbf{y}^t \mathbf{K}^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} f(\mathbf{x}, \boldsymbol{\theta}) \mathbf{K} - f(\mathbf{x}, \boldsymbol{\theta})^t \mathbf{K}^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} f(\mathbf{x}, \boldsymbol{\theta}) \mathbf{K}) \end{aligned} \quad (24)$$

respectively. Therefore, equation (22) becomes

$$\begin{aligned} \Gamma(\boldsymbol{\theta}) &= \mathbf{C} - \frac{1}{2} \ln|\mathbf{K}^t \mathbf{V} \mathbf{K}| - \frac{1}{2} (\mathbf{y}^t \mathbf{K}^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \mathbf{y} \mathbf{K} - f(\mathbf{x}, \boldsymbol{\theta})^t \mathbf{K}^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} \mathbf{y} \mathbf{K}) \\ &\quad + \frac{1}{2} (\mathbf{y}^t \mathbf{K}^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} f(\mathbf{x}, \boldsymbol{\theta}) \mathbf{K} - f(\mathbf{x}, \boldsymbol{\theta})^t \mathbf{K}^t (\mathbf{K}^t \mathbf{V} \mathbf{K})^{-1} f(\mathbf{x}, \boldsymbol{\theta}) \mathbf{K}) \end{aligned} \quad (25)$$

linear combinations of $\mathbf{y}$ which have expectation to be zero rather than $\mathbf{y}$ itself. To obtain these vectors in the error space the linear approximation of the residuals is used $\mathbf{z}_0 = \mathbf{D}_0 \boldsymbol{\delta} + \boldsymbol{\varepsilon}$ as shown in (10).

To estimate the variance components from the nonlinear functions of $\mathbf{y}$ that would not involve $\hat{\theta}^*$, vectors of the form $\mathbf{k}^t \mathbf{y}$ are formed whereby $\mathbf{k}$ is selected so that $\mathbf{k}^t \mathbf{D}_0 = 0$ which falls in the linear estimate to the error space. $\mathbf{k}^t \mathbf{y}$ is called the error contrasts, that is, the part of the data that is orthogonal to the fixed effects (not dependent on the values of the fixed effect estimates), $\mathbf{k}$ is a vector from a full rank matrix $\mathbf{K}$ and maximizing the likelihood on $\mathbf{K}^t \mathbf{y}$, the function of the log likelihood on $\mathbf{K}^t \mathbf{y}$, is

$\mathbf{V} = \sigma^2 \mathbf{I} = \mathbf{K} \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K}^t$ and $\mathbf{V} \mathbf{V}^{-1}$ can be expressed as given below.

$$\mathbf{V} \mathbf{V}^{-1} = (\mathbf{K} \mathbf{V} \mathbf{K}^t) (\mathbf{K} \mathbf{V} \mathbf{K}^t)^{-1} = \mathbf{V} (\mathbf{K} (\mathbf{K} \mathbf{V} \mathbf{K}^t)^{-1} \mathbf{K}^t) = \sigma^2 (\mathbf{Q}_h \mathbf{V}_j)$$

Inserting $\mathbf{V}$ into equation (25) we have

$$\begin{aligned} \Gamma(\boldsymbol{\theta}) = & \mathbf{C} - \frac{1}{2} \ln \left| \mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right| \\ & - \frac{1}{2} \left(\mathbf{y}^t \mathbf{K}^t \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} \mathbf{y} \mathbf{K} - f(x, \theta)^t \mathbf{K}^t \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} \mathbf{y} \mathbf{K} \right) \\ & + \frac{1}{2} \left(\mathbf{y}^t \mathbf{K}^t \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} f(x, \theta) \mathbf{K} \right. \\ & \left. - f(x, \theta)^t \mathbf{K}^t \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} f(x, \theta) \mathbf{K} \right) \end{aligned} \quad (26)$$

Differentiate partially equation (26) w.r.t. σ_i^2 and equate to zero. By transformation all other terms in the equation becomes zero since $\mathbf{K}^t D_0 = \mathbf{K}^t f(x, \theta) = \mathbf{K} f(x, \theta)^t = 0$. Hence, we have

$$\begin{aligned} \frac{\partial \Gamma(\boldsymbol{\theta})}{\partial \sigma_i^2} = & -\frac{1}{2} \frac{1}{\left| \mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right|} (\mathbf{K}^t \mathbf{I}_i \mathbf{I}_i' \mathbf{K}) \\ & + \frac{1}{2} \mathbf{y}^t \mathbf{K}^t \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} (\mathbf{K}^t \mathbf{I}_i \mathbf{I}_i' \mathbf{K}) \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} \mathbf{y} \mathbf{K} \quad (27) \\ & \frac{1}{2} \left(\frac{1}{\left| \mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right|} (\mathbf{K}^t \mathbf{I}_i \mathbf{I}_i' \mathbf{K}) \right) = \frac{1}{2} \mathbf{y}^t \mathbf{K}^t \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} (\mathbf{K}^t \mathbf{I}_i \mathbf{I}_i' \mathbf{K}) \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} \mathbf{y} \mathbf{K} \end{aligned}$$

Let $\mathbf{Q}_h = \mathbf{K}^t \left(\mathbf{K}^t \sum_{i=0}^2 \sigma_i^2 \mathbf{I}_i \mathbf{I}_i' \mathbf{K} \right)^{-1} \mathbf{K}$ the equation (26) becomes

$$\frac{1}{2} (tr(\mathbf{Q}_h \mathbf{V}_i)) = \frac{1}{2} (\mathbf{y}^t \mathbf{Q}_h \mathbf{V}_i \mathbf{Q}_h \mathbf{y}) \quad (28)$$

Multiply the left hand side of equation (28) by $\mathbf{V} \mathbf{V}^{-1}$ we have

$$\frac{1}{2} (tr(\mathbf{Q}_h \mathbf{V}_i)) \sigma_{j(h+1)}^2 (\mathbf{Q}_h \mathbf{V}_j) = \frac{1}{2} (\mathbf{y}^t \mathbf{Q}_h \mathbf{V}_i \mathbf{Q}_h \mathbf{y}) \quad (29)$$

$$\langle tr(\widehat{\mathbf{Q}}_{(h)} \widehat{\mathbf{V}}_i \widehat{\mathbf{Q}}_{(h)} \widehat{\mathbf{V}}_j) \rangle \times \langle (\widehat{\sigma}_{j(h+1)}^2) \rangle = \langle (\mathbf{y}^t \widehat{\mathbf{Q}}_{(h)} \widehat{\mathbf{V}}_i \widehat{\mathbf{Q}}_{(h)} \mathbf{y}) \rangle \quad (30)$$

$$\langle (\widehat{\sigma}_{j(h+1)}^2) \rangle = \langle tr(\widehat{\mathbf{Q}}_{(h)} \widehat{\mathbf{V}}_i \widehat{\mathbf{Q}}_{(h)} \widehat{\mathbf{V}}_j) \rangle^{-1} \times \langle (\mathbf{y}^t \widehat{\mathbf{Q}}_{(h)} \widehat{\mathbf{V}}_i \widehat{\mathbf{Q}}_{(h)} \mathbf{y}) \rangle \quad (31)$$

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

3.3 Kronecker product

Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be $m \times n$ and $p \times q$ matrices, respectively. Then the Kronecker product

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij} \mathbf{B})$$

is an $mp \times nq$ matrix expressible as a partitioned matrix with $\mathbf{a}_{ij} \mathbf{B}$ as the (i, j) th partition, $i = 1, \dots, m$ and $j = 1, \dots, n$.

The following results are consequences of the definition above:

- (i) $\mathbf{0} \otimes \mathbf{A} = \mathbf{A} \otimes \mathbf{0} = \mathbf{0}$
- (ii) $(\mathbf{A}_1 + \mathbf{A}_2) \otimes \mathbf{B} = (\mathbf{A}_1 \otimes \mathbf{B}) + (\mathbf{A}_2 \otimes \mathbf{B})$
- (iii) $\mathbf{A} \otimes (\mathbf{B}_1 + \mathbf{B}_2) = (\mathbf{A} \otimes \mathbf{B}_1) + (\mathbf{A} \otimes \mathbf{B}_2)$
- (iv) $a\mathbf{A} \otimes b\mathbf{B} = ab\mathbf{A} \otimes \mathbf{B}$
- (v) $\mathbf{A}_1 \mathbf{A}_2 \otimes \mathbf{B}_1 \mathbf{B}_2 = (\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2)$
- (vi) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$, if the inverses exist.
- (vii) $(\mathbf{A} \otimes \mathbf{B})^- = \mathbf{A}^- \otimes \mathbf{B}^-$, using any g -inverses.
- (viii) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$
- (ix) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}) = \mathbf{1}$

3.4 Hadamard product

If $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be each $m \times n$ matrices, then their Hadamard product is the $m \times n$ matrix of element-wise products $\mathbf{A} \bullet \mathbf{B} = (a_{ij} b_{ij})$.

We have the following results involving Hadamard products:

- (i) $\mathbf{A} \bullet \mathbf{0} = \mathbf{0}$
- (ii) $\mathbf{A} \bullet \mathbf{e} \mathbf{e}' = \mathbf{A} = \mathbf{e} \mathbf{e}' \bullet \mathbf{A}$
- (iii) $\mathbf{A} \bullet \mathbf{B} = \mathbf{B} \bullet \mathbf{A}$
- (iv) $(\mathbf{A} + \mathbf{B}) \bullet \mathbf{C} = \mathbf{A} \bullet \mathbf{C} + \mathbf{B} \bullet \mathbf{C}$
- (v) $\text{tr}(\mathbf{A} \mathbf{B}) = \mathbf{e}'(\mathbf{A} \bullet \mathbf{B})\mathbf{e}$

3.5 The Khari-Rao product

Let $\mathbf{A} = (\mathbf{A}_1 : \dots : \mathbf{A}_k)$ and $\mathbf{B} = (\mathbf{B}_1 : \dots : \mathbf{B}_k)$ be two partitioned matrices with the same number of partitions. Khari and Rao, in 1968, defined a new product:

$$(\mathbf{A} \odot \mathbf{B}) = (\mathbf{A}_1 \otimes \mathbf{B}_1 : \dots : \mathbf{A}_k \otimes \mathbf{B}_k)$$

where $\otimes$ denotes the Kronecker product. It is easy to verify, among several properties, that:

- (i) $(\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C})$
- (ii) $(\mathbf{T}_1 \otimes \mathbf{T}_2)(\mathbf{A} \odot \mathbf{B}) = \mathbf{T}_1 \mathbf{A} \odot \mathbf{T}_2 \mathbf{B}$

3.6 Matrix exponential

Matrix exponential is mathematically defined as a matrix function on matrices analogous to the ordinary exponential function. It is mostly used to solve systems of linear differential equations. For an $m \times n$ real or complex matrix $\mathbf{A}$, the exponential of $\mathbf{A}$, denoted by

$e^{\mathbf{A}}$ or $\exp(\mathbf{A})$, is the $m \times n$ matrix given by the power series:

$$e^{\mathbf{A}} = \sum_{u=0}^{\infty} \frac{1}{u!} \mathbf{A}^u$$

where $\mathbf{A}^0$ is defined as the identity matrix $\mathbf{I}$ with the same dimensions as $\mathbf{A}$. The above

series always converges, so the exponential of $\mathbf{A}$ is well-defined.

3.7 Maximum likelihood estimation

Let X be a random sample with probability density function $f(x; \theta)$. Then the likelihood function is defined as:

$$l(x; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

Let X be a random sample with probability density function $f(x; \theta)$. Then the maximum likelihood estimate (MLE) of the parameter θ are the values of $\hat{\theta}$ that maximizes the likelihood function, where $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ and $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$. Based on the equation above, the maximum likelihood estimates of θ are the likelihood of $\hat{\theta}$ such that:

$$l(x; \hat{\theta}) = \max_{\theta \in \Omega} l(x; \theta)$$

where Ω is the parameter space.

The determination of $\hat{\theta}$ becomes relatively simple if the likelihood function $l(x; \theta)$ is twice differentiable in the range of definition of θ . In other words, $\hat{\theta}$ is the solution of the following k equations:

$$\frac{\partial l(x; \theta)}{\partial \theta_j} = 0, \quad j = 1, 2, 3, \dots, k$$

But the values of θ that maximizes $l(x; \theta)$ also maximizes its common logarithm $\log l(x; \theta)$ or natural logarithm $\ln l(x; \theta)$, and in most cases it is easier to solve the following k equations rather than the one above.

$$\frac{\partial \ln l(x; \theta)}{\partial \theta_j} = 0, \quad j = 1, 2, 3, \dots, k$$

3.8 Procedure for the proposed alternative

The proposed alternative is attempted under the following steps:

Step 1: Transform the non-linear split-plot model to a probability density function.

Step 2: Apply the method of MLE to obtain the parameter estimates of the mean function.

Step 3: Apply the method of MLE to obtain the parameter estimates of the variance-covariance components.

Step 4: Discuss findings via comparison of the existing and proposed procedures.

4.0 IMPLEMENTATION

4.1 Development of the proposed procedure

Our development commences with a theorem and an accompanying proof that the multiple linear regression model can be transformed to a probability distribution (or density) function. Now we consider the theorem below.

Now, the nonlinear split-plot model which has WPE and SPE is a special case of a nonlinear model with random effects (that is, nonlinear model with variance components WPE and SPE). The formulated model is given as follows.

Let

$$y = f(x, \theta) + w + \varepsilon$$

Inserting the levels of the factors to be investigated makes this equation to become.

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

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

Theorem 3:

Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ be the multiple regression model, such that $\mathbf{Y}_{n \times 1}$ is a response matrix, $\mathbf{X}_{n \times (p+1)}$ is a design matrix, $\boldsymbol{\beta}_{(p+1) \times 1}$ parameter matrix, and $\boldsymbol{\varepsilon}_{n \times 1}$ error matrix. If there exists a non-unique transformation matrix $\mathbf{F}_{n \times (p+1)}$ of arbitrary constants then there also exists a function

$\mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) \sim \exp(\boldsymbol{\beta})$ such that $\mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) \geq 0$ and $\int_{-\infty}^{\infty} \mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) \mathbf{d}\mathbf{x} = 1$.

Proof:

Now,

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \Rightarrow \mathbf{Y} - \boldsymbol{\varepsilon} &= \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

$$\Rightarrow \boldsymbol{\varepsilon} - \mathbf{Y} = -\mathbf{X}\boldsymbol{\beta} \quad (32)$$

Let $\mathbf{F}_{n \times (p+1)}$ be a matrix of arbitrary constants. Then pre-multiplying equation (34) by $\mathbf{F}^T$ gives

$$\begin{aligned} \Rightarrow \mathbf{F}^T (\boldsymbol{\varepsilon} - \mathbf{Y}) &= \mathbf{F}^T (-\mathbf{X}\boldsymbol{\beta}) \\ \Rightarrow \mathbf{F}^T (\boldsymbol{\varepsilon} - \mathbf{Y}) &= -\mathbf{F}^T (\mathbf{X}\boldsymbol{\beta}) \\ \Rightarrow \mathbf{F}^T (\boldsymbol{\varepsilon} - \mathbf{Y}) &= -(\mathbf{F}^T \mathbf{X})\boldsymbol{\beta} \end{aligned}$$

Let $\mathbf{X}^* = \mathbf{F}^T \mathbf{X}$ be the new design matrix, then taking the exponent of both sides of the last equation gives

$$\Rightarrow \exp\{\mathbf{F}^T (\boldsymbol{\varepsilon} - \mathbf{Y})\} = \exp\{-\mathbf{X}^* \boldsymbol{\beta}\}$$

Multiplying (Hadamard product) both sides of the last equation by $\boldsymbol{\beta}$ gives

$$\Rightarrow \mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) = \boldsymbol{\beta} \exp\{\mathbf{F}^T (\boldsymbol{\varepsilon} - \mathbf{Y})\} = \boldsymbol{\beta} \exp\{-\mathbf{X}^* \boldsymbol{\beta}\}$$

That is,

$$\Rightarrow \mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) = \boldsymbol{\beta} \exp\{-\mathbf{X}^* \boldsymbol{\beta}\} \quad (33)$$

$$\Rightarrow \mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) = \boldsymbol{\beta} \exp\{\mathbf{F}^T (\boldsymbol{\varepsilon} - \mathbf{Y})\} \quad (34)$$

Equation (35) shows that $\mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) \sim \exp(\boldsymbol{\beta})$ and as such $\mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) \geq 0$ and $\int_{-\infty}^{\infty} \mathbf{h}(\mathbf{x}; \boldsymbol{\beta}) \mathbf{d}\mathbf{x} = 1$.

4.2 Parameter estimation via the proposed procedure

Here, we consider the non-linear split-plot model in its matrix form

$$\mathbf{Y} = \mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) + \mathbf{W} + \boldsymbol{\varepsilon} \quad (35)$$

Applying theorem (3), we have

$$\mathbf{Y} - (\mathbf{W} + \boldsymbol{\varepsilon}) = \mathbf{f}(\mathbf{x}; \boldsymbol{\theta})$$

$$\Rightarrow (\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y} = -\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) \quad (36)$$

Let $\mathbf{F}$ be a matrix of arbitrary constants. Then pre-multiplying equation (38) by $\mathbf{F}^T$ gives

$$\begin{aligned} \Rightarrow \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y} &= \mathbf{F}^T \{-\mathbf{f}(\mathbf{x}; \boldsymbol{\theta})\} \\ \Rightarrow \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y} &= -\mathbf{F}^T \{\mathbf{f}(\mathbf{x}; \boldsymbol{\theta})\} \end{aligned}$$

Now, we assume that that the non-linear component may be written as product of the design matrix $\mathbf{X}$ and the parameter matrix $\boldsymbol{\theta}$, in which case, $\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}$. Hence,

$$\begin{aligned} \Rightarrow \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y} &= -\mathbf{F}^T (\mathbf{X}\boldsymbol{\theta}) \\ \Rightarrow \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y} &= -(\mathbf{F}^T \mathbf{X})\boldsymbol{\theta} \end{aligned}$$

Let $\mathbf{X}^* = \mathbf{F}^T \mathbf{X}$ be the new design matrix, then taking the exponent of both sides of the last equation gives

$$\Rightarrow \exp\{\mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\}\} = \exp\{-\mathbf{X}^* \boldsymbol{\theta}\}$$

Multiplying (Hadamard product) both sides of the last equation by $\boldsymbol{\theta}$ gives

$$\Rightarrow \mathbf{h}(\mathbf{x}; \boldsymbol{\theta}) = \boldsymbol{\theta} \exp\{\mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\}\} = \boldsymbol{\theta} \exp\{-\mathbf{X}^* \boldsymbol{\theta}\}$$

That is,

$$\Rightarrow \mathbf{h}(\mathbf{x}; \boldsymbol{\theta}) = \boldsymbol{\theta} \exp\{-\mathbf{X}^* \boldsymbol{\theta}\} \sim \exp(\boldsymbol{\theta}) \quad (37)$$

$$\Rightarrow \mathbf{h}(\mathbf{x}; \boldsymbol{\theta}) = \boldsymbol{\theta} \exp\{\mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\}\} \quad (38)$$

In order to estimate the parameters of the model, we now apply the maximum likelihood estimation method to equation (37) beginning with obtaining the likelihood function

$$\begin{aligned}\Rightarrow \mathbf{h}(\mathbf{x}; \boldsymbol{\theta}) &= \boldsymbol{\theta} \exp \{-\mathbf{X}^* \boldsymbol{\theta}\} \\ \Rightarrow L(\mathbf{x}; \boldsymbol{\theta}) &= \prod_{i=1}^{p+1} \boldsymbol{\theta} \exp(-\mathbf{X} \boldsymbol{\theta}) \\ \Rightarrow L(\mathbf{x}; \boldsymbol{\theta}) &= \prod_{i=1}^{p+1} \boldsymbol{\theta} e^{-\mathbf{X} \boldsymbol{\theta}} \\ \Rightarrow L(\mathbf{x}; \boldsymbol{\theta}) &= \boldsymbol{\theta}^{p+1} e^{-\boldsymbol{\theta} \sum \mathbf{X}} \\ \Rightarrow L(\mathbf{x}; \boldsymbol{\theta}) &= \boldsymbol{\theta}^{p+1} \exp(-\boldsymbol{\theta} \sum \mathbf{X})\end{aligned}$$

Taking the natural logarithm of both sides of the last equation gives

$$\begin{aligned}\Rightarrow \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} &= \ln \{\boldsymbol{\theta}^{p+1} \exp(-\boldsymbol{\theta} \sum \mathbf{X})\} \\ \Rightarrow \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} &= \ln \{\boldsymbol{\theta}^{p+1}\} + \ln \{\exp(-\boldsymbol{\theta} \sum \mathbf{X})\} \\ \Rightarrow \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} &= (p+1) \ln \boldsymbol{\theta} - \boldsymbol{\theta} \sum \mathbf{X}\end{aligned}$$

Differentiating the last equation partially with respect to each parameter θ_i of the non-linear split-plot model gives:

$$\begin{aligned}\Rightarrow \frac{\partial}{\partial \theta_i} \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} &= \frac{\partial}{\partial \theta_i} \{(p+1) \ln \boldsymbol{\theta} - \boldsymbol{\theta} \sum \mathbf{X}\} \\ \Rightarrow \frac{\partial}{\partial \theta_i} \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} &= (p+1) \frac{\partial}{\partial \theta_i} \{\ln \boldsymbol{\theta}\} - \sum \mathbf{X} \frac{\partial}{\partial \theta_i} \{\boldsymbol{\theta}\}\end{aligned}$$

$\forall i = 1, 2, 3, \dots, p+1$. In order to obtain the parameter estimates $\hat{\boldsymbol{\theta}}$ for the non-linear split-plot model, we equate the each of the $p+1$ partial derivatives above to zero and solve to get each $\hat{\theta}_i$. That is,

$$\Rightarrow (p+1) \frac{\partial}{\partial \theta_i} \{\ln \boldsymbol{\theta}\} - \sum \mathbf{X} \frac{\partial}{\partial \theta_i} \{\boldsymbol{\theta}\} = 0 \quad (39)$$

$$\Rightarrow \frac{\frac{\partial}{\partial \theta_i} \{\ln \boldsymbol{\theta}\}}{\frac{\partial}{\partial \theta_i} \{\boldsymbol{\theta}\}} = \frac{\sum \mathbf{X}}{(p+1)} \quad (40)$$

4.3 Variance component estimation via the proposed procedure

For the variance component estimation, we shall obtain the maximum likelihood estimates of the random error effects using equation (38). Of course, equation (38) is a probability density function especially because it is a variant of equation (37) which has been established to be an exponential distribution.

$$\begin{aligned}\mathbf{h}(\mathbf{x}; \boldsymbol{\theta}) &= \boldsymbol{\theta} \exp \{\mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\}\} \\ \Rightarrow L(\mathbf{x}; \boldsymbol{\theta}) &= \prod_{i=1}^r \{\boldsymbol{\theta} \exp \{\mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\}\}\} \\ \Rightarrow L(\mathbf{x}; \boldsymbol{\theta}) &= \boldsymbol{\theta}^r \exp \left\{ \sum_{i=1}^r \{\mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\}\} \right\}\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} = \ln \left\{ \boldsymbol{\theta}^r \exp \left\{ \sum_{i=1}^r \left\{ \mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\} \right\} \right\} \right\} \\
&\Rightarrow \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} = \ln \boldsymbol{\theta}^r + \ln \left\{ \exp \left\{ \sum_{i=1}^r \left\{ \mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\} \right\} \right\} \right\} \\
&\Rightarrow \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} = r \ln \boldsymbol{\theta} + \sum_{i=1}^r \left\{ \mathbf{F}^T \{(\mathbf{W} + \boldsymbol{\varepsilon}) - \mathbf{Y}\} \right\} \\
&\Rightarrow \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} = r \ln \boldsymbol{\theta} + \sum_{i=1}^r \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) - \sum_{i=1}^r \mathbf{Y}
\end{aligned}$$

Differentiating the last equation partially with respect to each w_i in $\mathbf{W}$ of the variance component gives:

$$\begin{aligned}
&\Rightarrow \frac{\partial \{ \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} \}}{\partial w_i} = \frac{\partial \left\{ r \ln \boldsymbol{\theta} + \sum_{i=1}^r \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) - \sum_{i=1}^r \mathbf{Y} \right\}}{\partial w_i} \\
&\Rightarrow \frac{\partial \{ \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} \}}{\partial w_i} = \frac{\partial \left\{ \sum_{i=1}^r \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) \right\}}{\partial w_i} \\
&\Rightarrow \frac{\partial \{ \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} \}}{\partial w_i} = \sum_{i=1}^r \frac{\partial \{ \mathbf{F}^T \mathbf{W} + \mathbf{F}^T \boldsymbol{\varepsilon} \}}{\partial w_i} \\
&\Rightarrow \frac{\partial \{ \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} \}}{\partial w_i} = \sum_{i=1}^r \frac{\partial \{ \mathbf{F}^T \mathbf{W} \}}{\partial w_i} \tag{41}
\end{aligned}$$

Differentiating the same equation partially with respect to each ε_i in $\boldsymbol{\varepsilon}$ of the variance component gives:

$$\begin{aligned}
&\Rightarrow \frac{\partial \{ \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} \}}{\partial \varepsilon_i} = \frac{\partial \left\{ r \ln \boldsymbol{\theta} + \sum_{i=1}^r \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) - \sum_{i=1}^r \mathbf{Y} \right\}}{\partial \varepsilon_i} \\
&\Rightarrow \frac{\partial \{ \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} \}}{\partial \varepsilon_i} = \frac{\partial \left\{ \sum_{i=1}^r \mathbf{F}^T (\mathbf{W} + \boldsymbol{\varepsilon}) \right\}}{\partial \varepsilon_i} \\
&\Rightarrow \frac{\partial \{ \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} \}}{\partial \varepsilon_i} = \sum_{i=1}^r \frac{\partial \{ \mathbf{F}^T \mathbf{W} + \mathbf{F}^T \boldsymbol{\varepsilon} \}}{\partial \varepsilon_i} \\
&\Rightarrow \frac{\partial \{ \ln \{L(\mathbf{x}; \boldsymbol{\theta})\} \}}{\partial \varepsilon_i} = \sum_{i=1}^r \frac{\partial \{ \mathbf{F}^T \boldsymbol{\varepsilon} \}}{\partial \varepsilon_i} \tag{42}
\end{aligned}$$

In order to obtain the estimates $\hat{\mathbf{W}}$ and $\hat{\boldsymbol{\varepsilon}}$ of the random effects for the variance component of the non-linear split-plot model, we equate the each of the r partial derivatives in equation (41) and equation (42) to zero and solve accordingly.

4.4 DISCUSSION OF FINDINGS

The results in this work have shown that, rather than use the EGLS technique for estimating the parameters of the mean function in a non-linear split-plot model, a conversion of the said model to a probability density function could be a useful (but non-complex) procedure for obtaining the said

estimates of the parameters. More so, rather than use the MLE-approach of Marcia and John (1990), as well as David, Asiribo and Dikko (2018; 2019), the MLE approach may yet be used but this time with a variant of the probability density function with which estimates of the parameters of the mean function were obtained.

4.5 Suggestions for further studies

As suggestions for further studies, attempts should be made at estimating the parameters of a non-linear split-plot model via other estimation procedures aside the previously used ones as this could lead to deeper theoretical and practical results. And attempts should also be made at remodelling the non-linear split-plot model to any other distribution, with which a distribution-adaptive estimation procedure could be used to estimate both the parameters of the mean function and the variance components.

REFERENCES

- Bradley, J., Christopher, J. N. (2009). Split-plot designs: What, why, and how. *Journal of Quality Technology*, 41 (4), 340-361.
- Bruce, B. F. (2018). Split-Plot Design. USA: The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation.
- Caroline, M. M. (1990). Maximum Likelihood Methods for Nonlinear Regression Models with Compound-symmetric Errors Covariance. A Doctor of Philosophy Dissertation Submitted to the Department of Biostatistics, University of North Carolina.
- Cochran, W. G., Cox. G. M. (1992). *Experimental Designs* (2nd Ed.). USA: Wiley Classic Library.
- Cox, D. R., Reid, N. (2000). *The Theory of the Design of Experiments*. Boca Raton: Chapman and Hall/CRC.
- Chang-Yun, L. (2016). Robust split-plot designs for model misinterpretation. *Stat Co*, 3(2), 1 – 21.
- David, J. M. (2009). *Methods of System Identification, Parameter Estimation and Optimization Applied to Problems of Modeling and Control in Engineering and Physiology*. A Thesis Submitted for the Degree of Doctor of Science in Engineering at the University of Glasgow.
- David, I. J., Asiribo, O. E., Dikko, H. G. (2018). Non-linear split-plot design model in parameters estimation using estimated generalized least square – maximum likelihood estimation. *Computer, Mathematics and Engineering Applications*, 9 (2), 65 – 71.
- David, I. J., Asiribo, O. E., Dikko, H. G. (2019). Parameter estimation of non-linear split-plot design models: A theoretical framework. *Journal of Reliability and Statistical Studies*, 12 (1), 117-129.
- Douglas, M. B., Jose, C. P. (1998). Linear and non-linear mixed-effect models. *10th Annual Conference Proceedings on Applied Statistics in Agriculture*.
- Glenn, L. (2015). *Curve Fitting and Parameter Estimation*. USA: Spring Press.
- Gumpertz, M. L., Rawlings, J. O. (1992). Nonlinear regression with variance components: Modeling effects of ozone on crop yield. *Crop Science*, 32, 219 – 224.
- Howard, J. S. (2018). *Analysis of Split-Plot Designs*. *Experimental Design and Analysis*. London: Oxford Press.
- Johannes, L. (2010). Split-plot designs: Discussions and examples. *International Journal of Quality Engineering and Technology*, 1 (4), 441 – 457.
- Katarzyna A., Iwona, M., Stanislaw, M. (2014). On the relative efficiency of split-split-plot design to split-plot by split-block design. *Colloquium Biometrician*, 44 (2), 70 – 78.
- Kay, W. V., Laura, P. S., Randall, M. R., Nicholas, J. S., Sean, P. S. (2005). *Confidence Region Estimation Techniques for Non-linear Regression: Three Case Studies*. USA: Sandia National Laboratories.
- Marcia, L. G., John, O. R. (1990). Nonlinear regression for split-plot experiments.

- Conference on Applied Statistics in Agriculture.*
- Montgomery, D. C. (2013). Design and Analysis of Experiment (8th Ed.). USA: John Wiley and Sons, Inc.
- Mukesh, S., Rinku, A. (2003). Maximum likelihood method for parameter estimation in non-linear models with below detection data. *Environmental and Ecological Statistics*, 10 (2), 445-454.
- Naes, T., Aastveit, A. H., Sahni, N. S. (2007). Analysis of split-plot designs: An overview and comparison of methods. *Quality and Reliability Engineering International*, 23 (2), 801-820.
- Oehlert, W. O. (2010). A First Course in Design and Analysis of Experiments. USA: Library of Congress.
- Oluwole, N., Amahia, G. N. & Fakorede, A. (2014). Construction of split-plot design for estimating variance components. *European Scientific Journal*, 10(18), 408 – 418.
- Sourish, D. (2008). Generalized Linear Models and Beyond – An Innovative Approach from Bayesian Perspective. A PhD Dissertation from University of Connecticut.
- Scott, M. K., John, A. C. & Geoffrey G. V. (2002). Split-plot designs and estimation methods for mixture experiments with process variables. *Technometrics*, 44(1), 72 – 79.
- Teresa, G. E., James, R. Simpson, Drew, L. Peter, A. P. (2012). An efficient split-plot approach for modeling aerodynamic effects. *Journal of Quality Engineering*, 24 (4).
- Toth, B. A. (2011). Computational Methods for Parameter Estimation in Nonlinear Models. A Doctor of Philosophy Dissertation in Physics with Specialization in Computational Physics Submitted to the University of California, San Diego.
- Xiao, Z., Feng, D., Ling, X., Ahmed, A., Tasawar, H. (2019). A hierarchical approach for joint parameter and state estimation of a bilinear system with autoregressive noise. *Mathematics*, 7 (4), 356 – 363.