Model Fitness and Predictive Accuracy in Linear Mixed-Effects Models with Latent Clusters

Waheed B. Yahya[a], Yusuf Bello[b,*], Abdulrazaq AbdulRaheem[c]

[a]Department of Statistics, , University of Ilorin, P.M.B. 1515, Ilorin, Kwara State, Nigeria.
[b]Department of Statistics, Federal University, Dutsin-Ma, P.M.B. 5001, Dutsin-Ma, Katsina State, Nigeria.
[c]Department of Statistics and Mathematical Sciences, Kwara State University, Malete, P.M.B. 1530, Ilorin, Kwara State, Nigeria.

Abstract

In clustered data, observations within a cluster show similarity between themselves because they share common features different from observations in the other clusters. In a given population, different clustering may surface because correlation may occur across more than one dimension. The existing multilevel analysis techniques of the primal linear mixed-effect models are limited to natural clusters which are often not realistic to capture in real-life situations. Therefore, this paper proposes dual linear mixed models (DLMMs) for modeling unobserved latent clusters when such are present in data sets to yield appreciable gains in model fitness and predictive accuracy. The methodology explored the development and analysis of the dual linear mixed models (DLMMs) based on the derived latent clusters from the natural clusters using multivariate cluster analysis. A published data set on political analysis was used to demonstrate the efficiency of the proposed models. The proposed DLMMs have yielded minimum values of the models' assessment criteria (Akaike information criterion, Bayesian information criterion, and root mean squared error), and hence, outperformed the classical PLMMs in terms of model fitness and predictive accuracy.

Communicated by: O. Adeyeye

1. Introduction

The multilevel modeling technique follows a similar process involved when fitting the Generalized Linear Model [1]. In particular, a Linear mixed model (LMM) is one of the approaches in modeling normally distributed clustered data [2]. In clustered data, observations within a cluster show similarity between themselves because they share common features different from observations in the other clusters. In a given population, different clustering may surface because correlation may occur across more than one dimension [3]. They further argued that clustering is in essence a design problem, either a sampling design or an experimental design issue. Even if data is collected in an unclustered way, there is still natural clustering in the population.

As an illustration from Nigeria's crime analysis, the initial dataset comprised 36 states grouped into six geo-political zones and 12 Police Zonal Commands that share spatial and socio-ethnic similarities. However, the optimal number of clusters provided new structure classifications based on crime rates sim-

*Corresponding author tel. no: +2348134983581
Email address:* ybello@fudutsinma.edu.ng (Yusuf Bello)

ilarities different from the initial spacial and socio-ethnic similarities [4]. It is posited here that the observations in the newly formed clusters based on multivariate clustering similarities are more correlated than in the natural clusters based on sampling and experimental design similarities. The former would better account for the differences between the clusters and improve model fitness and predictive accuracy.

The natural clusters and latent clusters are respectively described as 'primal clusters' and 'dual clusters'. The linear mixed-effects models (LMEMs or simply LMMs) on the primal clusters and dual clusters are respectively described as primal linear mixed models (PLMMs) and dual linear mixed models (DLMMs). This paper proposes efficient DLMMs for modeling data with latent clusters with appreciable gains in model fitness and predictive accuracy.

2. The Concept of LMEMs on Latent Clusters and Model Assessment Criteria

The general concept of LMEMs on latent clusters is to maximize correlation of observations within clusters, model fitness and predictive accuracy. The latent clusters were formed from natural clusters using the multivariate cluster analysis. Both the natural and dual clusters contain the same observations, although the cluster structures differ. The argument for comparing models formed from same data set with differing data structures was demonstrated by [14]. Agglomerative algorithm is a common approach in cluster analysis for classifying observations that share common properties into groups. The algorithm starts by calculating the distances between all pairs of observations followed by stepwise agglomeration of close observations into groups. Euclidean distance is the most commonly used distance measure in numerical data, while Ward method is the most frequently used linkage method [5].

LMM is a linear model with an extension of accounting for dependency among clustered observations. In biological and social sciences, a model-based cluster analysis utilizes LMM in the grouping of individuals into one of two or more clusters according to their longitudinal behaviour similarities. [6, 7]. In contrast to those studies that utilize Expectation-Maximization algorithms in cluster formations, this study conjoins LMEMs and multivariate cluster analysis to develop efficient techniques for modeling unobserved latent groupings in a data set.

The degree of clustering in a data set is measured by the intraclass correlation (ICC). The ICC is the proportion of total variance in the data that is due to the clusters [8]. The argument in the multilevel analysis on latent clusters is that the increasing clustering in DLMM would simultaneously reduce the indices of the model assessment criteria. Many indices are available to measure the performances of competing models [9], however, the models' assessment criteria in this work are the root mean squared error (RMSE), Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

The RMSE indicates the absolute fit of the model to the data. Smaller values of RMSE indicate better fit results. The LMM improves model fitness and predictive performance, and this is because a multilevel model produces fitted values, $\hat{y}$, that are on the average closer to the observed y than those obtained by fitting simply the fixed part of the model. Again, even in the multilevel analysis, when the estimated random effects tend to be biased towards zero; it pulls the fitted values in the direction of those of the fixed part of the model that results in bias estimates. Furthermore, simpler models such as random intercept models produce larger bias relatively than the complex models such as random intercept and slope models [10]. Therefore, by analogy, the more the estimated random effects tend to be larger, the more the $\hat{y}$ moves closer to y, and hence the lower the RMSE.

The simplest information criterion widely applicable to nonnested models is the AIC [11, 12]. This traditional AIC is not appropriate in clustered data, and therefore marginal AIC (mAIC) is the most widely used in model selection in LMMs [12]. A related criterion to the mAIC in their marginal likelihoods is the BIC [13]. The presence of random effects in LMMs results in smaller AIC and BIC than in the LMs [14].

3. The Linear Mixed Model

Consider a vector y of data from J clusters, the LMM as define by [2, 15] is

$$y_j = X_j\beta + Z_j\upsilon_j + \epsilon_j \tag{1}$$

where y_j is the n_j vector for cluster j, where $j = 1, 2, \cdots, J$ is the cluster index, β is the p-vector of fixed effects, υ_j is the q-vector of random effects for cluster j, X_j and Z_j are respectively the $n_j \times p$ and $n_j \times q$ matrices of covariates for the fixed and random effects of full rank. It is assumed that υ_j and ϵ_j follow independent and multivariate Gaussian distributions such that [16]:

$$\begin{bmatrix} \upsilon_j \\ \epsilon_j \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} T & 0 \\ 0 & \sigma^2 I \end{bmatrix} \right) \tag{2}$$

were T is $q \times q$ positive definite covariance matrix of the random effects, υ_j is assumed independently for each j, ϵ_j associated with different clusters are assumed independent of each other, and that ϵ_j is assumed independent of υ_j [15].

In a marginal model

$$y_j = N(X_j\beta, V_j) \quad V_j = Z_j T Z_j' + \sigma^2 I \tag{3}$$

with a marginal likelihood given as

$$l(y_j/\hat{\beta}, \hat{\theta}) = -\frac{1}{2}log|\hat{V}_j| - \frac{1}{2}(y_j - X_j\hat{\beta})'\hat{V}_j^{-1}(y_j - X_j\hat{\beta}) \tag{4}$$

3.1. Cluster Effects

The cluster effect or the dependency among clustered observations is measured by the ICC, and is defined as

$$ICC = \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \sigma_e^2} \tag{5}$$

where σ_υ^2 and σ_e^2 are random effects variance and random error variance respectively.

To show that the cluster effect is higher in the dual clusters than in the primal clusters, that is, $ICC_{(D)} > ICC_{(P)}$:

Let P and D describe PLMM and DLMM respectively, $\sigma^2_{e(P)} = \sigma^2_e$, $\sigma^2_{e(D)} = \sigma^2_e - \delta_1$, $\sigma^2_{v(P)} = \sigma^2_v$, $\sigma^2_{v(D)} = \sigma^2_v + \delta_2$, where $\delta_2 \geq \delta_1$, δ_2 and δ_1 are small increment of random effects variance and decrement random error variance, respectively,.

Case 1: If $\delta_2 = \delta_1$, such that $\delta_2 - \delta_1 = 0$, then

$$
\begin{aligned}
ICC_{(D)} - ICC_{(P)} &= \frac{\sigma^2_v + \delta_2}{(\sigma^2_v + \delta_2) + (\sigma^2_e - \delta_1)} - \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_e} \\
&= \frac{\sigma^2_v + \delta_2}{(\sigma^2_v + \sigma^2_e) + (\delta_2 - \delta_1)} - \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_e} \\
&= \frac{\sigma^2_v + \delta_2}{(\sigma^2_v + \sigma^2_e)} - \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_e} \\
&= \frac{\delta_2}{(\sigma^2_v + \sigma^2_e)}
\end{aligned}
\tag{6}
$$

Since (6) results in a positive difference, then

$$ICC_{(D)} > ICC_{(P)}$$

Case 2: If $\delta_2 > \delta_1$, such that $\delta_2 - \delta_1 = \delta_3 > 0$, then

$$
\begin{aligned}
ICC_{(D)} - ICC_{(P)} &= \frac{\sigma^2_v + \delta_2}{(\sigma^2_v + \delta_2) + (\sigma^2_e - \delta_1)} - \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_e} \\
&= \frac{\sigma^2_v + \delta_2}{(\sigma^2_v + \sigma^2_e) + (\delta_2 - \delta_1)} - \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_e} \\
&= \frac{\sigma^2_v + \delta_2}{(\sigma^2_v + \sigma^2_e + \delta_3)} - \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_e} \\
&= \frac{\sigma^2_v + \delta_2}{(\Sigma + \delta_3)} - \frac{\sigma^2_v}{\Sigma}, \qquad \text{where } \Sigma = \sigma^2_v + \sigma^2_e \\
&= \frac{\delta_2 \sigma^2_e + \delta_1 \sigma^2_v}{\Sigma(\Sigma + \delta_3)}
\end{aligned}
\tag{7}
$$

Since (7) results in a positive difference, then

$$ICC_{(D)} > ICC_{(P)}$$

3.2. Root Mean Squared Error

The RMSE is the difference between observed data and the predicted values from the model, and it is defined as

$$
RMSE = \sqrt{\frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_{ij} - \hat{y}_{ij})^2}{\sum_{j=1}^{J} n_j}}
\tag{8}
$$

where J is the number of clusters, n_j is the number of observations in the jth cluster, y_{ij} and $\hat{y}_{ij}$ are the ith observed and estimated y in jth cluster, respectively [17].

3.3. Marginal Akaike Information Criterion

The commonly used information criterion is the AIC [11]. This criterion which is based on Kullback-Leibler distance is defined as

$$AIC = -2log[f(y/\hat{\psi}(y))] + 2k$$

where $f(y/\hat{\psi}(y))$ is the maximized likelihood, and k is the number of parameters. This AIC is not appropriate in clustered data, and hence the mAIC is widely used in the clustered data [12].

The mAIC in the LMM uses the likelihood of the implied marginal model $y \sim N(X\beta, V)$ with $V = I_n + ZTZ'$. The number of estimable parameters then is $p + q$, with $\beta = (\beta_1, \cdots, \beta_p)$ and q the number of unknown parameters θ in V. Thus, the mAIC is defined as

$$mAIC = -2log[f(y/\hat{\beta}, \hat{\theta})] + 2(p + q) \tag{9}$$

where $f(y/\hat{\beta}, \hat{\theta})$ is the maximized marginal likelihood. However, the mAIC is positively biased, and favours smaller models without random effects [18].

3.4. Bayesian Information Criterion

The is obtained by taking the mAIC (9) and replacing the constant 2 in the penalty by log(n) to obtain

$$BIC = -2log[f(y/\hat{\beta}, \hat{\theta})] + log(n)(p + q) \tag{10}$$

This definition ensures that BIC bears the same relationship to mAIC for model (1) as BIC bears to AIC in regression and so should inherit some of its properties [13].

3.5. Cluster Effects on Model Assessment Criteria

Cluster effects in mixed models are explained by the random effects variance of the models, and including the random effects has an effect on the covariance matrix, V_j. As an illustration, consider a random intercept model from a data where five observations are taken on each cluster, so that $n_j = 5$ for all j. Therefore, Z_j is a matrix of dimension 5×1 and $R_j = \sigma^2 \times I_{5 \times 5}$. Then

$$
\begin{aligned}
V_j &= \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \times \sigma^2_v \times \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix} \\
&\quad + \sigma^2 \times \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & & & \vdots \\ \vdots & & 1 & & \vdots \\ \vdots & & & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \sigma^2 + \sigma^2_v & \sigma^2_v & \sigma^2_v & \sigma^2_v & \sigma^2_v \\ \sigma^2_v & \sigma^2 + \sigma^2_v & \sigma^2_v & \sigma^2_v & \sigma^2_v \\ \sigma^2_v & \sigma^2_v & \sigma^2 + \sigma^2_v & \sigma^2_v & \sigma^2_v \\ \sigma^2_v & \sigma^2_v & \sigma^2_v & \sigma^2 + \sigma^2_v & \sigma^2_v \\ \sigma^2_v & \sigma^2_v & \sigma^2_v & \sigma^2_v & \sigma^2 + \sigma^2_v \end{pmatrix}
\end{aligned}
\tag{11}
$$

The elements in the diagonal, $\sigma^2_v + \sigma^2$, are correlations between two observations from the same cluster, and the elements at off diagonal, σ^2_v, are the covariances between any two units on the same cluster. By relating the two terms, the intraclass correlation between two observations from the same cluster is $\sigma^2_v/(\sigma^2_v + \sigma^2)$ [14].

Denoting V_j as $V_{j(P)}$ and $V_{j(D)}$ for the PLMM and DLMM respectively, therefore, if the $ICC_{(D)} > ICC_{(P)}$ as in (6) and (7), then $V_{j(D)} > V_{j(P)}$.

From (4), we respectively ascribe the marginal likelihood for the PLMM and DLMM as $l(y_j/\hat{\beta}, \hat{\theta})_{(P)}$ and $l(y_j/\hat{\beta}, \hat{\theta})_{(D)}$, such that

$$l(y_j/\hat{\beta}, \hat{\theta})_{(P)} = -\frac{1}{2}log|\hat{V}_{j(P)}| - \frac{1}{2}(y_j - X_j\hat{\beta})' \hat{V}_{j(P)}^{-1}(y_j - X_j\hat{\beta}) \quad (12)$$

and

$$l(y_j/\hat{\beta}, \hat{\theta})_{(D)} = -\frac{1}{2}log|\hat{V}_{j(D)}| - \frac{1}{2}(y_j - X_j\hat{\beta})' \hat{V}_{j(D)}^{-1}(y_j - X_j\hat{\beta}) \quad (13)$$

Similar to the basic concept of fraction that a negative fraction increases with the increase in the denominator, the second term in (13), $-\frac{1}{2}(y_j - X_j\hat{\beta})' \hat{V}_{j(D)}^{-1}(y_j - X_j\hat{\beta})$, is relatively higher than in (13) because of the increase of the inverse of the matrix $\hat{V}_{j(D)}^{-1}$ relative to $\hat{V}_{j(P)}^{-1}$. Again, in the basic concept of logarithm that a negative logarithm decreases with the increase of a number, the first term in (13), $-\frac{1}{2}log|\hat{V}_{j(D)}|$, is relatively lower than in (12) because of the increase of the inverse of $\hat{V}_{j(D)}$ relative to $\hat{V}_{j(P)}$. Although, the first term and second term in (13) decrease and increase respectively, the increase outweights the decrease, such that

$$l(y_j/\hat{\beta}, \hat{\theta})_{(D)} > l(y_j/\hat{\beta}, \hat{\theta})_{(P)} \quad (14)$$

The presence of negative sign in the $-2log[f(y/\hat{\beta}, \hat{\theta})]$ for the information criteria in (9) and (10) has changed the direction of the inequality in (14), such that $l(y_j/\hat{\beta}, \hat{\theta})_{(D)} < l(y_j/\hat{\beta}, \hat{\theta})_{(P)}$. The p and q are the same in both the PLMM and DLMM, and therefore

$$\begin{aligned} mAIC_{(D)} &< mAIC_{(P)} \qquad \text{and} \\ mBIC_{(D)} &< mBIC_{(P)} \end{aligned} \quad (15)$$

The LMM improves model fitness and predictive performances because it incorporates clustering effects when estimating the fixed parameter. This adjustment enhances it to produce fitted values, $\hat{y}$, that are on the average closer to the observed y than those obtained by fitting simply the fixed part of the model [10]. In our proposal, DLMM has higher clustering effect than PLMM that enhances it to produce fitted values, $\hat{y}$, that are on the average closer to the observed y than those produced by the PLMM.

4. Cluster Algorithm: The Agglomerative Algorithm

The agglomerative procedure depends on the definition of the distance between two clusters. For a particular case where metric $A = (S_{X_1X_1}^{-1}, \cdots, S_{X_pX_p}^{-1})$ is used for the standardization of the variables, the Euclidean distance d_{ij} between two cases i and j with variable values $x_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$, $x_j = (x_{j1}, x_{j2}, \cdots, x_{jp},)$ is defined by

$$d_{ij} = \left[\sum_{k=1}^{p} \frac{(x_{ik} - x_{jk})^2}{S_{X_kX_k}} \right]^{\frac{1}{2}}$$

where $S_{X_kX_k}$ is the variance of the kth component [19].

Ward algorithm computes the distance between groups and joins the ones that do not increase a given measure of heterogeneity "too much" so the resulting groups are as homogeneous as possible. If two objects or groups say, P and Q, are united, one computes the distance between this new group (object) $P + Q$ and group R using the following distance function

$$\begin{aligned} d(R, P + Q) &= \frac{n_R + n_P}{n_R + n_P + n_Q} d(R, P) \\ &+ \frac{n_R + n_Q}{n_R + n_P + n_Q} d(R, Q) - \\ &\frac{n_R}{n_R + n_P + n_Q} d(P, Q) \end{aligned} \quad (16)$$

The heterogeneity of group R is measured by the inertia inside the group. This inertia is defined as $I_R = \frac{1}{n_R} \sum_{i=1}^{n_R} d^2(x_i, \bar{x}_R)$ [20].

5. Illustration and Analysis

A published data set on political analysis was used to demonstrate the efficiency of the proposed models. The dataset dcese provided with the ceser R package came from [21]. It contains information on 299 ($i = 1, 2, \cdots, 299$) observations across 47 countries ($j = 1, 2, \cdots, 47$). The outcome variable is the effective number of electoral parties (enep). The explanatory variables are the number of presidential candidates (enpc), the proximity of presidential and legislative elections (proximity); the effective number of ethnic groups (eneg), the logarithm of average district magnitudes (logmag), and an interaction term between the logarithm of the district magnitude and the number of ethnic groups (logmag_eneg = logmag × eneg).

5.1. Comparison between Primal and Dual Linear Mixed Models

We begin with a preliminary comparison of the PLMM and DLMM using primal and dual cluster data sets with $J = 47$ number of groups, and subsequently test the significance of the comparison. The comparison is in terms of the variance-covariance components and their impact on model fitness and predictive accuracy. The summary outputs of the models are presented in Table 1.

It reveals from the summary in Table 1 that while σ_e^2 is higher under PLMM, the σ_v^2 and ICC are relatively higher under DLMM. There is a 61 percent decrease of σ_e^2 from PLMM to DLMM, and respectively 64 and 38 percent increase in σ_v^2 and ICC from PLMM to DLMM. The AIC, BIC and RMSE are lower in DLMM than under PLMM by 18, 17 and 38 percent, respectively. Hence, the proposed DLMM has increased the homogeneity of the observations within clusters and the heterogeneity of the clusters, which in turn increased the model fitness and predictive accuracy.

The PLMM and DLMM in Table 1 are described as 'full models' because they compose of significant and non-significant explanatory variables. We shall now determine if we can obtain similar gains in the model assessment criteria when only significant variables are included in the models. The

Table 1. Summary outputs for the LM, PLMM and DLMM

Clustering Effect	LM	PLMM	DLMM	Covariate	Estimate LM	PLMM	DLMM
σ_v^2		1.8460	5.0900	Intercept	1.2374	3.2081	4.5972
σ_e^2		1.4790	0.5823	enpc	0.8636	0.5092	0.1995
ICC		0.5552	0.8973	proximity	-0.0173	-0.1921	0.0190
AIC	1168.80	1073.09	881.98	eneg	-0.1208	-0.1764	-0.3091
BIC	1194.70	1102.69	911.58	logmag	-0.1982	-0.0956	0.0413
RMSE	1.6689	1.1345	0.7024	logmag_eneg	0.3663	0.0652	0.0274

models with only significant variables are described as 'reduced models'. The enpc is the only significant variable in both the PLMM and DLMM. The summary of the reduced models is in Table 2.

The ICC in DLMM has increased by 38 percent from PLMM, and this increase is the same as it was in the full model. The AIC, BIC, and RMSE have smaller values under DLMM than under PLMM by 18, 18, and 38 percent, respectively. Similarly, the percentage decrease is almost the same as it was in the full model. Although the magnitudes of the AIC and BIC have reduced when non-significant explanatory variables are excluded in both the full PLMM and DLMM; however, the percentage difference between the PLMM and DLMM is almost the same in both the full and reduced models.

The PLMM and DLMM in Table 2 are random intercept models, we recast them to random intercept and slope models to assess the effects of increasing complexity in DLMMs. The summary of the random intercept and slope models is in Table 3.

The ICC in DLMM has increased by 20 percent from PLMM, which is lower than in the random intercept model. The AIC, BIC and RMSE have lower values in DLMM than under PLMM by 17, 17 and 32 percents, respectively. A similar percentage differences are recorded between the PLMM and DLMM as were in the random intercept models; however, the difference is smaller in RMSE.

The comparison reveals a superiority of random intercept and slope DLMM over random intercept DLMM in terms of model fitness. The comparison reveals a superiority of random intercept and slope DLMM over random intercept DLMM in terms of model fitness. This coincides with the work of [14] when random intercept and slope model has smaller value of AIC than in the random intercept model. The model predictive accuracy is higher in DLMM than in PLMM, and also it is higher in random intercept and slope DLMM than in random intercept DLMM. Higher predictive accuracy signifies smaller RMSE.

The above comparison used single sample outcome, $J = 47$, and hence it has not satisfied statistical testing procedure. Therefore, we obtained fifteen sample combinations of the PLMMs and the corresponding DLMMs and compared their respective outcomes. Some sample combinations were replicated to explore possible outcome variability.
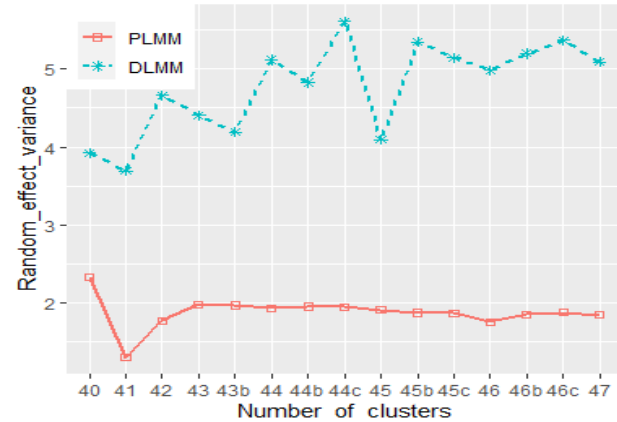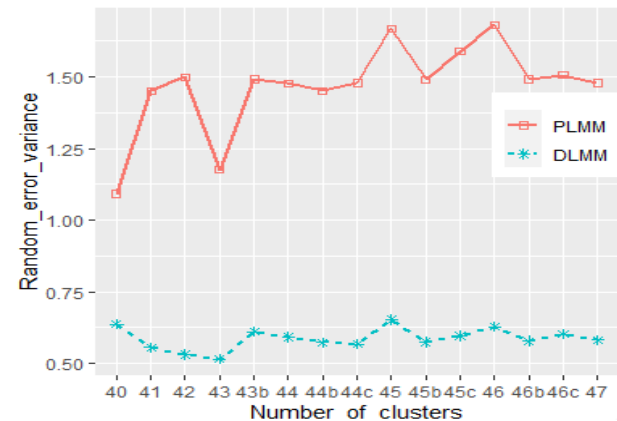

Figure 1. Random Effects Variance

Figure 2. Random Error Variance

5.2. Assessing Clustering Effects between the PLMMs and DLMM

The ICC is a function of σ_v^2 and σ_e^2, and they are presented in Table 4 and Figures 1 and 2.

It shows that σ_e^2 decreases and σ_v^2 increases significantly from PLMM to DLMM. The decrease in the σ_e^2 signifies a homogeneity of observations; that is, the increase of correlations/ dependency of observations within the dual clusters. The increase in the σ_v^2 signifies heterogeneity of clusters; that is, the between cluster variations. The two variance components have greatly affected the ICC, which is significantly higher in the DLMMs. This signifies higher grouping structure in the dual clusters, and higher clustering effects in the DLMMs.

Table 2. Significant Explanatory Variable in Random Intercept Models

ICC		AIC		BIC		RMSE	
PLMM	DLMM	PLMM	DLMM	PLMM	DLMM	PLMM	DLMM
0.5603	0.8972	1066.84	876.36	1081.64	891.16	1.1360	0.7053

Table 3. Significant Explanatory Variable in Random Intercept and Slope Models

ICC		AIC		BIC		RMSE	
PLMM	DLMM	PLMM	DLMM	PLMM	DLMM	PLMM	DLMM
0.7340	0.9149	1052.11	869.71	1074.31	891.91	0.9570	0.6472

Table 4. Random effects variance, random error variance and ICC from PLMMs and DLMMs

| Cluster	σ_v^2		σ_e^2		ICC	
	PLMM	DLMM	PLMM	DLMM	PLMM	DLMM
40	2.3330	3.7956	1.0920	0.6377	0.6811	0.8561
41	1.3060	3.6943	1.4520	0.5564	0.4734	0.8691
42	1.7770	4.6620	1.4990	0.5310	0.5425	0.8978
43	1.9850	4.4058	1.1760	0.5158	0.6280	0.8952
43b	1.9740	5.4642	1.4920	0.6113	0.5695	0.8994
44	1.9340	4.8178	1.4780	0.5924	0.5668	0.8905
44b	1.9570	4.8289	1.4510	0.5736	0.5743	0.8938
44c	1.9490	5.6101	1.4790	0.5677	0.5685	0.9081
45	1.9130	5.1335	1.6680	0.6533	0.5343	0.8871
45b	1.8770	5.3469	1.4900	0.5747	0.5574	0.9030
45c	1.8700	5.1142	1.5880	0.5965	0.5407	0.8956
46	1.7570	4.9810	1.6820	0.6280	0.5110	0.8881
46b	1.8580	5.1988	1.4920	0.5783	0.5546	0.9000
46c	1.8800	5.3696	1.5060	0.6018	0.5552	0.8992
47	1.8460	5.0900	1.4790	0.5823	0.5552	0.8973
Mean	1.8811	4.9008	1.4683	0.5867	0.5608	0.8920
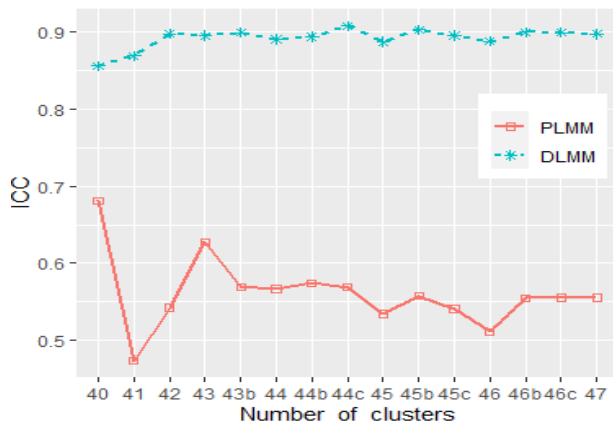
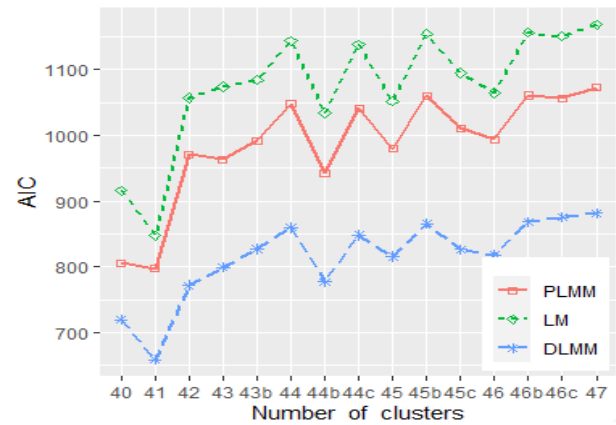
Figure 3. The plot of ICC for the Model assessment

Figure 4. The plot of AIC for the Model assessment

5.3. Assessing Model Fitness and Predictive Accuracy between PLMMs and DLMMs

The model assessment criteria are presented in Table 5 and Figures 4, 5 and 6.

The DLMMs have smaller AIC and BIC than PLMMs, this indicates a significant gain in model fitness in the DLMMs over PLMMs. In addition to the relative selection of the best-fitted model carried out using the AIC and BIC, we supplemented the selection with the assessment of the model's predictive accuracy. The RMSE is significantly lower in the DLMMs, and this

Table 5. Model Assessment Criteria from PLMMs and DLMMs

	AIC			BIC			RMSE		
Cluster	LM	PLMM	DLMM	LM	PLMM	DLMM	LM	PLMM	DLMM
40	915.9	806.2	720.0	940.2	834.0	747.8	1.609	0.965	0.735
41	847.5	798.0	659.4	871.4	825.2	686.6	1.568	1.117	0.677
42	1057.1	971.1	772.1	1082.3	999.9	800.9	1.670	1.144	0.671
43	1074.0	963.8	799.4	1099.5	993.0	828.6	1.564	1.011	0.663
43b	1084.2	991.8	827.6	1109.5	1020.7	856.5	1.694	1.139	0.720
44	1143.8	1048.0	860.0	1169.5	1077.5	889.5	1.675	1.137	0.711
44b	1034.4	943.0	777.4	1059.4	971.5	805.9	1.696	1.118	0.693
44c	1138.1	1041.8	848.7	1163.8	1071.0	878.1	1.681	1.136	0.696
45	1052.5	979.8	816.0	1077.5	1008.4	844.5	1.743	1.198	0.738
45b	1154.6	1060.3	866.4	1180.4	1089.8	895.9	1.672	1.141	0.699
45c	1094.9	1011.3	826.9	1120.2	1040.3	855.9	1.715	1.174	0.709
46	1064.9	993.7	818.4	1090.0	1022.4	847.1	1.732	1.205	0.723
46b	1156.4	1060.9	868.5	1182.2	1090.4	898.0	1.678	1.140	0.701
46c	1150.6	1056.9	875.8	1176.3	1086.3	905.3	1.683	1.145	0.714
47	1168.8	1073.1	882.0	1194.7	1102.7	911.6	1.669	1.135	0.702
Mean	1075.8	986.6	814.6	1101.1	1015.5	843.5	1.670	1.127	0.704
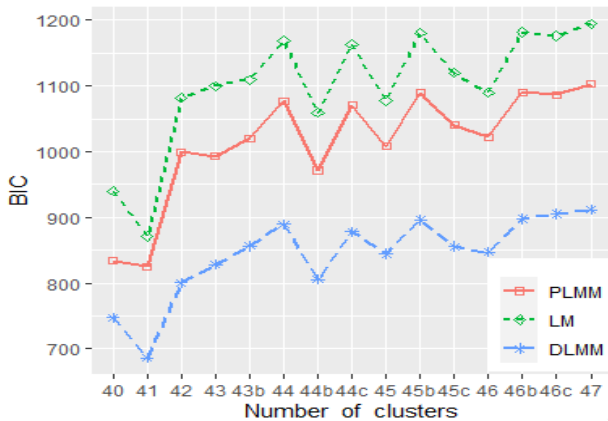
Figure 5. The plot of BIC for the Model assessment
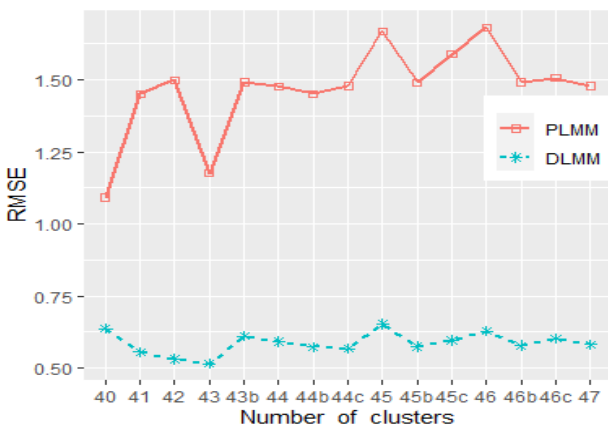
Figure 6. The plot of RMSE for the Model assessment

explains the significant gain in model predictive accuracy in the proposed DLMM over the existing PLMMs.

6. Conclusion

The paper proposed the development and analysis of DLMM on the dual clusters derived from the primal clusters. The clustering similarity in the dual clusters was based on the commonly occurring phenomenon or the experimental designs, and the similarity in the dual clusters was based on the multivariate clustering algorithms. Findings revealed that observations in the dual clusters are more correlated than in the primal clusters. The proposed DLMM is relatively more efficient than the classical PLMM based on the results of the models' assessment criteria (AIC, BIC, and RMSE) in which the DLMM yielded minimum values of the assessment criteria. Therefore, the proposed DLMM outperformed the classical PLMM in terms of model fitness and predictive accuracy.

References

[1] O. S. Adesina, "Bayesian Multilevel Models for Count Data", Journal of the Nigerian Society of Physical Sciences **3** (2021) 224. doi:10.46481/jnsps.2021.168

[2] N. M. Laird & J. H. Ware, "Random-effects models for longitudinal data", Biometrika (1982) 963.

[3] A. Abadie, S. Athey, G. W. Imbens & J. Wooldridge, "When should you adjust standard errors for clustering?" Available at https://economics.mit.edu/files/13927 Deposited (2017).

[4] Y. Bello, S.U. Gulumbe & S. A. Yelwa, "Simultaneous application of agglomerative algorithms on interval measures for better classification of crime Data across the ctates in Nigeria", Research Journal of Applied Science **7** (2012) 41. doi: 10.3923/rjasci.2012.41.47

[5] W. Härdle & Z. Hlávka, *Multivariate statistics: Exercises and solutions*, Springer-Verlag, New York, 2007.

[6] B. Villarroel, G. Marshall & Baron, A., "Cluster analysis using multivariate mixed effects models", Statistics in Medicine **28** (2009) 2552. doi: 10.1002/sim.3632

[7] G. Celeux, O. Martin & C. Lavergne, "Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments", Statistical Modelling **5** (2005) 1. doi.org/10.1191/1471082X05st096oa

[8] R. L. Wears, "Advanced statistics: Statistical methods for analyzing cluster and cluster-randomized Data", Academic Emergency Medicine **9** 2002, doi.org/10.1197/aemj.9.4.330

[9] O. J. Ibidoja, F. P. Shan, Mukhtar, J. Sulaiman & M. K. M. Ali "Robust M-Estimators and Machine Learning Algorithms for Improving the Predictive Accuracy of Seaweed Contaminated Big Data", Journal of the Nigerian Society of Physical Sciences **9** (2023) 1137. doi:10.46481/jnsps.2022.1137

[10] R. Lehtonen, C. Särndal & A. Veijanen, "The Effect of model choice in estimation for domains, including small domains", Statistics Canada **29** (2003) 33.

[11] H. Akaike, "Information theory and an extension of the maximum likelihood principle", In International Symposium on Information Theory, Ed. B. N. Petrov and F. Csaki, pp. 267-81. Budapest: Akademia Kiado (1973).

[12] F. Vaida & S. Blanchard, "Conditional Akaike information for mixed-effects models", Biometrika **92** (2005) 351. doi.org/10.1093/biomet/92.2.351

[13] S. Müller, J. Scealy & A. Welsh, "Model selection in linear mixed models", Statistical Science **28** (2013) 135. doi.org/10.1007/s10182-019-00359-z

[14] A. F. Zuur, E. N. Ieno, N.J. Walker, A. A. Saveliev & G.M. Smith, *Mixed effects models and extensions in Ecology with R*, Springer Science+Business Media New York, 2009.

[15] B. T. West, K. B. Welch & A. T. Galecki, *Linear Mixed Models: A Practical guide using statistical software*, Chapman & Hall/CRC Boca Raton, 2007.

[16] J. De Leeuw, E. Meijer & H. Goldstein, *Handbook of multilevel analysis*, Springer New York (2008).

[17] I. Ercanli, A. Gunlu & E. Z. Baṣkent, "Mixed effect models for predicting breast height diameter from stump diameter of Oriental beech in Göldağ", Scientia Agricola (2014).

[18] S. Greven & T. Kneib, "On the behavior of marginal and conditional Akaike information criteria in linear mixed models", Johns Hopkins University, Department of Biostatistics Working Papers, Paper 179. http://www.bepress.com/jhubiostat/paper179/ Deposited (2008).

[19] W. Härdle & L. Simar, *Applied multivariate statistical analysis*, 2nd edition, Springer-Verlag New York (2003).

[20] J. H.Ward, "Hierarchical grouping methods to optimize an objective function", Journal of the American Statistical Association **58** (1963) 236. doi.org/10.2307/2282967

[21] R. Elgie, C. Bucur, B. Dolez & A. Laurent, "Proximity, candidates, and presidential power: How directly elected presidents shape the legislative party system", Political Research Quarterly **67** (2014) 467. doi.org/10.1177/1065912914530514