

Nonlinear mixed effects model for cumulative tree  
bole volume

Project work report

Likelihood-based inference for hierarchical/mixed  
statistical models

Nordic informatics network

Tune Landboskole, Greve, Denmark

Leakemariam Berhe

August 17, 2005

## **Acknowledgements**

I would like to thank the tireless effort of the teachers: Dr Henrik Stryhn, Dr William John Brown and Dr Rasmus Waagepetersen for diligently delivering the course and sharing their knowledge and experience including in the intensive lab work. Moreover, Henrik's technical support and encouragement for the completion of this work was wonderful for me. Finally, I would also like to thank for the excellent administrative management, food and accomadation facilities made possible by professor Anders Ringgaard Kristensen and the sponsor NOVA University network.

## Introduction

Tree growth and yield modeling data usually exhibits either a repeated measures or longitudinal structure. In such nested structure data, several or repeated measurements within sampling unit (tree) are correlated and should be taken into account in the modeling effort for validity of inferences and improved estimation. Generally, in the clustered forestry data setting, it may be appropriate to distinguish three components of random variability in the modeling process; mainly, the random effects that stems from heterogeneity between individual (tree) profiles; serial correlation between observations within sampling unit; and measurement error.

To draw valid scientific inferences, longitudinal data require special statistical methods to account correlation between set of observations in one subject (Diggle et al., 2002). Schabenberger (1995) indicated that random effects does not remove serial correlation completely and both random terms and direct modeling of the within subject correlation is required. Chi and Reinsel (1989) proposed modeling of longitudinal data by including of the AR(1) or other time series specification in addition to possible random effects. On the other hand, Gregoire and Schabenberger (1996) argues that mixed-effects model effectively annihilates the within subject correlation. Diggle et al(2002), argued that in specific applications the effects of serial correlation may be dominated by the combination of random effects and measurement errors.

Therefore, in a grouped data type where several unbalanced measurements are taken within a subject such as in a tree merchantable volume models, modeling and estimation procedures should be flexible enough to accommodate quite general intraindividual covariance structure.

Application of mixed effects models is a promising approach for improving forest estimation and valid inferences. While mixed effects models are widely used in many other areas such as in health sciences, they are not yet developed in forestry practice. Accordingly, the purpose of this work is to develop nonlinear mixed effects models for estimation of tree merchantable volume.

Gregoire and Schabenberger (1996) developed a nonlinear mixed effects model for the cumulative tree bole volume (variable top merchantable volume). The nonlinear mixed effects model form used by Gregoire and Schabenberger was (1):

$$V_{d_{ij}} = (\beta_1 + \beta_2 X_i) \exp(-\beta_3 t'_{ij} \exp(\beta_4 t_{ij})) + \varepsilon_{ij} \quad (1)$$

where  $V_{d_{ij}}$  = cumulative bole volume on the  $i$ th tree ( $i = 1, \dots, n_j$ ;  $j = 1, \dots, m_i$ ) to the upper diameter  $d_j$  at the  $j$ th location;  $t_{ij} = \frac{d_{ij}}{D_i}$ ;  $t'_{ij} = t_{ij}/1000$ ;  $D_i$  = dbh of the  $i$ th tree  $X_i = D_i^2 H_i / 1000$ .

In this work also model(1) is used to develop a nonlinear mixed effects model for cumulative tree bole volume of *Cupressus lusitanica*.

## Data and methods

Cumulative bole volume data from 20 trees of *Cupressus lusitanica* at Shashemene Forest Enterprise, Ethiopia, is used for this work. Normally, each tree is felled and diameters and other measurements are measured at short intervals to generate tree cumulative bole volume. Measurement points are at 0.20(stump height), 0.35, 0.50, 0.65, 0.80, 1.0 and 1.3 meters in the lower section and at every one meter interval from the diameter at breast height (1.3). The volume of all sections to every measurement point from the ground are computed and accumulated to determine respective cumulative volume. A total of 562 observations with an average of 28 cumulative bole volume observations from each tree are used for this study.

The nlme method in R statistical software is used to analyze the data.

## Results and Discussion

Cumulative tree bole volume plots shown in Fig 1 indicate strong differences among the trees. Residuals distribution within a tree (Fig2) also pronounces considerable difference between trees.

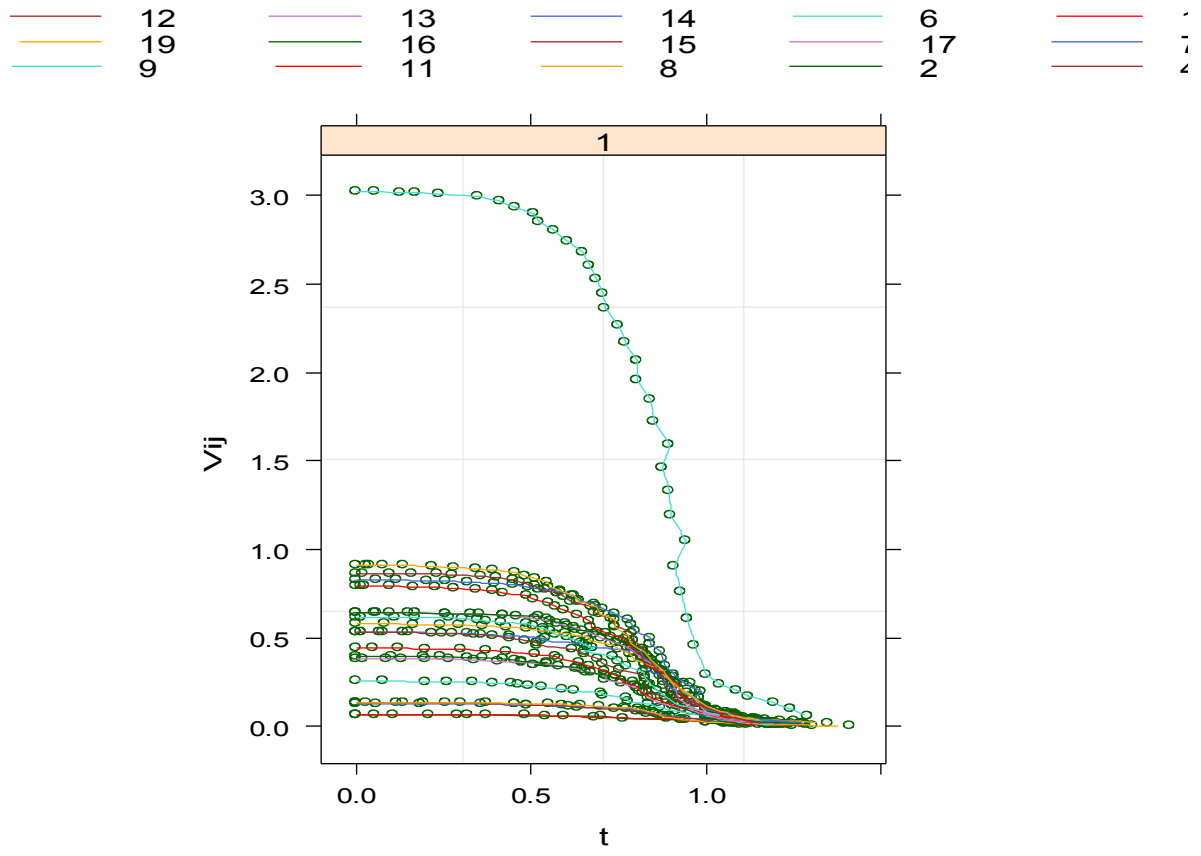


Fig 1: Cumulative volume ( $V_{ij}$ ) of each tree versus  $t$  ( $d/D$ ) plots

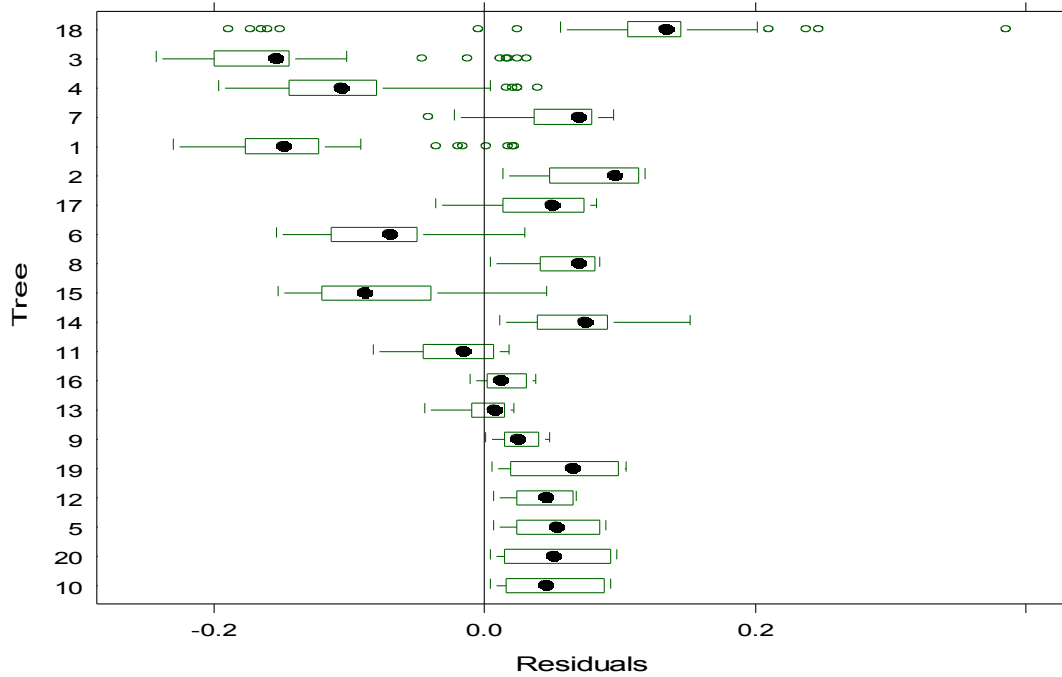


Fig 2: Non linear least squares fit residuals distribution within tree

Residuals for some trees are positive and for some negative only pronouncing individual variation.

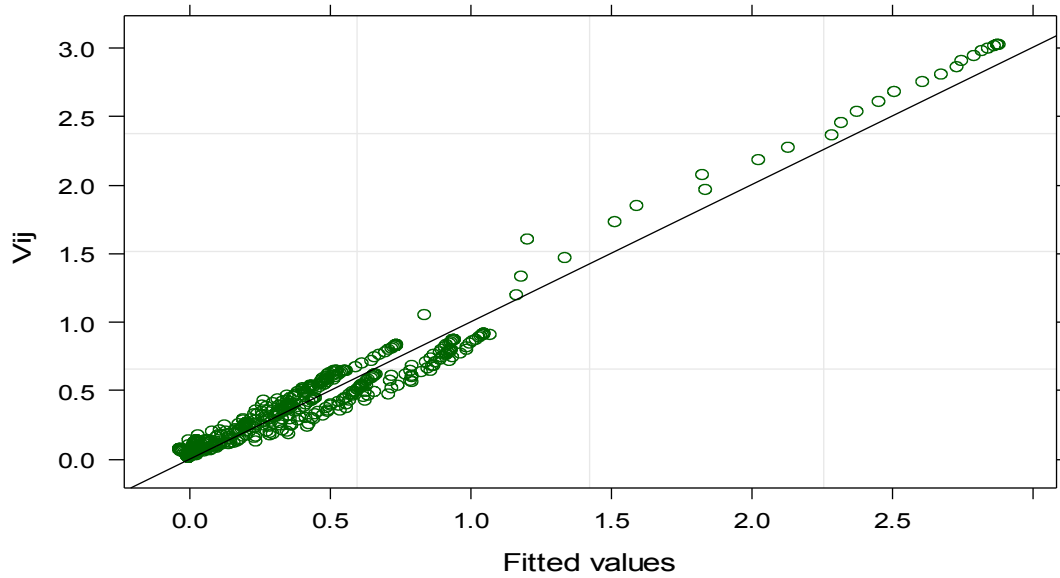


Fig3: Observe ( $V_{ij}$ ) versus fitted plot from the nls model

It is also evident from Fig3 that the nls model is not a good fit.

For the selection of the best nlme model , different combination of random effects with different covariance structure and correlation structure of errors are compared in this study. The result is summarized in Table 1.

Table 1: Observed AIC, BIC and LogLik when different combination of random effects , covariance structure ( $\Delta$ ) and correlation structure are used in model(1)

Random effect(s)	AIC	BIC	LogLik	$\Delta$	Correlation structure
None (nls)	-1077.16	-1059.84	542.58	-	-
$\gamma_1$	-1864.94	-1838.95	938.47	General	-
$\gamma_2$	-1890.03	-1864.05	951.02	General	-
$\gamma_3^*$	-	-	-	-	-
$\gamma_4$	-1265.51	-1239.52	638.75	General	-
$\gamma_1, \gamma_2$	-1887.33	-1852.67	951.66	“	-
$\gamma_1, \gamma_3$	-20008.68	-1974.03	1012.34	“	-
$\gamma_1, \gamma_4$	-1984.94	-1950.29	1000.47	“	-
$\gamma_2, \gamma_3$	-2031.96	-1996.64	1023.65	“	-
$\gamma_2, \gamma_4$	-2008.48	-1973.83	1012.24	“	-
$\gamma_3, \gamma_4$	-1523.74	-1493.40	768.87	“	-
$\gamma_1, \gamma_2, \gamma_3$	-2026.92	-1979.28	1024.46	“	-
$\gamma_1, \gamma_2, \gamma_4$	-2003.94	-1956.30	1012.97	“	-
$\gamma_2, \gamma_3, \gamma_4$	-2094.47	-2046.82	1058.23	“	-
$\gamma_1, \gamma_3, \gamma_4$	-2004.68	-1955	1012.34	“	-
$\gamma_1, \gamma_2, \gamma_3, \gamma_4$	-2084.18	-2021.21	1058.09	“	-
$\gamma_1, \gamma_2, \gamma_3, \gamma_4$	-2306.86	-2259.21	1164.43	diagonal	ARMA(p=0,q=2)
$\gamma_2, \gamma_3, \gamma_4$	-2307.82	-2260.18	1164.91	diagonal	ARMA(p=0,q=3)
$\gamma_2, \gamma_3, \gamma_4$	-2308.86	-2265.55	1164.43	diagonal	ARMA(p=0,q=2)
$\gamma_2, \gamma_3, \gamma_4$	-2140.62	-2101.63	1079.31	diagonal	ARMA(p=0,q=1)

\* fail to converge

\*\* for any model with AR(1) or  $p \geq 1$  in ARMA failed to converge

On the bases of Table 1, we have compared the nls and three best nlme models with likelihood ratio test in Table 2.

Table2: Likelihood ratio test of the nls and three best nlme models.

	Model*	df	AIC	BIC	logLik	Test	L.Ratio	p-value
fnlme234	1	11	-2094.467	-2046.821	1058.2335			
fnlme234c2	2	10	-2308.862	-2265.547	1164.4308	1 vs 2	212.3946	<.0001
fnlme234c3	3	11	-2307.823	-2260.177	1164.9117	2 vs 3	0.9617	0.3268
nls1	4	4	-1077.164	-1059.838	542.5819	3 vs 4	1244.6596	<.0001

\*

fnlme234 =  $\gamma_2, \gamma_3, \gamma_4$  + general positive definite covariance matrix

fnlme234c2 =  $\gamma_2, \gamma_3, \gamma_4$  + simple diagonal matrix + ARMA(p=0, q=2)

fnlme234c3 =  $\gamma_2, \gamma_3, \gamma_4$  + simple diagonal matrix + ARMA(p=0, q=3)

From Table 2, we clearly see fnlme234c2 and fnlme234c3 are the best models. However, in this work the parsimonious model fnlme234c2 is selected as best and further exploration is made on this model.

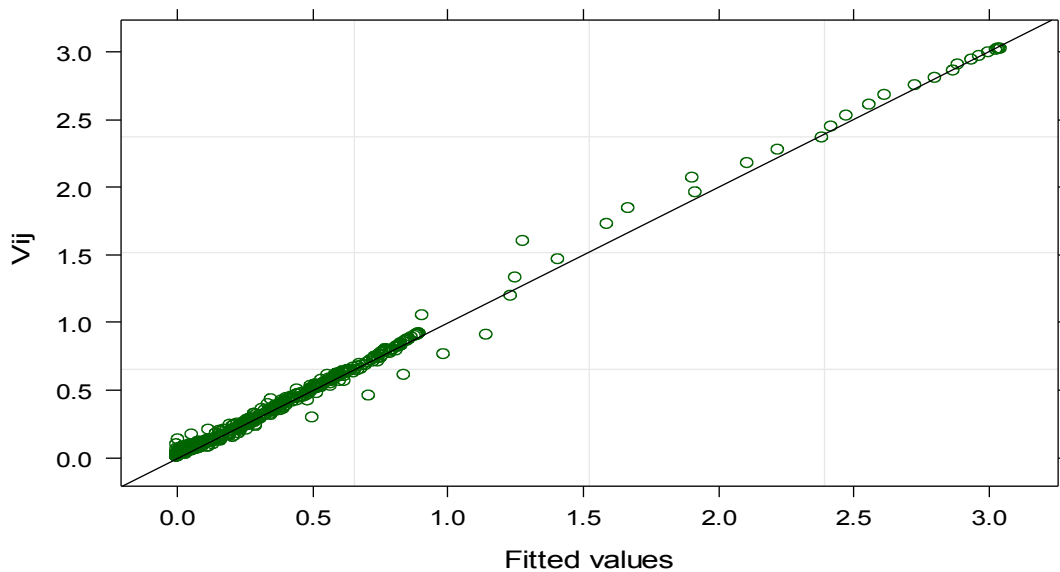


Fig4: Observed ( $V_{ij}$ ) versus fitted from fnlme234c2 model

As compared to the nls plot in Fig 3, the nlme model (Fig 4) is best fitted.

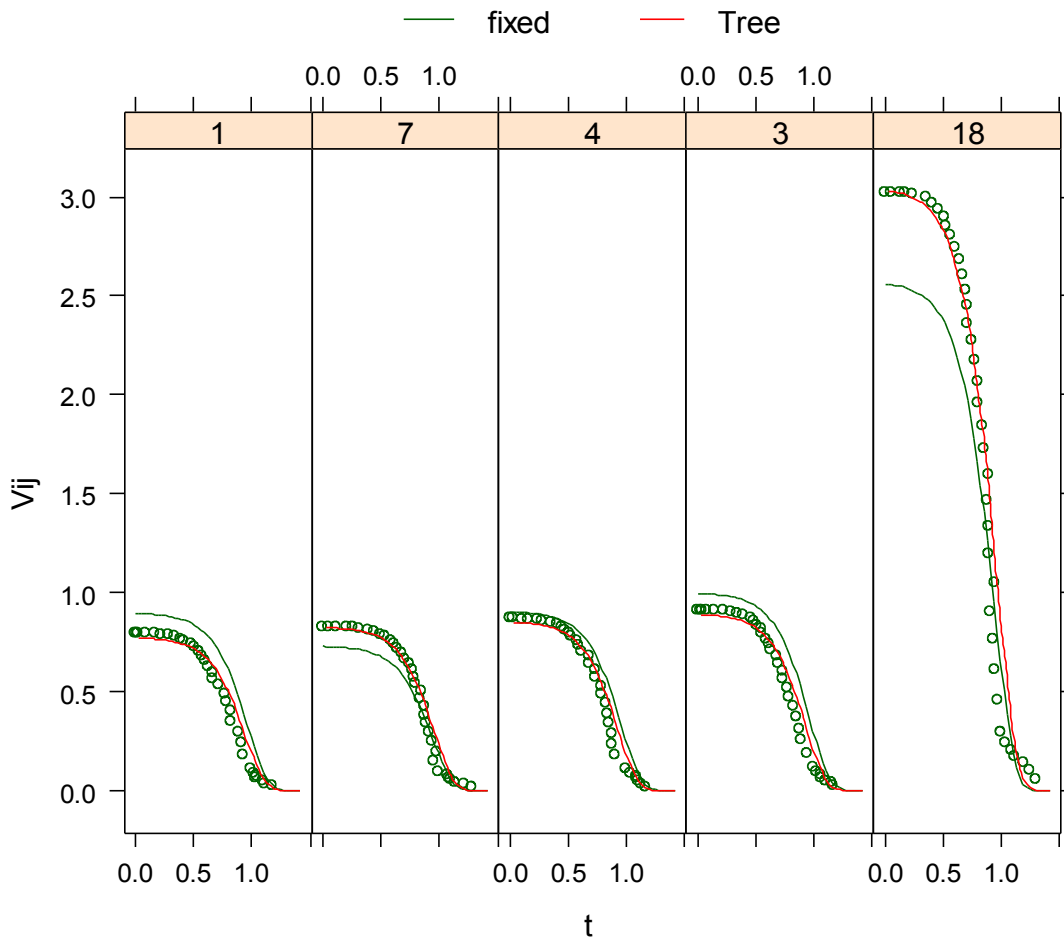


Fig 5: Population predictions (fixed), within group predictions (Tree), and observed cumulative tree bole volume (circles) versus  $t(di/D)$  for fnlme234c2 model.

Further more, Fig 5 demonstrated the within tree predictions are in close agreement with the observed cumulative tree bole volume illustrating the nlme model can accommodate individual effects. Therefore, we may conclude that the fnlme234c2 model provides a good representation of the tree cumulative bole volume data.

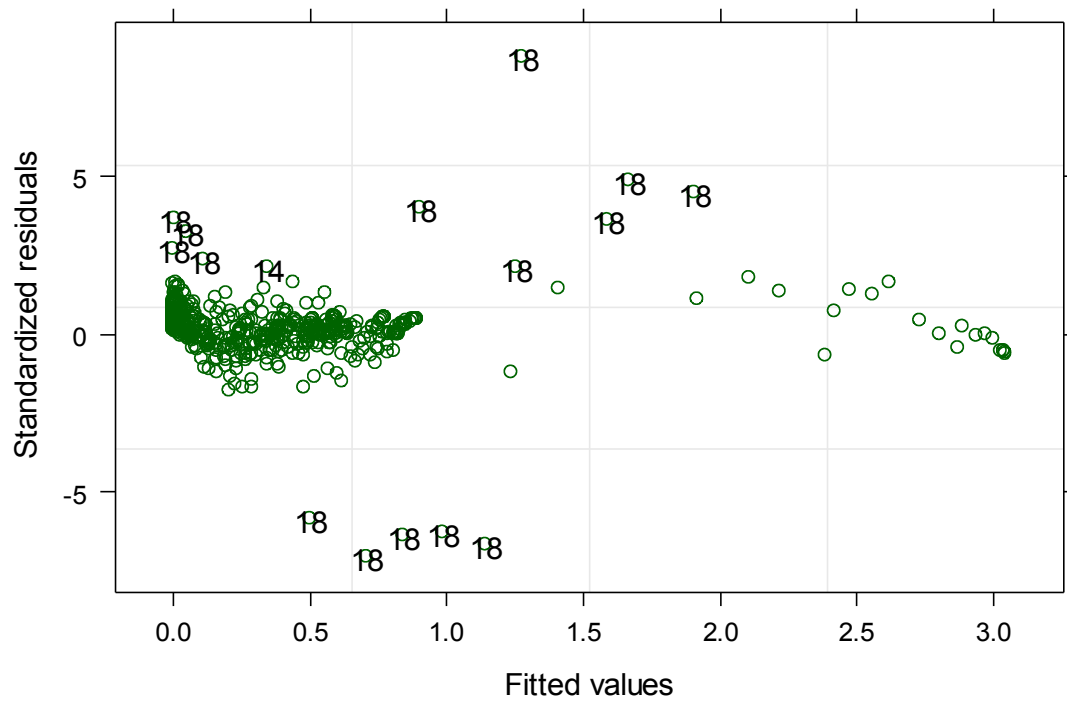
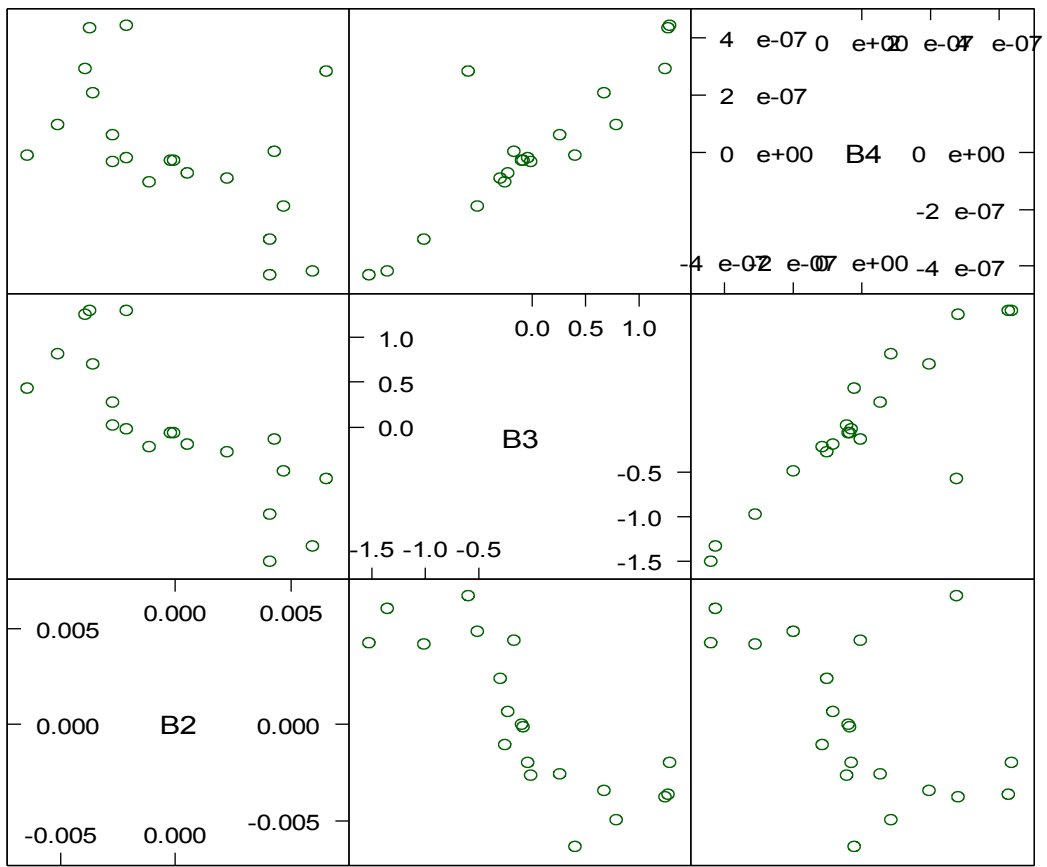


Fig 6: Standardized residuals versus fitted values from fnlme234c2.

With the exception to the biggest tree (18) in the data, the residuals are within  $\pm 3$  standard deviation.



Scatter Plot Matrix

Fig7: Pairs plot for the random effects estimates corresponding to fnlme234c2

The scatter plots in Fig 7 may suggest that there seems to appear considerable correlations between the individual parameter estimates. This connotes further study in this regard.

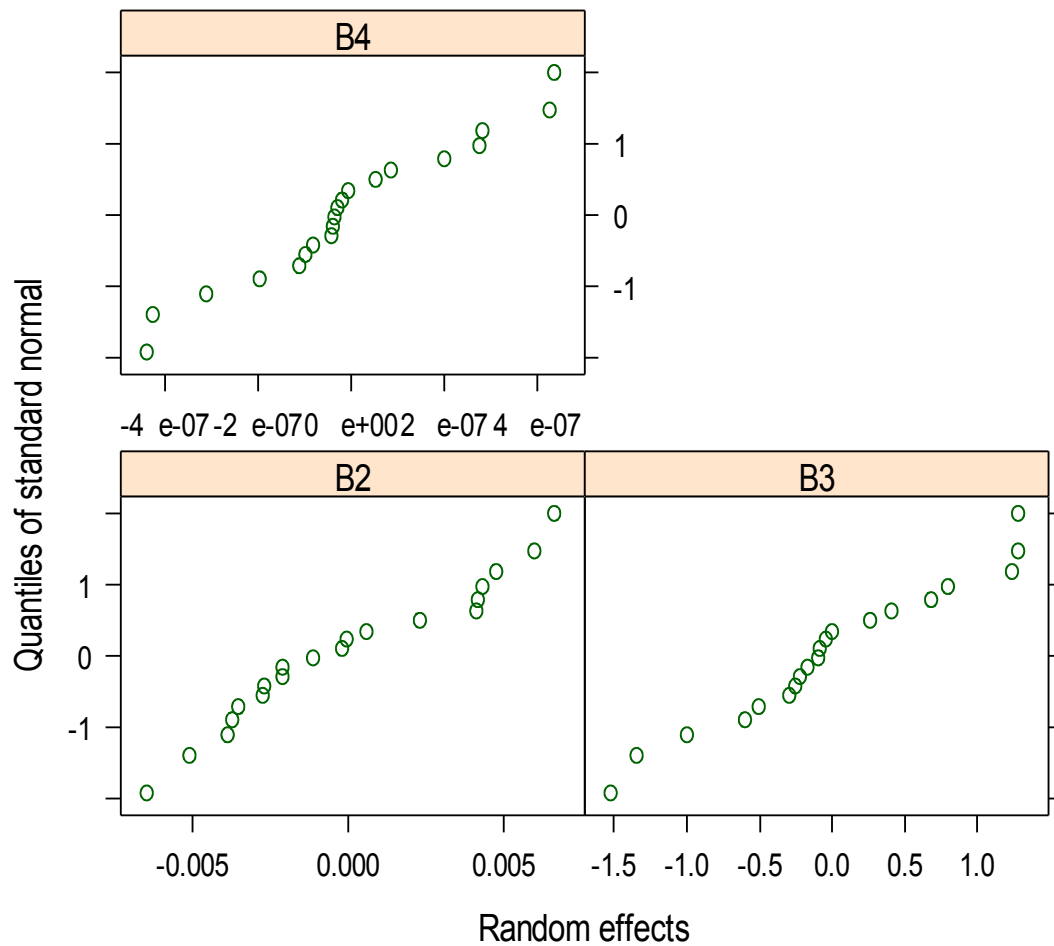


Fig9: Normal plot of estimated random effects for the fnlme234c2 nlme model fit.

The assumption of normality seems reasonable for all three random effects

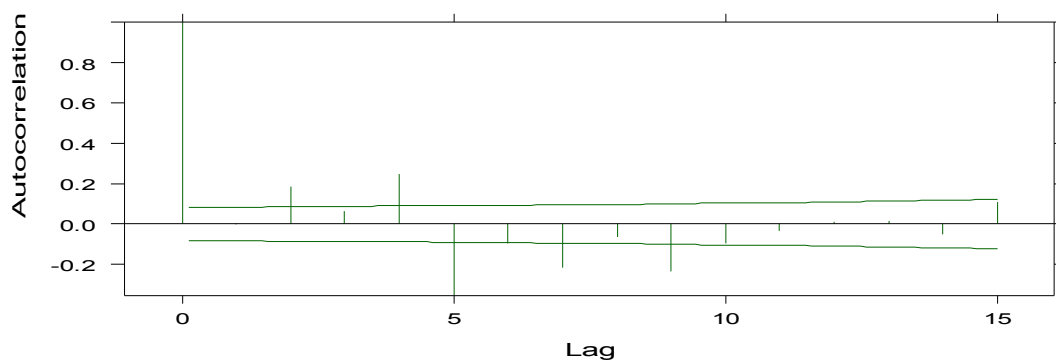


Fig10: Empirical autocorrelation function plot corresponding to the normalized residuals of the fnlme234c2 fit.

Despite the incorporation of the ARMA( $p=0,q=2$ ) in the model for the correlation structure within tree, it failed to completely remove. Probably this correlation structure is not the appropriate. Since convergence failed for  $p \geq 1$ , choices of the correlation structure were limited in this study.

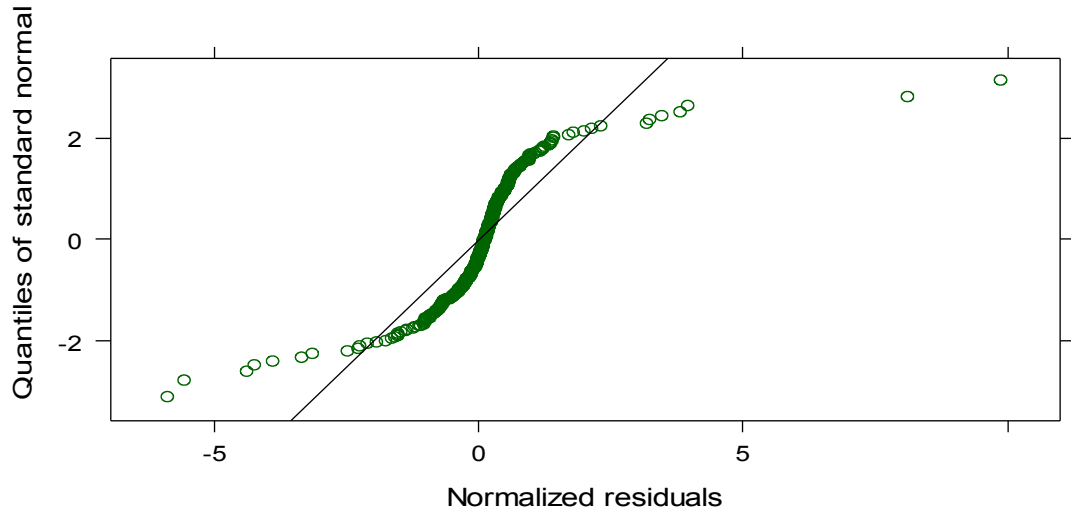


Fig 11: Normal plots of normalized residuals for the fnlme234c2 nlme model fit.

Fig 11 may warrant further study with the model construction, choices of random effects, covariance matrix and correlation structure.

The parameter estimates and correlation matrix of the random effects for the fnlme234c2 model was shown below.

Structure: Diagonal  
 B2            B3            B4            Residual  
 StdDev: 0.004367804 1.048924    0.0002770453    0.03598995

Parameter estimate(s):  
 Theta1    Theta2  
 0.6438964 0.4716641

	Value	Std.Error	DF	t-value	p-value
B1	0.008917	0.0127456	539	0.69958	0.4845
B2	0.036016	0.0015718	539	22.91483	0.0000
B3	6.065473	0.8147044	539	7.44500	0.0000
B4	5.796626	0.1389599	539	41.71437	0.0000

Correlation:  
 B1    B2    B3  
 B2 -0.697  
 B3 0.016 0.064  
 B4 0.000 -0.061 -0.921

As we can learn from Fig 6, most of residuals of tree number 18 are outliers. Accordingly, since tree number 18 is very big tree as compared to others, the nlme analysis is repeated after removing this outlier tree.

Table3: The observed AIC, BIC and LogLik for the best models with and without the outlier tree no 18.

Model	df	AIC	BIC	LogLik	Test	L.Ratio	p-value
fnlme234r	1	-2646.198	-2599.343	1334.099			
fnlme234rc	2	-2803.216	-2760.621	1411.608	1 vs 2	155.0183	<.0001
fnlme234c2	3	-2308.862	-2265.547	1164.4308			

fnlme234r =  $\gamma_2, \gamma_3, \gamma_4$  +general positive definite covariance matrix +removed tree no 18

fnlme234rc =  $\gamma_2, \gamma_3, \gamma_4$  + simple diagonal matrix +ARMA(p=0,q=2) +removed tree no 18

fnlme234c2 =  $\gamma_2, \gamma_3, \gamma_4$  +simple diagonal matrix +ARMA(p=0,q=2)+ included tree no 18

After deleting the outlier tree, the AIC, BIC and -2loglik of the models are decreased very much as compared to the nlme models with the outlier tree (Table3) indicating the improvement of the models performance.

Accordingly, Further analysis of the best model (fnlme234rc) is carried out below.

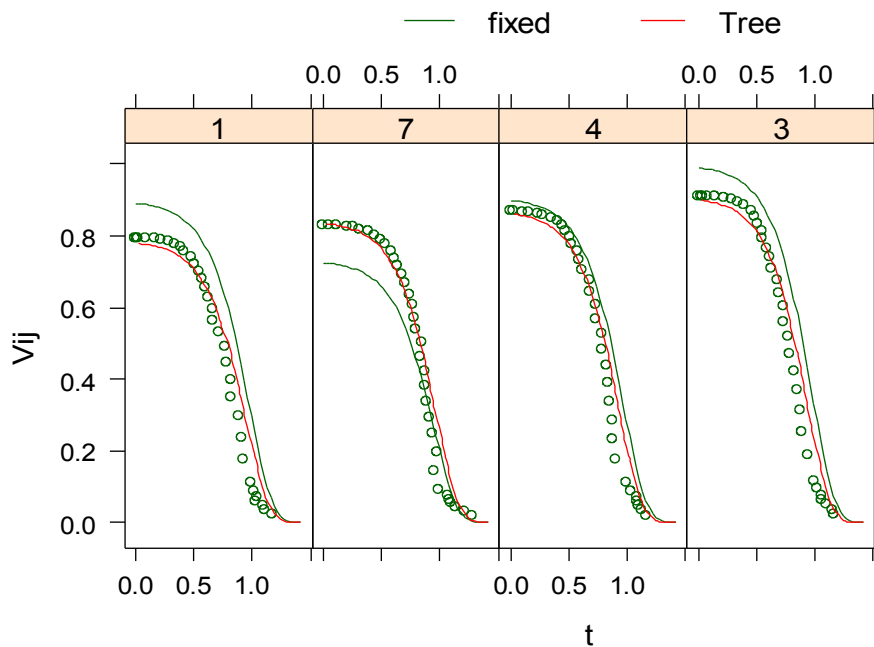


Fig 12: Population predictions (fixed), within group predictions(Tree), and observed cumulative tree bole volume (circles) versus  $t(di/D)$  for fnlme234rc model.

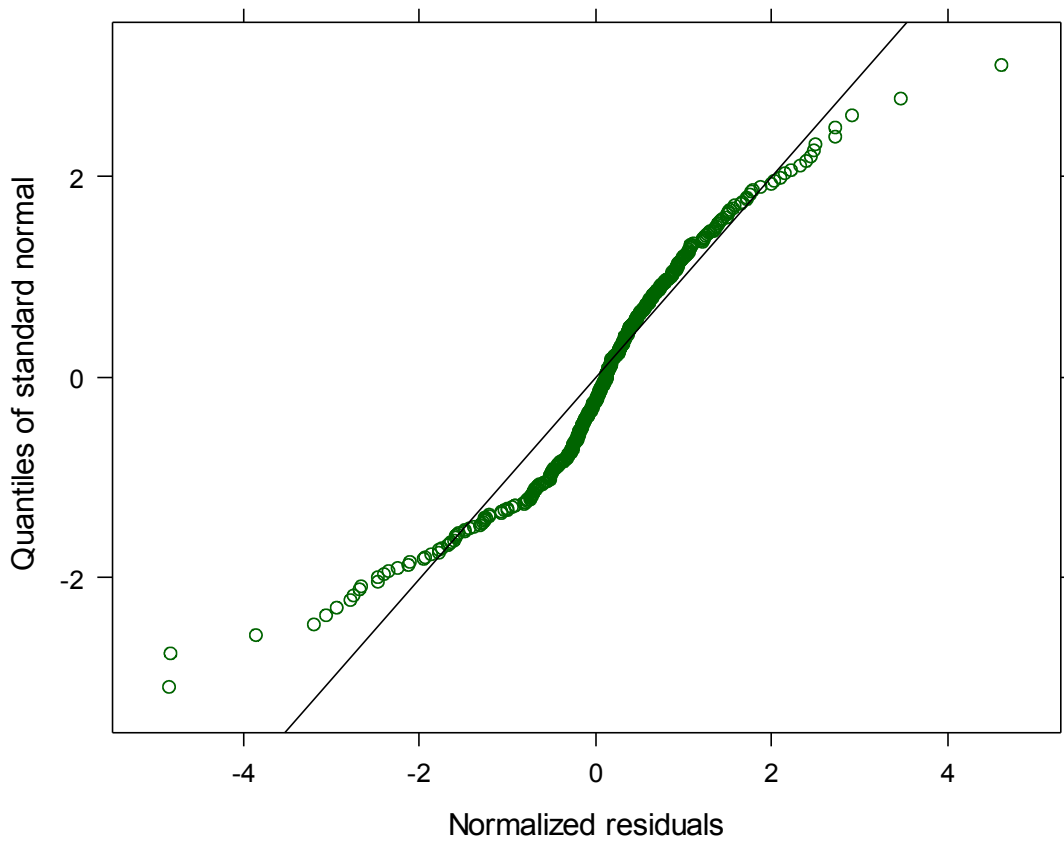


Fig 13: Normal plots of normalized residuals for the fnlme234rc nlme model fit.

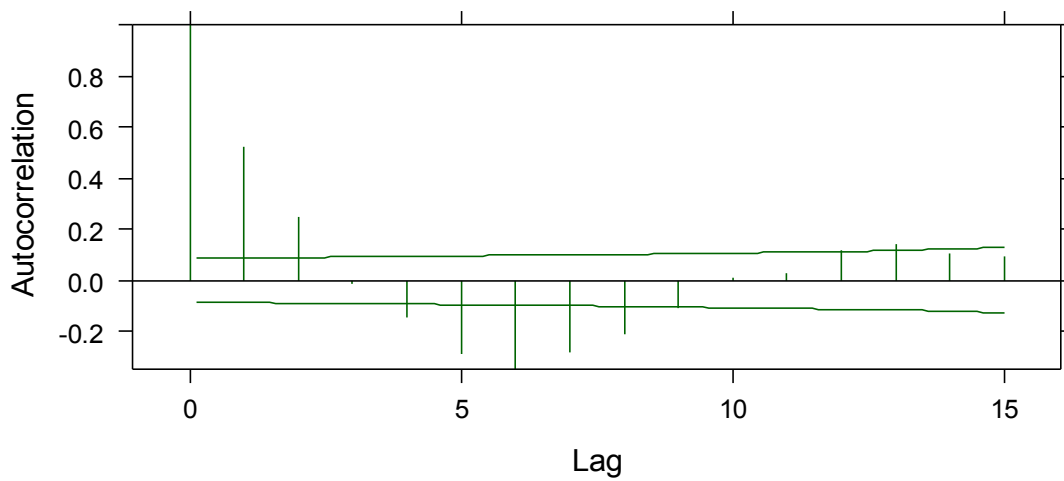


Fig14: Empirical autocorrelation function plot corresponding to the normalized residuals of the fnlme234r fit (without correlation structure).

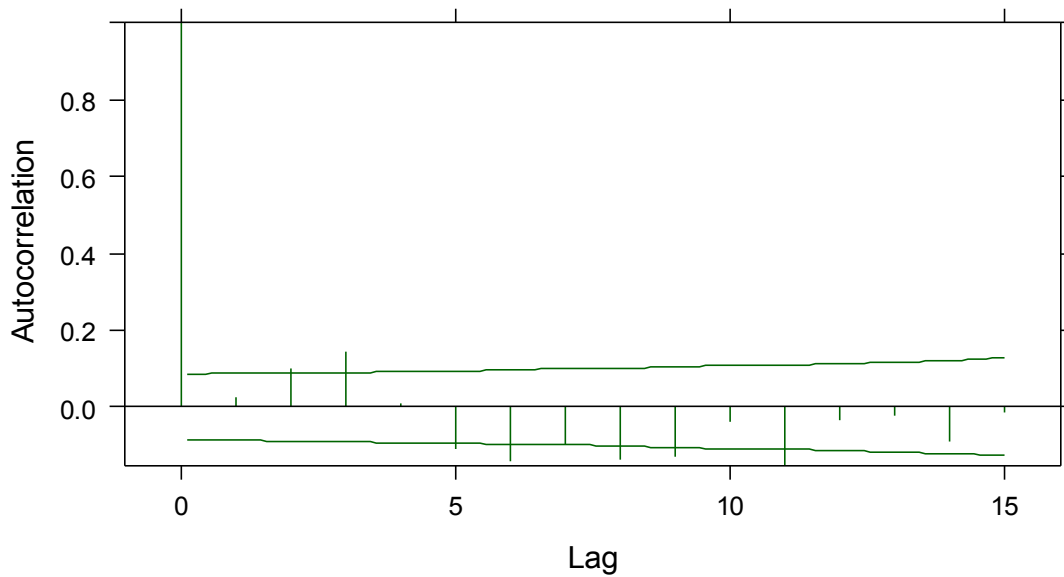


Fig15: Empirical autocorrelation function plot corresponding to the normalized residuals of the fnlme234rc fit

The correlation structure (ARMA( $p=0,q=2$ )) in model fnlme234rc, reasonably removed the serial correlation within tree (Fig 15) as compared to the fnlme234r in fig14.

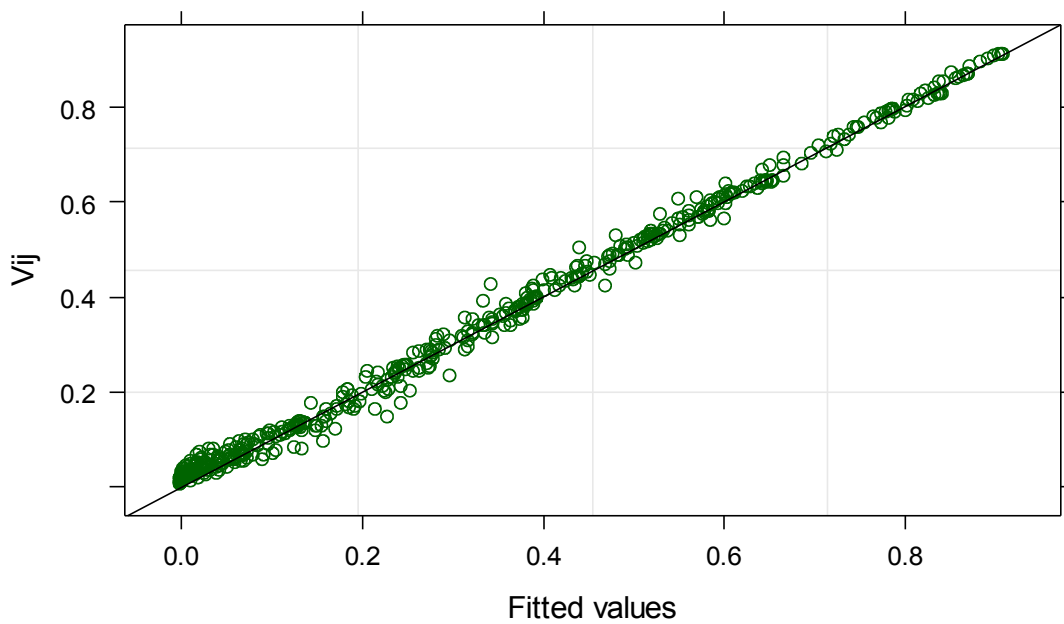


Fig16: Observed ( $V_{ij}$ ) versus fitted from fnlme234rc model

The fit of the model `fnlme234rc` (fig 16) is also considerably improved compared to the fit of model `fnlme234c2` (fig4).

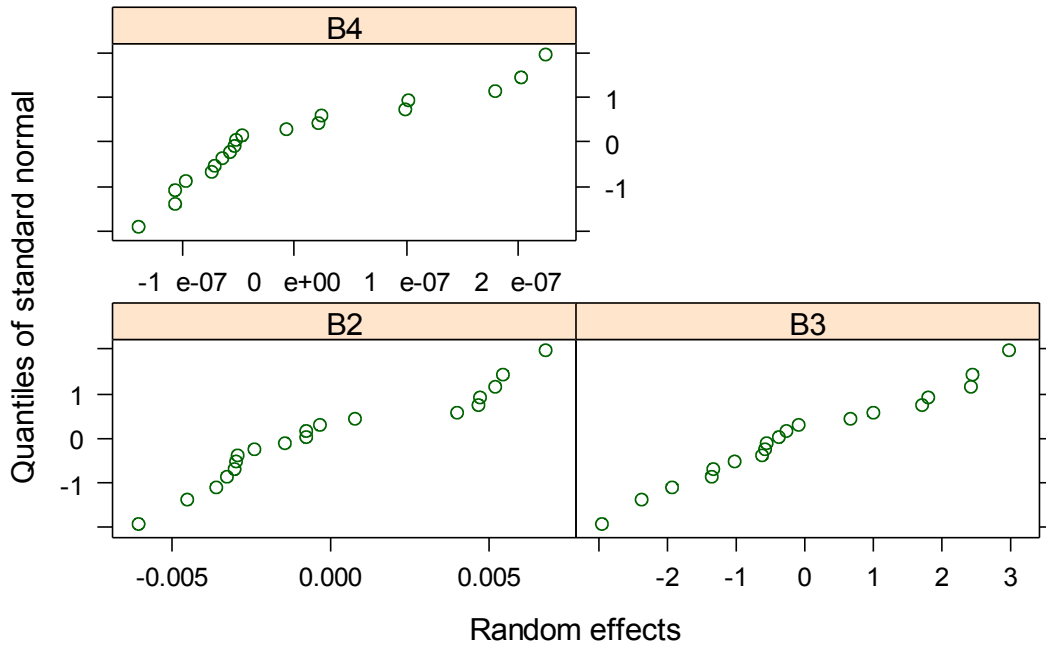


Fig17: Normal plot of estimated random effects for the `fnlme234c2` nlme model fit.

As shown in Fig 17, the estimated random effects are also quite reasonable to say normally distributed.

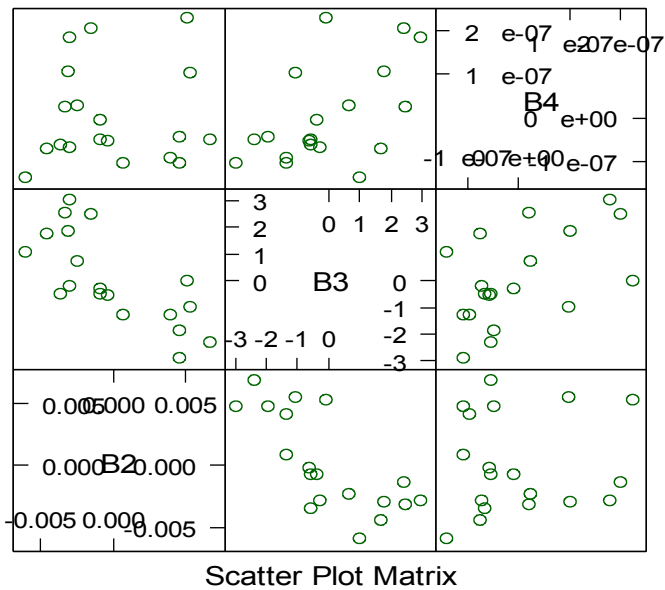


Fig17: Pairs plot for the random effects estimates from `fnlme234rc`.

As compared to the models before deletion of the outlier the correlation between the random effects seem to be decreased from model fnlme234rc (compare Fig7 and Fig 17).

The parameter estimates and the correlation matrix of the model fnlme234rc is shown below.

Nonlinear mixed-effects model fit by maximum likelihood

Model:  $V_{ij} \sim (B1 + B2 * X) * \exp(-B3 * t1 * \exp(B4 * t))$

Data: volume.gr

AIC	BIC	logLik
-2803.216	-2760.621	1411.608

Random effects:

Formula: list(B2 ~ 1, B3 ~ 1, B4 ~ 1)

Level: Tree

Structure: Diagonal

	B2	B3	B4	Residual
StdDev:	0.004149638	1.960221	0.0001778805	0.01796155

Correlation Structure: ARMA(0,2)

Formula: ~1 | Tree

Parameter estimate(s):

	Theta1	Theta2
	0.6171154	0.3545187

Fixed effects: B1 + B2 + B3 + B4 ~ 1

	Value	Std.Error	DF	t-value	p-value
B1	0.012278	0.0084231	501	1.45765	0.1456
B2	0.035834	0.0014154	501	25.31704	0.0000
B3	11.818115	1.2236330	501	9.65822	0.0000
B4	5.041845	0.1022755	501	49.29672	0.0000

Correlation:

	B1	B2	B3
B2	-0.679		
B3	0.028	0.053	
B4	-0.006	-0.059	-0.900

## Conclusion

It is very apparent from this study that the nonlinear mixed effects models for cumulative tree bole volume performs well than its counterpart nonlinear least squares model.