Analyzing the Linear Mixed Effects (LME) Models

Vignesh J Muralidharan

Introduction: Production of Biodiesel is getting easier day by day. But, maintaining the quality of a batch of biodiesel depends on the variables maintaining in each experiment level. Here at University of Idaho we produce almost 350 gallons of biodiesel once in every two weeks from various types of feedstocks like Vegetable, Canola, Waste, Mustard and even switchgrass. In this scenario using different enzyme to catalyze the reaction of production of biodiesel is one other major part in comparing the load production for each week.

So first, we get different types of oil to be used in the process around 7 in this experiment with different enzyme used for reaction depending on the availability in the laboratory. We have total of 2 reactors Flash and the Normal bioreactors. The temperature is the major variable which directly affects the production of biodiesel and here based on the reactor type and enzyme type the temperature is set to either High (75°C) or Low (45°C).

In this experiment, I would like to check if temperature is making significant effect on the production of biodiesel or not with each batch set of enzymes. Based on these assumptions I am considering the enzyme or batch number and scenario of using different oils as the random effect for the model and temperature and the reactor type as the fixed effect of the model having amount of production of biodiesel (gal) as a response variable.

Data Visualization: in the mixed model we add one or more random effects to the fixed effects. These random effects essentially give structure to the error term "e". In the case of my model here, I add a random effect for the subject and this characterizes idiosyncratic variation that is due to individual difference. Here in the graph with respect to the low and high temperatures the total biodiesel production is almost like all the batches of production. And regarding the boxplot here for each scenario the biodiesel load of production is a total variability.

The temperature with high and low have different variance in the biodiesel load. Here we can clearly say that whenever temperature goes down the biodiesel load goes down.
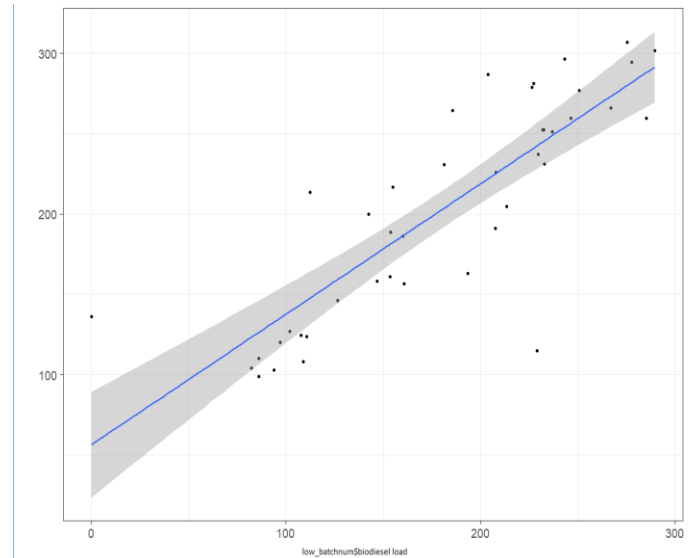
Results & Discussion - *Model Equation:*

Yij = αi +β1Xj +bi1Z1ij+bi2Z2ij+εij

- αi – reactor types either Flash or normal
- β1Xj – temperature (high / low)
- bi1Z1ij - lipase enzymes in each batches (random)
- bi2Z2ij - types of oil used in experiment (random)

With only one fixed effect "temperature" in the mixed model to predict the biodiesel load,

> Biodiesel load~temperature+(1|batchnum)+(1|scenario)

controlling for by batch number and by scenario variability. We can see that the scenario has much less variability than the batch number. Regarding the fixed effect the batch load is lower in the lower temperature than the high temperature with maximum difference of 15 gallons. Secondly, the reactor type was added in the model as a fixed effect because the relationship between the reactor and the biodiesel load is systematic and

> biodieselload~temperature+reactortype+(1|batchnum)+(1|scenario)

predictable which is different from the random effects batch number and scenario, where the relationship between the biodiesel load is much more unpredictable and "random". Here the general error

term is still contained. This is because even if I have accounted for each batch variation, there is still going to be random differences between different utterances from same batch. Compared to the earlier model without the fixed effect reactor type the variation that's associated with the random effect "batchnum" dropped considerably. This is because the variation that's due to reactor type was confounded with the variation that's due to batch number. The model didn't know that flash and normal reaction and so its predictions were relatively more off, creation relatively larger residuals. We can see that flash and normal reactor differ about 111 gallons.
Statistical Significance: p-values for mixed models can be focused though likelihood ratio test. Having null and full model difference with chi-square value comparison we could analyzed the p-value

Nullmodel:
Biodieselload~reactortype+(1|batchnum)+(1|scenario)
Fullmodel: **Biodieselload~reactortype+temperature**
                     +(1|batchnum)+(1|scenario)

Here the temperature affected biodiesel production with (chi-square value 5.407, p= 0.02005), lowering it by about 15 gallons ± 6.498 (standard errors).
The adding and the interaction methods doesn't significantly improve on the model fit.

Random slopes and Random intercept:
The model I used was the random intercept model. Here it has been accounted for the baseline-differences in biodiesel load, but if whatever the effect of temperature is its going to be the same for all the batch number and scenarios. We may not say that it's a valid assumption because it is expected that some scenarios would elicit temperatures. That is, the effect of temperature might be different for different scenarios. Likewise, the effect of temperatures might be different for different batch numbers. It can be expected that some batches of lipase are more less reacted with the temperatures than others. So, it's better to have a random slope model, where batch numbers and scenarios are not only allowed to have different intercepts but allowing different slopes for the effect for temperatures.
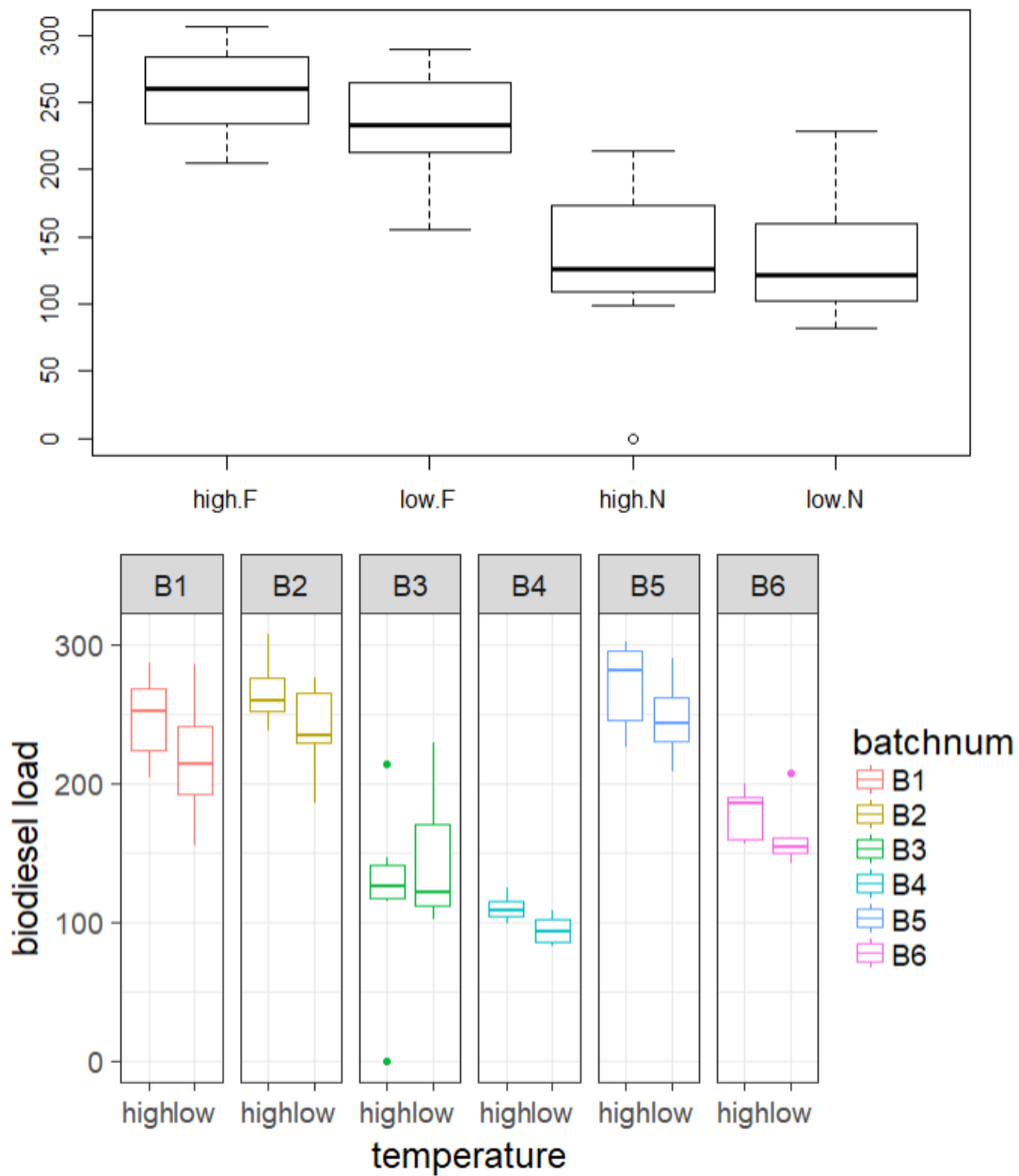
Biodieselload~temperature+reactortype+(1+temperature| batchnum)+(1|scenario)

The coefficients of the model will change by batch number and scenario for the effect of temperatures is different for each batch number and scenario. However, its negative and that many of the values are quite similar for each batch. This means that despite individual variation there is also consistency in how biodiesel load for all the batch numbers, the production of biodiesel tends to go down when temperatures are low, but sometimes it goes down slightly more so than the other batches depending upon the fatty acid content in each random batch samples. Despite individual batch variation, there is also consistency in how the temperatures affects the production for all the batch scenarios. But for all this depends even on the reactor type and how it is run at that time with the oils available separately. The coefficients do not change. That is because I didn't specify random slopes for the by batch numbers or by scenario for effect of reactor type. So, it appears that we can see that the model with random intercept for batch numbers is significantly better than the model with just random slope for this condition.

Conclusion:
Since this model is solely based on temperatures by not on the reactor differences we have only random slopes for effect of temperatures but not for reactor Researchers, still have argument having mixed models without random slopes creates relatively high type 1 error rate. Conceptually, it makes a lot of sense to include random slopes along with random intercepts. After all, we can almost always expect each batch to be different with how they react to an experimental manipulation! And likewise we can almost always expect that the effect of an experimental manipulation is not going to be the same for all experiments. Thus, here the mixed model makes sense with the variables available in this format. It's better to include all random slopes that are justified by our experimental design and having all fixed effects are important for the overall interpretation in the study.

GRAPHS

CODES

```
Classes 'tbl_df', 'tbl' and 'data.frame':    84 obs. of  5 variables:
 $ batchnum      : chr  "B1" "B1" "B1" "B1" ...
 $ reacter type  : chr  "F" "F" "F" "F" ...
 $ scenario      : num  1 1 2 2 3 3 4 4 5 5 ...
 $ temperature   : chr  "low" "high" "low" "high" ...
 $ biodiesel load: num  213 204 285 260 204 ...

Linear mixed model fit by REML ['lmerMod']
Formula: `biodiesel load` ~ temperature + (1 | batchnum) + (1 | scenario)
```

Data: bio

REML criterion at convergence: 826.5

Scaled residuals:
    Min   1Q Median   3Q   Max
-4.1625 -0.4781 -0.0219 0.4862 3.1429

Random effects:
 Groups   Name   Variance Std.Dev.
 scenario (Intercept) 286.3   16.92
 batchnum (Intercept) 4214.4   64.92
 Residual   876.7   29.61
Number of obs: 84, groups: scenario, 7; batchnum, 6

Fixed effects:
              Estimate Std. Error t value
(Intercept)   199.077   27.664   7.196
temperaturelow -15.349   6.498 -2.362

Correlation of Fixed Effects:
            (Intr)
temperatrlw -0.123
Linear mixed model fit by REML ['lmerMod']
**Formula: `biodiesel load` ~ temperature + `reacter type` + (1 | batchnum) +   (1 |
scenario)**
    Data: bio

REML criterion at convergence: 808.2

Scaled residuals:
    Min   1Q Median   3Q   Max
-4.1352 -0.5269 -0.0304 0.4347 3.1697

Random effects:
 Groups   Name   Variance Std.Dev.
 scenario (Intercept) 286.3   16.92
 batchnum (Intercept) 588.3   24.25
 Residual   876.7   29.61
Number of obs: 84, groups: scenario, 7; batchnum, 6

Fixed effects:
               Estimate Std. Error t value
(Intercept)   255.017   16.415 15.536
temperaturelow -15.332   6.497 -2.360
`reacter type`N -111.898   20.831 -5.372

Correlation of Fixed Effects:
            (Intr) tmprtr
temperatrlw -0.207
`rectrtyp`N -0.635 0.000
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: `biodiesel load` ~ `reacter type` + (1 | batchnum) + (1 | scenario)
    Data: bio

     AIC   BIC   logLik deviance df.resid
   843.6   855.7   -416.8   833.6   79

Scaled residuals:
    Min   1Q Median   3Q   Max
-3.7583 -0.4244 -0.0473 0.6277 2.8514

Random effects:
 Groups   Name   Variance Std.Dev.

```
 scenario (Intercept) 251.4    15.85
 batchnum (Intercept) 390.2    19.75
 Residual             935.2    30.58
Number of obs: 84, groups:  scenario, 7; batchnum, 6

Fixed effects:
             Estimate Std. Error t value
(Intercept)      246.99      13.72  18.002
`reacter type`N  -111.90      17.45  -6.411

Correlation of Fixed Effects:
           (Intr)
`rectrtyp`N -0.636
Linear mixed model fit by maximum likelihood  ['lmerMod']
```

Formula: `biodiesel load` ~ `reacter type` + temperature + (1 | batchnum) +   (1 |
scenario)

```
   Data: bio

     AIC      BIC   logLik deviance df.resid
   840.2    854.8   -414.1    828.2       78

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.1716 -0.5302 -0.0456  0.4385  3.1930

Random effects:
 Groups    Name        Variance Std.Dev.
 scenario (Intercept) 263.6     16.24
 batchnum (Intercept) 397.3     19.93
 Residual             865.7     29.42
Number of obs: 84, groups:  scenario, 7; batchnum, 6

Fixed effects:
             Estimate Std. Error t value
(Intercept)     255.008     14.217  17.937
`reacter type`N -111.898     17.495  -6.396
temperaturelow  -15.315      6.455  -2.373

Correlation of Fixed Effects:
           (Intr) `typ`N
`rectrtyp`N -0.615
temperatrlw -0.238  0.000
Data: bio
```

Models:
biodiesel.nullmodel1: `biodiesel load` ~ `reacter type` + (1 | batchnum) + (1 | scenario)
biodiesel.fullmodel2: `biodiesel load` ~ `reacter type` + temperature + (1 | batchnum) +
biodiesel.fullmodel2:   (1 | scenario)

```
                   Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
biodiesel.nullmodel1  5 843.58 855.73 -416.79   833.58
biodiesel.fullmodel2  6 840.17 854.76 -414.09   828.17 5.4077      1    0.02005 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] -5.407741
[1] 1
[1] 0
Data: bio
Models:
```
biodiesel.fullmodel2: `biodiesel load` ~ `reacter type` + temperature + (1 | batchnum) +
biodiesel.fullmodel2:   (1 | scenario)
biodiesel.fullmodel3: `biodiesel load` ~ `reacter type` * temperature + (1 | batchnum) +
biodiesel.fullmodel3:   (1 | scenario)
```
                   Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
biodiesel.fullmodel2  6 840.17 854.76 -414.09   828.17
biodiesel.fullmodel3  7 839.68 856.69 -412.84   825.68 2.4965      1     0.1141
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: `biodiesel load` ~ `reacter type` * temperature + (1 | batchnum) +        (1 |
scenario)
   Data: bio

     AIC       BIC    logLik deviance df.resid
   839.7     856.7    -412.8    825.7        77

Scaled residuals:
    Min       1Q  Median       3Q      Max
-4.0412 -0.4933   0.0714   0.5142   3.0927

Random effects:
 Groups    Name         Variance Std.Dev.
 scenario (Intercept) 256.2      16.01
 batchnum (Intercept) 400.0      20.00
 Residual             838.2      28.95
Number of obs: 84, groups:  scenario, 7; batchnum, 6

Fixed effects:
                                Estimate Std. Error t value
(Intercept)                      260.313     14.562  17.876
`reacter type`N                 -122.506     18.734  -6.539
temperaturelow                   -25.443      8.987  -2.831
`reacter type`N:temperaturelow   20.253     12.715   1.593
Correlation of Fixed Effects:
            (Intr) `typ`N tmprtr
`rectrtyp`N -0.643
temperatrlw -0.323  0.252
`rctrtyp`N:  0.229 -0.356 -0.707
$scenario
   (Intercept) temperaturelow `reacter type`N
1    242.6733      -15.33208       -111.8976
2    262.6037      -15.33208       -111.8976
3    267.3904      -15.33208       -111.8976
4    277.5569      -15.33208       -111.8976
5    254.1522      -15.33208       -111.8976
6    234.5670      -15.33208       -111.8976
7    246.1742      -15.33208       -111.8976

$batchnum
    (Intercept) temperaturelow `reacter type`N
B1    241.1753      -15.33208       -111.8976
B2    259.0659      -15.33208       -111.8976
B3    254.7901      -15.33208       -111.8976
B4    224.9437      -15.33208       -111.8976
B5    264.8093      -15.33208       -111.8976
B6    285.3166      -15.33208       -111.8976

attr(,"class")
[1] "coef.mer"
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: `biodiesel load` ~ temperature + `reacter type` + (1 + temperature |
    batchnum) + (1 + temperature | scenario)
   Data: bio

     AIC       BIC    logLik deviance df.resid
   848.0     872.3    -414.0    828.0        74

Scaled residuals:
    Min       1Q  Median       3Q      Max
-4.1037 -0.5249 -0.0676   0.4322   3.2437

Random effects:
 Groups    Name             Variance  Std.Dev. Corr
```

```
  scenario (Intercept)    323.37576 17.9827
           temperaturelow  10.64335  3.2624  -1.00
  batchnum (Intercept)    400.70630 20.0176
           temperaturelow   0.03651  0.1911  -1.00
  Residual                862.14296 29.3623
Number of obs: 84, groups:  scenario, 7; batchnum, 6
Fixed effects:
                 Estimate Std. Error t value
(Intercept)        254.84      14.53  17.543
temperaturelow     -15.35       6.56  -2.340
`reacter type`N   -111.44      17.48  -6.376
Correlation of Fixed Effects:
            (Intr) tmprtr
temperatrlw -0.323
`rectrtyp`N -0.602  0.000
$scenario
   (Intercept) temperaturelow `reacter type`N
1    241.2378      -12.88572       -111.4448
2    262.1051      -16.67146       -111.4448
3    269.2640      -17.97024       -111.4448
4    279.3436      -19.79887       -111.4448
5    253.8354      -15.17118       -111.4448
6    232.1726      -11.24110       -111.4448
7    245.8874      -13.72925       -111.4448

$batchnum
    (Intercept) temperaturelow `reacter type`N
B1    241.7103      -15.22726       -111.4448
B2    258.8175      -15.39056       -111.4448
B3    254.6020      -15.35032       -111.4448
B4    225.6655      -15.07409       -111.4448
B5    264.4438      -15.44427       -111.4448
B6    283.7718      -15.62877       -111.4448

attr(,"class")
[1] "coef.mer"
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: `biodiesel load` ~ temperature + `reacter type` + (1 | batchnum) +      (1 |
scenario)
   Data: bio

     AIC      BIC   logLik deviance df.resid
   840.2    854.8   -414.1    828.2       78

Scaled residuals:
    Min     1Q  Median     3Q     Max
-4.1716 -0.5302 -0.0456  0.4385  3.1930

Random effects:
 Groups    Name         Variance Std.Dev.
 scenario (Intercept) 263.6     16.24
 batchnum (Intercept) 397.3     19.93
 Residual             865.7     29.42
Number of obs: 84, groups:  scenario, 7; batchnum, 6

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       255.008     14.217  17.937
temperaturelow    -15.315      6.455  -2.373
`reacter type`N  -111.898     17.495  -6.396
Correlation of Fixed Effects:
            (Intr) tmprtr
temperatrlw -0.238
`rectrtyp`N -0.615  0.000
```

```
$scenario
  (Intercept) temperaturelow `reacter type`N
1    242.8432       -15.31471        -111.8976
2    262.4850       -15.31471        -111.8976
3    267.2025       -15.31471        -111.8976
4    277.2206       -15.31471        -111.8976
5    254.1559       -15.31471        -111.8976
6    234.8543       -15.31471        -111.8976
7    246.2923       -15.31471        -111.8976

$batchnum
     (Intercept) temperaturelow `reacter type`N
B1    241.7557       -15.31471        -111.8976
B2    258.8837       -15.31471        -111.8976
B3    254.7899       -15.31471        -111.8976
B4    226.2149       -15.31471        -111.8976
B5    264.3837       -15.31471        -111.8976
B6    284.0182       -15.31471        -111.8976

attr(,"class")
[1] "coef.mer"
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: `biodiesel load` ~ `reacter type` + temperature + (1 + temperature |
    batchnum) + (1 + temperature | scenario)
   Data: bio

     AIC       BIC    logLik deviance df.resid
   848.0     872.3    -414.0    828.0        74

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.1037 -0.5249 -0.0676  0.4322  3.2437
Random effects:
 Groups   Name            Variance  Std.Dev. Corr
 scenario (Intercept)     323.37593 17.9827
          temperaturelow   10.64336  3.2624  -1.00
 batchnum (Intercept)     400.70648 20.0177
          temperaturelow    0.03651  0.1911  -1.00
 Residual                 862.14291 29.3623
Number of obs: 84, groups:  scenario, 7; batchnum, 6

Fixed effects:
                Estimate Std. Error t value
(Intercept)       254.84      14.53  17.543
`reacter type`N  -111.44      17.48  -6.376
temperaturelow    -15.35       6.56  -2.340

Correlation of Fixed Effects:
            (Intr) `typ`N
`rectrtyp`N -0.602
temperatrlw -0.323  0.000


Data: bio
Models:
biodiesel.anothermodel1: `biodiesel load` ~ temperature + `reacter type` + (1 | batchnum)
+
biodiesel.anothermodel1:      (1 | scenario)
biodiesel.anothermodel2: `biodiesel load` ~ `reacter type` + temperature + (1 +
temperature |
biodiesel.anothermodel2:      batchnum) + (1 + temperature | scenario)
                        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
biodiesel.anothermodel1  6 840.17 854.76 -414.09   828.17
biodiesel.anothermodel2 10 847.95 872.26 -413.98   827.95 0.2171      4     0.9945
```