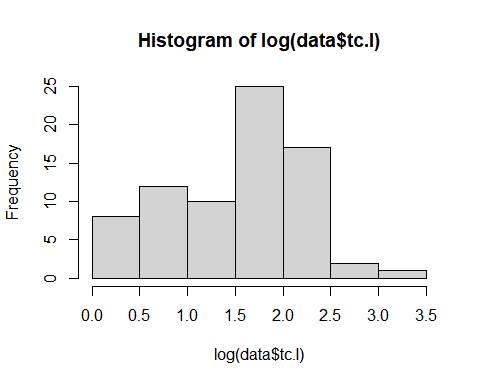
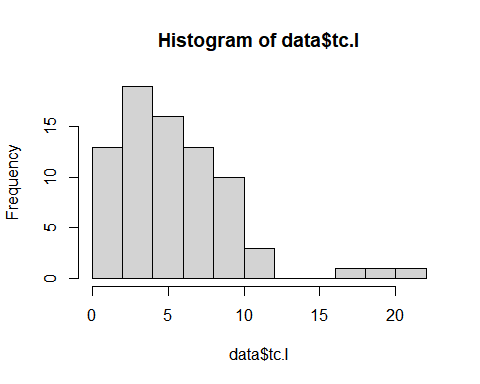
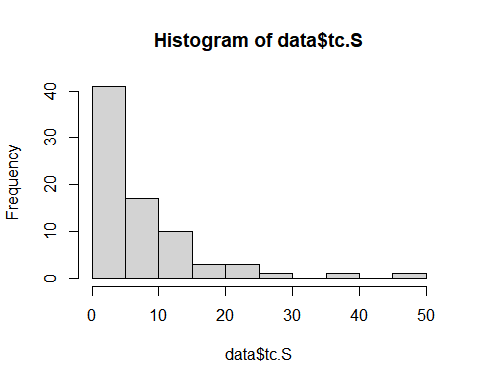
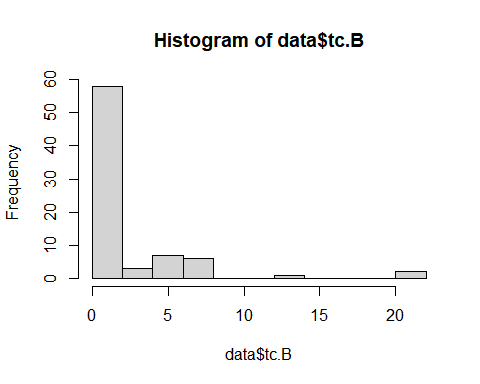
Simple and Mixed Model Analysis

Alexis Swanson

2/4/2021

#Histograms of the 3 Approaches for tc



## [1] 0.00 20.41

## [1] 0.00 45.69

## [1] 0.00 21.63

The log of the Iterative Solution has the most normal distribution.

The ranges are for Blaisdell, Scour Depth and the Iterative solutions respectively.

After our discussion on 2/4 the range of expected critical shear stress was estimated using the min and max median grain size and the shield’s diagram. This range was 0.97 - 27.5 Pa. Based upon this and the normal distribution of the log of the Iterative Solution it was chosen to move forward with the Iterative Solution for this analysis.

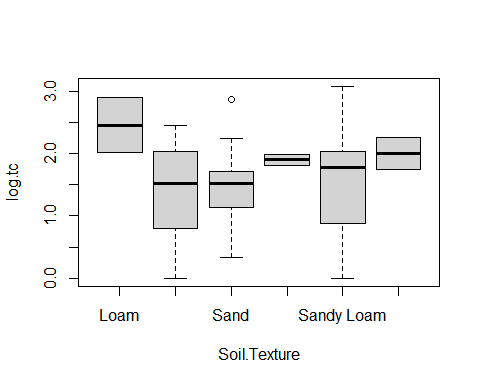
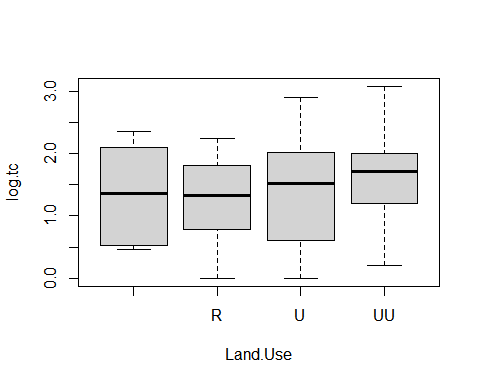
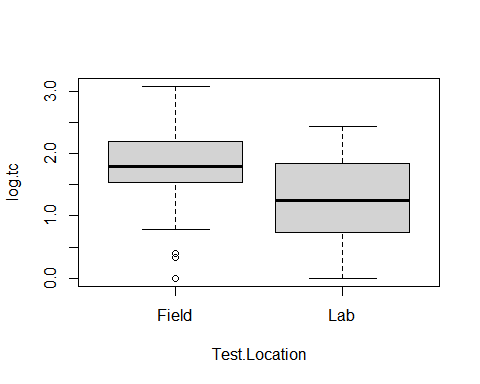
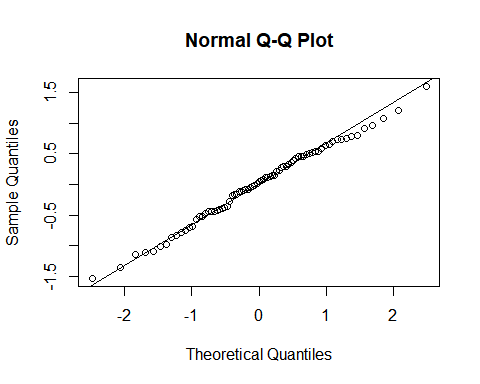
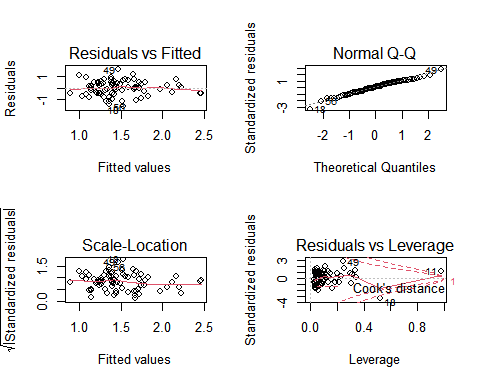
I finished the calibration and will now go back to adjust the JET Spreadsheets to account for the change in the coefficients. This will change the values of my critical shear stresses but I do not anticipate a large change.

# First Linear Model

# All Predictors

##   
## Call:  
## lm(formula = log.tc ~ Temp + EC + Cor.Moist + D50 + Cu + Cc +   
## GSD + Pass200, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.53372 -0.43737 0.04092 0.45957 1.59830   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.265979 0.505511 2.504 0.0147 \*  
## Temp -0.021094 0.019337 -1.091 0.2792   
## EC -1.974880 0.860560 -2.295 0.0248 \*  
## Cor.Moist 1.875193 0.993972 1.887 0.0635 .  
## D50 0.021848 0.662896 0.033 0.9738   
## Cu -0.001070 0.015961 -0.067 0.9467   
## Cc -0.050448 0.224724 -0.224 0.8231   
## GSD 0.056161 0.041924 1.340 0.1848   
## Pass200 0.008258 0.009123 0.905 0.3686   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6713 on 68 degrees of freedom  
## Multiple R-squared: 0.2076, Adjusted R-squared: 0.1144   
## F-statistic: 2.227 on 8 and 68 DF, p-value: 0.03574

## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced  
  
## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced

 The significant predictors in this model include: the intercept, EC and Moisture content. These variables will be used in the next linear model. The R-Squared is low at 0.2076 but the p-value shows the model significant at 0.03574.

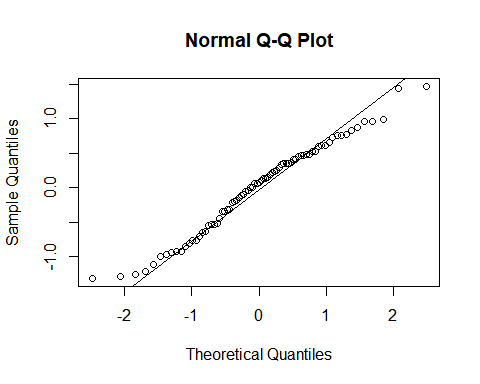
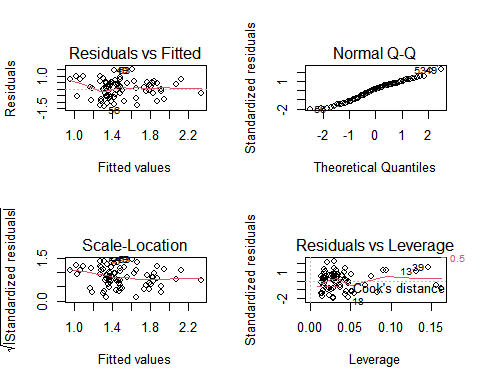
The residuals for the model are pretty good. They follow the predicted versus sample line pretty closely only varying at the upper limit.

Here I also plotted box plots to see if there was a significant difference in the categorical variables. The field and lab almost show a statistical difference but are still overlapping slightly. The rural, urban and ultra urban are not showing much of a difference. The soil texture however seems to have something going on where the critical shear stress could be statiscially different for each soil texture.

# Linear Model with just EC and Moist

# Based on the results those were the only significant factors

##   
## Call:  
## lm(formula = log.tc ~ EC + Cor.Moist, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.31003 -0.53433 0.07384 0.46611 1.46941   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.0455 0.2596 4.028 0.000135 \*\*\*  
## EC -2.1410 0.7269 -2.945 0.004309 \*\*   
## Cor.Moist 2.4273 0.8604 2.821 0.006143 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6704 on 74 degrees of freedom  
## Multiple R-squared: 0.1399, Adjusted R-squared: 0.1167   
## F-statistic: 6.019 on 2 and 74 DF, p-value: 0.003784

 Here the model shows that all factors are significant. The R-squared went down a little bit but the p-value is better. When looking at the residuals of this model we do see a little bit more variation. My conclusion from these first two linear models would be that the “best” simple linear model is probably somewhere in between. (i.e. adding a few more variables back that were close to significance in the first model but still limiting some of the variables.)

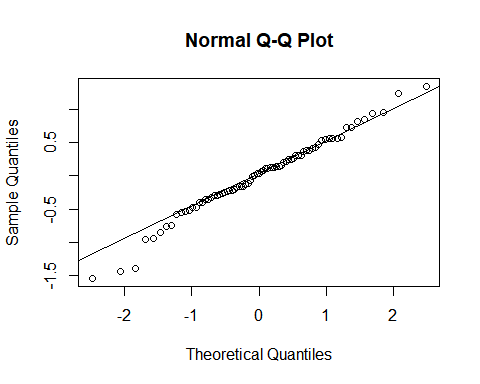
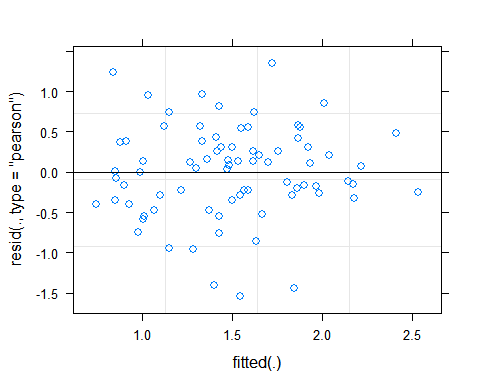
# Mixed Model with 2 Random Factors

# Test Location and Land Use

## Linear mixed model fit by REML ['lmerMod']  
## Formula:   
## log.tc ~ (1 | Test.Location) + (1 | Land.Use) + Temp + EC + Cor.Moist +   
## D50 + Cu + Cc + GSD + Pass200  
## Data: data  
##   
## REML criterion at convergence: 170  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.53995 -0.47473 0.07259 0.60873 2.22245   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Land.Use (Intercept) 0.01976 0.1406   
## Test.Location (Intercept) 0.13601 0.3688   
## Residual 0.37069 0.6088   
## Number of obs: 77, groups: Land.Use, 4; Test.Location, 2  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 0.915238 0.547195 1.673  
## Temp -0.009168 0.017891 -0.512  
## EC -1.740470 0.810111 -2.148  
## Cor.Moist 2.386143 0.922638 2.586  
## D50 0.325736 0.622505 0.523  
## Cu -0.004759 0.014902 -0.319  
## Cc -0.134570 0.209892 -0.641  
## GSD 0.041251 0.039985 1.032  
## Pass200 0.008379 0.008399 0.998  
##   
## Correlation of Fixed Effects:  
## (Intr) Temp EC Cr.Mst D50 Cu Cc GSD   
## Temp -0.664   
## EC -0.221 0.285   
## Cor.Moist -0.289 0.009 -0.429   
## D50 0.140 0.056 0.280 0.009   
## Cu -0.097 -0.058 -0.333 0.008 -0.946   
## Cc -0.226 -0.023 -0.206 -0.029 -0.934 0.845   
## GSD -0.214 -0.092 0.118 -0.070 -0.342 0.128 0.425   
## Pass200 0.155 -0.096 0.329 -0.355 0.677 -0.617 -0.641 -0.281

## Analysis of Variance Table  
## npar Sum Sq Mean Sq F value  
## Temp 1 0.2875 0.2875 0.7757  
## EC 1 0.9358 0.9358 2.5244  
## Cor.Moist 1 4.6752 4.6752 12.6123  
## D50 1 0.3394 0.3394 0.9156  
## Cu 1 0.7931 0.7931 2.1395  
## Cc 1 0.0869 0.0869 0.2345  
## GSD 1 0.6925 0.6925 1.8680  
## Pass200 1 0.3690 0.3690 0.9954

## $Land.Use  
## (Intercept)  
## 0.03323668  
## R -0.06775524  
## U -0.07009103  
## UU 0.10460960  
##   
## $Test.Location  
## (Intercept)  
## Field 0.2509691  
## Lab -0.2509691  
##   
## with conditional variances for "Land.Use" "Test.Location"

 This is the first mixed model where I used two random factors, test location and land use. The random effects of this model account for about 30% of the variance observed in the data. This is broken into about 26% for test location and 4% for land use. This implies that test location is actually contributing to the variance more than the land use. This is also justified in the box plots from the beginning of the report. Based on this we may consider using soil texture as a random variable since there seemed to be difference between the different categories in the box plots.

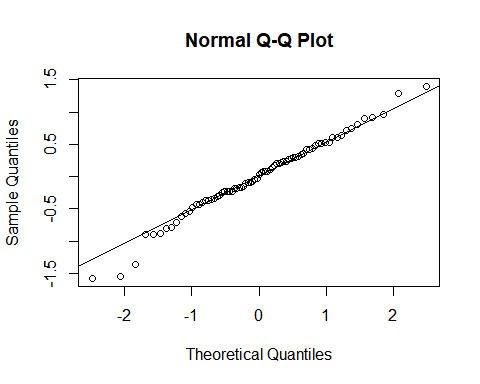
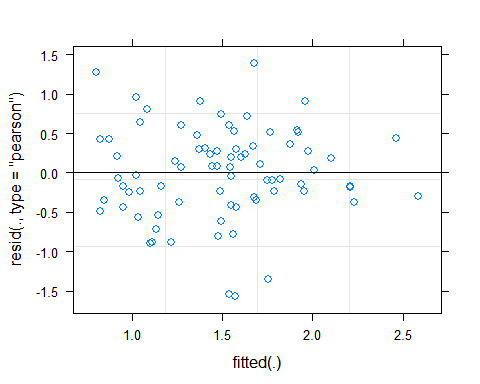
The residuals of this model are similar to the previous simple linear model expect now we are seeing more variation at the lower end more the upper end.

# Mixed Model with only Test Location as Random Factor

## Linear mixed model fit by REML ['lmerMod']  
## Formula: log.tc ~ (1 | Test.Location) + Temp + EC + Cor.Moist + D50 +   
## Cu + Cc + GSD + Pass200  
## Data: data  
##   
## REML criterion at convergence: 170.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.54653 -0.55711 0.05352 0.57924 2.25861   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Test.Location (Intercept) 0.1362 0.3690   
## Residual 0.3821 0.6181   
## Number of obs: 77, groups: Test.Location, 2  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 0.937677 0.541858 1.730  
## Temp -0.010559 0.018060 -0.585  
## EC -1.504632 0.803797 -1.872  
## Cor.Moist 2.128370 0.918154 2.318  
## D50 0.215173 0.612926 0.351  
## Cu -0.004057 0.014722 -0.276  
## Cc -0.094222 0.207314 -0.454  
## GSD 0.058677 0.038612 1.520  
## Pass200 0.008313 0.008401 0.989  
##   
## Correlation of Fixed Effects:  
## (Intr) Temp EC Cr.Mst D50 Cu Cc GSD   
## Temp -0.668   
## EC -0.228 0.299   
## Cor.Moist -0.295 -0.002 -0.407   
## D50 0.122 0.069 0.332 -0.018   
## Cu -0.074 -0.076 -0.368 0.019 -0.947   
## Cc -0.212 -0.034 -0.257 -0.002 -0.932 0.846   
## GSD -0.240 -0.081 0.061 -0.010 -0.323 0.118 0.412   
## Pass200 0.138 -0.084 0.352 -0.371 0.670 -0.602 -0.635 -0.292

## Analysis of Variance Table  
## npar Sum Sq Mean Sq F value  
## Temp 1 0.2295 0.2295 0.6006  
## EC 1 0.8273 0.8273 2.1652  
## Cor.Moist 1 4.1098 4.1098 10.7556  
## D50 1 0.1728 0.1728 0.4522  
## Cu 1 0.8208 0.8208 2.1480  
## Cc 1 0.0528 0.0528 0.1381  
## GSD 1 1.3663 1.3663 3.5758  
## Pass200 1 0.3741 0.3741 0.9791

## $Test.Location  
## (Intercept)  
## Field 0.2508506  
## Lab -0.2508506  
##   
## with conditional variances for "Test.Location"

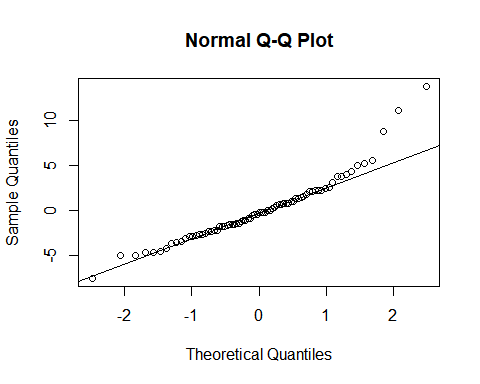
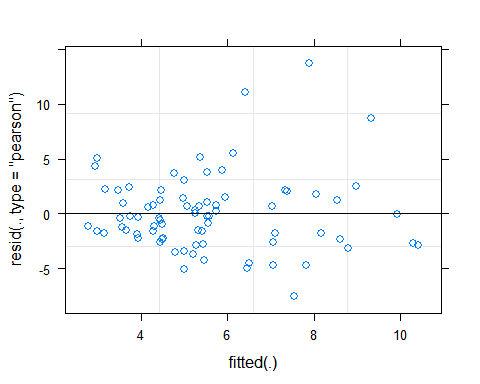
 This mixed model only has Test Location as a random factor. Test location is still accounting for 26% of the variance but the model’s residuals do look better than the first mixed model.

# Mixed Model with only Land Use as Random Factor

## Linear mixed model fit by REML ['lmerMod']  
## Formula: tc.I ~ (1 | Land.Use) + Temp + EC + Cor.Moist + D50 + Cu + Cc +   
## GSD + Pass200  
## Data: data  
##   
## REML criterion at convergence: 411.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.0432 -0.6133 -0.0750 0.4086 3.7170   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## Land.Use (Intercept) 0.1589 0.3986   
## Residual 13.6484 3.6944   
## Number of obs: 77, groups: Land.Use, 4  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 1.858119 2.803405 0.663  
## Temp -0.023569 0.106615 -0.221  
## EC -12.041740 4.773348 -2.523  
## Cor.Moist 10.769406 5.510792 1.954  
## D50 -4.103923 3.683281 -1.114  
## Cu 0.108672 0.088572 1.227  
## Cc 0.977389 1.247065 0.784  
## GSD 0.275172 0.235052 1.171  
## Pass200 0.008875 0.050431 0.176  
##   
## Correlation of Fixed Effects:  
## (Intr) Temp EC Cr.Mst D50 Cu Cc GSD   
## Temp -0.753   
## EC -0.231 0.273   
## Cor.Moist -0.327 -0.011 -0.436   
## D50 0.165 0.052 0.306 -0.016   
## Cu -0.105 -0.062 -0.353 0.020 -0.947   
## Cc -0.263 -0.021 -0.234 -0.007 -0.932 0.845   
## GSD -0.264 -0.090 0.079 -0.034 -0.333 0.123 0.418   
## Pass200 0.167 -0.089 0.349 -0.367 0.675 -0.608 -0.637 -0.287

## Analysis of Variance Table  
## npar Sum Sq Mean Sq F value  
## Temp 1 21.905 21.905 1.6050  
## EC 1 38.054 38.054 2.7882  
## Cor.Moist 1 117.266 117.266 8.5919  
## D50 1 0.150 0.150 0.0110  
## Cu 1 50.715 50.715 3.7158  
## Cc 1 8.708 8.708 0.6380  
## GSD 1 22.189 22.189 1.6258  
## Pass200 1 0.423 0.423 0.0310

## $Land.Use  
## (Intercept)  
## 0.018841958  
## R -0.174957830  
## U -0.003291652  
## UU 0.159407524  
##   
## with conditional variances for "Land.Use"



In this model land use is only accounting for 1% of the variance and therefore I conclude that this model is not the best. This conclusion is also supported by the residuals as the points show a lot of variation from the predicted versus sample line.

# Conclusion

From this point I am going to adjust the values for tc based upon the change in the coefficient for the different water temperatures. After those are adjusted I will re-run all the models to confirm the conclusions to this point will not change.

Next Steps: - Can we get a better model by adding some of the variables back in to the simple linear model. Things to consider in this step are making a better model with as little variables as possible - easier to use. - Consider Soil Texture as a random variable. - Consider clustering of data by soil texture. - Consider clustering of data by test location.

\*One thing that I think we need to keep in mind after reading the materials more is that the random variable isn’t neccessarily a variable that is random in nature but that its effect on the response variable is unknown.

*This is only for my reference when writing the code in R*  
Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.