**3D Visualization and Prediction of Spine Fractures**

**Progress Report #2**

**February 14, 2016**

*Jiayuan Shi, Jing Liu,*

*Qianyun Zou, Su Yang*

**Background**

In the project “3-D Visualization and Prediction of Spine Fractures”, Professor Morgan and her group have developed a method of measuring the deformations that occur throughout a human vertebra (one of the bones in the spine) as they compress the vertebrae to failure in the laboratory. The prior studies of human vertebrae failure are using numerical simulation with Finite Elements (FE) methods, whose prediction accuracy has not been assessed until now. The goal of Professor Morgan and her group is to perform this assessment by comparing the FE simulations predictions to the measurements on an actual crushed vertebra.

Figure A and B below show the experimental result for one vertebra, and Figure 3C-J show FE predictions for the same vertebra, for eight different types of simulations. These eight simulations can be divided into two types and four methods. There are two types of tissues used in the simulations, Crushable Foam and von Mises. Under each type of tissue, there are four simulation methods applied: Experimentally Matched FE, Idealized FE, IVD-Generic FE, and IVD-Specific FE.

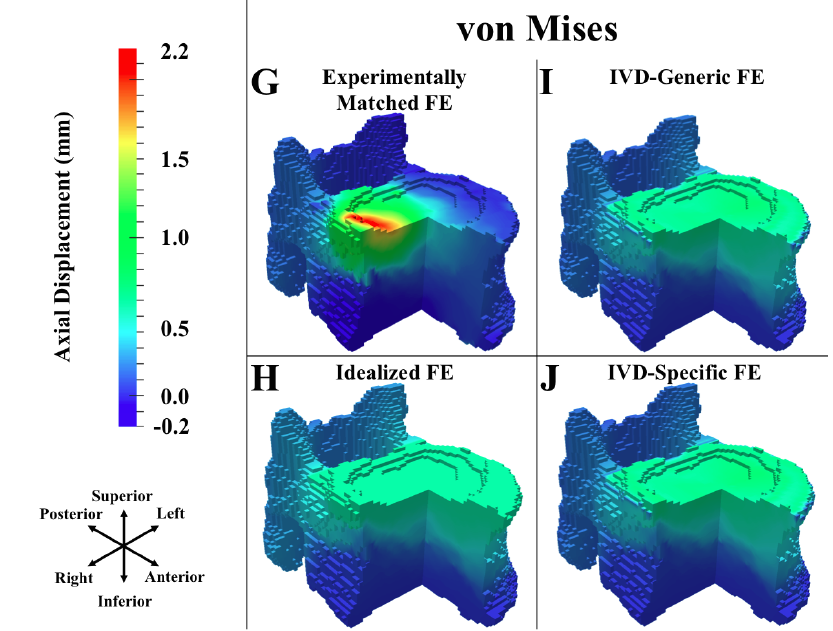
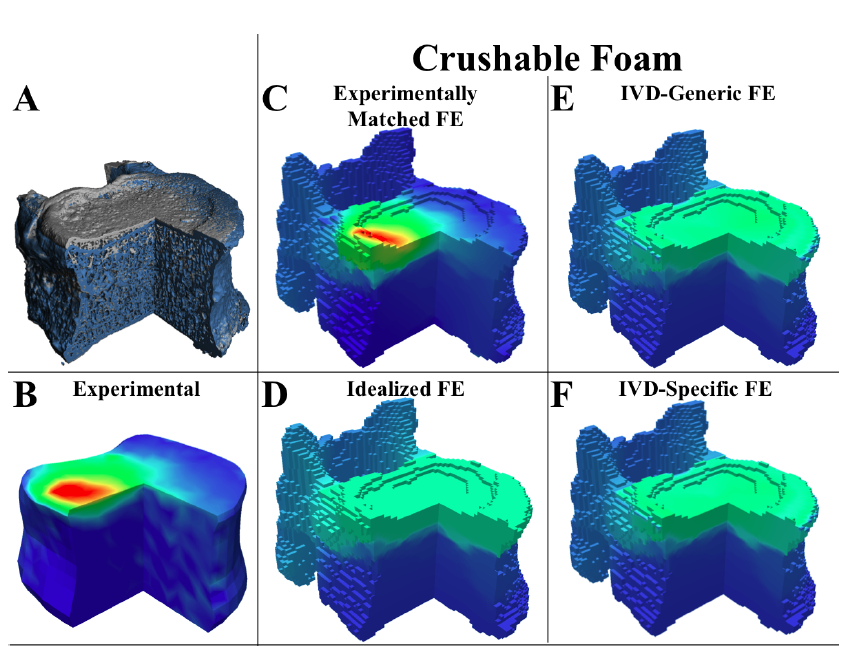


Figure A-J from Prof. Elise Morgan

**Research Questions:**

Based on these results, Professor Morgan wants to know:

1. How similar is each of the simulation results to the experimental results?
2. Are some of the eight simulations more similar than others to the experimental measurement?

**Dataset**

There are two types of compressions. The first is called Union, which means compression only, while the second is called Flexion, which is a combination of flexion and compression. Each compression contains data from 14 specimens, so the whole data set contains data from 28 specimens. The specimen has been divided into several grids, and one single number is provided as the deformation of one grid using region-averaging method.

Below is displayed a part of the dataset. The first line is the ID for one specimen, under which are names for simulations. There are eight simulations under each specimen (two tissue types X four models = eight distinct simulation settings). For each simulation, there are three columns: the first column is grid ID, the second column is the experimental result, and the third column is the simulation result.

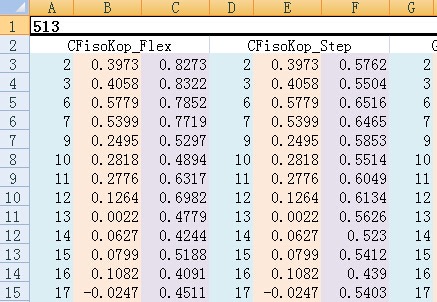


Figure of data set provided by Prof. Elise Morgan

**Analysis**

* Visualization

For global comparison, we started by looking into distributions of differences between simulation and experimental results. At this stage, our analysis is ignoring the 3-D structure of the data, which means that we are only comparing the values of deformation of grids, regardless of where they are on the vertebra. Below are the plots of Flexion 513 and Union 581.

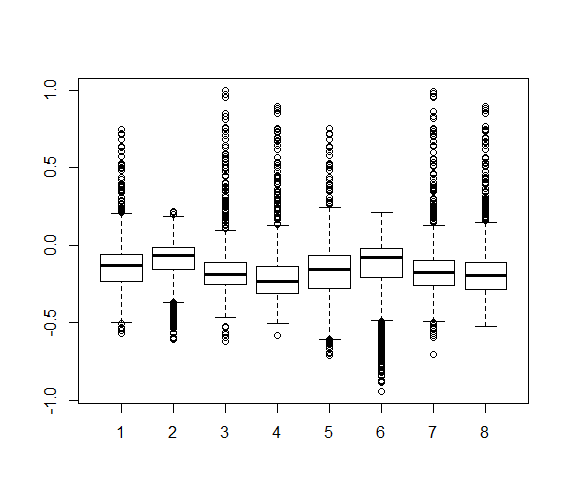
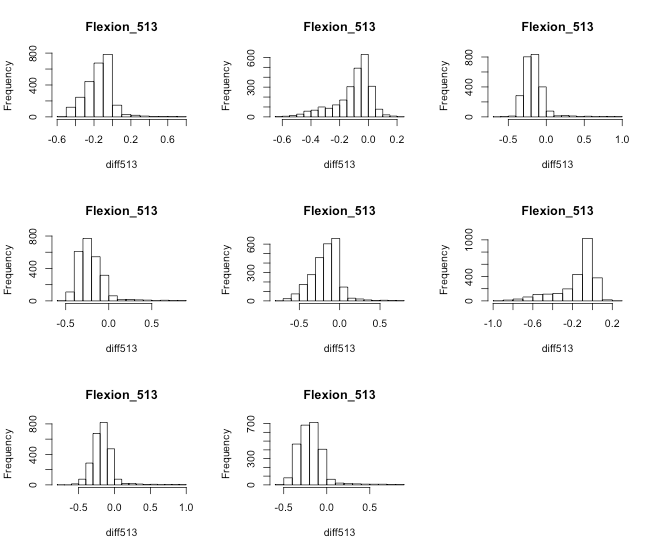


Figure K. Histograms of Vertebra Sample 513 Figure L. Boxplots of Vertebra Sample 513

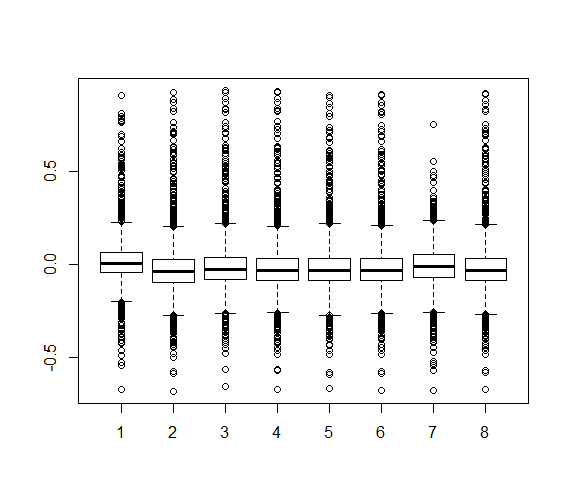
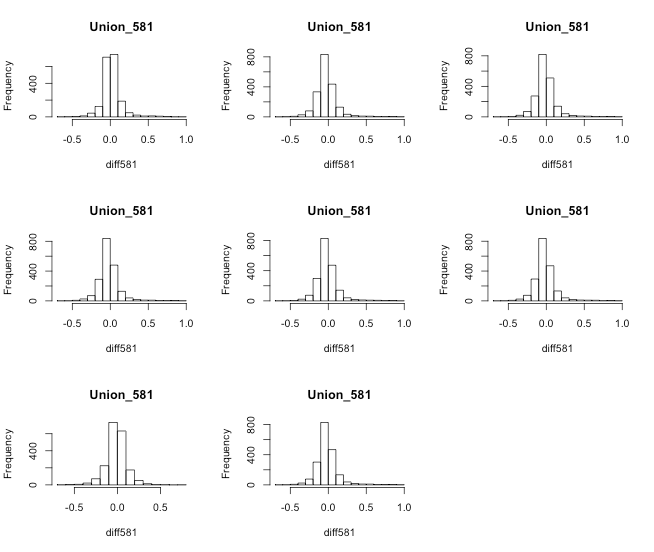


Figure M. Histograms of Vertebra Sample 581 Figure N. Boxplots of Vertebra Sample 581

Results in Flexion 513 shows that the difference are not normally distributed, instead they are either left-skewed or right-skewed and there seems no patterns between. We also notice that in most cases, simulation results have higher value than experimental results. For Union 581, distributions of the difference are approximately normal. Thus, based on the above plots, we might anticipate that the performances of the eight simulations are different within the Flexion group, and similar in the Union group. To confirm this, we then move on to two-way ANOVA.

* Two-Way ANOVA

As we mentioned before, the eight simulations differ in type and method, thus these two factors may have impact on their performances. Therefore, a two-way ANOVA is conducted to statistically test this hypothesis.

We use the mean difference as one measure of the distance between experimental and simulation results for displacements, which is,

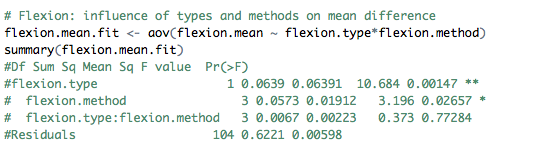
Screen Shot 2016-02-03 at 下午2.19.26.png

Another measure of distance we consider is the squared error, which is,

Screen Shot 2016-02-03 at 下午2.12.59.png

In order to examine whether the types of tissue and methods of simulation have influence on mean difference and squared error between the experimental and simulation results, we conduct two-way analysis of variance (ANOVA) for both Flexion group and Union group. The two-way ANOVA not only aims at assessing the main effect of type and method but also if there is any interaction between them. The two-way ANOVA procedure, conducts a hypothesis tests to check if the factors type, method and their interactions have significant influence on the mean difference or squared error. We have a null hypothesis of non-significance, and if the resulting p-value of one factor is less than 0.05, then with 95% confidence, we can reject the null hypothesis, and conclude that factor has significant influence on the mean difference or squared error.

Below we display the output of the two-way ANOVA procedure for Flexion group, to test the influence of type of tissue on mean difference. The procedure is based on a linear regression model which predicts the mean difference from type. The null hypothesis is that there is no significant influence of type on the mean difference between experimental and simulation results. The procedure outputs in the last column of “flexion.type” a p-value of 0.00147, which is less than 0.05, so with with 95% confidence, we can reject the above null hypothesis. Hence, type in the Flexion group has significant influence on its mean difference. Similarly, we can see in the Flexion group, not only the type, but method also has significant influence on its mean difference. However, their interaction does not influence the mean difference significantly because of its large p-value of 0.77284.



For each of Flexion and Union group, we conduct two two-way ANOVA, one is for test influence on mean difference and the other is for test influence on squared error. We get the following results in the table below,

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Flexion Group | | | Union Group | | |
|  | Type | Method | Interaction | Type | Method | Interaction |
| Mean Difference | **Significant** | **Significant** | Non-significant | Non-significant | Non-significant | Non-significant |
| Squared Error | **Significant** | Non-significant | Non-significant | Non-significant | Non-significant | Non-significant |

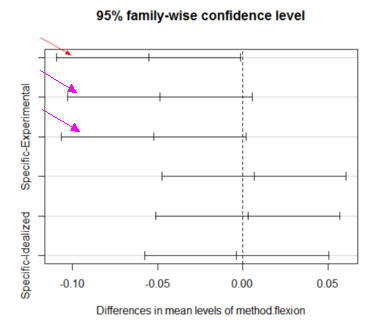
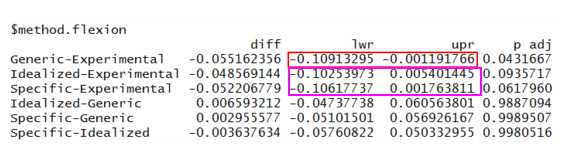
Table A. Significance Results from ANOVA

From the above results, for the three significant ones (in Flexion group, the effect of both type and method on mean difference, and the effect of type on squared error), we can also use one-way ANOVA procedure to re-confirm that they are indeed significant. The results from the one-way ANOVA are the same as those from the two-way ANOVA considering only main effects.

The results from the ANOVA procedure are not surprising given our initial exploratory data visualizations. Back to the graphs above, for Flexion group, distributions of mean differences between simulation and experimental results are not normal, instead they are either left-skewed or right-skewed away from zero (zero would indicate no difference). However, for Union group, distributions are approximately normal and centered at zero. In conclusion, via the ANOVA procedure we confirm that the levels of the factors type and method have a significant effect on the mean difference and the levels of the factor type have a significant effect on the squared error for Flexion group, but non-significant effect of type or method for Union group.

* Tukey HSD Test

Elaborating on the ANOVA output, which shows that there is significant difference between the four methods in Flexion group on the mean difference measure of distance between simulation and experimental observations, we next want to find out where the difference lies. That is, we want to determine which methods in the group differ significantly from each other. To do this in a statistically rigorous manner, we conduct Tukey HSD Test, a post-hoc test of the ANOVA procedure.

The four levels of the factor variable method are Experimentally Matched FE, IVD-Generic FE, Idealized FE and IVD-Specific FE. The output of the Turkey HSD Test is six confidence intervals, six since we want to consider all possible pairwise comparisons between our four levels (i.e. Experimentally Matched FE compared to IVD-Generic FE, Experimentally Matched FE compared to Idealized FE, etc.). Each confidence interval is used to perform a hypothesis test. The null hypothesis is that there is no significant difference between the two levels. If the confidence intervals does not include zero, we would reject the null hypothesis. We run this multiple testing procedure keeping the overall confidence level at 95%. The comparison between IVD-Generic FE and Experimentally Matched FE rejects the null hypothesis and reaches the conclusion that there is significant difference between them. Since the difference has a negative value, we conclude that the mean difference of Experimentally Matched FE is larger than IVD-Generic FE. Below, to the right of the confidence interval output, is a plot visualizing this conclusion. If we set the confidence level to be a little bit lower than 95%, two more groups of comparison results show up to be significant - the comparisons between Idealized FE and Experimentally Matched FE, and IVD-Specific FE and Experimentally Matched FE. The rest of the six pairwise tests are not significant. In conclusion, the mean difference of Experimentally Matched FE is always larger than the one from the other three methods while there is no significant difference between the other methods.Figure O. Confidence Intervals of Tukey Results for Methods in Flexion Group

**Conclusions**

We have compared the simulation results to the experimental results by separating them into Flexion group and Union group. Visualization and tests show that under union compression, simulation results are very close to the experimental results and there is no statistically significant difference between either mean difference or squared error for both methods and types, which means that we cannot conclude that there is a type or a method under Union compression which is significantly better than the others. However, for Flexion group, two-way ANOVA shows that both type and method have significant influence on the mean difference. After further comparing the four methods under flexion compression using Turkey HSD Test, we conclude that since the mean difference of Experimentally Matched FE is always larger than the other methods, IVD-Generic FE, Idealized FE and IVD-Specific FE, then the latter three are closer to the experimental results than Experimentally Matched FE simulation method. In other word, Experimentally Matched FE performs worse than the other three methods.

**Possible Next Steps**

Our current analysis does not take into account the 3D structure of the grid, thus spatial comparison is a natural next step. In order to do so, we can use the spatial coordinates of each grid point on the vertebra samples, but this would require a much more sophisticated model.

**Appendix**

* Visualization

Flexion\_513 <- read.csv("Flexion\_513.csv",header=T)

Union\_581 <- read.csv("Union\_581.csv",header=T)

# Flexion\_513

par(mfrow=c(3,3))

matrix.diff513 <- matrix(rep(0,19944), nrow=2493)

i <- 1

k <- 0

while(i < 24){

diff513 <- Flexion\_513[,i+1] - Flexion\_513[,i+2]

k <- k+1

matrix.diff513[,k] <- diff513

hist(diff513, main = "Flexion\_513")

i <- i+3

}

df513 <- data.frame(matrix.diff513)

mean(df513[,1])

median(df513[,2]) # the 2nd is closest to 0 one

mean(df513[,3])

mean(df513[,4])

mean(df513[,5])

median(df513[,6])

mean(df513[,7])

mean(df513[,8])

# Union\_581

par(mfrow=c(3,3))

matrix.diff581 <- matrix(rep(0,15512), nrow=1939)

i <- 1

k <- 0

while(i < 24){

diff581 <- Union\_581[,i+1] - Union\_581[,i+2]

k <- k+1

matrix.diff581[,k] <- diff581

hist(diff581, main = "Union\_581")

i <- i+3

}

df581 <- data.frame(matrix.diff581)

mean(df581[,1])

mean(df581[,2])

mean(df581[,3])

mean(df581[,4])

mean(df581[,5])

mean(df581[,6])

mean(df581[,7])

mean(df581[,8])

# all of them are quite close to each other, the 2nd and 7th one seems to be even better

* Two-Way ANOVA

flexion\_data <- read.csv("Flexion.csv",header = T)

union\_data <- read.csv("uni.csv",header = T)

names(flexion\_data)

flexion.difference <- matrix(rep(0,279216),nrow=2493)

i <- 1

j <- 1

flexion.mean <- NULL

flexion.se <- NULL

while (i < 225) {

flexion.difference[,j] <- flexion\_data[,i]-flexion\_data[,i+1]

flexion.mean[j] <- mean(flexion.difference[,j],na.rm = T)

flexion.se[j] <- sum((flexion.difference[,j])^2,na.rm = T)

i <- i+2

j <- j+1

}

flexion.mean # mean difference for Flexion group

flexion.se # squared error for Flexion group

union.difference <- matrix(rep(0,320096),nrow=2858)

i <- 1

j <- 1

union.mean <- NULL

union.se <- NULL

while (i < 225) {

union.difference[,j] <- union\_data[,i]-union\_data[,i+1]

union.mean[j] <- mean(union.difference[,j],na.rm = T)

union.se[j] <- sum((union.difference[,j])^2,na.rm = T)

i <- i+2

j <- j+1

}

union.mean # mean difference for Union group

union.se # squared error for Union group

flexion.type <- c(rep(c(rep("CF",4),rep("vM",4)), 14))

flexion.method <- c(rep(c("Idealized","Experimental","Specific","Generic"),28))

flexion.data.mean <- cbind(flexion.type, flexion.method, flexion.mean )

flexion.data.se <- cbind(flexion.type, flexion.method, flexion.se)

# Flexion: influence of types and methods on mean difference

flexion.mean.fit <- aov(flexion.mean ~ flexion.type\*flexion.method)

summary(flexion.mean.fit)

#Df Sum Sq Mean Sq F value Pr(>F)

#flexion.type 1 0.0639 0.06391 10.684 0.00147 \*\*

# flexion.method 3 0.0573 0.01912 3.196 0.02657 \*

# flexion.type:flexion.method 3 0.0067 0.00223 0.373 0.77284

#Residuals 104 0.6221 0.00598

# Flexion: influence of types and methods on squared error

flexion.se.fit <- aov(flexion.se ~ flexion.type\*flexion.method)

summary(flexion.se.fit)

#Df Sum Sq Mean Sq F value Pr(>F)

#flexion.type 1 75975 75975 5.108 0.0259 \*

# flexion.method 3 18659 6220 0.418 0.7403

#flexion.type:flexion.method 3 11397 3799 0.255 0.8573

#Residuals 104 1546770 14873

type.union <- c(rep(c(rep("CF",4),rep("vM",4)), 14))

union.method <- c(rep(c("Idealized","Experimental","Specific","Generic"),28))

union.data.mean <- cbind(type.union, union.method, union.mean )

union.data.se <- cbind(type.union, union.method, union.se )

# Union: influence of types and methods on mean difference

union.mean.fit <- aov(union.mean ~ type.union\*union.method)

summary(union.mean.fit)

#Df Sum Sq Mean Sq F value Pr(>F)

#type.union 1 0.003 0.00306 0.094 0.760

#union.method 3 0.015 0.00505 0.156 0.926

#type.union:union.method 3 0.023 0.00779 0.240 0.868

#Residuals 104 3.378 0.03248

# Union: influence of types and methods on squared error

union.se.fit <- aov(union.se ~ type.union\*union.method)

summary(union.se.fit)

#Df Sum Sq Mean Sq F value Pr(>F)

#type.union 1 9964 9964 0.263 0.609

#union.method 3 141505 47168 1.243 0.298

#type.union:union.method 3 159942 53314 1.405 0.245

#Residuals 104 3945671 37939

* Tukey HSD Test

flexion.mean.Tukey <- TukeyHSD(flexion.mean.fit, conf.level=0.95)

flexion.mean.Tukey

#Tukey multiple comparisons of means

#95% family-wise confidence level

#

#Fit: aov(formula = flexion.mean ~ flexion.type \* flexion.method)

#

#$flexion.type

#diff lwr upr p adj

#vM-CF -0.0477743 -0.07675817 -0.01879043 0.0014652

#

#$flexion.method

#diff lwr upr p adj

#Generic-Experimental -0.055162356 -0.10913295 -0.001191766 0.0431667

#Idealized-Experimental -0.048569144 -0.10253973 0.005401445 0.0935717

#Specific-Experimental -0.052206779 -0.10617737 0.001763811 0.0617960

#Idealized-Generic 0.006593212 -0.04737738 0.060563801 0.9887094

#Specific-Generic 0.002955577 -0.05101501 0.056926167 0.9989507

#Specific-Idealized -0.003637634 -0.05760822 0.050332955 0.9980516

plot(flexion.mean.Tukey)