# Warming Model-Species & Soil

*Jinfei Xue, Janhao Yan, Si Chen*

*Oct 27 2018*

## Introduction

The green leaf dataset mainly includes the following variables:

- Year: The year when sample collected
- Month: The month when sample collected
- Species: Tree species of leaf. ACRU = Acer rubrum (red maple); QURU = Quercus rubra (red oak)
- Treatment: Experimental treatment (Heated or Control)
- X.wt.Bsi: Percent biogenic silica in leaves by dry weight (0-100), indicating the silica level

| Model 1 : Tree Species |
| --- |

## Exploratory Data Analysis (EDA)

In order to explore the relationship between the silica levels in leaves and some factors, including tree species(ACRU/QURU) and treatments(Heated/Control), and how the silica levels in tree leaves changed as time went by, we first make exploratory data analysis as followings.

### 1.1 Compare silica levels under different treatments

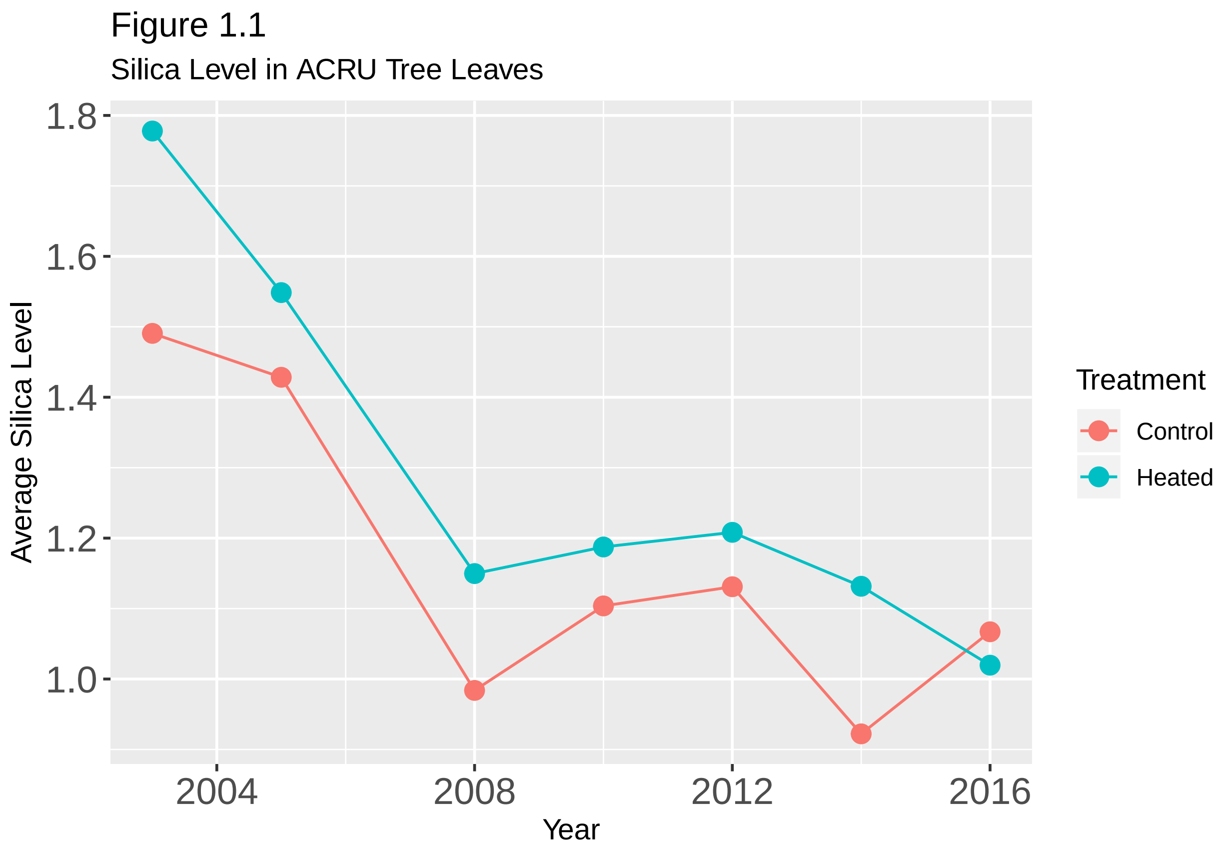
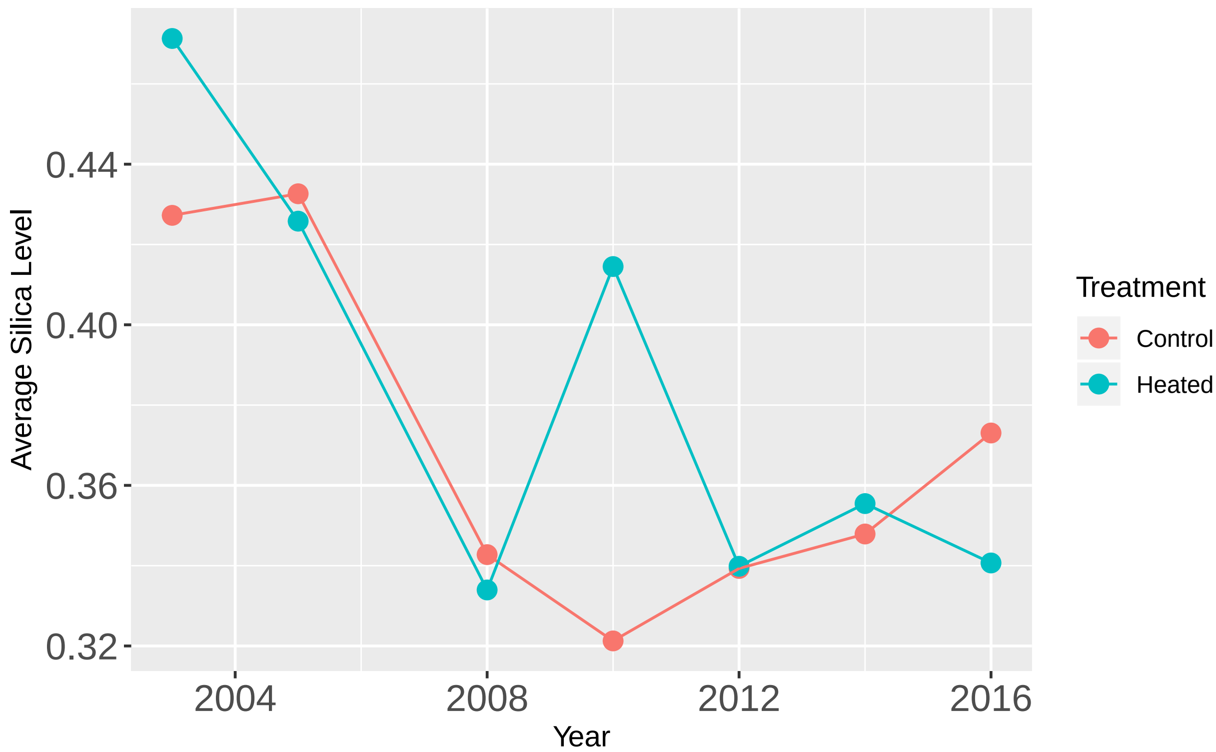


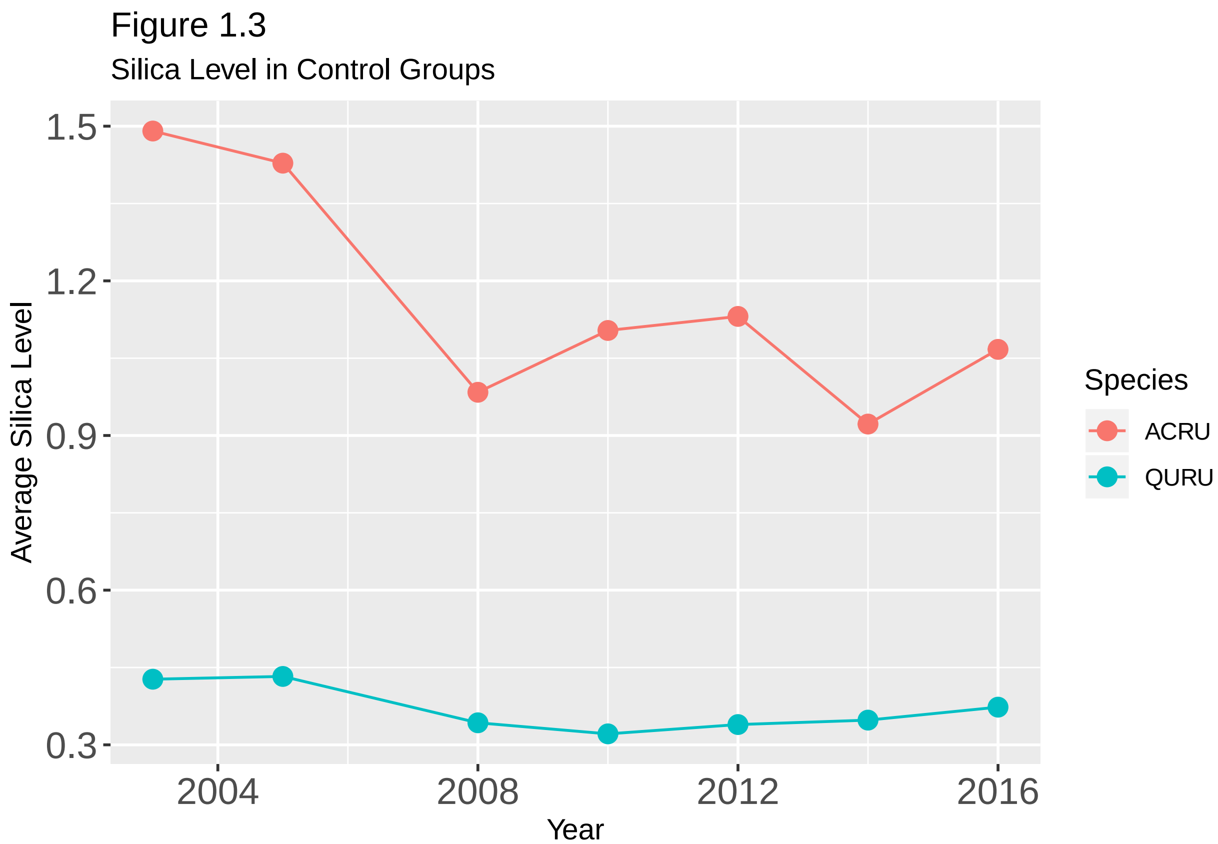
Figure 1.2

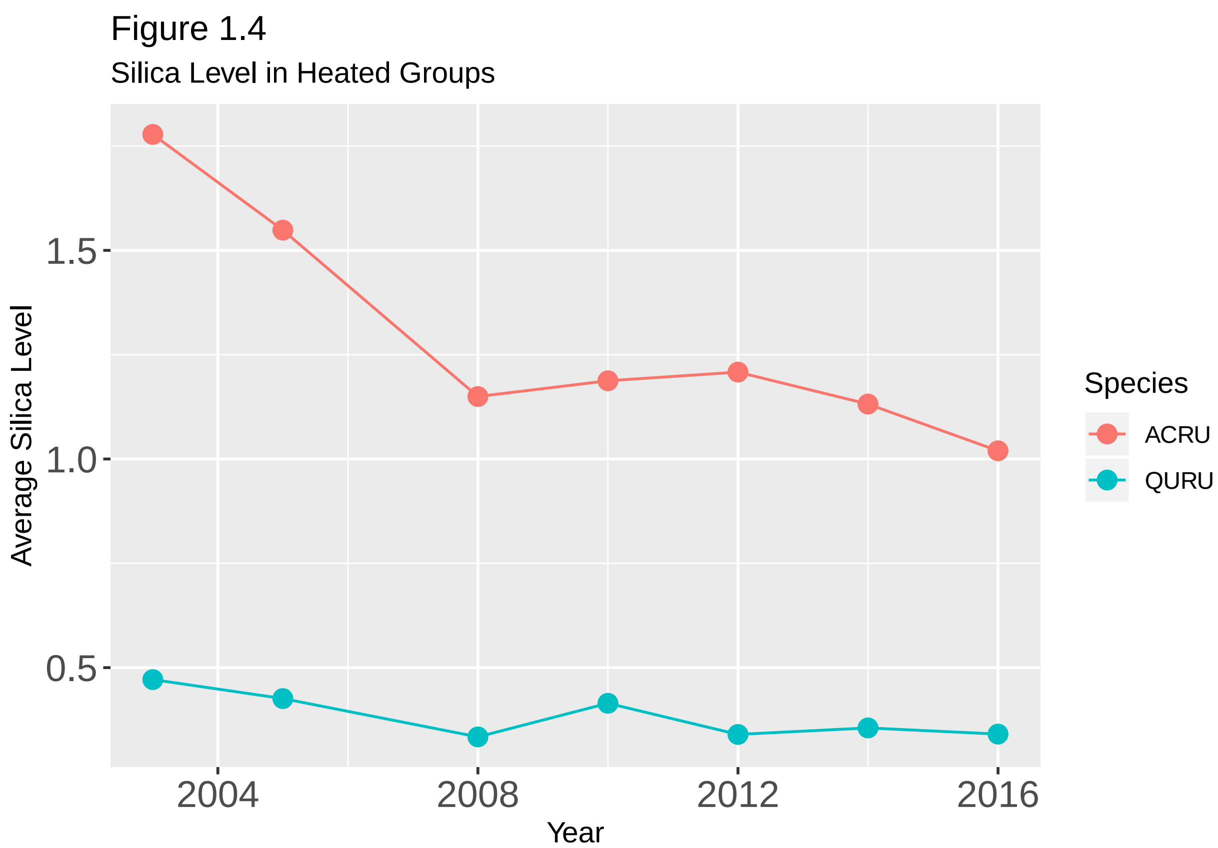
Silica Level in QURU Tree Leaves




By comparing the two plots above, we can clearly see the silica levels in QURU tree leaves of heated group are higher than that in QURU tree leaves of control group.

**Compare Silica levels in different tree species**






By comparing the two plots above, we can clearly see the silica levels in ACRU tree leaves are much higher than that in QURU tree leaves.

## Modeling

In the model, X.wt.Bsi is the dependent variable; Species, Treatment, Months and Month are selected as independent variables.

Months indicate the number of months from the time when the experiment started to the time when the samples were collected. We make log transformation for X.wt.Bsi and standardize Months.

```
##
## Call:
## lm(formula = log(X.wt.Bsi) ~ factor(Species) + factor(Treatment) +
##     Months + factor(Month) + factor(Species) * Months, data = glbsi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.14741 -0.05672 -0.00257  0.04226  0.33510
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.11221    0.02585   4.341 4.26e-05 ***
## factor(Species)QURU        -1.17285    0.01723 -68.085  < 2e-16 ***
## factor(Treatment)Heated     0.06759    0.01723   3.924 0.000188 ***
## Months                     -0.03731    0.01849  -2.018 0.047119 *
```

```
## factor(Month)7              -0.05227    0.02725  -1.918 0.058837 .
## factor(Month)8               0.24314    0.03545   6.859 1.53e-09 ***
## factor(Species)QURU:Months  0.06586    0.01733   3.800 0.000287 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07894 on 77 degrees of freedom
## Multiple R-squared:  0.9845, Adjusted R-squared:  0.9833
## F-statistic: 817.7 on 6 and 77 DF,  p-value: < 2.2e-16
```

From the regression result, we can see that:

1) R-squared is 0.9845(near 1), which means that 98.45% variations in dependent variable can be explained by independent variables in the model, indicating that our regression model fits the data very well.
2) p-value of F-statistic is very small, indicating that all of independent variables in the model can together explain the data very well.
3) Most of coefficients are statistically significant, indicating that each independent variable can explain the data very well.

Now, we look at the marginal model plots and residual plots to check the fitted model

```
##                (Intercept)        factor(Species)QURU
##                       0.11                      -1.17
##    factor(Treatment)Heated                     Months
##                       0.07                      -0.04
##             factor(Month)7             factor(Month)8
##                      -0.05                       0.24
## factor(Species)QURU:Months
##                       0.07
```

So the regression model is:

$log(X.wt.Bsi) = 0.11 - 1.17 \times QURU + 0.07 \times Heated - 0.04 \times Months - 0.05 \times Month_7 + 0.24 \times Month_8 + 0.07 \times QURU \times Months$

Model 2 : Soil Layers

# Introduction

The soil dataset mainly includes the following variables:

- Date: The date when sample collected
- Layer: Layer of soil core (organic or mineral)
- Treatment: Experimental treatment (Heated or Control)
- X.wt.Bsi: Percent biogenic silica in leaves by dry weight (0-100), indicating the silica level

# Exploratory Data Analysis (EDA)

In order to explore the relationship between the silica levels in soil and some factors, including the layer of soil core (organic/mineral) and treatments(Heated/Control), and how the silica levels in soil changed as time went by, we first make exploratory data analysis as followings.

**Compare silica levels under different treatments**

Figure 2.1

Silica Level in Soil Mineral Level

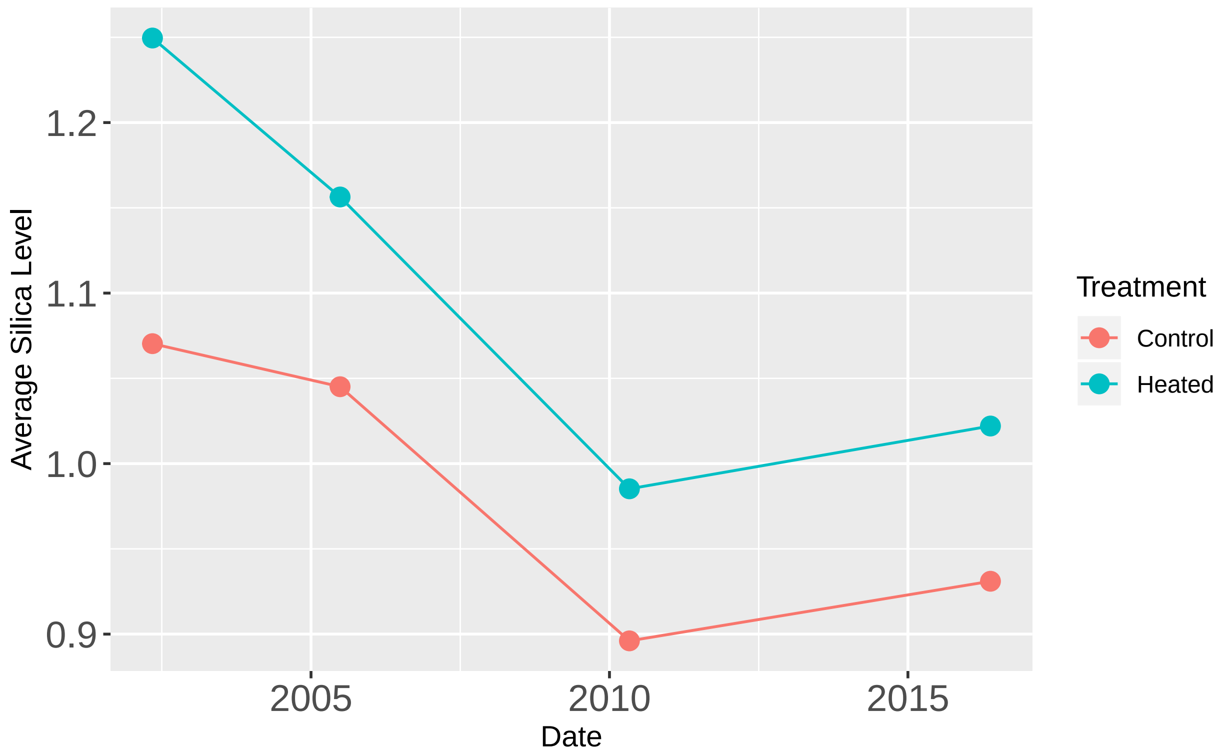
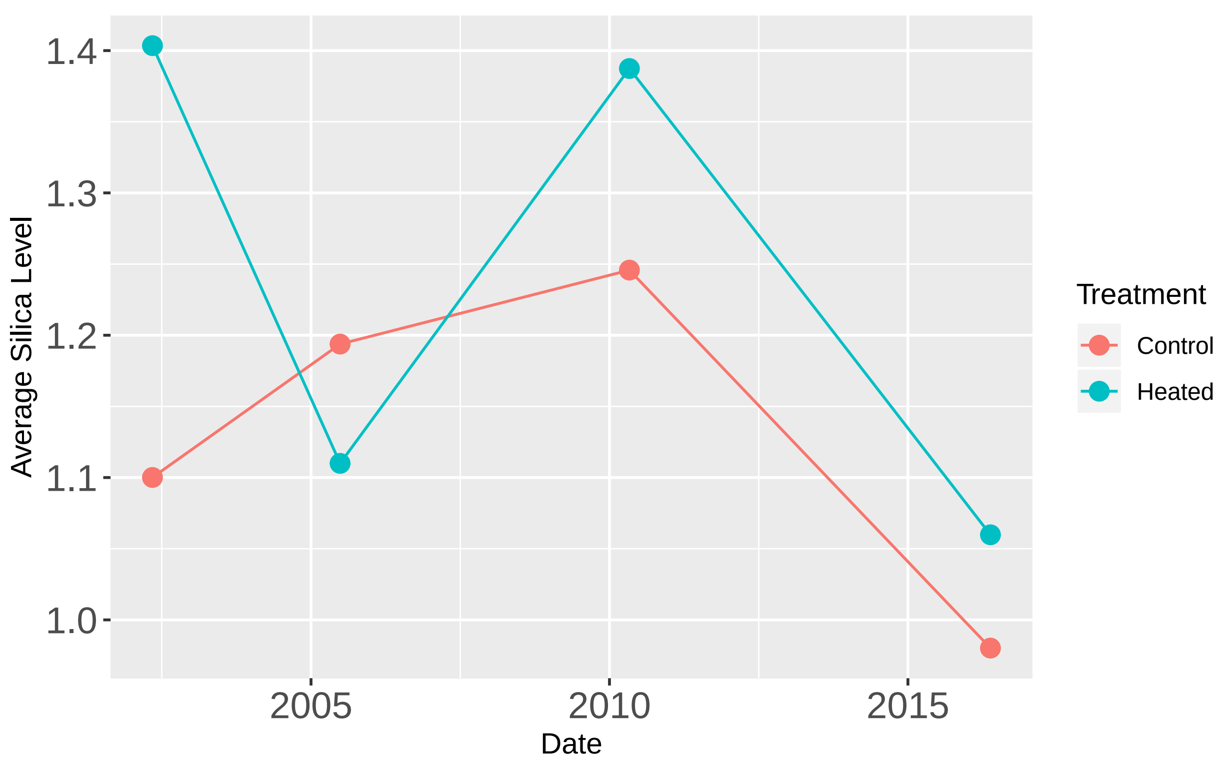



Figure 2.2

Silica Level in Soil Organic Level



By comparing the two plots above, we can clearly see almost all silica levels in control group(except one point) are higher than that in heated group.

# Modeling

In the model, X.wt.Bsi(Silicalevel) is the dependent variable; Layer, Treatment, deployedtime and season are selected as independent variables.

"deployedtime" indicates the number of days from the time when the experiment started(2002/05/07) to the time when the samples were collected. "season" indicates the month of the date when the samples were collected.

```
##
## Call:
## lm(formula = Silicalevel ~ factor(Layer) + factor(Treatment) +
##     deployedtime + factor(season), data = soil_data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48545 -0.10855  0.01391  0.16089  0.35609
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1.090e+00  7.527e-02  14.485   <2e-16 ***
## factor(Layer)Organic     1.406e-01  6.414e-02   2.192   0.0339 *
## factor(Treatment)Heated  1.139e-01  6.414e-02   1.776   0.0827 .
## deployedtime            -3.976e-05  1.764e-05  -2.255   0.0293 *
## factor(season)6         -4.559e-02  7.885e-02  -0.578   0.5661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2222 on 43 degrees of freedom
## Multiple R-squared:  0.2333, Adjusted R-squared:  0.162
## F-statistic: 3.271 on 4 and 43 DF,  p-value: 0.01989
```

From the regression result, we can see that:

1) p-value of F-statistic is smaller than 0.05, indicating that all of independent variables in the model can together explain the data well.
2) Most of coefficients are statistically significant, indicating that each independent variable can explain the data well.

```
##             (Intercept)    factor(Layer)Organic factor(Treatment)Heated
##                1.090276                0.140582                0.113948
##            deployedtime         factor(season)6
##               -0.000040               -0.045594
```

So the regression model is:

$X.wt.Bsi = 1.090276 + 0.140582 \times Organic + 0.113948 \times Heated - 0.000040 \times deployedtime - 0.045594 \times season_6$

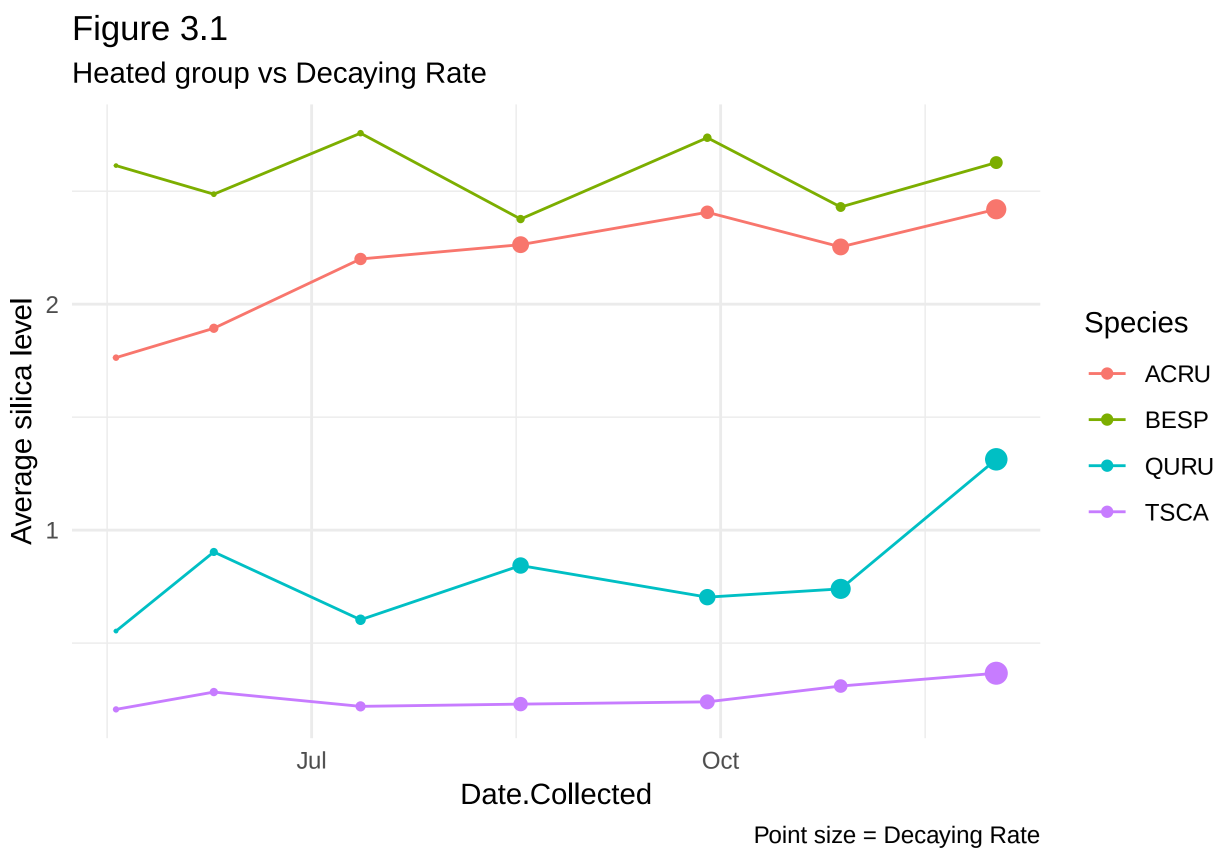
| Model 3: Silica Level VS Decaying Rate |
|---|

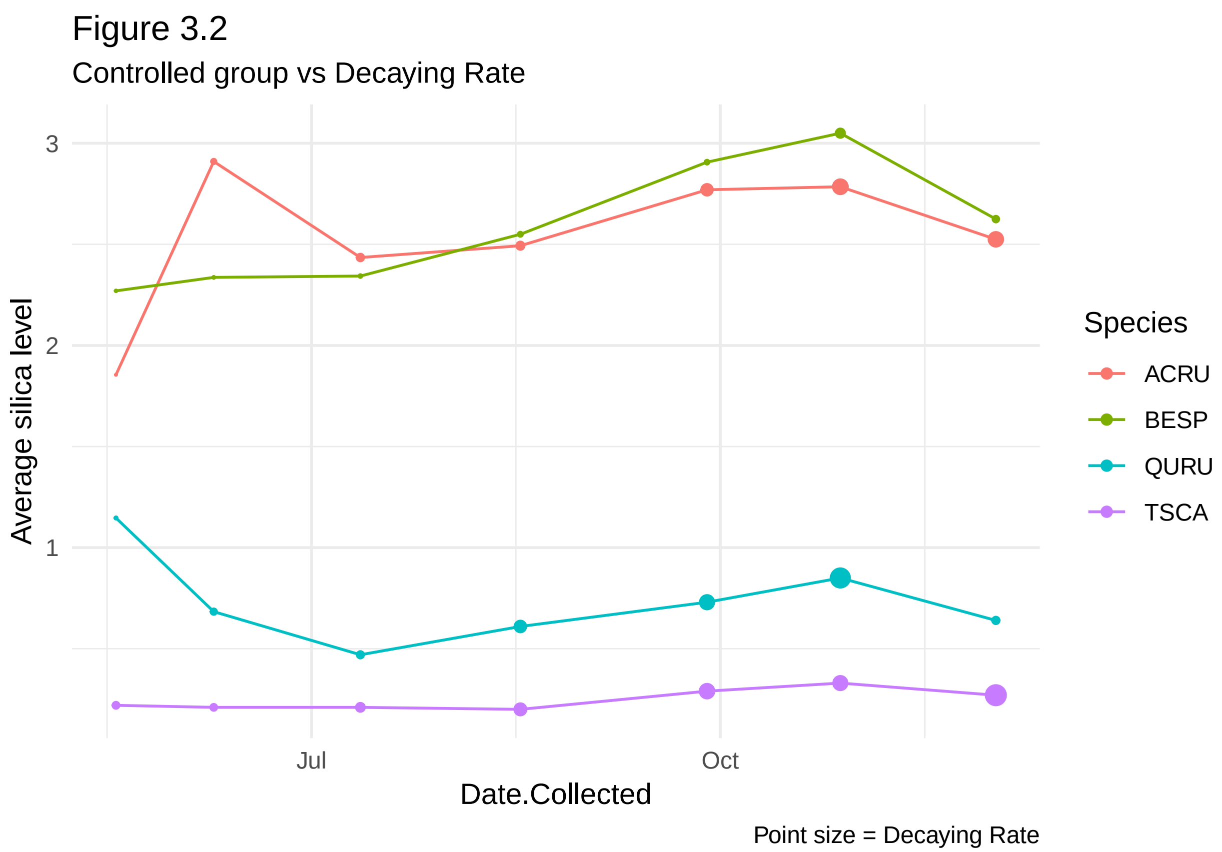
## Introduction

This dataset mainly includes these variables -Species: The species of the leaves that we collected -Treantment: Heated group and Controlled group -Time.Deployed: The length of time from the deployed date to collected date -Mass.Loss: The weight of the leaves caused by decomposition -X.wt.Bsi: The silica level in leaves

## Exploratory Data Analysis (EDA)

In order to explore the relationship between the silica levels in leaves and decaying rate under different species and treatments, we draw two plots.






By comparingh these two plots ,we can find that the heated treatment can cause faster decaying rate than the controlled group. What's more, we should notice that the decaying rate increases with the time increases. But we should also notice that it seems like that silica level may not have such a strong relationship with decaying rate. And it is obvious that the silica level in TSCA is the lowest both in heated and controlled group.

# Modelling

In this section, our group want to figure out what is the relationship between silica level in leaves and decaying rate. We draw the two plots. And we fit the linear regression model. In this model, we do z-scores to the silica level and log transformation. The reponsive variable is silica level, and independent variables are deploy time, mass.loss, tree species and treatment.

```
##
## Call:
## lm(formula = X.wt.Bsi ~ factor(Species) + factor(Treatment) +
##     Time.Deployed + Mass.Loss, data = datalb1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78085 -0.21964 -0.04029  0.14841  1.80911
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.2393264  0.1055297  21.220   <2e-16 ***
```

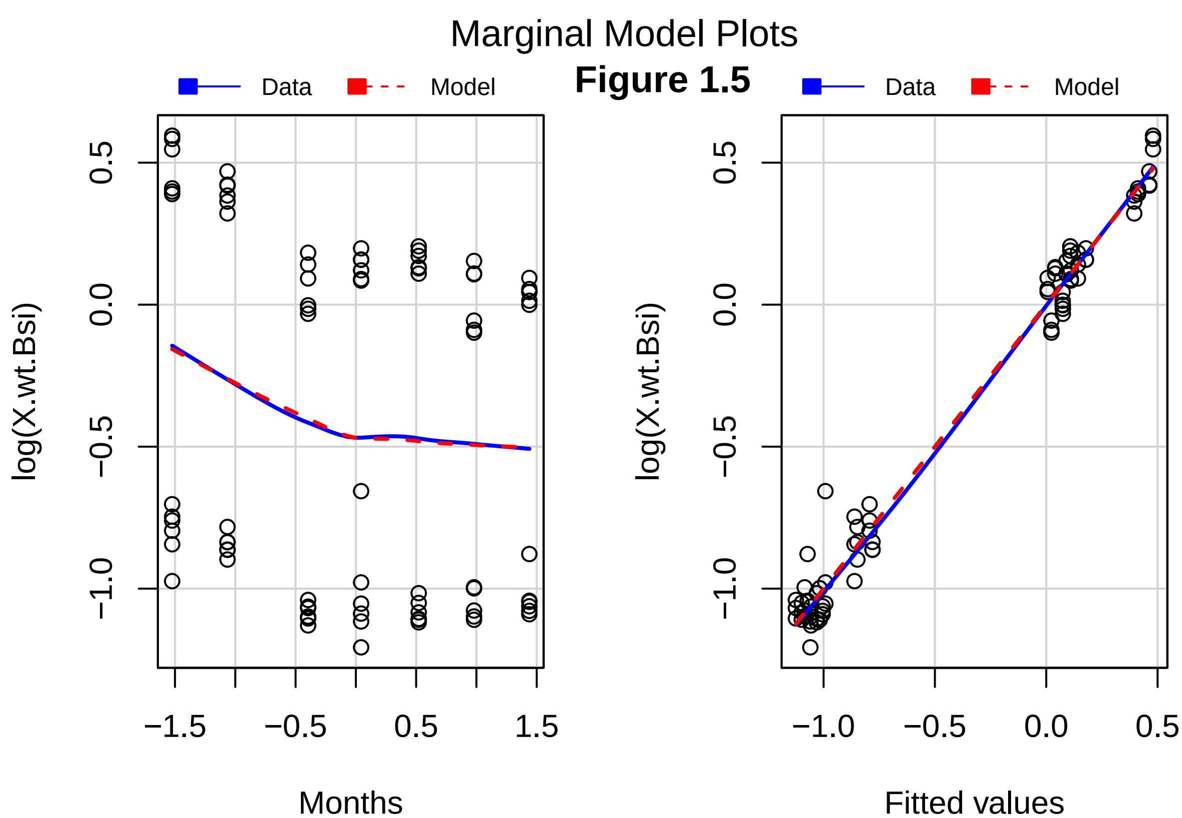
```
## factor(Species)BESP       0.2361733  0.1201926   1.965   0.0513 .
## factor(Species)QURU      -1.5711491  0.1049745 -14.967   <2e-16 ***
## factor(Species)TSCA      -2.0925649  0.1085454 -19.278   <2e-16 ***
## factor(Treatment)Warming -0.0869626  0.0767125  -1.134   0.2588
## Time.Deployed             0.0011800  0.0009257   1.275   0.2044
## Mass.Loss                 0.0212869  0.0714187   0.298   0.7661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4577 on 144 degrees of freedom
## Multiple R-squared:  0.8299, Adjusted R-squared:  0.8228
## F-statistic: 117.1 on 6 and 144 DF,  p-value: < 2.2e-16
```

The overview of this model gives us a good first impression. The p-value is small and r squared is near to 1. What's more, the coeficients of intercept, QURU and TSCA are significant. And we combine our model with our plots. From the plots, we can find with the deployed time increasing, the decaying rate tends to increase too. From the model, the coeeficient of Time.deployed is 0.0011800 which is positve, and this correspond what we find in our plots.

# Appendix

## Model 1

## Marginal Model Plots



From the marginal model plots above, we can see the red dash line for the model almost cover the blue line for data, indicating that the model fit the data well.

### Residual Plots

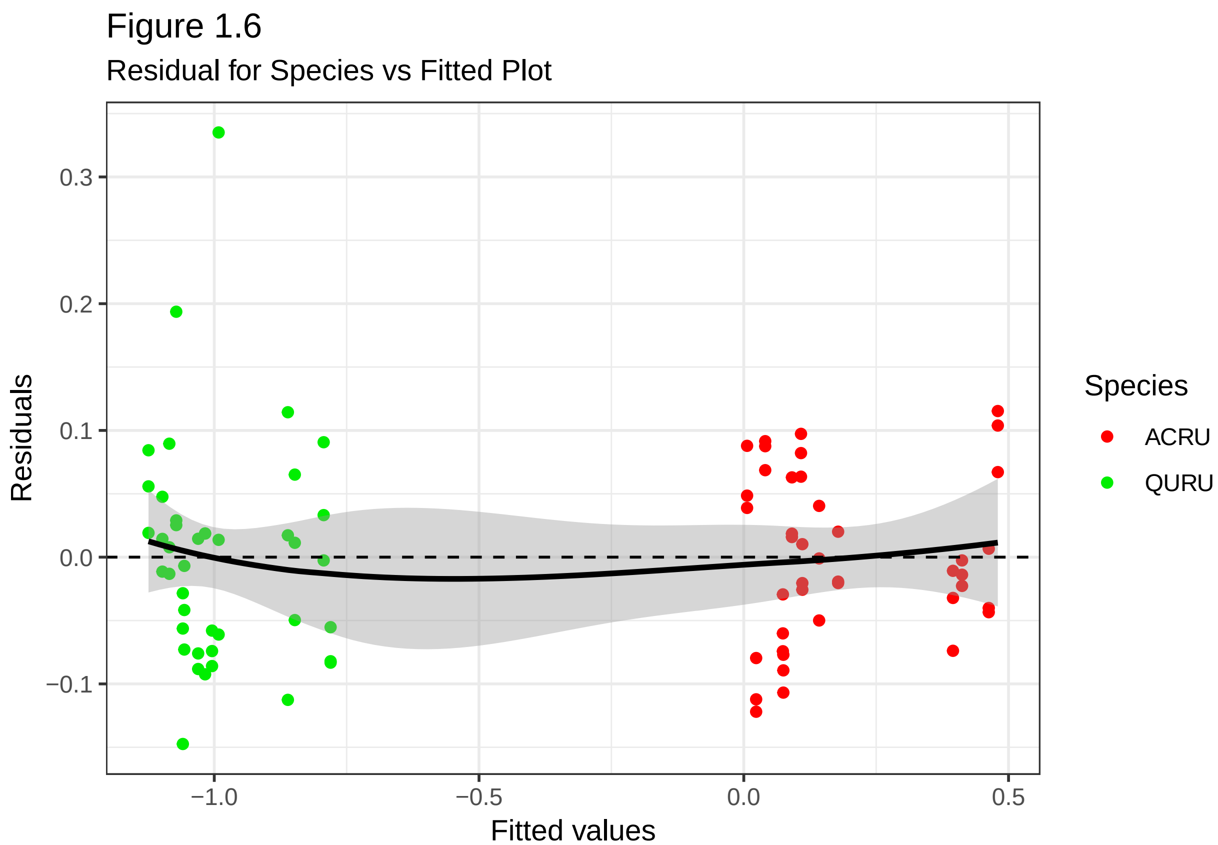
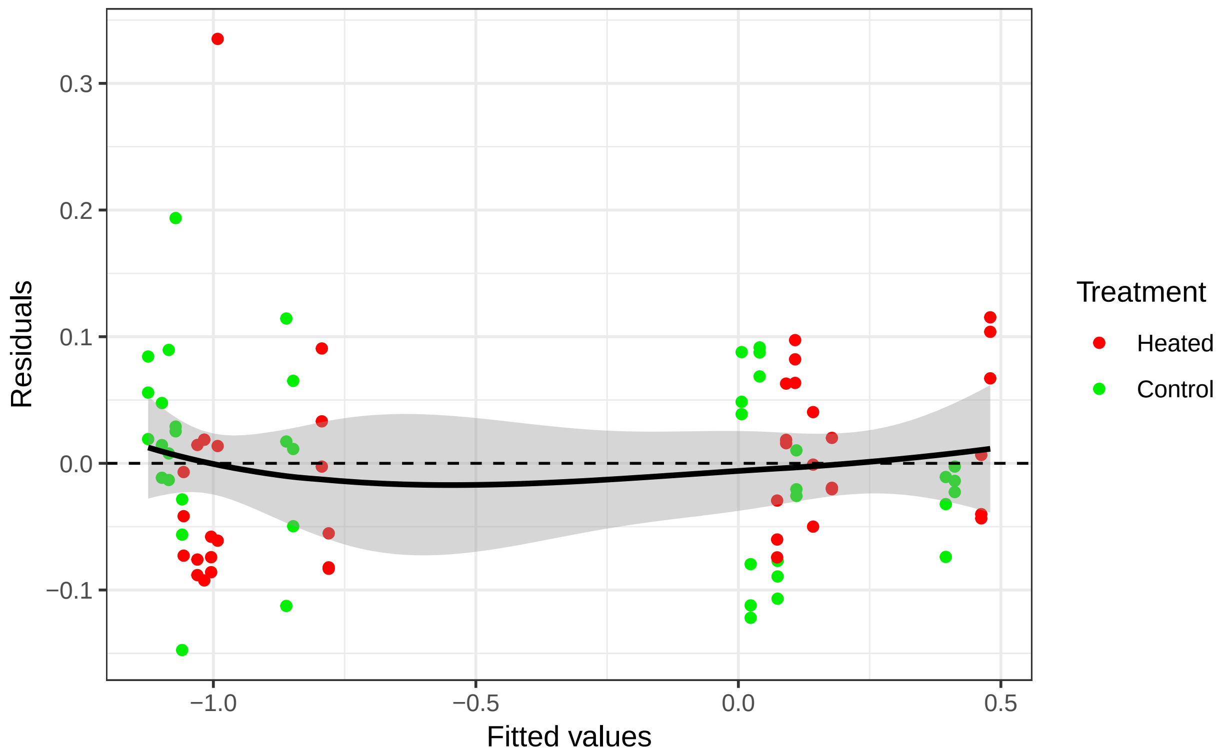
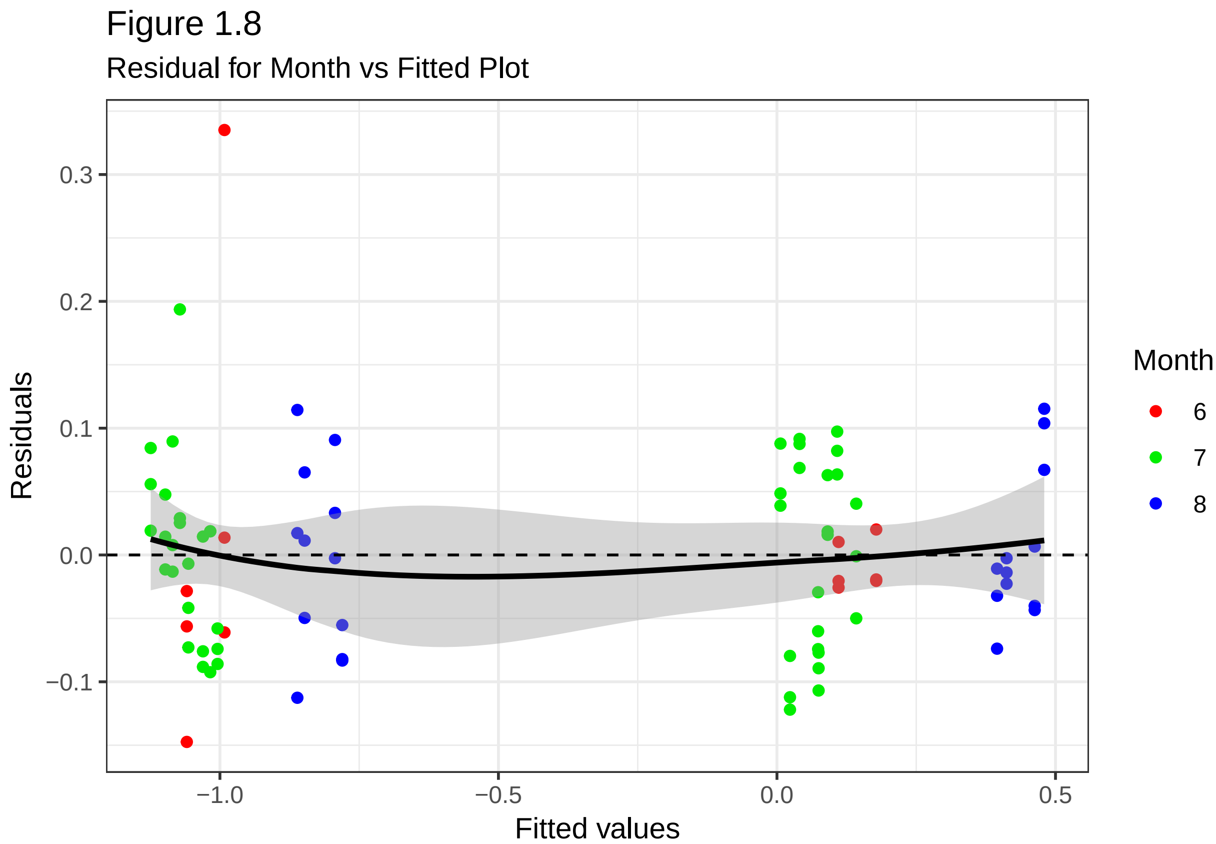


Figure 1.7

Residual for Treatment vs Fitted Plot



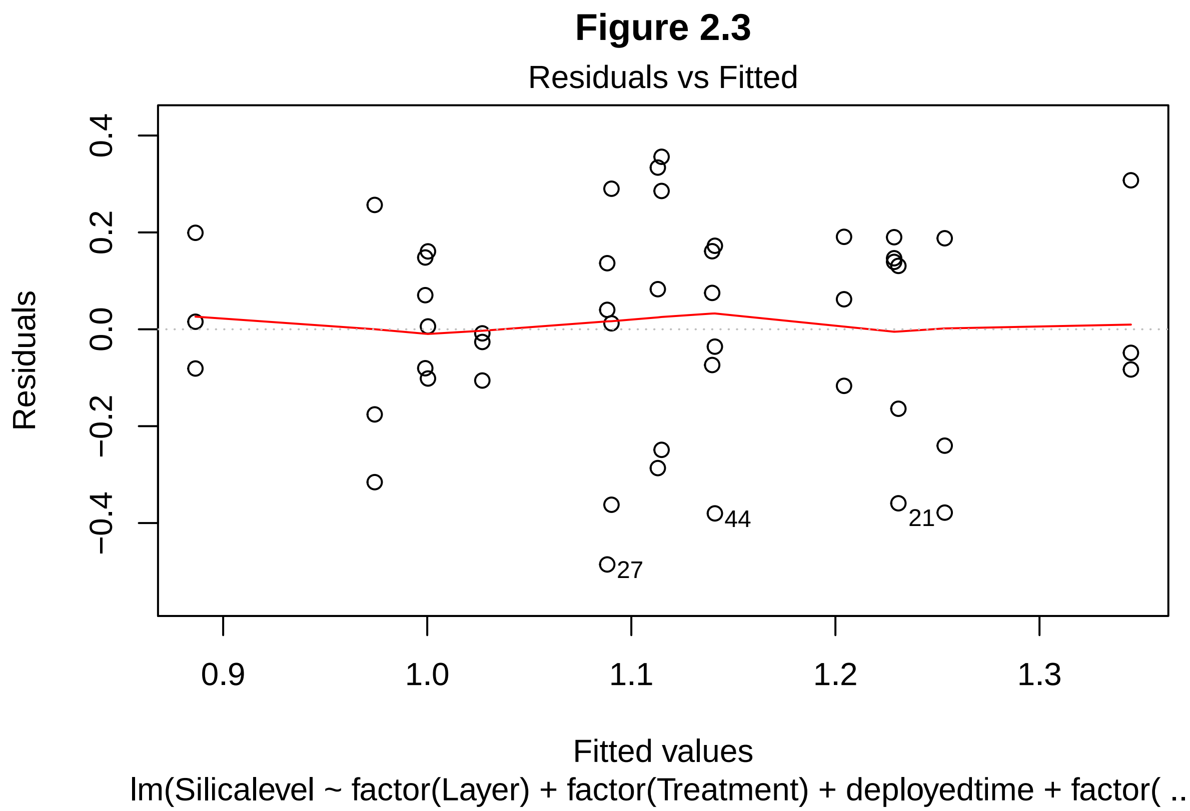



## summarise From the residual plots above, we can conclude that the model fits the data well.

### Model 2

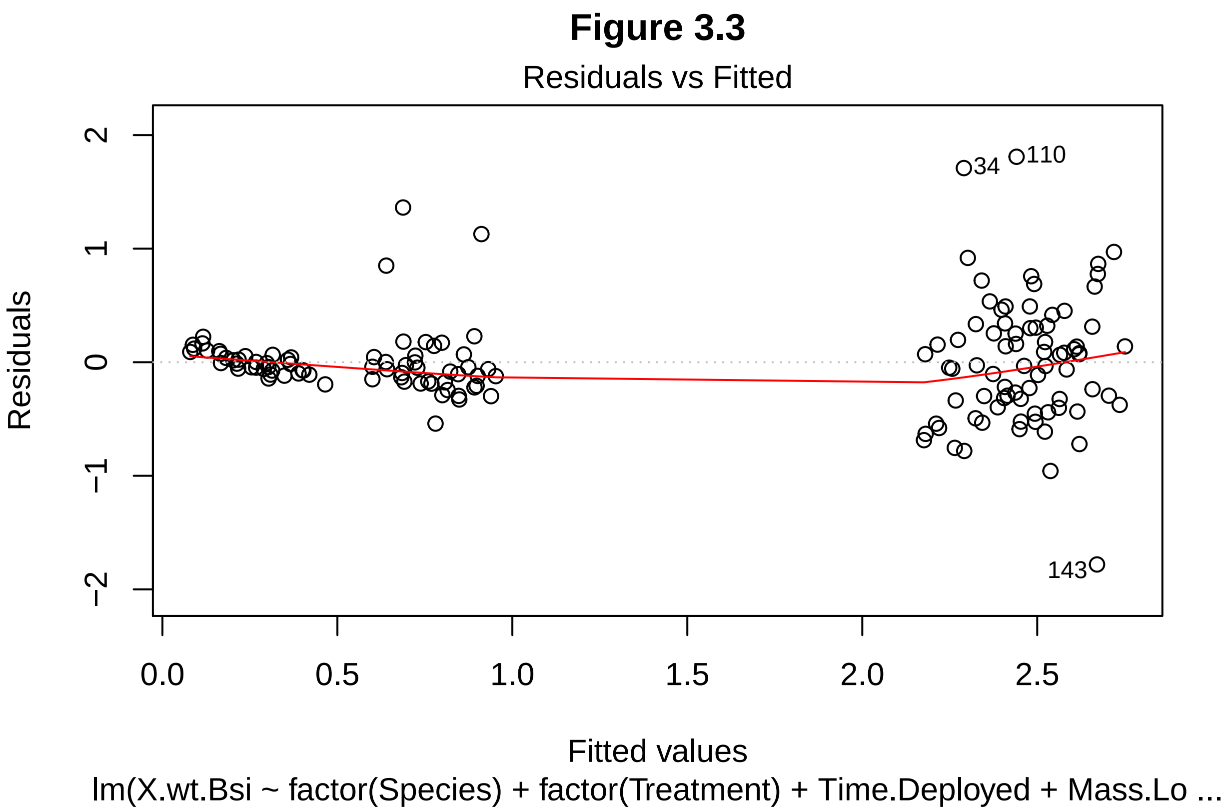
Then we look at the residual plot to check the fitted model

**Residual Plots**



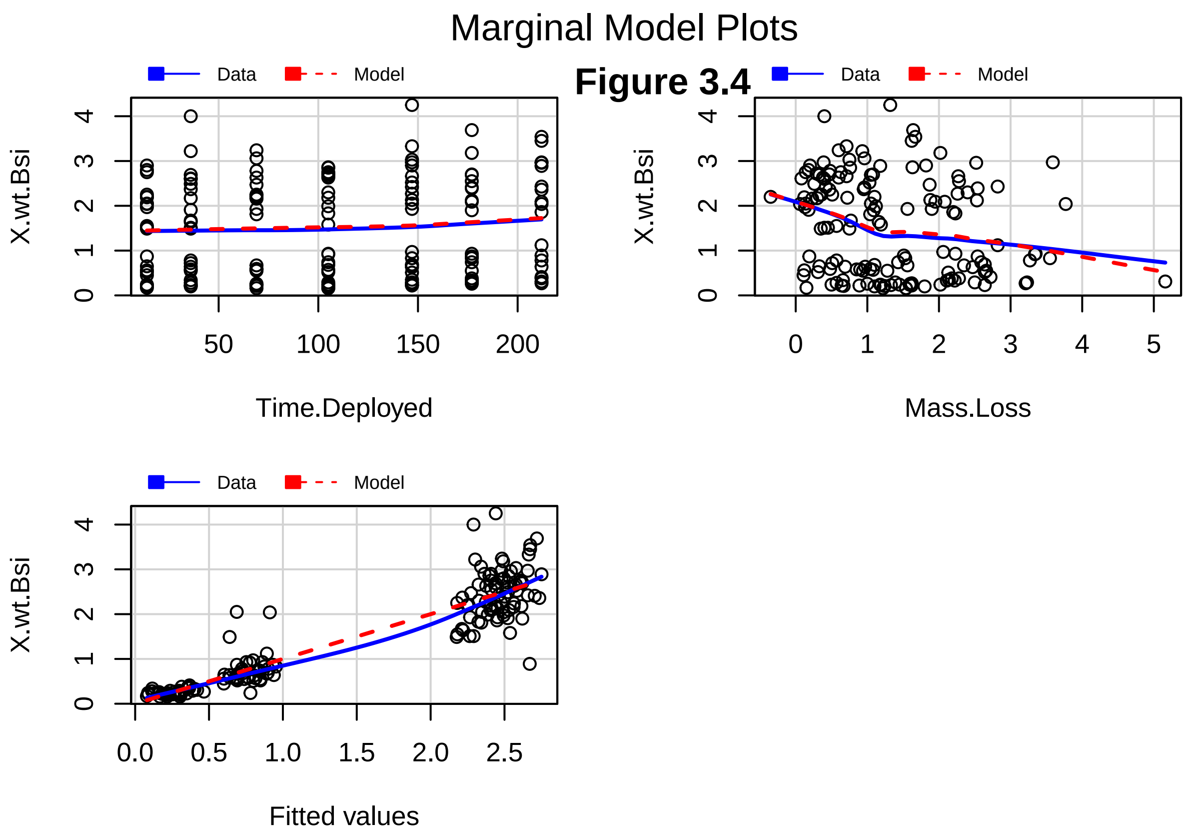
**Model 3**

**residual plot**




we can find that the residual is not that bad.

### marginal model plots




we can find marginal model plots is good.

### summarise

This model can do a great job to describe our data.