Warming Model-Species & Soil

Si Chen, Jinfei Xue, Jianhao Yan

Nov 20 2018

1. Introduction

1.1 Overview

Global warming has become a significant topic in the ecosystem field. It's worthwhile to study the effects global warming has brought about. Therefore, a long-term ecosystem warming experiment was established at Harvard Forest (Petersham, MA) in 2002. In this experiment, two 30 x 30 meter plots were created. One plot is continuously heated to 5 degrees Celsius above ambient temperature and we will refer to this as the treatment, while the other plot is left as a control. A wide variety of ecological response variables have been measured over the course of the experiment. The data were collected from 2002 to 2016 in different time intervals.

In this study, silica (silicon dioxide; SiO2) concentration was measured in the two ecosystem pools. Measurements can be divided into two broad categories including single-time point measurements of silica pools, specifically, green leaves, leaf litter, and soil, and measurement of silica loss from decaying leaves over time, using a litterbag experiment.

Our Client is interested in finding the relationship between the treatment and the silica level in different tree species and soil types, and the relationship between temparature and decomposition rate given the factors including tree species, temperature and time.

1.2 Outline

There are three seperate datasets corresponding to the three relationship. Therefore, we consider each relationship seperately. For each one, we first conduct exploratory data analysis to explore the relationship among variables in each dataset. Next, we fit models appropriate to each corresponding dataset. Lastly, we interpret our model and discuss conclusions.

2. Model1 : Tree Species

2.1 Introduction

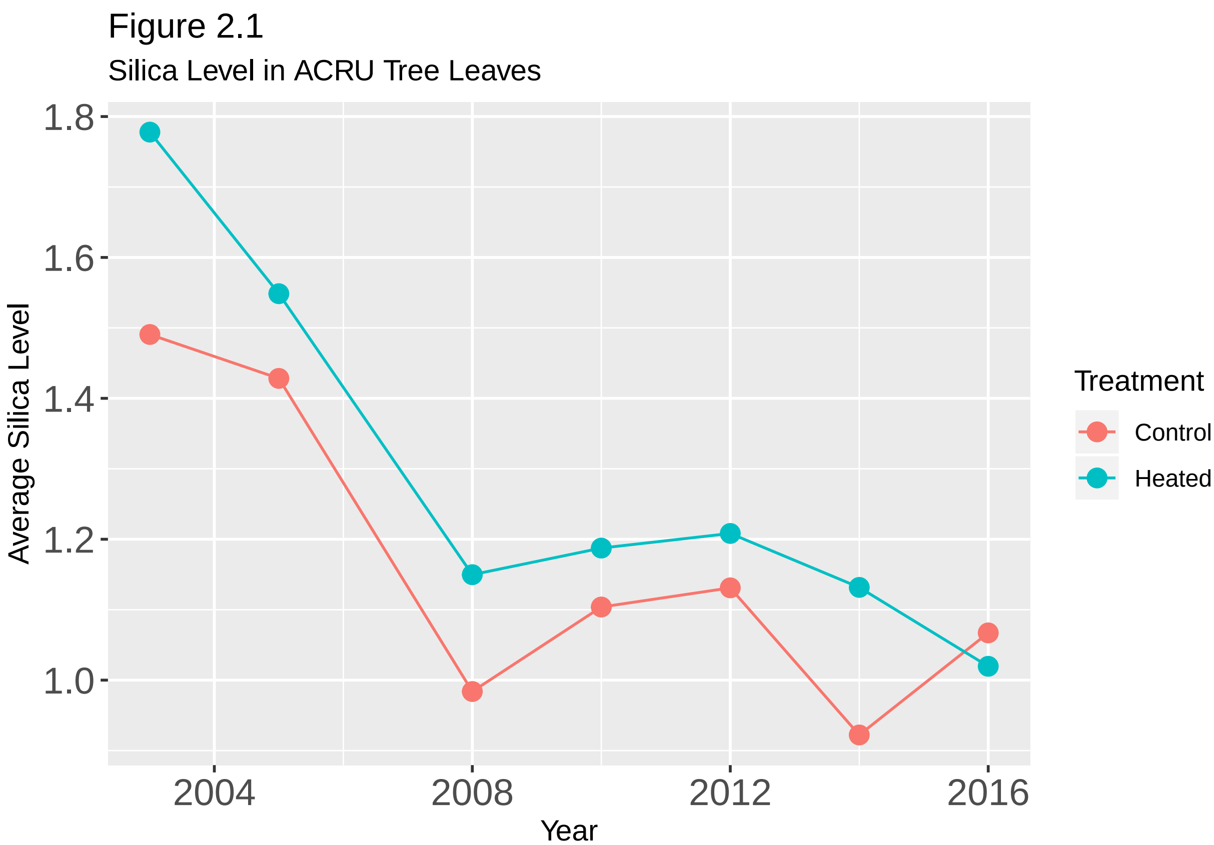
The green leaf dataset mainly has variables including Year (the year when sample collected), month (the month when sample collected), species (tree species of leaf, ACRU = Acer rubrum (red maple), QURU = Quercus rubra (red oak)), treatment(experimental treatment: Heated or Control) and X.wt.Bsi (percent biogenic silica in leaves by dry weight (0-100), indicating the silica level).

2.2 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.

Figure 2.1 & 2.2: Compare silica levels in leaves under different treatments

```
ggplot(data = glbsi_avg_ACRU) +
  aes(x = Year, y = avg, color = Treatment) +
  geom_point(size = 3) +
  labs(title = "Figure 2.1",
       subtitle = "Silica Level in ACRU Tree Leaves",
       x = "Year",
       y = "Average Silica Level")+
  theme(axis.text.x = element_text(size = 14))+
  theme(axis.text.y = element_text(size = 14))+
  geom_line()
```



2.2.1 Silica Level in ACRU Tree Leaves

In figure 2.1, avarage silica level is the reponsive variable, and we wat figure out how the tratment affects the silica level in ACRU(Red Maple). Adding lines connecting the points, we can see more clearly that heated treatment can bring a positive effect on silica level in ACRU tree leaves since the heated group line is mostly above the control group line. This follows our expectation if we supposed to see heat having positive effects. In 2016, there is an exception that heated group has lower average silica level than control group, but the overall trends of silica level that increases with the tempreture increasing is still obvious. What's more, we can also see a decreasing trend overtime for both groups

```
ggplot(data = glbsi_avg_QURU) +
  aes(x = Year, y = avg, color = Treatment) +
  geom_point(size = 3) +
```

```r
labs(title = "Figure 2.2",
     subtitle ="Silica Level in QURU Tree Leaves",
     x = "Year",
     y = "Average Silica Level")+
theme(axis.text.x = element_text(size = 14))+
theme(axis.text.y = element_text(size = 14))+
geom_line()
```

2.2.2 Silica Level in QURU Tree Leaves

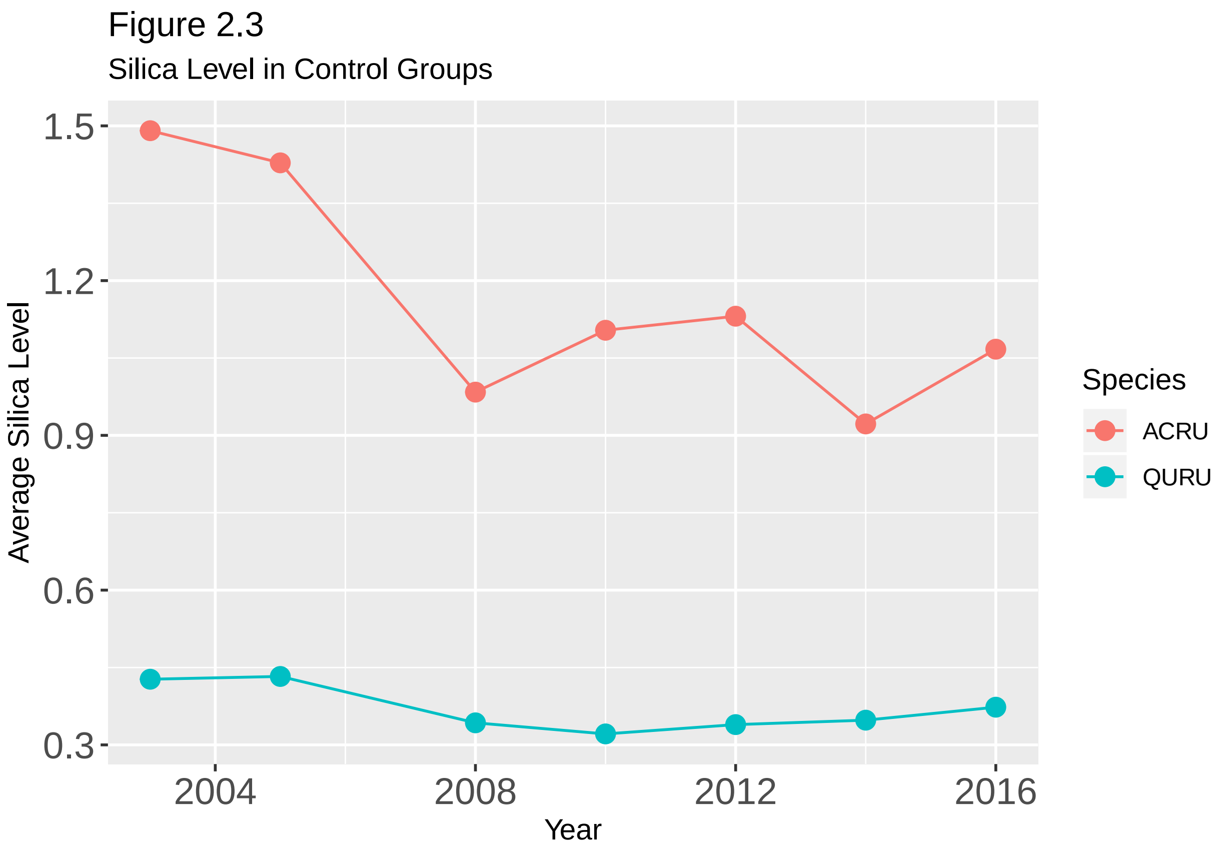
Like the ACRU plot, we move to the other tree species QURU, which stands for Red Oak. We again compare the difference between the treatment groups. This time the trend overtime and difference are not as clear as in ACRU case. There is a pike for heated group and a low point for controlled group in 2010 where a possible explaination for the sudden change might be the extreme weather in 2010. In other years, there is no explicit difference saying that one group has highr silica level than the other. Also, we might not be able to say that there is a declining trend overtime because of the peak of heated group in 2010.

2.2.3 Summery

By comparing the two cases above, we can clearly see that in ACRU species, heat may have positive effect on silica level while in QURU species, heat is not a varaible with much significance. Overall, silica level declines overtime, but there might exist exception such as heated QURU group is 2010.

Figure 2.3 & 2.4: Compare silica levels in different tree species

```
ggplot(data = glbsi_avg_control) +
  aes(x = Year, y = avg, color = Species) +
  geom_point(size = 3) +
  labs(title = "Figure 2.3",
       subtitle ="Silica Level in Control Groups",
       x = "Year",
       y = "Average Silica Level")+
  theme(axis.text.x = element_text(size = 14))+
  theme(axis.text.y = element_text(size = 14))+
  geom_line()
```

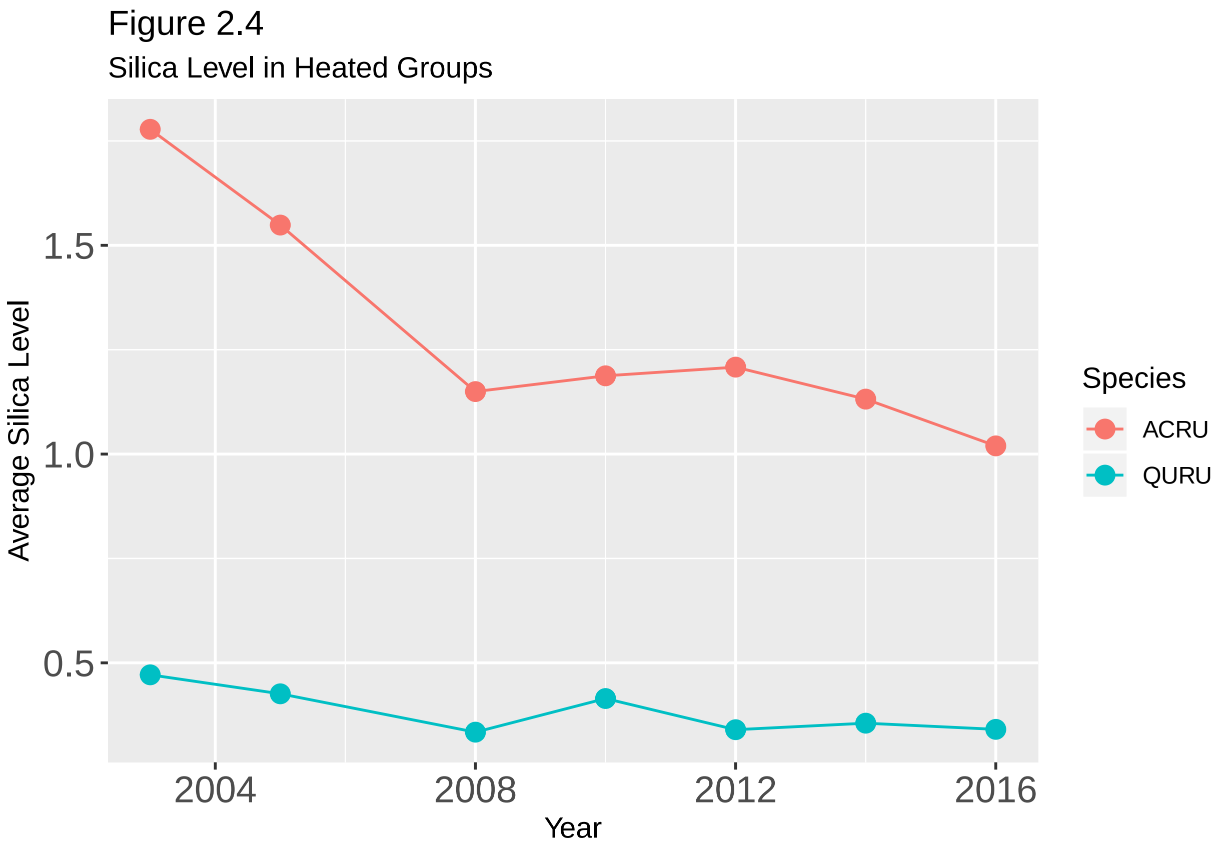


2.2.4 Silica Level in Control Groups

After compare difference between treatment groups, we now compare in another set up. In this case, we change the target variable from treatment to tree species. In order to compare difference between tree species, we neet to put different species under same treatment. In contrel group, we can conclude that silica level in ACRU species is higher than in QURU since the ACRU line is way above the QRUR line. Overtime, silica level in ACRU is slightly decreasing and has some fluctuation. Silica level in QURU is relatively stable.

```
ggplot(data = glbsi_avg_heated) +
  aes(x = Year, y = avg, color = Species) +
  geom_point(size = 3) +
  labs(title = "Figure 2.4",
       subtitle = "Silica Level in Heated Groups",
```

```r
      x = "Year",
      y = "Average Silica Level")+
  theme(axis.text.x = element_text(size = 14))+
  theme(axis.text.y = element_text(size = 14))+
  geom_line()
```



2.2.5 Silica Level in Heated Groups

Similar result shows in heated group: ACRU line is totally above QURU and has a decreasing trend overtime; QURU line is stable.

2.2.6 Summary

Since under both groups, we derived same conclusion that QURU has lower silica level than ACRU. So we may conclude that tree species do affect the silica level in tree leaves. For example, ACRU is higher than QURU.

2.3 Modeling

In the model, X.wt.Bsi is the dependent variable; Species, Treatment, Months and Month are selected as independent variables. Months indicates the number of months from the time when the experiment started to the time when the samples were collected. Month indicates the month when the samples were selected.

In order to make the model fit the data better, we make log transformation for X.wt.Bsi. From the previous EDA, we can find out that the effect of different species on silica level depends on the value of Months. That is to say, the two variables (Species and Months) interact with each other. Therefore, we add the interaction term for these two variables in the model.

```
glbsi <- read.csv("green_leaf_bsi.csv", encoding = "UTF-8")

glbsi$Year <- as.character(glbsi$Year)
glbsi$Month <- as.character(glbsi$Month)
glbsi$Date <- as.yearmon(paste0(glbsi$Year, "-0", glbsi$Month))
glbsi$Months <- time_length(interval(as.yearmon("2002-05"),glbsi$Date), "month")

r_species <- lm(log(X.wt.Bsi) ~
                  factor(Species) + factor(Treatment) + Months +
                  factor(Month)+factor(Species)*Months, data = glbsi)
```

2.4 Discussions

```
summary(r_species)
```

```
##
## 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)                0.1797475  0.0427875   4.201 7.08e-05 ***
## factor(Species)QURU       -1.2920691  0.0357889 -36.103  < 2e-16 ***
## factor(Treatment)Heated    0.0675925  0.0172263   3.924 0.000188 ***
## Months                    -0.0007130  0.0003534  -2.018 0.047119 *
## factor(Month)7            -0.0522699  0.0272543  -1.918 0.058837 .
## factor(Month)8             0.2431411  0.0354478   6.859 1.53e-09 ***
## factor(Species)QURU:Months  0.0012587  0.0003312   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 R-squared is 0.9845(near 1), which means that 98.45% of the variations in dependent variable can be explained by independent variables in the model.

Besides, p-value of F-statistic is very small, indicating that the linear regression model provides a better fit to the data than a model that contains no independent variables.

The p-value for each independent variable tests the null hypothesis that the variable has no correlation with the dependent variable. In the regression result, p-values of most of coefficients in the model are smaller than the usual significance level of 0.05, so we should reject the null hypothesis, which means those coefficients are statistically significant and we should not remove them.

The coefficient of **factor(Species)QURU** is negative, indicating that with the same level of the other variables, the silica level in QURU tree leaves is lower than that in ACRU tree leaves. This result is consistent with the comparison between figure 2.3 and 2.4. The coefficient of **factor(Treatment)Heated** is positive, indicating that with the same level of the other variables, the silica level under heated treatment is lower than that under control treatment. This result is consistent with the comparison between figure 2.1 and 2.2. The coefficient of **Months** is negative, indicating that with the same level of the other variables, the silica level in tree leaves decreased as time went by. This result is also consistent with the overall trends in figures in part 3.2.

In order to check the fitted model, we make the marginal model plots and residual plots in the appendix, which shows the model fit the data well.

```
##                 (Intercept)        factor(Species)QURU
##                    0.179747                  -1.292069
##    factor(Treatment)Heated                     Months
##                    0.067593                  -0.000713
##              factor(Month)7             factor(Month)8
##                   -0.052270                   0.243141
## factor(Species)QURU:Months
##                    0.001259
```

So the linear regression model is:

$log(X.wt.Bsi) = 0.179747 - 1.292069 \times QURU + 0.067593 \times Heated - 0.000713 \times Months - 0.052270 \times Month_7 + 0.243141 \times Month_8 + 0.001259 \times QURU \times Months$

QURU equals to 1 when samples are from QURU tree leaves; **QURU** equals to 0 when samples are from ACRU tree leaves.

Heated equals to 1 when samples are from heated group; **Heated** equals to 0 when samples are from control group.

If samples are collected in June, then **factor(Month)7** and **factor(Month)8** are both equal to 0; If samples are collected in July, then **factor(Month)7** equal to 1 but **factor(Month)8** equals to 0; If samples are collected in August, then **factor(Month)8** equal to 1 but **factor(Month)7** equals to 0.

QURU*Months equals to the value of **Months** when samples are from QURU tree leaves(because **QURU** equals to 1 in this case). However, **QURU*Months** equals to 0 when samples are from ACRU tree leaves(because **QURU** equals to 0 in this case).

The interaction indicates that the effect of **Months** on log **X.wt.Bsi** is different for different species of trees. The coefficient on the interaction term is 0.0012587, which represents the difference in the slope of the lines predicting log **X.wt.Bsi** on **Months**, comparing QURU and ACRU species.

An equivalent way to understand the interaction term is to look at the separate linear regressions for different species:

QURU tree leaves (QURU = 1) : $log(X.wt.Bsi) = 0.179747 - 1.292069 \times 1 + 0.067593 \times Heated - 0.000713 \times Months - 0.052270 \times Month_7 + 0.243141 \times Month_8 + 0.001259 \times 1 \times Months = -1.112322 + 0.067593 \times Heated + 0.000546 \times Months - 0.052270 \times Month_7 + 0.243141 \times Month_8$

ACRU tree leaves (QURU = 0) : $log(X.wt.Bsi) = 0.179747 - 1.292069 \times 0 + 0.067593 \times Heated - 0.000713 \times Months - 0.052270 \times Month_7 + 0.243141 \times Month_8 + 0.001259 \times 0 \times Months = 0.179747 + 0.067593 \times Heated - 0.000713 \times Months - 0.052270 \times Month_7 + 0.243141 \times Month_8$

3. Model 2 : Soil Layers

3.1 Introduction

The soil dataset mainly includes the following variables:

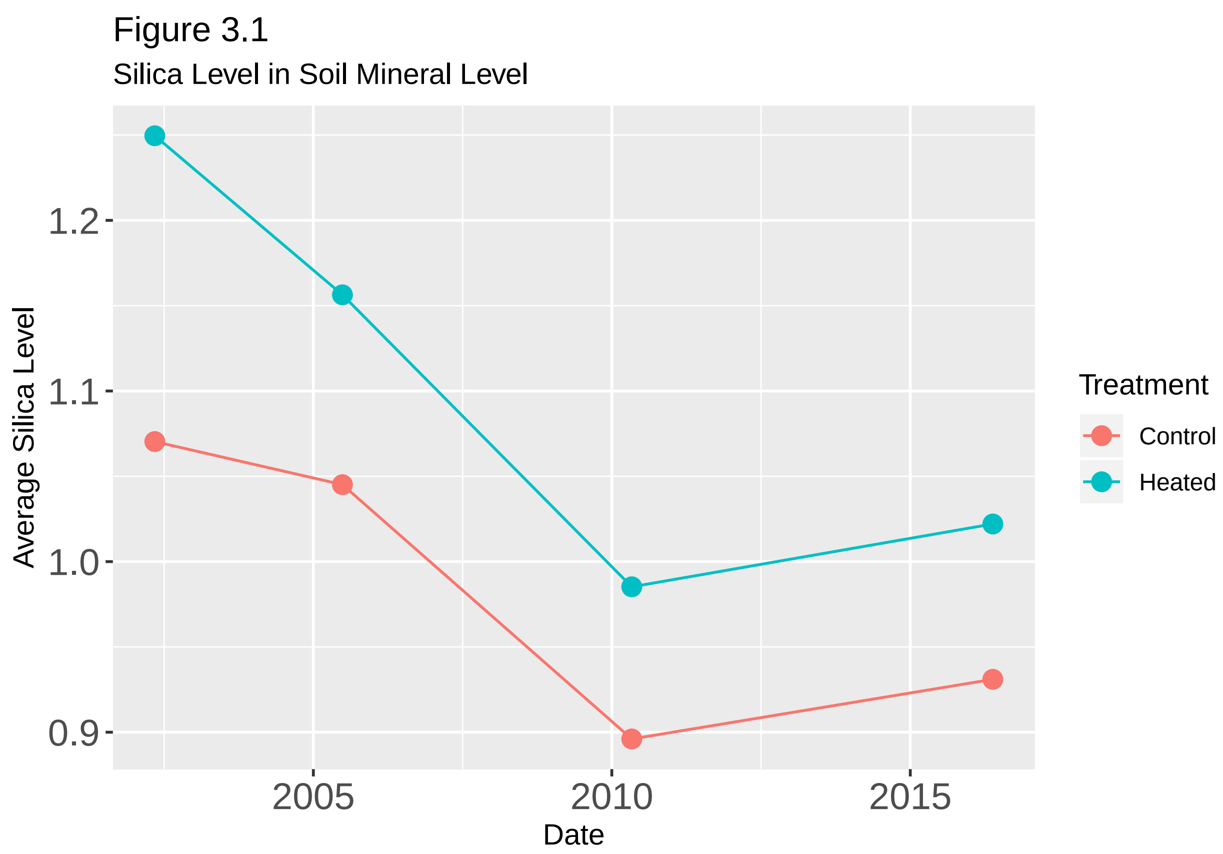
- Date: The date when sample collected
- Layer: Layer of soil core (organic or mineral)
- Treatment: Experimental treatment (Heated or Control)
- X.wt.Bsi: Percent biogenic silica in leaves by dry weight (0-100), indicating the silica level

3.2 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.

Figure 3.1 & 3.2: Compare silica levels under different treatments

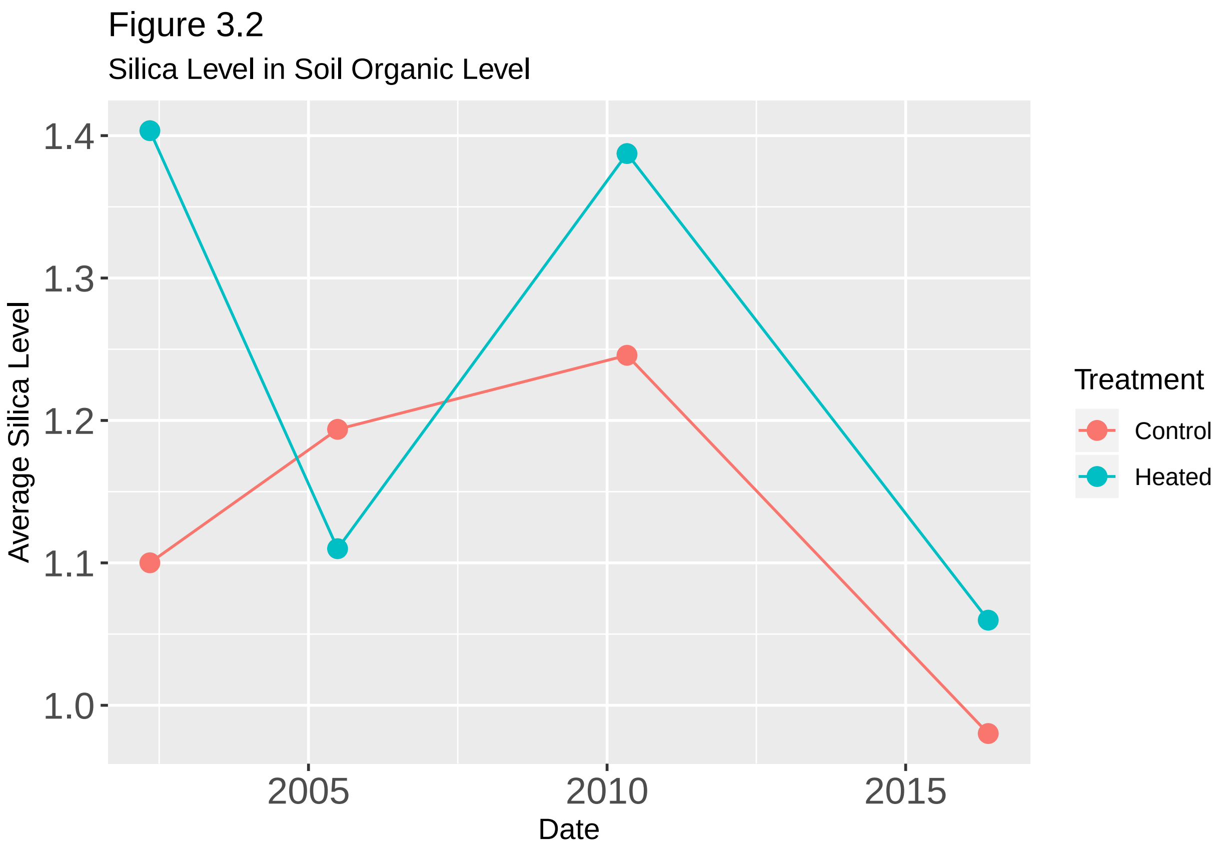
```
ggplot(data = SM) +
  aes(x = Date, y = avg, color = Treatment) +
  geom_point(size = 3) +
  labs(title = "Figure 3.1",
       subtitle ="Silica Level in Soil Mineral Level",
       x = "Date",
       y = "Average Silica Level")+
  theme(axis.text.x = element_text(size = 14))+
  theme(axis.text.y = element_text(size = 14))+
  geom_line()
```



3.2.1 Silica Level in Soil Mineral Level

We compare difference of silica level between treatment groups in soil mineral level. Heated group has higher silica level than control group, and overall they have decreaing trend overtime. In 2010, both groups have low silica level, so it might be affected by factor affecting both groups equally such as weather.

```
ggplot(data = SM2) +
  aes(x = Date, y = avg, color = Treatment) +
  geom_point(size = 3) +
  labs(title = "Figure 3.2",
       subtitle ="Silica Level in Soil Organic Level",
       x = "Date",
       y = "Average Silica Level")+
  theme(axis.text.x = element_text(size = 14))+
  theme(axis.text.y = element_text(size = 14))+
  geom_line()
```



3.2.2 Silica Level in Soil Organic Level

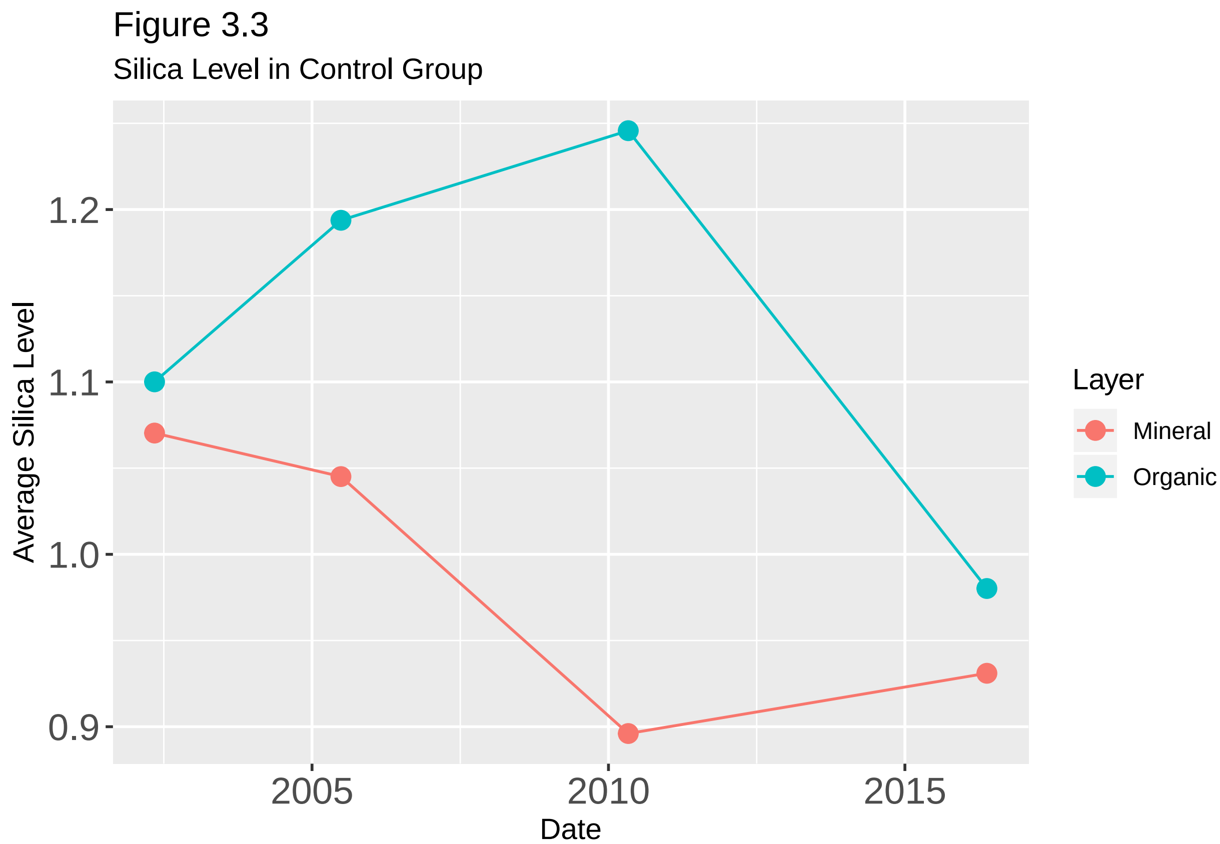
Now we compare silica level in soil organic level. In this case, heated group is not higher than control group all the time any more. It has drastic fluctuations. Also, control group is not decreasing along the time as well. So this case may need further analysis.

3.2.3 Summary

In Mineral level, we may conclude that heat is affecting the silica level but in organic level we cannot make same conclusion. Also, in mineral level, we see decreasing trend while in organic level we do not.

Figure 3.3 & 3.4: Compare silica level in different soil layers

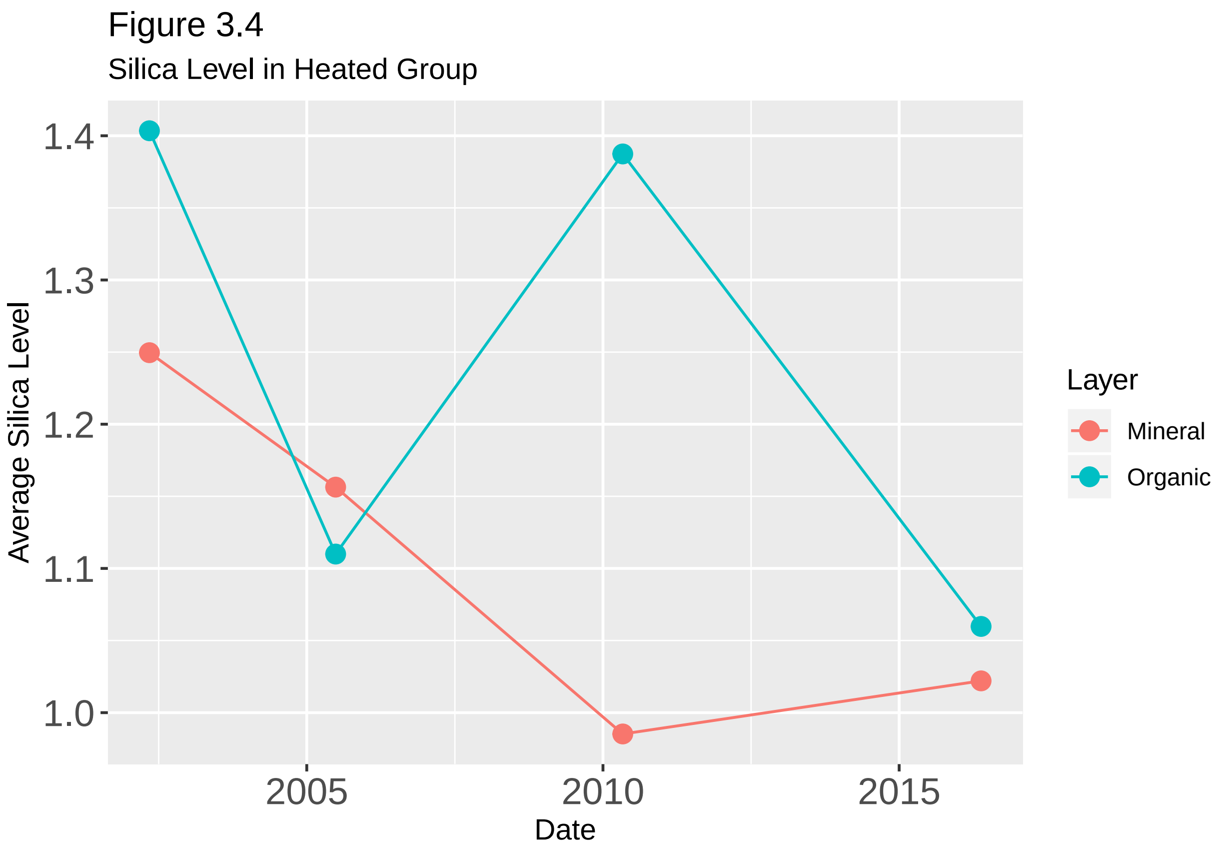
```
ggplot(data = SC) +
  aes(x = Date, y = avg, color = Layer) +
  geom_point(size = 3) +
  labs(title = "Figure 3.3",
       subtitle ="Silica Level in Control Group",
       x = "Date",
       y = "Average Silica Level")+
  theme(axis.text.x = element_text(size = 14))+
  theme(axis.text.y = element_text(size = 14))+
  geom_line()
```



3.2.4 Silica Level in Control Group

Under control treatment, organic level has silica increasing and then decreasing where in mineral level, silica is basically deceasing overtime. overall, organic has higher cilica level than mineral level does, since the organic line is above the mineral line.

```
ggplot(data = SH) +
  aes(x = Date, y = avg, color = Layer) +
  geom_point(size = 3) +
  labs(title = "Figure 3.4",
       subtitle ="Silica Level in Heated Group",
       x = "Date",
       y = "Average Silica Level")+
  theme(axis.text.x = element_text(size = 14))+
  theme(axis.text.y = element_text(size = 14))+
  geom_line()
```



3.2.5 Silica Level in Heated Group

Under heated treatment, mineral level still has decreasing silica level but organic level starts to fluctuate. Mostly, organic level still has relatively higher silica level except in 2005.

3.2.6 Summary

We may conclude that organic level has higher silica level, which means layer of soil does affect the silica level. However, wo cannot conclude any trend overtime.

3.3 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.

```
soil_bsi <- read.csv("soil_bsi.csv")

soil_bsi$Date<-mdy(soil_bsi$Date)
soil_data<-na.omit(soil_bsi[,c("Date","Layer","Treatment","X.wt.Bsi")])
soil_data<-soil_data
standaeddate<-c(ymd(020507))
```

```r
soil_data1<-soil_data %>%
  mutate(deployedtime=difftime(Date,standaeddate,units = "days"))
soil_data1$deployedtime<-as.numeric(soil_data1$deployedtime)
colnames(soil_data1)[4]<-"Silicalevel"
soil_data1$Layer<-factor(soil_data1$Layer)
soil_data1$Treatment<-factor(soil_data1$Treatment)
soil_data2<-soil_data1%>%
  mutate(season=month(Date))

soil_model<-lm(Silicalevel~factor(Layer)+factor(Treatment)+
                 deployedtime+factor(season),data = soil_data2)
```

3.4 Discussion

```r
summary(soil_model)
```

```
##
## 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 R-squared is 0.162, which means that 16.2% of the variations in dependent variable can be explained by independent variables in the model.

Besides, p-value of F-statistic is very small, indicating that the linear regression model provides a better fit to the data than a model that contains no independent variables.

The p-value for each independent variable tests the null hypothesis that the variable has no correlation with the dependent variable. In the regression result, p-values of most of coefficients in the model are smaller than the usual significance level of 0.05, so we should reject the null hypothesis, which means those coefficients are statistically significant and we should not remove them.

The intercept means that when the mineral layer in May and in controlled group with 0 deployed tike, the silica level is 1.09e. The coefficient of Layer Organic is positive, so it tells us that, with other variables stay the same, the silica level in organic level is higher than mineral level. When we look at the coefficient of treatment, we can know that the the silica level in heated group is higher than controlled group with other variables staying the same. And we can justify our conclusion by figure 3.1&figure 3.2. We can find that the green line(heated) is on the top of red line(controlled). From the model, we can know that the deployed time has negative effects on the silical level with other variables staying the same, which is consistent with the figure 3.1&3.2. What's more, it seems like that silical level in May is higher than the silical level in June with other variables staying the same.

In order to check the fitted model, we make the marginal model plots and residual plots in the appendix, which shows the model fit the data well.

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

The regression model is $silicallevel = 1.09 + 0.14 \times Layer + 0.1139 \times treatment - 0.000040 \times deployedtime - 0.045594 \times season$

4. Model 3: Silica Level VS Decaying Rate

4.1 Introduction

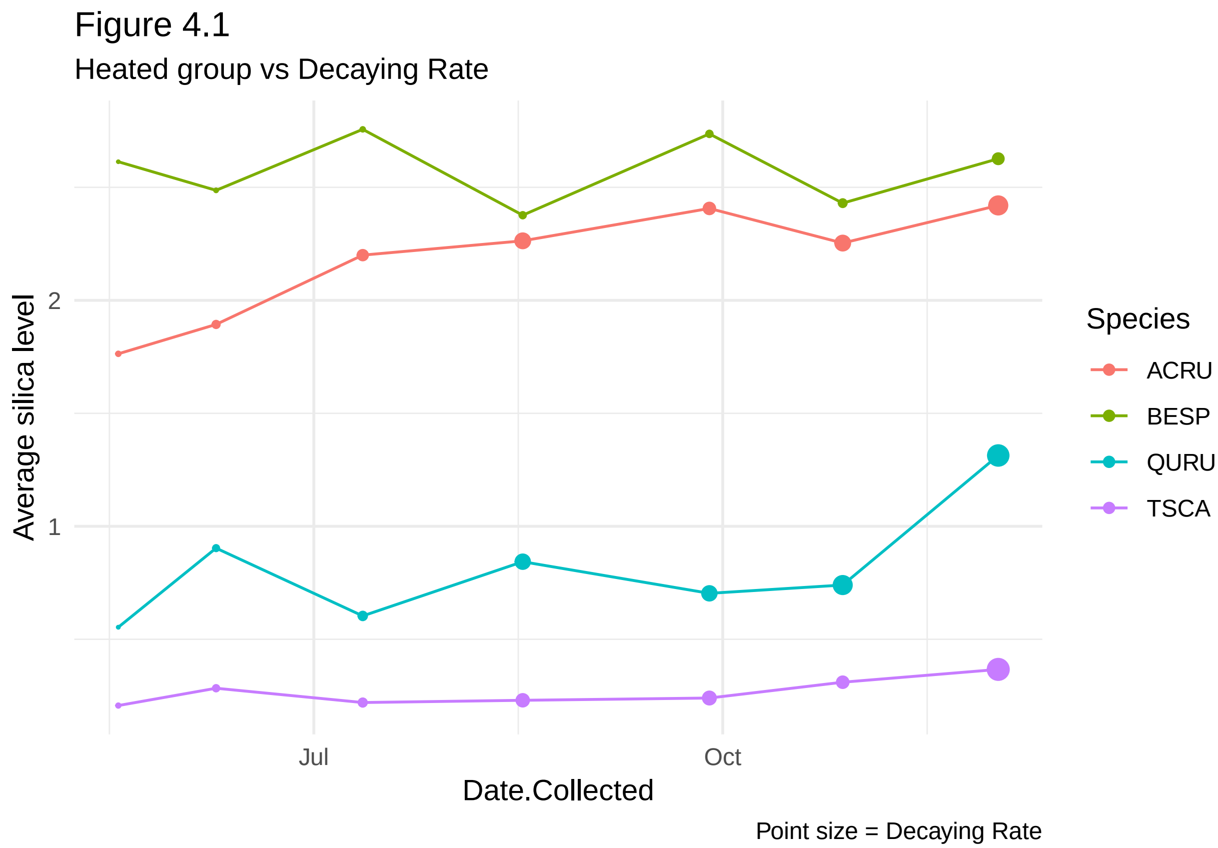
This dataset mainly includes these variables -Species: The species of the leaves that we collected -Treantment: Heated group and Controlled group -Time.Deployed: The length of time from the deployed date to collected date -Mass.Loss: The weight of the leaves caused by decomposition -X.wt.Bsi: The silica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

4.2 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 the following two plots.

Figure 4.1 & 4.2: Compare silica level and dacaying rate under different treaments

```
datalb_heated<-datalb%>%
  filter(Treatment=="Warming")%>%
  group_by(Species,Date.Collected)%>%
  summarise(mean=mean(X.wt.Bsi),Mass.Loss=mean(Mass.Loss))
ggplot(data = datalb_heated) +
  aes(x = Date.Collected, y = mean, color = Species) +
  geom_point(size = datalb_heated$Mass.Loss)+
  geom_line() +
  labs(title = "Figure 4.1",
       subtitle = "Heated group vs Decaying Rate",
       caption = "Point size = Decaying Rate",
    y = 'Average silica level') +
  theme_minimal()
```



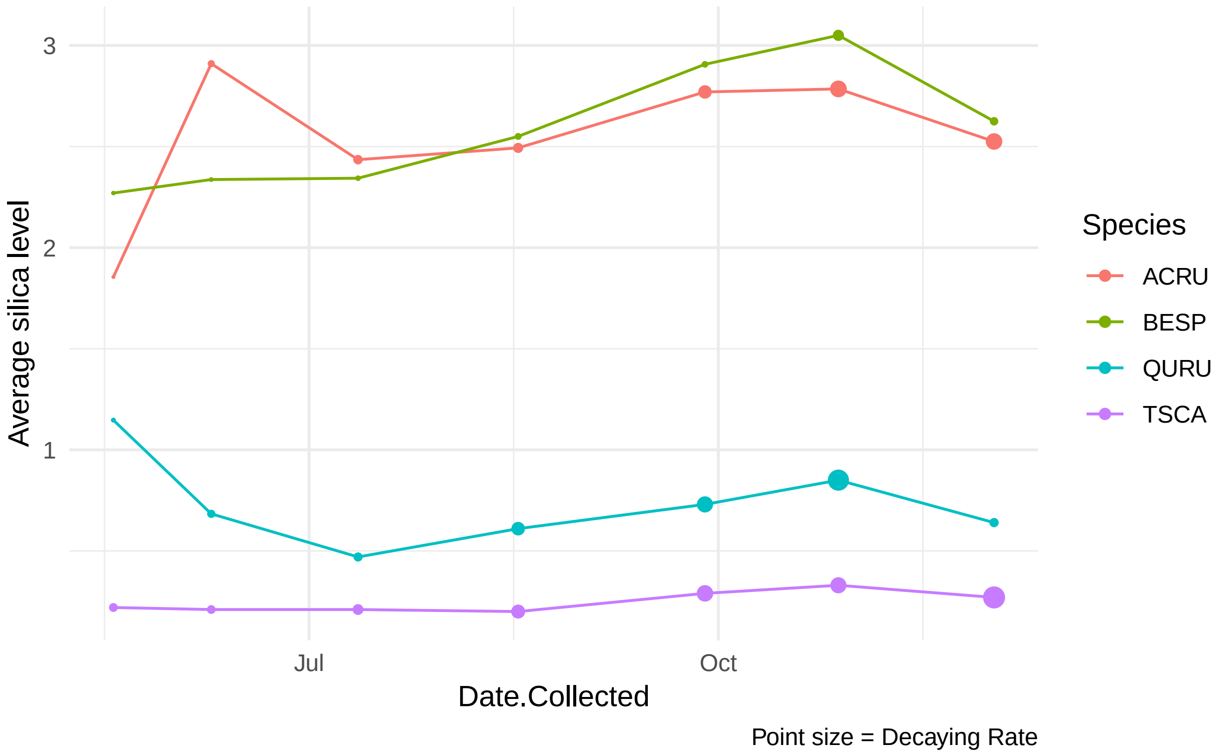
4.2.1 Heated group vs Decaying Rate

In this case, we can see that four different species have different silica levels: BESP > ACRU > QURU > TSCA. Also, decaying rate is denoted by point size. It is increasing as time goes. BESP and TCSA have relatively flat lines so that we do not have any decreasing or increasing trend overtime. However, ACRU and QURU have slightly upward lines, so they may have increasing trend. Most importantly, in all lines the decaying rate is incresing, but some of the lines go up, some do not chage. So we cannot conclude if there is effect to silica level by decaying rate.

```
datalb_normal<-datalb%>%
  filter(Treatment=="Control")%>%
  group_by(Species,Date.Collected)%>%
  summarise(mean=mean(X.wt.Bsi),Mass.Loss=mean(Mass.Loss))
ggplot(data = datalb_normal) +
  aes(x = Date.Collected, y = mean, color = Species) +
  geom_point(size = datalb_normal$Mass.Loss)+
  geom_line() +
  labs(title = "Figure 4.2",
         subtitle = "Controlled group vs Decaying Rate",
       caption = "Point size = Decaying Rate",
       y = 'Average silica level') +
  theme_minimal()
```

Figure 4.2

Controlled group vs Decaying Rate



4.2.2 Controlled group vs Decaying Rate

In control group, ACRU and BESP are not clearly separated any more. So we do not have clear difference among species. The other results are similar as in heated group.

4.2.3 Summary

By comparingh these two plots, we may conclude that decaying rate do not have severe effect on silica level. Also species do not have clear effect on silica level, at least not in ACRU and BESP.

4.3 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.

```
datalb$Time.Deployed<-as.numeric(datalb$Time.Deployed)
datalb1<-datalb%>%
  mutate(stdecaying=scale(Mass.Loss,center = TRUE,scale = TRUE),
         stBsi=scale(X.wt.Bsi,center = TRUE,scale = TRUE))
model_litterbags<-lm(data = datalb1,
                     X.wt.Bsi~factor(Species)+factor(Treatment)+
                       Time.Deployed+Mass.Loss)
summary(model_litterbags)
```

```
##
## 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
```

4.4 Discussion

From the regression result, we can see that R-squared is 0.8228, which means that 82 ?28% of the variations in dependent variable can be explained by independent variables in the model.

Besides, p-value of F-statistic is very small, indicating that the linear regression model provides a better fit to the data than a model that contains no independent variables.

The p-value for each independent variable tests the null hypothesis that the variable has no correlation with the dependent variable. In the regression result, p-values of most of coefficients in the model are smaller than the usual significance level of 0.05, so we should reject the null hypothesis, which means those coefficients are statistically significant and we should not remove them.

The intercept means that, the speciese of ACRU in controlled group that have 0 deployed time and mass loss will have2.2393 silical level. The coefficent of species gives us the information that sillical level differes from species. The BESP has the highest silical level compared with other species with other variables staying the same.And this is consistent with our figure that the line of BESP is on the top of other lines no matter what treatment is. The coefficient of treatment tells us that, with other variables staying the same, the silical level in heated grouop may be lower than controlled group. What's more, the deployed time can have positive effects on the silical level with other variables staying the same which we can find from the EDA that, with time increasing, the silical level tends ti increase. Finally, the Mass.Loss of litter leaves may cause the increase of the silical level in leaves with other variables staying the same.

In order to check the fitted model, we make the marginal model plots and residual plots in the appendix, which shows the model fit the data well.

```
##              (Intercept)      factor(Species)BESP      factor(Species)QURU
##                 2.239326                 0.236173                -1.571149
##      factor(Species)TSCA factor(Treatment)Warming            Time.Deployed
##                -2.092565                -0.086963                 0.001180
##                Mass.Loss
```

```
##                    0.021287
```

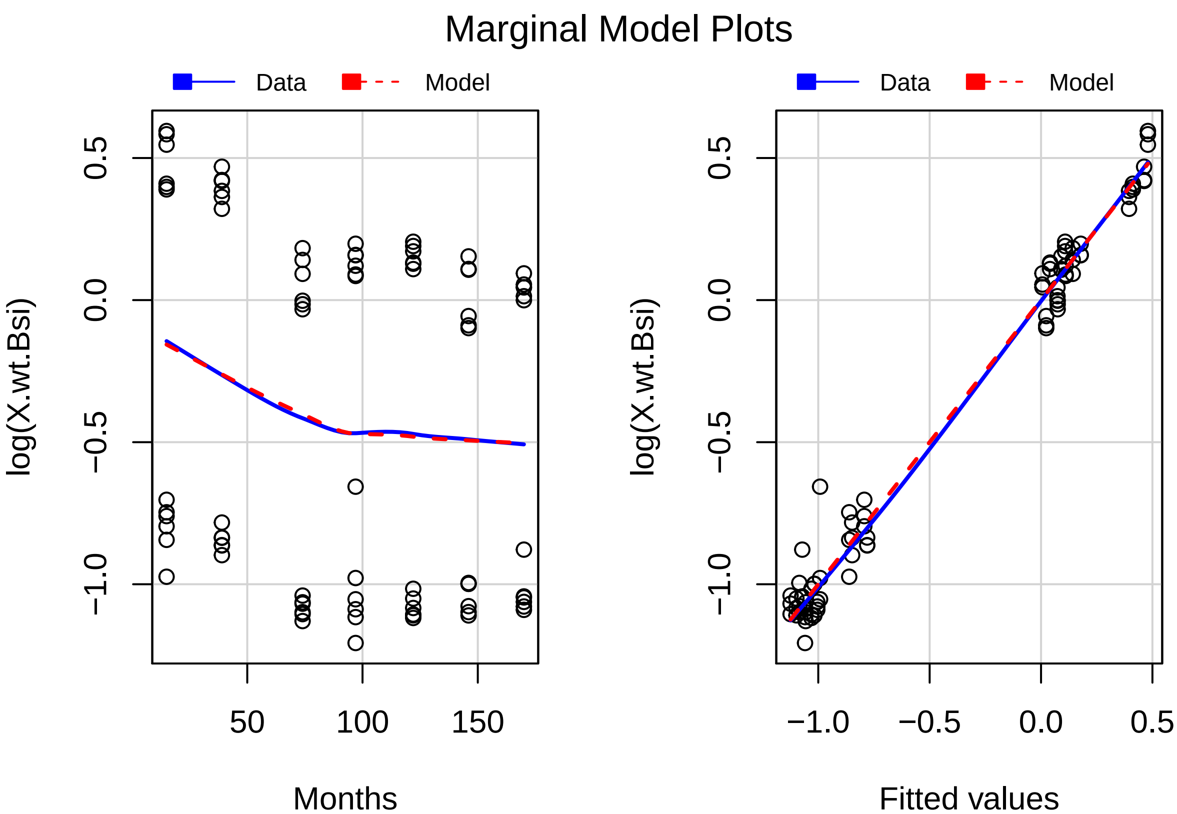
The model is $silicallevel = 2.23926 + 0.2361 \times BESP - 1.57 \times QURU - 2.09 \times TSCA - 0.08 \times warming + 0.00118 \times Time.Deployed + 0.0212 \times Mass.Lpss$

Appendix

Model 1

Marginal Model Plots

Figure 2.5



Marginal model plots for tree species data shows the dependent variable (log(X.wt.Bsi)) on the vertical axes and the horizontal axes denotes numeric independent values (plotted points) in the fitted linear regression model. Margin model plots provide a graphical representation of model fit by showing the marginal relationships between the dependent variable and independent variables. We fitted a regression function for each of the plots using a lowess smooth function for the data (solid blue line) and for the fitted values (dashed red line).

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

Residual Plots

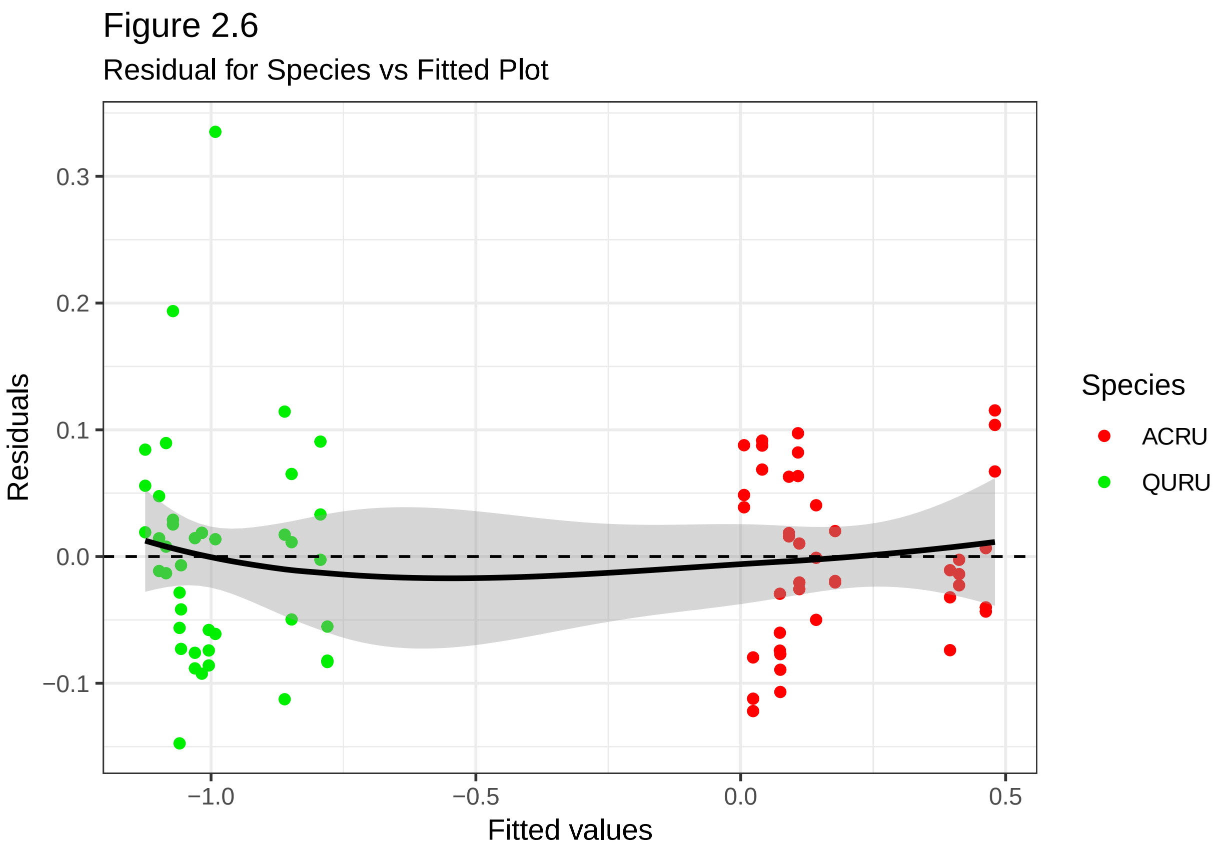


Figure 2.7

Residual for Treatment vs Fitted Plot

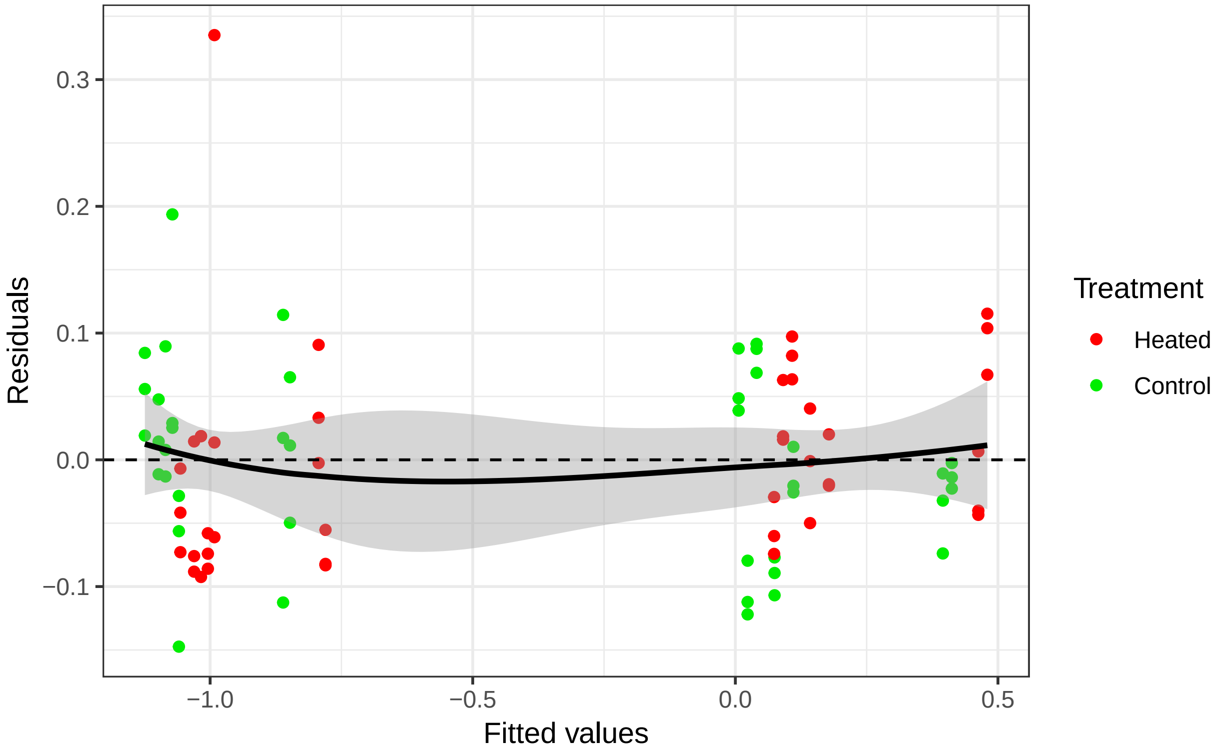
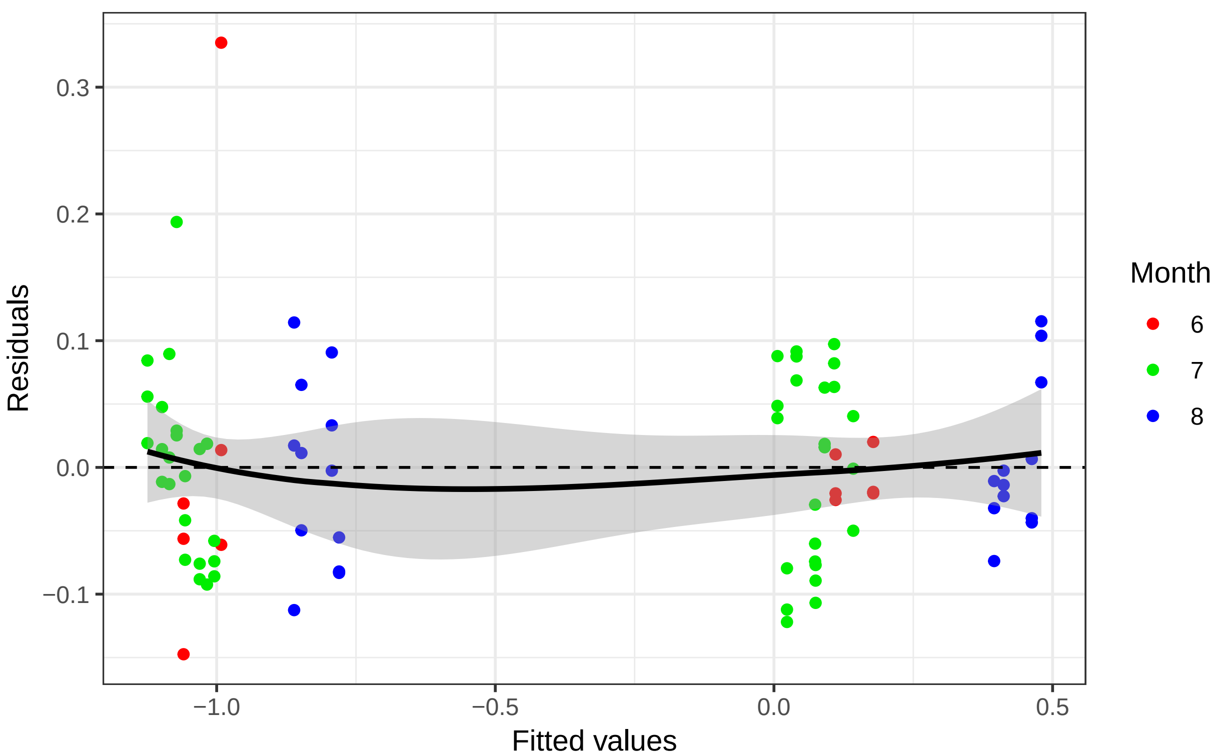


Figure 2.8

Residual for Month vs Fitted Plot



summarise

The difference between the observed value of the dependent variable and the fitted value is called the residual. A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. One of assumptions for residuals in linear regression model is that the mean of residuals is zero. Therefore, if the points in a residual plot are randomly dispersed around the horizontal line at zero(the dashed black line), the linear regression model is appropriate for the data; otherwise, a non-linear model is more appropriate.

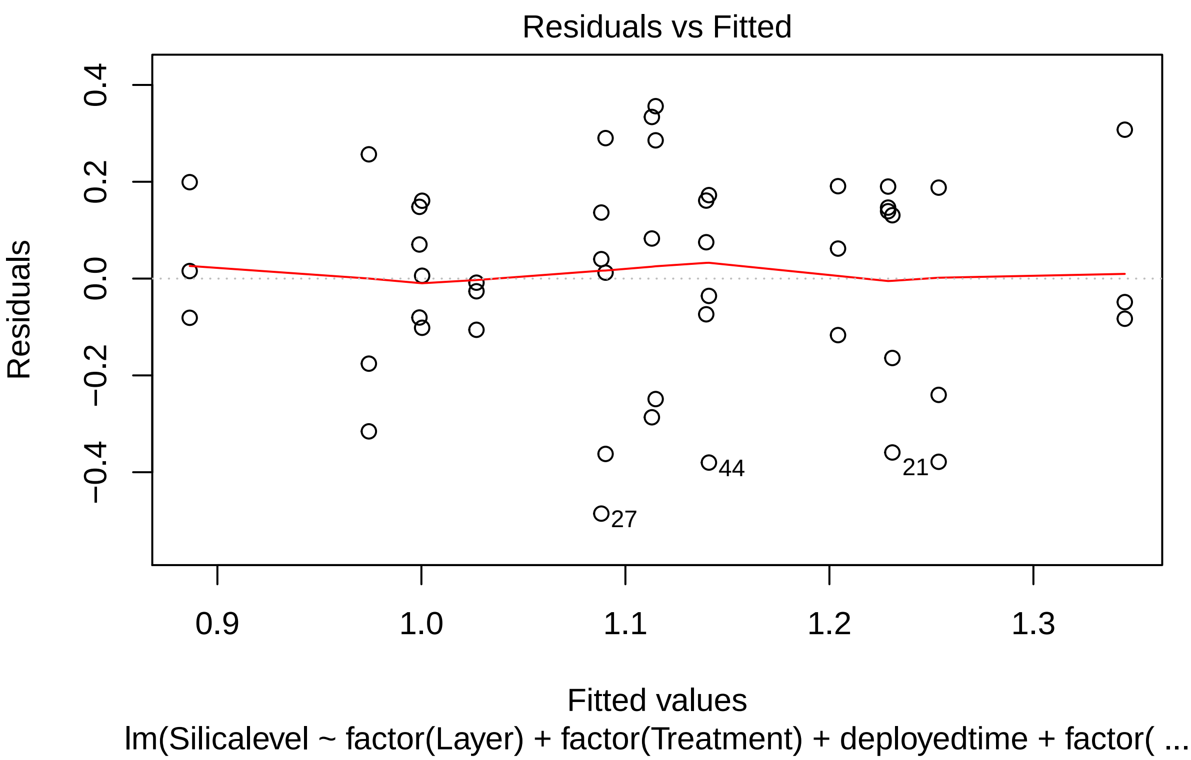
From the residual plots above, we can see that the points are dispersed randomly around the horizontal line at zero. Therefore, we can conclude that the linear model for tree species fits the data well.

Model 2

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

Residual Plots

Figure 3.5



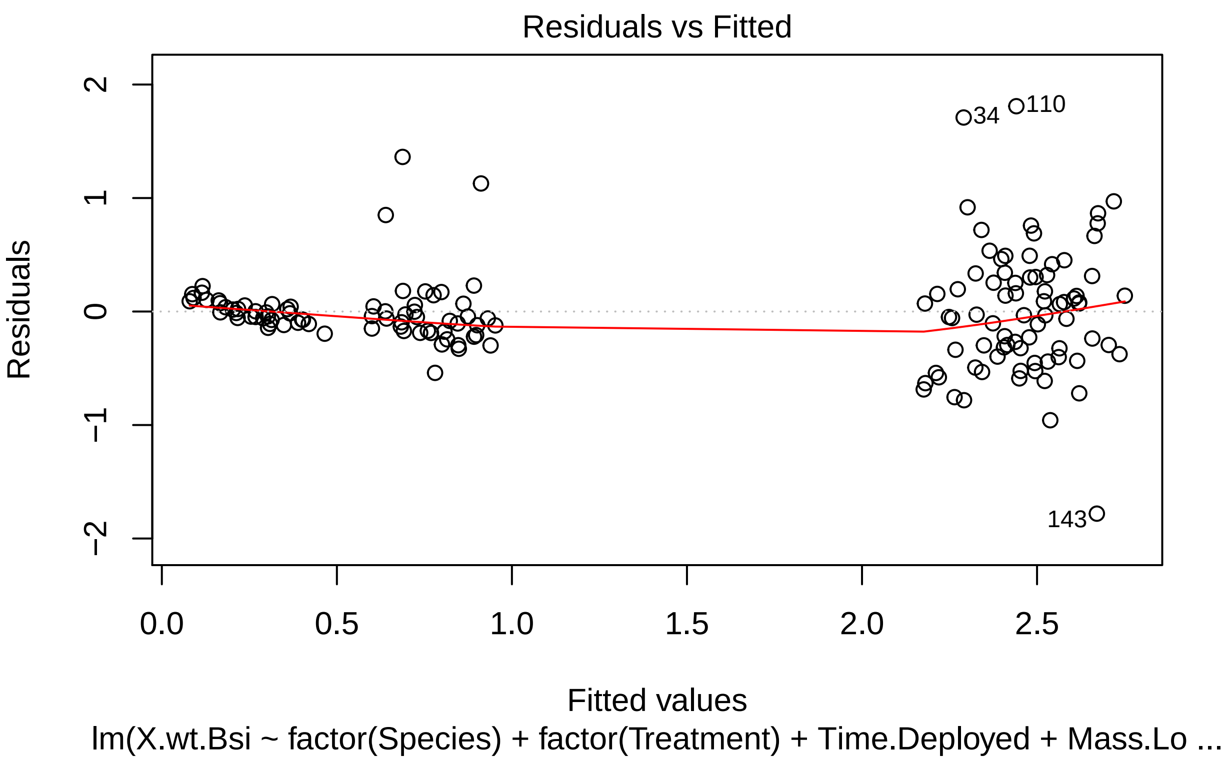
summarise

From the residual model, we can find that the dots are evenly distributed around both sides of x=0. And the red line is almost near the x=0 without too many fluctuations. So we can say the model fits well.

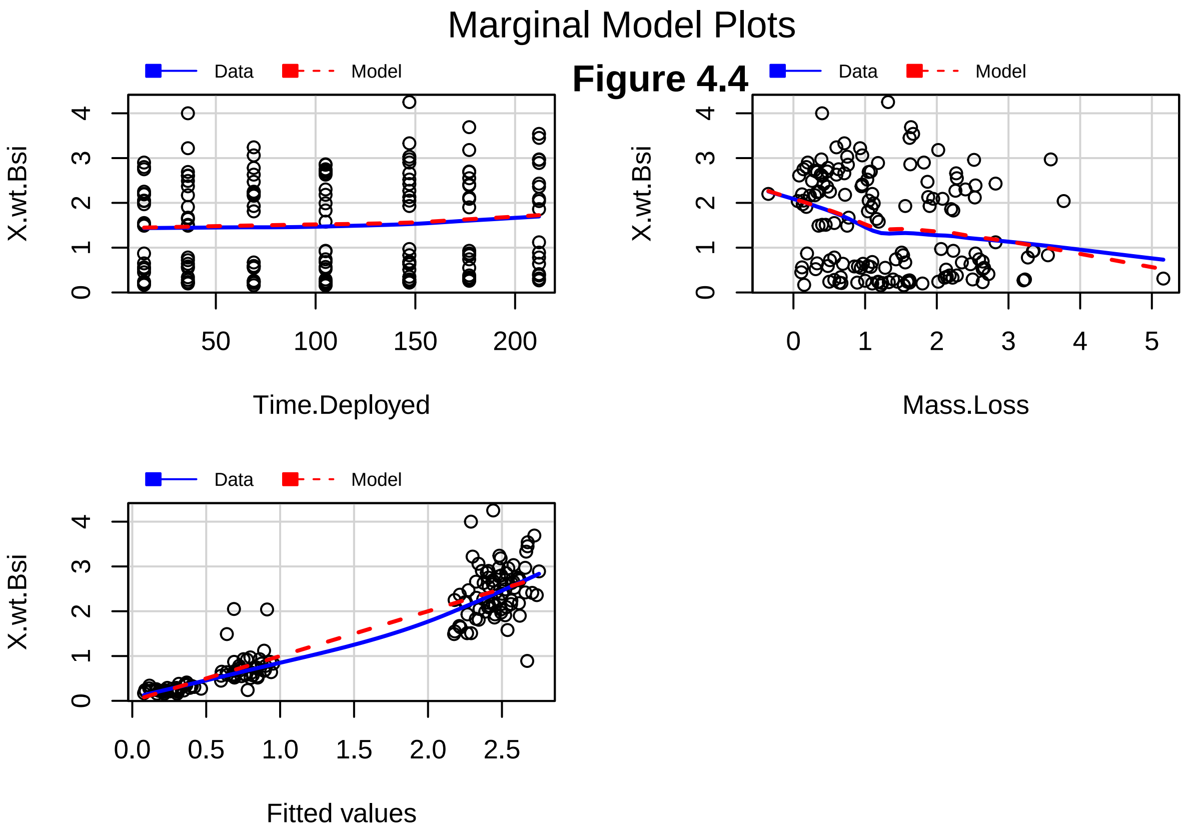
Model 3

residual plot

Figure 4.3



marginal model plots



summarise

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

Reason 1: The residual plot fits well. The dots evenly distributed, and the red line is almost the same with x=0.

Reason 2: The marginal model plots tell us that the model fits good.