Applied Analysis 4

Rohan Hore

August 2024

The data set for this analysis is taken from the paper *Effects of climate warming on photosynthesis in boreal tree species depend on soil moisture, Reich et al, Nature vol. 562, pages 263–267 (2018).*The data are provided in the file borealwarming.csv.

Experiment design The paper studies the effect of warmer temperatures on photosynthesis in trees. Two sites a large distance apart were split into three blocks each, and each block was split into four plots. In each plot, warming cables were laid in the soil with the capability of maintaining a temperature that is 3.4 degrees Celsius warmer than the ambient temperature. Each plot was then randomly assigned to one of two treatments—warmed, where the underground cables and above-ground heating lamps were turned on during the growing season, or ambient, where the electrical cables were not turned on. Within each plot, trees from 11 species were planted and measurements were taken for the photosynthesis activity level (measured in a single leaf). The soil water level was also measured.

variables The principal variables are:

- `site, block, plot id` denote the location of the plant being measured. There are two sites, six blocks (three per site), and 24 plots (four plots per block).
- `warming treatment` takes values warmed or ambient, for the treatment or control group. The randomization is at the level of the plot (i.e., all trees in the same plot are assigned the same treatment).
- `species` takes 11 values. Four species `(abiba, picgl, pinba, pinst)` are needle leaved (coniferous) trees, while seven species `(aceru, acesa, betpa, poptr, quema, queru, rhaca)` are broadleaf (deciduous) trees.
- year is the year when the measurement is taken $(2009, 2010, or 2011)$.
- doy is the day of the year (e.g., doy= 1 means that the measurement was taken on January 1st of that year). Note that two measurements may have the same doy value even if they occur in different years.
- `Asat` measures the photosynthesis activity of the leaf; 2049 measurements.
- `soil_water_VWC` measures the soil water content in the plot. (On any given day, this measurement is the same for all plants in the same plot.)
- You may ignore the other variables present in the data set.

Data Reading

```
boreal_data=read.csv("borealwarming.csv",header=T)
head(boreal_data)
```

```
##   site block warming_treatment plot_id species plant_id phylo year doy      Asat         gs
## 1  cfc     d           ambient      d4   betpa      403 angio 2009 162 15.264729 0.19485140
## 2  cfc     d           ambient      d4   aceru      383 angio 2009 162  7.606667 0.10300000
## 3  cfc     d           ambient      d5   aceru      503 angio 2009 162  9.492947 0.18150878
## 4  cfc     d           ambient      d5   betpa      519 angio 2009 162 14.800000 0.27833333
## 5  cfc     e           ambient      e3   betpa     1248 angio 2009 162 16.333333 0.25000000
## 6  cfc     e           ambient      e3   aceru     1222 angio 2009 162  5.084476 0.05591367
##         ci soil_water_VWC    tleaf      VPG percent_stomatal_limitation    Agmax  Vcmax25
## 1 237.1443      0.2193189 27.53724 2.107054                   0.2776526 21.13212 74.17029
## 2 259.0000      0.2193189 27.10000 1.846402                   0.3793262 12.25550 33.98207
## 3 287.5318      0.2341696 27.33566 2.149909                   0.2254133 12.25550 38.55557
## 4 280.6667      0.2341696 25.07000 1.407993                   0.2996443 21.13212 61.42398
## 5 258.6667      0.2234410 25.39333 1.530326                   0.2270849 21.13212 73.05544
## 6 236.7378      0.2234410 23.16409 1.771963                   0.5851269 12.25550 24.74640
```

1 Possible Questions

Problem 1

Before any trees were planted, a grid of underground electrical cables was laid down at a depth of 15 cm in each plot, with sufficiently small separation that the heating effect could be deemed uniform. The cables in the treated plots were used as heating elements. What was the purpose of the cables in the control plots?
Answer: For comparative purposes, it is essential that the treated plots be as similar as possible to the control plots. Site preparation and tillage affects the soil—its friability, its drainage capacity, and possibly its solar exposure. If the control plots were not prepared and cabled in the same way as the treated plots, it could legitimately be argued that any observed difference in photosynthesis or leaf conductance is not due to treatment but to differences in site preparation. Ideally, plots must be prepared prior to randomization, which implies that cables are in place in all plots. The purpose of cables in the control plots is to ensure comparability and to forestall objections about the lack thereof.

Problem 2

Use the data to reproduce the authors' plots in Fig 1 (plotting photosynthesis level against soil water content for two treatment levels) in a similar format. What, if anything, do these plots tell you about the effect of elevated temperature on the various species? (Note that the species can be grouped into deciduous vs. coniferous trees as given above.)
Answer:

```
 library(ggplot2)
plot_data=boreal_data[,c("warming_treatment","species","soil_water_VWC","Asat")]
ggplot(plot_data,aes(x=soil_water_VWC,y=Asat,color=warming_treatment))+
  geom_point(shape=21,size=0.8)+
  ylim(0,25)+
  geom_smooth(method = "lm", fill = NA)+
  ggtitle(paste("species",species))+
  theme(plot.title = element_text(hjust = 0.5))+
  scale_color_manual(values=c("ambient"="blue","warmed"="red"))+
  facet_wrap(.~species)+
  xlab("soil water content")+ylab("photosyntehsis level")
```

The five broadleaf species betpa, poptr, quema, queru and rhaca have similar slopes and spread while the rest two broadleaf species behave more like needle leaved ones; have lower scatterplots and less steeper slopes.

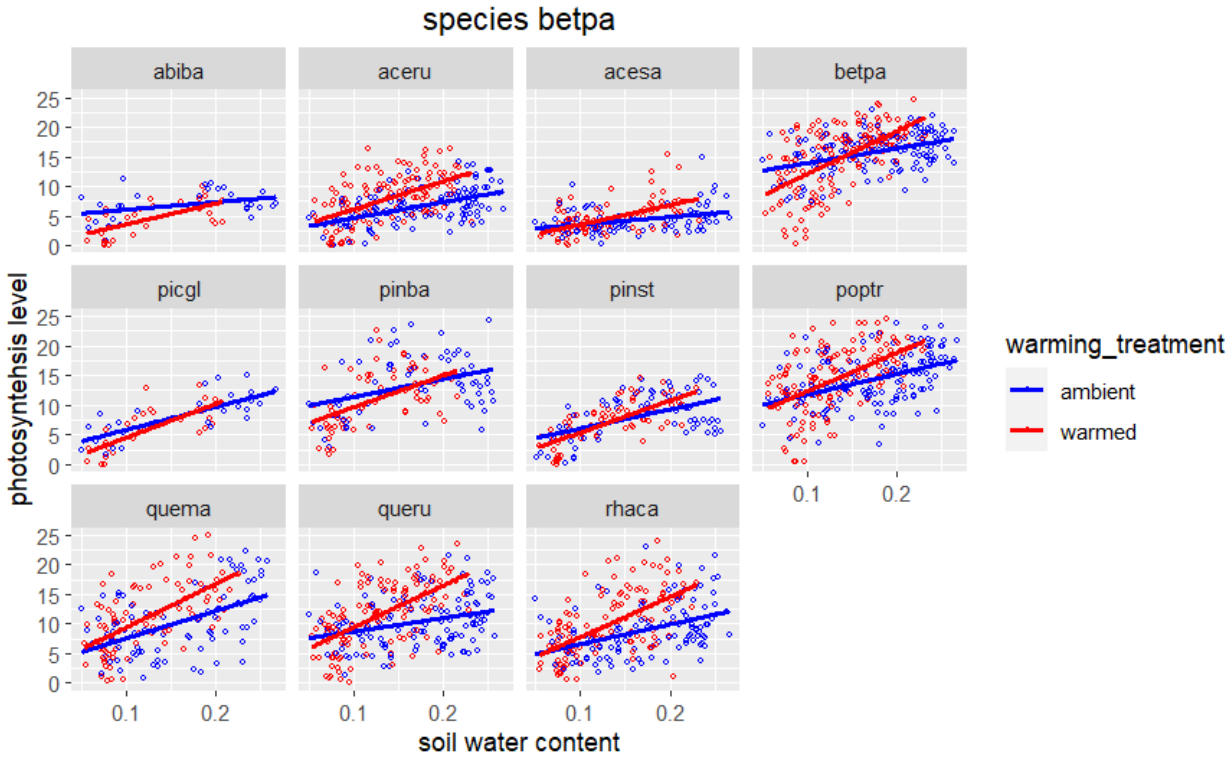

Figure 1: Fig 1 reconstructed

Problem 3

Soil water content is expected to vary from day to day with the most recent weather and from plot to plot depending on the topography, for example, plot elevation, exposure, drainage capacity of the sub-soil, and so on. We want to examine the effect of the treatment assignment (warming vs. ambient) on the soil water content (soil water VWC), using a linear gaussian model, taking appropriate account of such variations. Show any data pre-processing you need to do before deploying the model.
Answer:

```
water_data=unique( boreal_data[-339,c('plot_id','year','doy','soil_water_VWC','site','warming_treatment') ] )
summary(water_data)
water_data=water_data[is.na(water_data$soil_water_VWC)==FALSE,]
```

```
##    plot_id              year          doy        soil_water_VWC        site
##  Length:513         Min.   :2009   Min.   :162.0   Min.   :0.05027   Length:513
##  Class :character   1st Qu.:2009   1st Qu.:177.0   1st Qu.:0.10609   Class :character
##  Mode  :character   Median :2010   Median :210.0   Median :0.14819   Mode  :character
##                     Mean   :2010   Mean   :209.2   Mean   :0.15281
##                     3rd Qu.:2011   3rd Qu.:237.0   3rd Qu.:0.19988
##                     Max.   :2011   Max.   :269.0   Max.   :0.26746
##                                                    NA's   :12
##  warming_treatment
##  Length:513
##  Class :character
##  Mode  :character
##
```

```
##
##
##
```

```
hist(water_data$soil_water_VWC)
```

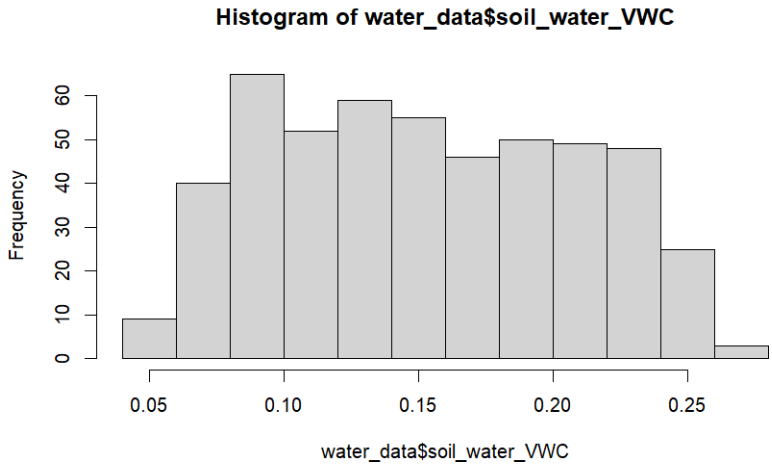

Figure 2: PLot 2

Problem 4

First, run an analysis assuming that the treatment effect is constant over sites and over years. You should fit a suitable linear Gaussian model with soil water content – or some suitable transformation thereof – as response. Explain how you fitted the model, show the parameter estimates with standard errors, and discuss conclusions about the effect of the treatment.
Answer:

```
 mod1=lmer(soil_water_VWC ~ warming_treatment + as.factor(year) + site + (1|plot_id) +
 (1|doy) ,data=water_data,REML=FALSE)
(ss1=summary(mod1))
mod2=lmer(soil_water_VWC ~  as.factor(year) + site + (1|plot_id) +(1|doy) ,data=water_data,REML=FALSE)
anova(mod2,mod1)
```

Linear mixed model on soil water content with fixed effects: treatement, year and site along with random effects of plot id and doy:

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: soil_water_VWC ~ warming_treatment + as.factor(year) + site +
##     (1 | plot_id) + (1 | doy)
##    Data: water_data
##
##      AIC      BIC   logLik deviance df.resid
##  -2579.8  -2546.0   1297.9  -2595.8      493
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.76702 -0.59633 -0.01681  0.60574  2.83367
```

```
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  doy      (Intercept) 0.0016100 0.04012
##  plot_id  (Intercept) 0.0002429 0.01559
##  Residual             0.0001917 0.01385
## Number of obs: 501, groups:  doy, 44; plot_id, 24
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)                0.152017   0.008590  17.696
## warming_treatmentwarmed -0.031568   0.005964  -5.293
## as.factor(year)2010        0.064356   0.009210   6.988
## as.factor(year)2011        0.031490   0.009532   3.304
## sitehwrc                  -0.029932   0.011158  -2.683
##
## Correlation of Fixed Effects:
##             (Intr) wrmng_ a.()2010 a.()2011
## wrmng_trtmn -0.346
## as.fc()2010 -0.161 -0.003
## as.fc()2011 -0.204 -0.003  0.928
## sitehwrc    -0.174  0.002 -0.744   -0.691
```

Is the treatment effect significant?

```
## Data: water_data
## Models:
## mod2: soil_water_VWC ~ as.factor(year) + site + (1 | plot_id) + (1 | doy)
## mod1: soil_water_VWC ~ warming_treatment + as.factor(year) + site + (1 | plot_id) + (1 | doy)
##      npar     AIC     BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod2    7 -2567.9 -2538.4 1291.0  -2581.9
## mod1    8 -2579.8 -2546.1 1297.9  -2595.8 13.834  1  0.0001997 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 5

You should have found a small negative treatment effect in previous question. Next, examine whether the treatment effect is indeed constant across sites, as assumed in the first part of the question. Similarly, examine whether it is constant across years. Explain how you might address these questions, and report your answers.

Answer:

```
 mod3=lmer(soil_water_VWC ~ warming_treatment*site + as.factor(year)+ (1|plot_id) +
           (1|doy) ,data=water_data,REML=FALSE)
summary(mod3)
anova(mod1,mod3)

mod4=lmer(soil_water_VWC ~ warming_treatment*as.factor(year) + site +
         +(1|plot_id) + (1|doy) ,data=water_data,REML=FALSE)
summary(mod4)
anova(mod1,mod4)
```

Model with interaction between treatment and site:

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: soil_water_VWC ~ warming_treatment * site + as.factor(year) +
```

```
##     (1 | plot_id) + (1 | doy)
##    Data: water_data
##
##      AIC      BIC   logLik deviance df.resid
##  -2589.5  -2551.5   1303.7  -2607.5      492
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.47597 -0.58988 -0.02528  0.59722  2.89871
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  doy      (Intercept) 0.0016110 0.04014
##  plot_id  (Intercept) 0.0001780 0.01334
##  Residual             0.0001896 0.01377
## Number of obs: 501, groups:  doy, 44; plot_id, 24
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                      0.163401   0.008660  18.868
## warming_treatmentwarmed         -0.054409   0.007884  -6.902
## sitehwrc                        -0.048676   0.011886  -4.095
## as.factor(year)2010              0.064317   0.009209   6.984
## as.factor(year)2011              0.031500   0.009528   3.306
## warming_treatmentwarmed:sitehwrc 0.037711   0.010559   3.572
##
## Correlation of Fixed Effects:
##             (Intr) wrmng_ sthwrc a.()2010 a.()2011
## wrmng_trtmn -0.454
## sitehwrc    -0.259  0.331
## as.fc()2010 -0.160  0.000 -0.698
## as.fc()2011 -0.202 -0.001 -0.649  0.929
## wrmng_trtm:  0.340 -0.747 -0.442 -0.002   -0.001
```

Is the treatment effect different across site? :

```
## Data: water_data
## Models:
## mod1: soil_water_VWC ~ warming_treatment + as.factor(year) + site + (1 | plot_id) + (1 | doy)
## mod3: soil_water_VWC ~ warming_treatment * site + as.factor(year) + (1 | plot_id) + (1 | doy)
##      npar     AIC     BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod1    8 -2579.8 -2546.1 1297.9  -2595.8
## mod3    9 -2589.5 -2551.5 1303.7  -2607.5 11.707  1  0.0006228 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model with interaction between treatment and year:

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: soil_water_VWC ~ warming_treatment * as.factor(year) + site +
##     +(1 | plot_id) + (1 | doy)
##    Data: water_data
##
##      AIC      BIC   logLik deviance df.resid
##  -2596.4  -2554.3   1308.2  -2616.4      491
```

```
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.92694 -0.59342  0.00538  0.55981  2.77100
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  doy      (Intercept) 0.0016160 0.04020
##  plot_id  (Intercept) 0.0002484 0.01576
##  Residual             0.0001826 0.01351
## Number of obs: 501, groups:  doy, 44; plot_id, 24
##
## Fixed effects:
##                                              Estimate Std. Error t value
## (Intercept)                                  0.153894   0.008636  17.820
## warming_treatmentwarmed                     -0.035704   0.006143  -5.812
## as.factor(year)2010                          0.063280   0.009336   6.778
## as.factor(year)2011                          0.025601   0.009617   2.662
## sitehwrc                                    -0.029777   0.011201  -2.658
## warming_treatmentwarmed:as.factor(year)2010  0.002385   0.003146   0.758
## warming_treatmentwarmed:as.factor(year)2011  0.012308   0.002771   4.442
##
## Correlation of Fixed Effects:
##             (Intr) wrmng_ a.()2010 a.()2011 sthwrc w_:.()2010
## wrmng_trtmn -0.354
## as.fc()2010 -0.165  0.026
## as.fc()2011 -0.207  0.025  0.919
## sitehwrc    -0.174  0.002 -0.735   -0.687
## wr_:.()2010  0.058 -0.175 -0.163   -0.059    0.002
## wr_:.()2011  0.068 -0.197 -0.068   -0.142    0.001  0.416
```

Is the treatment effect different across year? :

```
## Data: water_data
## Models:
## mod1: soil_water_VWC ~ warming_treatment + as.factor(year) + site + (1 | plot_id) + (1 | doy)
## mod4: soil_water_VWC ~ warming_treatment * as.factor(year) + site + +(1 | plot_id) + (1 | doy)
##      npar     AIC     BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod1    8 -2579.8 -2546.1 1297.9  -2595.8
## mod4   10 -2596.4 -2554.3 1308.2  -2616.4 20.645  2  3.288e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 6

Now, we will use the data from only the queru species (northern red oak). The photosynthesis level of a tree may depend on both the soil water content and the treatment (in addition to variation from day-to-day weather and effects of other factors).Check for any data pre-processing you need to do before making any inference.

Build a linear Gaussian model with photosynthesis (Asat) – or some suitable transformation thereof – as the response. Your analysis should accommodate block, plot, and/or temporal effects as needed. Explain how you fitted the model, and show the parameter estimates with standard errors. Give a brief summary of the conclusions reached on the basis of the models fitted, and report a conclusion testing the following hypothesis with a likelihood ratio test: the effect (slope) of soil water content on photosynthesis level is the

same for the two treatment groups.
Answer:

```
queru_data=boreal_data[boreal_data$species=="queru",]
hist(queru_data$Asat)
```

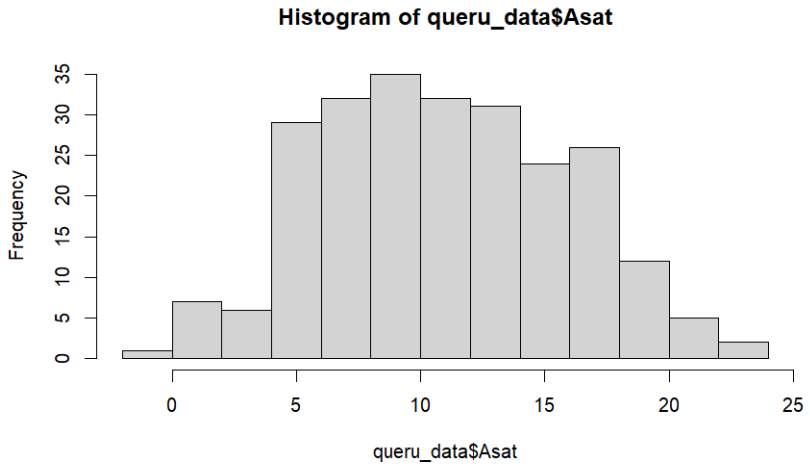

Figure 3: Plot 3

```
mod5=lmer(Asat ~ soil_water_VWC + warming_treatment + as.factor(year) +
            (1 | plot_id) + (1 |doy) ,data=queru_data,REML=FALSE)
summary(mod5)
mod6=lmer(Asat ~ soil_water_VWC * warming_treatment + as.factor(year) +
            (1 | plot_id) + (1 |doy) ,data=queru_data,REML=FALSE)
(ss6=summary(mod6))
anova(mod5,mod6)
mod7=lmer(Asat ~ soil_water_VWC*warming_treatment +site +soil_water_VWC:site + as.factor(year) +
            (1 | plot_id) + (1 |doy) ,data=queru_data,REML=FALSE)
anova(mod6,mod7)
```

Linear mixed model with random effects of plot id and doy, fixed effect of year, soil water content and warming treatment:

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Asat ~ soil_water_VWC + warming_treatment + as.factor(year) +
##     (1 | plot_id) + (1 | doy)
##    Data: queru_data
##
##      AIC      BIC   logLik deviance df.resid
##   1327.8   1355.7   -655.9   1311.8      234
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.35644 -0.68037 -0.05321  0.60493  2.83454
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  doy      (Intercept) 6.843    2.616
```

```
##  plot_id  (Intercept) 2.744    1.656
##  Residual             9.315    3.052
## Number of obs: 242, groups:  doy, 30; plot_id, 24
##
## Fixed effects:
##                         Estimate Std. Error t value
## (Intercept)               5.6948     1.9800   2.876
## soil_water_VWC           19.4884     9.7535   1.998
## warming_treatmentwarmed   1.8216     0.8673   2.100
## as.factor(year)2010       3.6636     1.4250   2.571
## as.factor(year)2011       0.5298     1.3398   0.395
##
## Correlation of Fixed Effects:
##             (Intr) s__VWC wrmng_ a.()2010
## sol_wtr_VWC -0.778
## wrmng_trtmn -0.515  0.429
## as.fc()2010 -0.285 -0.218 -0.084
## as.fc()2011 -0.464 -0.023 -0.007  0.749
```

Linear mixed model with random effects of plot id and doy, fixed effect of year and interaction between soil water content and warming treatment:

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Asat ~ soil_water_VWC * warming_treatment + as.factor(year) +
##     (1 | plot_id) + (1 | doy)
##    Data: queru_data
##
##      AIC      BIC   logLik deviance df.resid
##   1313.1   1344.5   -647.6   1295.1      233
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.72394 -0.60860 -0.07666  0.68757  2.77581
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  doy      (Intercept) 6.563    2.562
##  plot_id  (Intercept) 1.974    1.405
##  Residual             8.848    2.975
## Number of obs: 242, groups:  doy, 30; plot_id, 24
##
## Fixed effects:
##                                       Estimate Std. Error t value
## (Intercept)                             8.3153     1.9817   4.196
## soil_water_VWC                          4.4683     9.8740   0.453
## warming_treatmentwarmed                -4.0879     1.5977  -2.559
## as.factor(year)2010                     3.5684     1.3919   2.564
## as.factor(year)2011                     0.3929     1.3115   0.300
## soil_water_VWC:warming_treatmentwarmed 40.8006     9.6892   4.211
##
## Correlation of Fixed Effects:
##             (Intr) sl__VWC wrmng_ a.()2010 a.()2011
## sol_wtr_VWC -0.797
## wrmng_trtmn -0.500  0.505
## as.fc()2010 -0.295 -0.192  -0.026
```

```
## as.fc()2011 -0.463 -0.012   0.019  0.751
## sl_wt_VWC:_  0.297 -0.339  -0.874 -0.018   -0.026
```

Is the effect of soil water content different for different treatment? :

```
## Data: queru_data
## Models:
## mod5: Asat ~ soil_water_VWC + warming_treatment + as.factor(year) + (1 | plot_id) + (1 | doy)
## mod6: Asat ~ soil_water_VWC * warming_treatment + as.factor(year) + (1 | plot_id) + (1 | doy)
##      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mod5    8 1327.8 1355.7 -655.91   1311.8
## mod6    9 1313.1 1344.5 -647.57   1295.1 16.691  1  4.398e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Is the effect of soil water content different for site? :

```
## Data: queru_data
## Models:
## mod6: Asat ~ soil_water_VWC * warming_treatment + as.factor(year) + (1 | plot_id) + (1 | doy)
## mod7: Asat ~ soil_water_VWC * warming_treatment + site + soil_water_VWC:site + as.factor(year) +
(1 | plot_id) + (1 | doy)
##      npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## mod6    9 1313.1 1344.5 -647.57   1295.1
## mod7   11 1316.7 1355.1 -647.37   1294.7 0.386  2     0.8245
```

Problem 7

Imagine that in the year 2010 you measured Asat on a red oak leaf in a new plot at site hwrc that was assigned to the warming treatment. Under the model you estimated what would be the mean and SD of that measurement? Show code for your calculations, and explain how you arrived at your conclusions. How would you provide SE for these quantities?
Answer: We first compute the mean and variance of the water content based on the estimated model, taking into account the different variance components. Then we compute the prediction using the new water content mean and the variance plugging in the water content variance:

```
coef1 = ss1$coef
coef2 = ss6$coef
soil_water_VWC_new_mean = coef1[1] + coef1[2] + coef1[3]+coef1[5]
soil_water_VWC_new_var = ss1$varcor$doy[1]+ss1$varcor$plot_id[1]+ss1$sigma^2

pred=coef2[1] + soil_water_VWC_new_mean*(coef2[2]+coef2[6]) + coef2[3] + coef2[4]
pred_SD=sqrt(soil_water_VWC_new_var*(coef2[2]+coef2[6])^2 +
               ss6$varcor$doy[1]+ss6$varcor$plot_id[1]+ss6$sigma^2)
print(c(pred,pred_SD))
```

```
## [1] 14.806766  4.644961
```

Problem 8

Summarize the assumptions made in the models you fitted in previous questions, and discuss the validity and implications of these assumptions.
Answer: The most problematic assumption is the independence between observations on different days. Some days are consecutive and one would expect a strong correlation between measurements on such days - as can be seen from the analysis in the next problem.

Problem 9

Let Y be a Gaussian random vector with mean $X\beta$, covariance matrix $\Sigma = W^{-1}$, and let $R = QY$ be the residual projection in which $Q = I - X(X'WX)^{-1}X'W$. Compute the mean and variance of the residual quadratic form $R'\Delta R$ in which Δ is a given symmetric matrix. Your answers should be expressed in terms of the matrices Q, Σ, Δ.
Answer:

$$E(R'\Delta R) = tr(\Delta Q\Sigma) \quad \text{and } Var(R'\Delta R) = 2tr((\Delta Q\Sigma)^2)$$

Hint: If $\epsilon \sim N(0, \Sigma)$, then

$$E(\epsilon'\Delta\epsilon) = tr(\Delta\Sigma) \quad \text{and} \quad Var(\epsilon'\Delta\epsilon) = 2tr(\Delta\Sigma\Delta\Sigma)$$

Problem 10

Let Δ be the indicator function for consecutive calendar days, and let Y be the soil water content. Compare the observed residual quadratic form with its mean and variance as computed for the fitted model in problem 5 with two interactions. What does this calculation suggest?
Answer:

```
mod8 = lmer(soil_water_VWC ~ warming_treatment*(site+as.factor(year))+
              (1|plot_id)+(1|doy),data=water_data,REML=FALSE)
ss8=summary(mod8)
Z=getME(mod8,"Z")
D=diag(c(rep(ss8$varcor$doy[1],44),c(rep(ss8$varcor$plot_id[1],24))))
Sig=diag(rep(ss8$sigma^2,501))+ Z%*%D%*%t(Z)
W=solve(Sig)
X=getME(mod8,"X")
Wa=solve((t(X)%*%W%*%X))
Q=diag(501)-X%*%Wa%*%t(X)%*%W
R=Q%*%water_data$soil_water_VWC
datf=water_data$doy + (water_data$year - 2009) * 365
Delta <- abs(outer(datf, datf, "-"))==1



ST=t(R)%*%Delta%*%R
A=(Delta%*%Q%*%Sig)
MEAN=unlist(sum(diag(A)))
VAR=unlist(2*sum(diag(A%*%A)))
SD=sqrt(VAR)
print(c(ST[1,1],MEAN,VAR,(ST[1,1]-MEAN)/SD))
```

```
## [1]  5.0688026 -0.3847386  3.3422408  2.9830445
```

Shouldn't the residuals from consecutive days be uncorrelated?