Data Analysis Project - Long Nguyen

Abstract

Plants utilize sunlight to synthesize nutrients from carbon dioxide and water via a process called photosynthesis. Weather conditions such as cold temperature could potentially affect the photosynthesis of plants. In this assignment, Bayesian hierarchical model was employed for studying the effect of cold on the amount carbon uptake by Echinochloa crus-galli - a tropical wild grass. The results show that chilled condition reduces the CO_2 uptake from 5 to 10 µmol/m^2.sec and therefore potentially inhibits the photosynthesis of the grass.

1. Introduction

To survive plants need to ulilize sunlight to synthesize nutrients from carbon dioxide and water. This process is called photosynthesis and it could be affected by weather conditions such as cold temperature. This assignment investigated the effects of cold on photosynthesis of Echinochloa crus-galli - a tropical wild grass. Particularly, the amount CO_2 uptake by two types of Echinochloa crus-galli under chilled and nonchilled conditions was modelled using Bayesian herarchical models.

2. Data

The public CO2 dataset in R was used to study the problem. The dataset has 84 observations of five variables (Plant, Type, Treatment, conc and uptake). This was a results of an experiment on the cold tolerance of the grass species Echinochloa crus-galli (Potvin et. al. 1990). The experiment monitored the CO_2 uptake rate (µmol/m^2.sec) of two Echinochloa crus-galli types (Quebec and Mississippi) at seven levels of ambient CO_2 concentrations (mL/L). Each type consists of three individuals, of which each individual was treated and monitored at two weather conditions (nonchilled and chilled). The dataset was already clean and tidy. It contains no missing value so that it is readily to be exploited.

```r
summary(CO2)
```

```
##      Plant             Type         Treatment    conc       uptake
##  Qn1    : 7   Quebec     :42   nonchilled:42   95  :12   Min.   : 7.70
##  Qn2    : 7   Mississippi:42   chilled   :42   175 :12   1st Qu.:17.90
##  Qn3    : 7                                    250 :12   Median :28.30
##  Qc1    : 7                                    350 :12   Mean   :27.21
##  Qc3    : 7                                    500 :12   3rd Qu.:37.12
##  Qc2    : 7                                    675 :12   Max.   :45.50
##  (Other):42                                    1000:12
```

Figure 1 (left) shows that chilled treatment seems to reduce the carbon uptake rate from the same grass type, especially for Mississippi grasses (green). It also shows that Quebec grasses (red) have higher uptake rate than Mississippi grasses. Thus, type of grass should be included in the model beside treatment. Figure 1 (right) illustrates the similar trends along the gradient of ambient CO_2 concentration. Moreover, as ambient concentration increase, the uptake concentration increases quickly in the begining and slowdown toward higher ambient concentration. It seems like the grasses can only uptake CO_2 until a certain saturation point. Therefore, the uptake concentration is likely to correlate with log of ambient cocentration as shown in figure 1 (right).

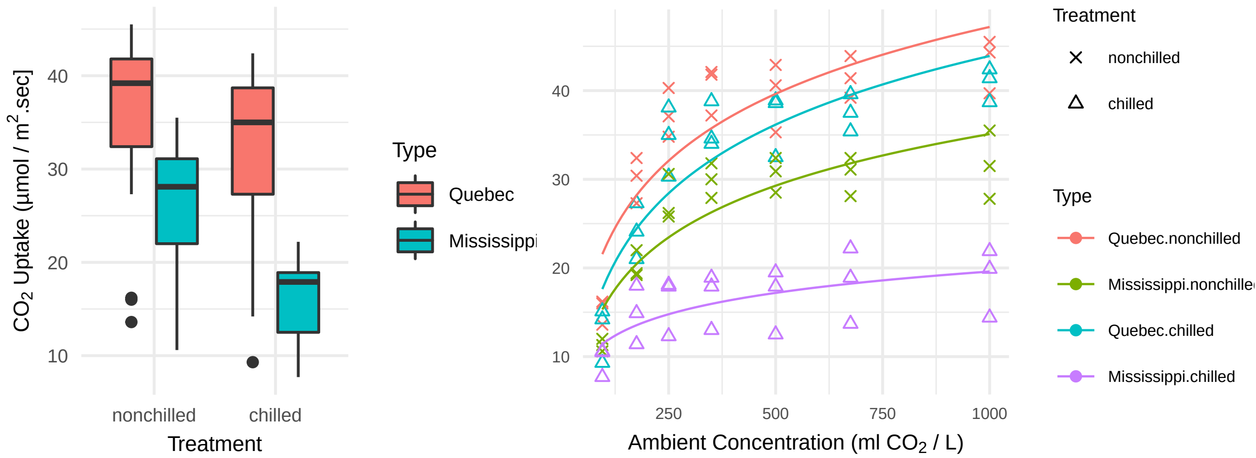

Figure 1: Exploratory plot. Left, boxplot of CO_2 uptake ~ Treament + Type (color). Right, CO_2 uptake ~ ambient concentration + Treatmen (shape) + Type (color). The least squared fits of uptake ~ log of ambient concentration for each case are also shown.

3. Model

With the reasons stated above, uptake CO_2 concentration will be predicted with log of ambient concentration, treatment (represented by a dummy variable is_chilled = {0,1} indicating the condition of nonchilled or chilled) and type of grass. Three Bayesian hierarchical models with increasing complexity and terms were employed. The individual grasses are conditionally independent, however, there are likely correlation among grasses having the same origin. Thus, hierarchical models are appropriate here to capture the correlation accross types of grass. In the models, grasses of the same type will share common parameters, which themselves share a common distribution accross the types. The three models are described below:

Model 1 (Type dependent, each type get it own intercept):

$$
\begin{aligned}
uptake_i \mid Type_i,\ \underset{\sim}{conc_i},\ \underset{\sim}{\alpha},\ \underset{\sim}{\beta},\ \sigma^2 &\overset{ind}{\sim} N(\alpha_{Type_i} + \beta_1 * log(conc_i) + \beta_2 * is\_chilled_i,\ \sigma^2) \\
Type_i &\in \{Quebec, Mississippi\};\ i = 1, ..., n \\
\alpha_{Type} \mid \mu,\ \tau^2 &\overset{iid}{\sim} N(\ \mu,\ \tau^2)
\end{aligned}
$$

Model 2 (similar to model 1 plus an interaction term between log of concentration and chilled):

$$
\begin{aligned}
uptake_i \mid Type_i,\ \underset{\sim}{conc_i},\ \underset{\sim}{\alpha},\ \underset{\sim}{\beta},\ \sigma^2 &\overset{ind}{\sim} N(\alpha_{Type_i} + \beta_1 * log(conc_i) + \beta_2 * is\_chilled_i \\
&\quad + \beta_3 * is\_chilled_i * log(conc_i),\ \sigma^2) \\
Type_i &\in \{Quebec, Mississippi\};\ i = 1, ..., n \\
\alpha_{Type} \mid \mu,\ \tau^2 &\overset{iid}{\sim} N(\ \mu,\ \tau^2)
\end{aligned}
$$

Model 3 (similar to model 2 but each type get it own intercept and slope):

$$
\begin{aligned}
uptake_i \mid Type_i,\ \underset{\sim}{conc_i},\ \underset{\sim}{\alpha},\ \underset{\sim}{\beta},\ \sigma^2 &\overset{ind}{\sim} N(\alpha_{Type_i} + \theta_{Type_i} * log(conc_i) + \beta_1 * is\_chilled_i \\
&\quad + \beta_2 * is\_chilled_i * log(conc_i),\ \sigma^2) \\
Type_i &\in \{Quebec, Mississippi\};\ i = 1, ..., n \\
\alpha_{Type} \mid \mu_\alpha,\ \tau_\alpha^2 \overset{iid}{\sim} N(\ \mu_\alpha,\ \tau_\alpha^2);\ \theta_{Type} &\mid \mu_\theta,\ \tau_\theta^2 \overset{iid}{\sim} N(\ \mu_\theta,\ \tau_\theta^2)
\end{aligned}
$$

The data was expected to drive the posterior distribution of the coeffcients. Therefore, non-informative normal priors ($\mu_0 = 0$, sigma$^2 = 10^6$) for mean and non-informative gamma priors (alpha = 1/2, beta = 10/2) for precision were used. The three models were fitted using JAGS and R. 6x10^5 iterations were conducted. The first 10^5 iterations were for burn-in process. The last 5x10^5 iterations were stored with thinning interval of 100. The models converge as validated by trace plots, Gelman and autocorrelation diagnostics. Modeling assumptions were also verified using residual analyses (figure 2). The residual variances are higher around 20 to 30 µmol CO_2 up take per m^2 per sec. The reason is that the linear models are not good at capturing the transition from a linear increasing phase in the beginning to a flat (saturated) phase around these uptake concentration. In the orgigin paper of Potvin et. al. 1990, a non linear model was employed which was better at modelling this kind of transition. However, non linear models are outside the scope of this course so that the linear Bayesian hierarchical models continue to be used.

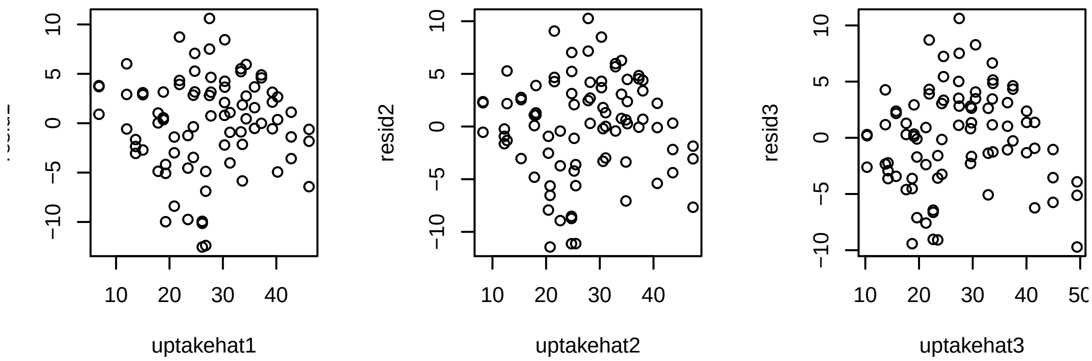

Figure 2: Residual analyses (predicted mean ~ residual) of the three models.

```
dic1
```

```
## Mean deviance:  507.5
## penalty 5.137
## Penalized deviance: 512.7
```

```
dic2
```

```
## Mean deviance:  505.8
## penalty 6.162
## Penalized deviance: 511.9
```

```
dic3
```

```
## Mean deviance:  493.5
## penalty 7.045
## Penalized deviance: 500.5
```

After comparing Deviance Information Criterion, model 3 was selected. Although this model has the highest number of parameters and thus the highest penalty, the penalized deviance is still the lowest amonong the three models. This indicate the effective of the parameters in the model.

4. Results

```
##              Mean         SD    Naive SE Time-series SE
## a[1]   -29.172979  6.5163066 0.053205420    0.095876988
## a[2]   -21.842555  6.4723994 0.052846920    0.092749475
## a0     -25.530560 14.2512719 0.116361148    0.135571502
## b[1]     6.672316  7.8727345 0.064280608    0.127546213
## b[2]    -2.325169  1.3409055 0.010948448    0.021759715
## sig      4.570100  0.3702156 0.003022798    0.003001993
## tau_a    8.378148 13.9691687 0.114057784    0.114057784
## tau_th   5.017267  6.8468758 0.055904507    0.055904507
```

```
## th[1]    11.377185  1.1102632 0.009065261     0.016327692
## th[2]     7.921262  1.1040489 0.009014521     0.015819997
## th0       9.612312  6.0794105 0.049638179     0.050279058
```

Figure 3 shows the prediction of the model (lines) for different combination of type and treatment. The model can explain above 80% of the variances in the dataset ($r^2 = 0.83$). In general, the uptake concentration increases quickly in the begining and slowdown toward higher ambient concentration. The posterior means of theta were significantly positive (around 11 and 8 with std.dev of around 1). However, the grasses from Quebec uptake 15 to 20 µmol CO_2/m^2.sec more than Mississippi in the same ambient CO_2 concentration and treatment. The slope of Quebec ($theta_1$) was higher than Mississippi ($theta_2$) 99% of the time. Considering grasses from the same type, chilled grasses uptake 5 to 10 µmol CO_2/m^2.sec less than non-chilled grasses in the same ambient CO_2 concentration. The slope ($beta_2$) was smaller than zero more than 95% of the time. Although the intercepts show the contradictions ($a_1 < a_2$ and $b_1 > 0$), the value of the slopes were large enough so that within the observation range (95 to 1000 ml CO_2/L) these trends and conclusions were still valid (figure 3). Moreover, the intercepts do not have a real interpretation because they correspond to the mean response for a grass that is not chilled and in a condition of 1 ml CO_2 per litre air, which is not the case for these grasses. Thus, we should not insist on the intercepts and extrapolate outside of the observation range.

```r
mean(mod_csim3[,"th[1]"] > mod_csim3[,"th[2]"])
```

```
## [1] 0.9999333
```

```r
mean(mod_csim3[,"b[2]"] < 0)
```

```
## [1] 0.9586667
```

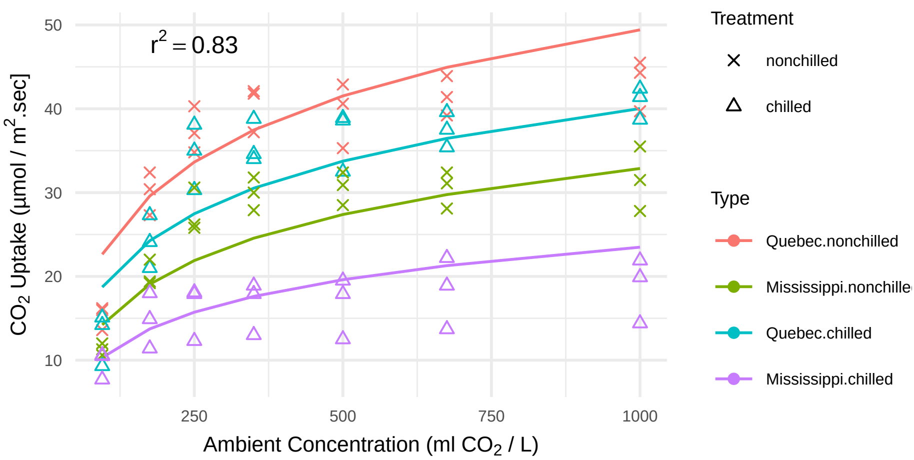

Figure 3: Model prediction. Lines: mean predicted CO_2 uptake ~ ambient concentration + Treatmen + Type (color). Points: observed CO_2 uptake ~ ambient concentration + Treatmen (shape) + Type (color).

5. Conclusions

In conclusion, three Bayesian hierarchical models with increasing complexity and terms were employed for studying the effect of cold condition on grass photosynthesis. The best performance model was the most complex with random intercepts and slopes for each type of grasses. This model can explain 83% of variance in the dataset. Based upon the selected model, it can be concluded that cold weather has negative effects on the photosynthesis of the grass.

References

1) Potvin, C., Lechowicz, M. J. and Tardif, S. (1990) "The statistical analysis of ecophysiological response curves obtained from experiments involving repeated measures", Ecology, 71, 1389–1400.

2) Pinheiro, J. C. and Bates, D. M. (2000) Mixed-effects Models in S and S-PLUS, Springer.