

植物 CO₂ 吸收量的影响因素探究

致理-数理 1 刘苏青* 2021013371

2024 / 06 / 28

摘 要

本文利用 R 语言中的经典 CO₂ 数据集,探讨了不同条件对植物 CO₂ 吸收量的影响,研究重点分析了植物种类、处理类型以及 CO₂ 浓度对 CO₂ 吸收量的主效应和交互效应。通过探索性数据分析、变量选择及 Fisher's LSD 检验,本研究发现植物种类和 CO₂ 浓度是显著影响 CO₂ 吸收量的主要因素。此外,本研究还提出了在中等偏高 CO₂ 浓度下种植来自 Quebec 地区植物的建议,并讨论了模型改进和实验设计的优化方向,为植物生理研究和生态系统管理提供了一定参考。

关键词: CO₂ 吸收量, 交互效应, 变量选择, Fisher's LSD 检验

*liu-sq21@mails.tsinghua.edu.cn

目录

1	引言	3
2	数据集及问题概述	3
3	模型选择与分析	4
3.1	探索性数据分析	4
3.2	变量选择	4
3.3	Fisher's LSD 检验	5
4	总结与反思	6
4.1	总结	6
4.2	模型改进	6
4.3	实验改进	6
5	附录	7
5.1	图片	7
5.2	R 代码	8
5.3	参考文献	8

1 引言

在生态学研究中，测量和分析植物对不同环境条件反应的能力是理解植物适应性和生长特性的关键。本文选用 CO₂ 数据集进行分析，CO₂ 数据集是 R 语言中自带的一个经典的数据集，其中记录了十二株植物在不同 CO₂ 浓度和处理条件下的 CO₂ 吸收量。这些数据为研究植物生理反应曲线提供了宝贵的资源。

本数据集背后的实验选用一种叫 *Echinochloa crus-galli* 的 C₄ 草种作为实验材料，它的种子分别来自魁北克 (Quebec) 和密西西比 (Mississippi)。这些种子在杜克大学植物温室生长室内生长，该生长室室内条件为 26°C 下 14 小时和 20°C 下 10 小时的光暗循环，光合光子通量密度为 1000 μmol · m⁻² · s⁻¹，空气湿度为 70%，CO₂ 浓度为 350 μL/L。实验人员将所有植物分为两组：控制组 (nonchilled) 和冷处理组 (chilled)，其中控制组植物白天 (14 小时部分) 保持在 26°C，冷处理组植物白天则受到 7°C 的冷处理。每株控制组植物需要随机配对一株将被冷处理的植物，以确保在冷处理前植物的生长情况非常均匀。经过一夜 (10 小时部分) 20°C 的恢复后，实验人员在七种不同的 CO₂ 浓度 (100、175、250、350、500、675、1000 μL/L) 下依次为冷处理和控制组植物连续测量 CO₂ 吸收量，并绘制 CO₂ 曲线。

因此，本实验的变量以及不变量可以总结如下：

- 变量：种子来源、白天室内温度、测量时的 CO₂ 浓度；
- 不变量：光暗循环周期、光合光子通量密度、空气湿度、预处理时的 CO₂ 浓度等。

2 数据集及问题概述

Table 1: CO₂ 数据集

变量	记号	含义	说明
Plant	τ	植物编号	共 12 株 (Qn1~3, Qc1~3, Mn1~3, Mc1~3)
Type	x_1	植物种类	共 2 种 (Quebec, Mississippi)
Treatment	x_2	处理类型	共 2 种 (nonchilled, chilled)
conc	x_3	CO ₂ 浓度	共 7 种 (100, 175, 250, 350, 500, 675, 1000 μL/L)
uptake	y	CO ₂ 吸收量	每株植物的 CO ₂ 吸收量

基于上文所述实验流程，实验人员获得了 CO₂ 数据集。该数据集共包含 5 个变量，其中 Plant 是植物编号，Type、Treatment 和 conc 是自变量，uptake 是因变量。Type

变量包括 Quebec 和 Mississippi 两类, Treatment 变量包括 nonchilled 和 chilled 两类, conc 变量包括 100, 175, 250, 350, 500, 675, 1000 μ L/L 七类, 并且对于任意的因子组合, 实验人员重复了三次 (平衡设计), 所以该数据集共包含了 $2 \times 2 \times 7 \times 3 = 84$ 个样本。

接下来本文将依据 CO_2 数据集, 探究以下问题:

- 哪些因素会显著影响植物 CO_2 吸收量?
- 不同因素之间是否存在显著的交互效应?
- 什么条件下的植物最有利于 CO_2 的吸收?

3 模型选择与分析

3.1 探索性数据分析

首先, 我们对数据集进行探索性数据分析, 通过绘制 CO_2 吸收量关于植物种类、处理类型以及 CO_2 浓度的散点图, 我们可以大致观察到 CO_2 吸收量随着 CO_2 浓度的增加而增加, 同时来自 Mississippi 的植株 CO_2 吸收量较低且容易受到冷处理的影响, 因此我们猜测 CO_2 吸收量与植物种类、处理类型、 CO_2 浓度以及它们的交互效应均有关。

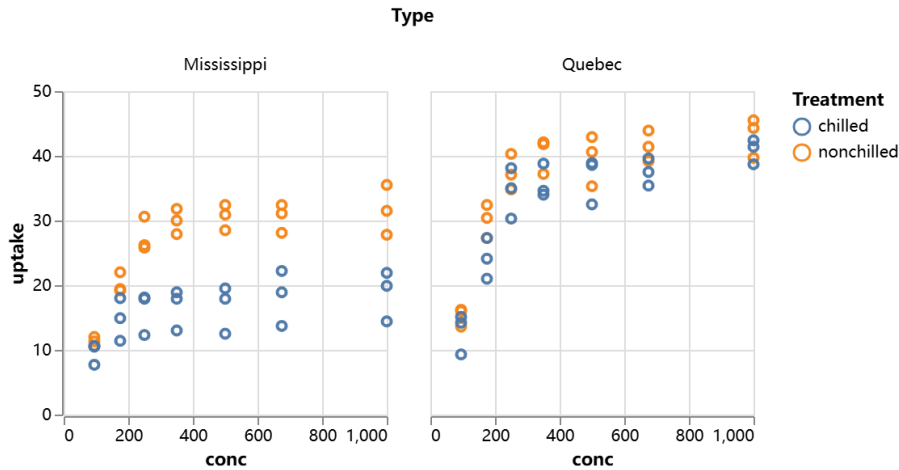


Figure 1: EDA

3.2 变量选择

然后, 我们考虑包括所有变量的主效应及它们的交互效应的全模型:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \varepsilon \quad (1)$$

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    25.585034   2.255256   11.345 < 2e-16 ***
TypeMississippi -7.131741   3.189414   -2.236  0.0283 *
Treatmentchilled -4.163993   3.189414   -1.306  0.1956
conc           0.022410   0.004295    5.218 1.52e-06 ***
TypeMississippi:Treatmentchilled -1.747509   4.510513   -0.387  0.6995
TypeMississippi:conc -0.005171   0.006074   -0.851  0.3973
Treatmentchilled:conc  0.001340   0.006074    0.221  0.8259
TypeMississippi:Treatmentchilled:conc -0.011057   0.008589   -1.287  0.2019
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figure 2: Full Model

由结果可知，该模型中只有植物种类和 CO_2 浓度的主效应显著，其余交互效应均不显著，这也恰好印证了附录交互效应图 (Figure 4) 中各组的平行关系。因此，我们重新考虑仅包括植物种类和 CO_2 浓度变量的简化模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon \quad (2)$$

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    25.830052   1.579918   16.349 < 2e-16 ***
conc           0.017731   0.002625    6.755 2.00e-09 ***
TypeMississippi -12.659524   1.544261   -8.198 3.06e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Analysis of Variance Table

```

Response: uptake
      Df Sum Sq Mean Sq F value    Pr(>F)
conc    1  2285.0   2285.0   45.627 1.997e-09 ***
Type    1  3365.5   3365.5   67.204 3.061e-12 ***
Residuals 81  4056.4     50.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figure 3: Reduced Model

由结果可知，该模型中植物种类和 CO_2 浓度变量的主效应均显著，且附录模型检验图 (Figure 5) 显示模型符合方差齐性和正态性假设。

3.3 Fisher's LSD 检验

前文的 ANOVA 中显示 conc 变量的 $p\text{-value} \ll 0.05$ ，证明 CO_2 浓度变量的主效应十分显著，但我们并不知道具体哪些组之间有显著差异。因此，接下来我们将通过 Fisher's LSD 检验进一步确认哪些组之间的均值差异是显著的。

由结果可知 (Figure 6), CO_2 浓度为 95 或 $175\mu\text{L/L}$ 的组和其他组之间的均值差异均显著, 其余组之间的均值差异均不显著, 这也符合前文探索性数据分析图中的结果: 无论其他变量取何水平, 我们均能观测到较低 CO_2 浓度组的 CO_2 吸收量显著小于其他组。

4 总结与反思

4.1 总结

综上所述, 本研究认为显著影响植物 CO_2 吸收量的因素为植物种类和 CO_2 浓度, 而处理类型对植物 CO_2 吸收量无显著影响, 并且不同因素之间并不存在显著的交互效应。此外, 结合 Fisher's LSD 检验以及探索性数据分析的图可知, 适当提升 CO_2 浓度可以显著提升植物 CO_2 吸收量, 但是过高的 CO_2 浓度不会对植物 CO_2 吸收量有太多改善。这个结果启示我们, 应当在中等偏高的 CO_2 浓度下种植来自 Quebec 地区的植物, 其耐寒性较强 (不易受冷处理操作的干扰), 且 CO_2 吸收量优于来自 Mississippi 地区的植物, 有利于植物自身的生长和生产情况以及改善生态系统健康。

4.2 模型改进

在前文建立的模型中, 我们默认实验人员在每种因子组合下选取三株植物进行的实验为三次重复。然而, 植物与植物之间可能存在天生的差异, 或许我们应当把植物编号作为一个随机效应加到原本的模型中, 并采用似然比检验来比较既有模型和改进模型:

$$y = \beta_0 + \tau + \beta_1 x_1 + \beta_3 x_3 + \varepsilon, \quad \tau \sim \mathcal{N}(0, \sigma_\tau^2) \quad (3)$$

由结果可知 (Figure 7), 改进模型显著优于既有模型, 但同时进行后续分析也更难。

4.3 实验改进

由于时间原因, 我没有自己设计实验, 而是选择已有的数据集。因此我认为本实验在设计上存在一定缺陷, 例如实验人员在七种不同的 CO_2 浓度下依次测量同一株植物 CO_2 吸收量时, 可能忽略了前面的实验对后面的实验造成的干扰 (如过低或过高的 CO_2 浓度对该植物造成了一定程度的创伤), 因此我提议增大实验的重复次数, 并在每一次重复中更换七种不同的 CO_2 浓度的测量顺序, 从而排除/发现该因素可能造成的影响。

5 附录

5.1 图片

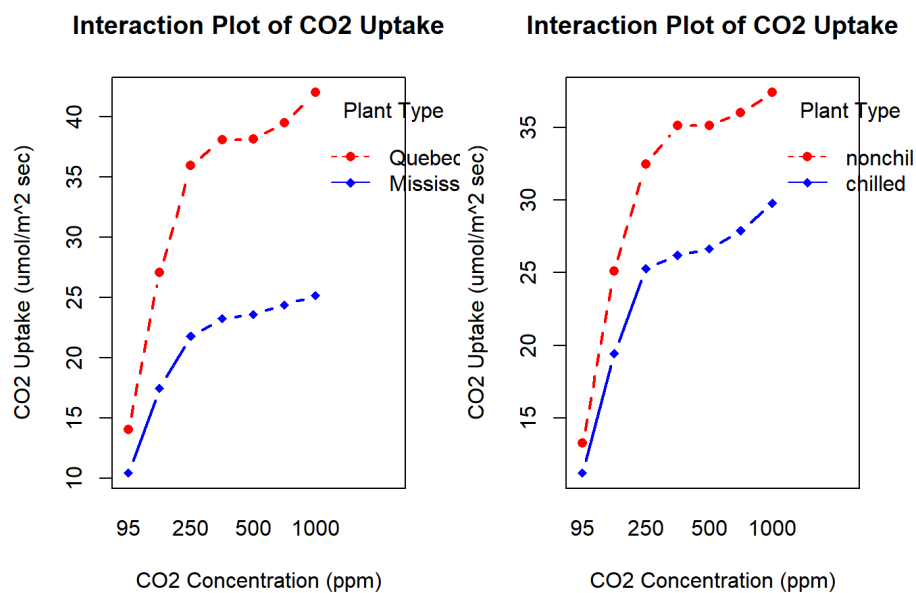


Figure 4: Interaction Plot

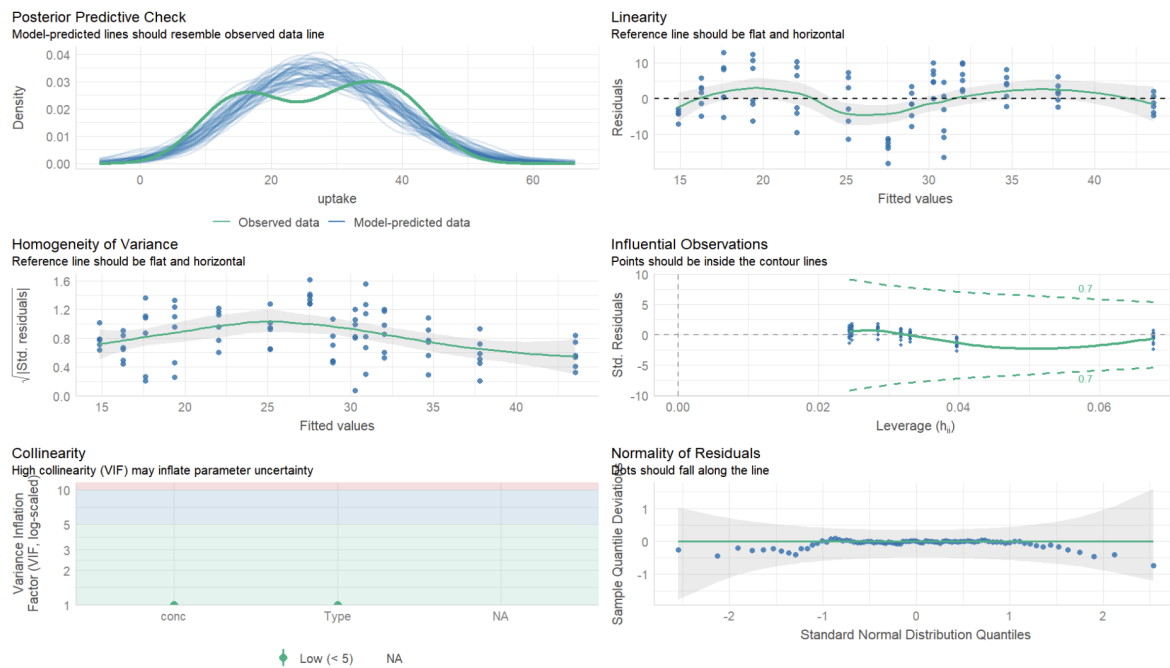


Figure 5: Check Model

```

Models:
reduced_model: uptake ~ conc + Type
reduced_model_ad: uptake ~ Type + conc + (1 | Plant)
              npar      AIC      BIC logLik deviance Chisq Df Pr(>Chisq)
reduced_model      4 572.07 581.79 -282.04  564.07
reduced_model_ad    5 563.41 575.56 -276.70  553.41 10.663  1  0.001093 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figure 7: LRT Test

```

$comparison
difference pvalue signif.      LCL      UCL
95 - 175   -10.0250000 0.0008   *** -15.773300 -4.2767004
95 - 250   -16.6166667 0.0000   *** -22.364966 -10.8683671
95 - 350   -18.4083333 0.0000   *** -24.156633 -12.6600337
95 - 500   -18.6166667 0.0000   *** -24.364966 -12.8683671
95 - 675   -19.6916667 0.0000   *** -25.439966 -13.9433671
95 - 1000  -21.3250000 0.0000   *** -27.073300 -15.5767004
175 - 250   -6.5916667 0.0251    *  -12.339966 -0.8433671
175 - 350   -8.3833333 0.0048   **  -14.131633 -2.6350337
175 - 500   -8.5916667 0.0039   **  -14.339966 -2.8433671
175 - 675   -9.6666667 0.0012   **  -15.414966 -3.9183671
175 - 1000 -11.3000000 0.0002   *** -17.048300 -5.5517004
250 - 350   -1.7916667 0.5369          -7.539966  3.9566329
250 - 500   -2.0000000 0.4907          -7.748300  3.7482996
250 - 675   -3.0750000 0.2903          -8.823300  2.6732996
250 - 1000  -4.7083333 0.1070         -10.456633  1.0399663
350 - 500   -0.2083333 0.9427          -5.956633  5.5399663
350 - 675   -1.2833333 0.6581          -7.031633  4.4649663
350 - 1000  -2.9166667 0.3157          -8.664966  2.8316329
500 - 675   -1.0750000 0.7108          -6.823300  4.6732996
500 - 1000  -2.7083333 0.3513          -8.456633  3.0399663
675 - 1000  -1.6333333 0.5734          -7.381633  4.1149663

```

Figure 6: LSD Test

5.2 R 代码

见 Code.qmd/Code.html 文件。

5.3 参考文献

[1] Potvin, Catherine & Lechowicz, Martin & Tardif, Serge. (1990). The Statistical Analysis of Ecophysiological Response Curves Obtained from Experiments Involving Repeated Measures. *Ecology*. 71. 1389-1400. 10.2307/1938898.