

混合效应模型-实际应用案例解析

张霜

中国科学院生态环境研究中心

2021-5-9



北京凌云翼数据科技有限公司

混合效应模型

- 多重比较
- 模型选择
- 多元回归
- 统计power计算
- 随机效应相关矩阵的自定义
- Meta分析
- 方差分解

混合效应模型--多重比较

- 解释变量有多个水平，如不同的处理
- 进行不同水平之间的两两比较。

##多重比较案例

```
multi <-
```

```
read.csv("multi_comparison.csv")
```

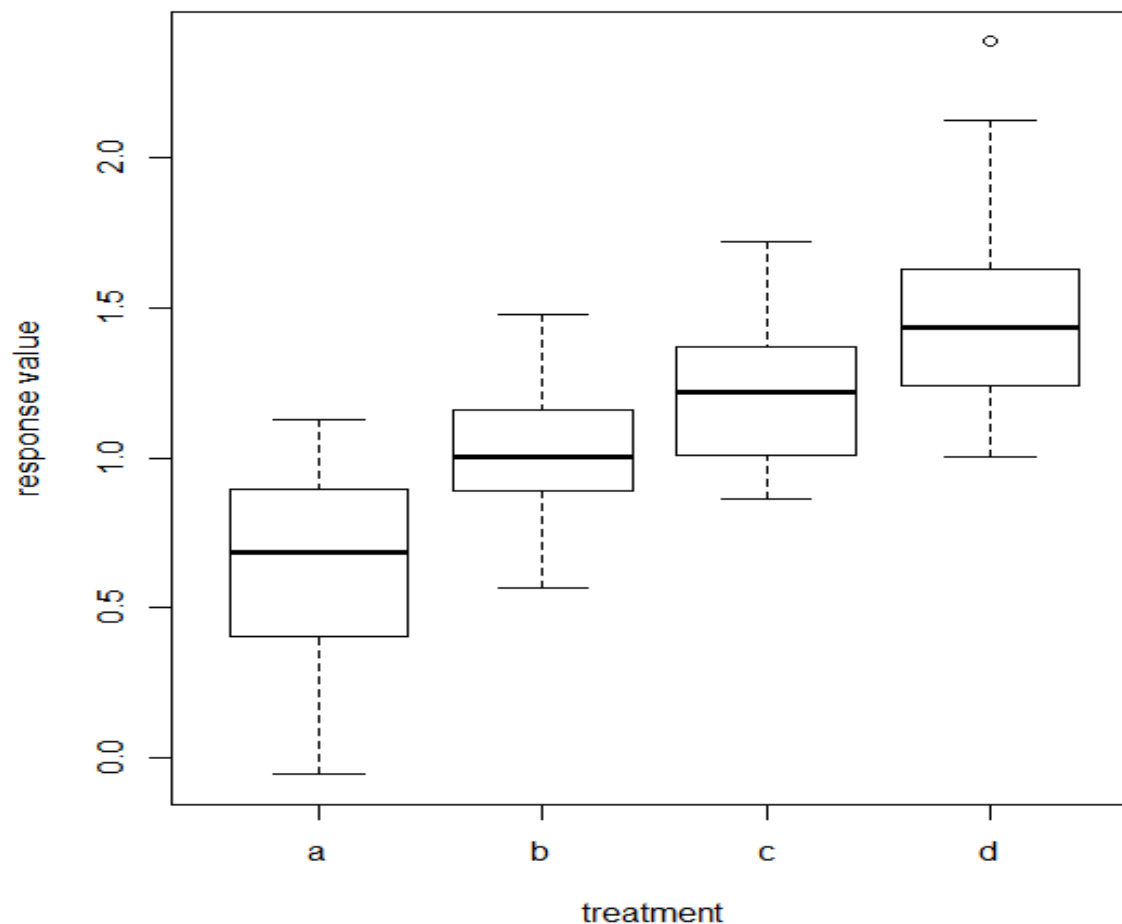
```
library(lme4)
```

```
library(MuMIn)
```

```
View(multi)
```

```
hist(multi$y)
```

```
boxplot(y~x3,data=multi)
```



混合效应模型--多重比较

```
mt1<-lmer(y~x3+(1|species),data=multi)
```

```
r.squaredGLMM(mt1)
```

```
summary(mt1)
```

```
anova(mt1)
```

到底是哪些水平之间有差异？

```
> summary(mt1)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: y ~ x3 + (1 | species)
Data: multi

REML criterion at convergence: 64.4

Scaled residuals:
    Min       1Q   Median       3Q      Max
-2.5453 -0.7844 -0.0671  0.6049  3.3877

Random effects:
Groups   Name              Variance Std.Dev.
species (Intercept)  0.00000    0.0000
Residual                    0.07556    0.2749
Number of obs: 199, groups: species, 156

Fixed effects:
              Estimate Std. Error    df t value Pr(>|t|)
(Intercept)   0.64407    0.05861 195.00000   10.990 < 2e-16 ***
x3b            0.36917    0.07963 195.00000    4.636 6.49e-06 ***
x3c            0.57553    0.07479 195.00000    7.695 6.86e-13 ***
x3d            0.81321    0.06392 195.00000   12.722 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) x3b      x3c
x3b  -0.736
x3c  -0.784  0.577
x3d  -0.917  0.675  0.718

> anova(mt1)
Type III Analysis of Variance Table with Satterthwaite's method
      Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
x3    14.49   4.8299     3    195   63.92 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

混合效应模型--多重比较

```
library(multcomp)
```

```
## 多重比较Tukey post-hoc test with Bonferroni correction for the P values
```

```
summary(glht(mt1, linfct=mcp(x3 = "Tukey")), test = adjusted("bon"))
```

```
> ## 多重比较Tukey post-hoc test with Bonferroni correction for the P values
> summary(glht(mt1, linfct=mcp(x3 = "Tukey")),
+         test = adjusted("bon"))
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lmer(formula = y ~ x3 + (1 | species), data = multi)

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)	
b - a == 0	0.36917	0.07963	4.636	2.13e-05	***
c - a == 0	0.57553	0.07479	7.695	8.53e-14	***
d - a == 0	0.81321	0.06392	12.722	< 2e-16	***
c - b == 0	0.20636	0.07117	2.900	0.0224	*
d - b == 0	0.44404	0.05965	7.445	5.84e-13	***
d - c == 0	0.23768	0.05301	4.484	4.40e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- bonferroni method)

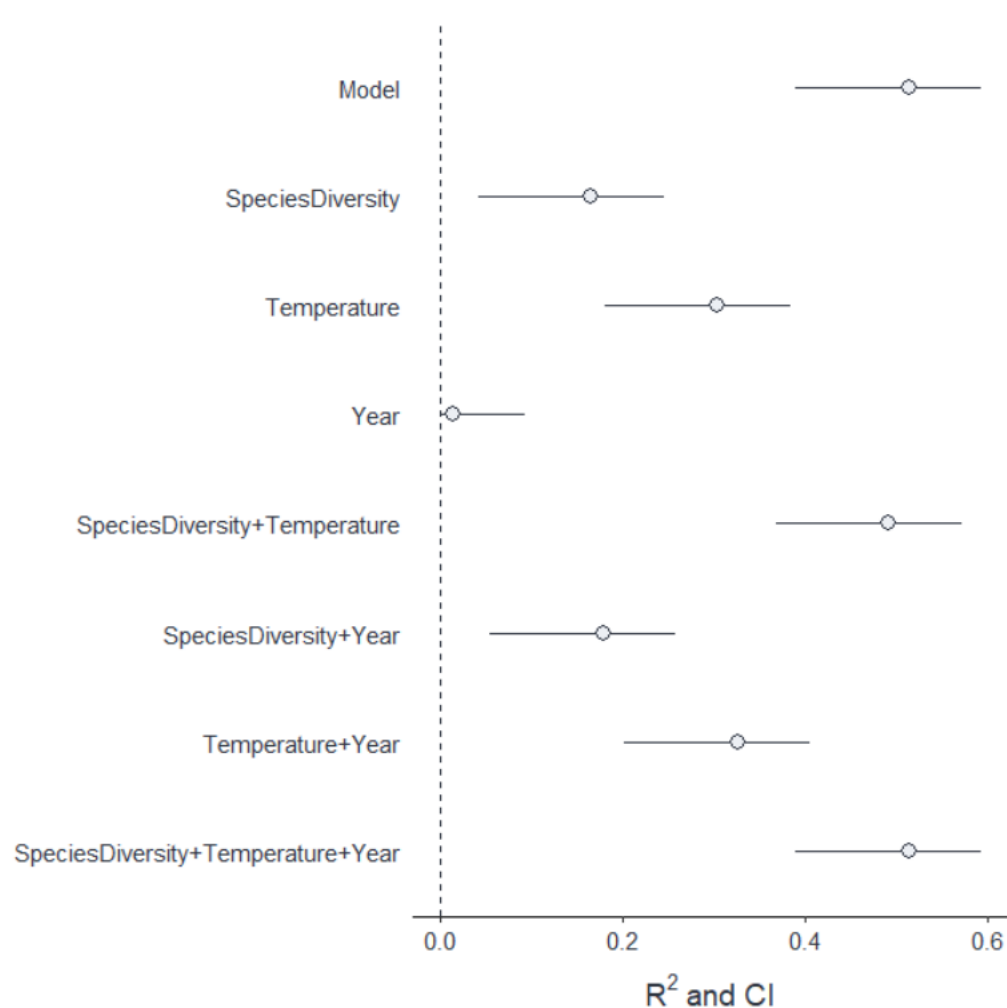
混合效应模型的方差分解partR2

```
library(partR2)
library(lme4)
data(biomass)
mod <- lmer(Biomass ~ Year + Temperature + SpeciesDiversity +
  (1|Population),
  data = biomass)
(R2 <- partR2(mod, partvars = c("SpeciesDiversity", "Temperature", "Year"),
  R2_type = "marginal", nboot = 100, CI = 0.95,
  data = biomass))
forestplot(R2, type = "R2", line_size = 0.7, text_size = 14, point_size = 3)
```

混合效应模型的方差分解partR2

Part (semi-partial) R²:

Predictor(s)	R ²	CI_lower	CI_upper	nboot	ndf
Model	0.5133	0.3891	0.5928	100	4
SpeciesDiversity	0.1653	0.0411	0.2448	100	3
Temperature	0.3038	0.1796	0.3833	100	3
Year	0.0130	0.0000	0.0925	100	3
SpeciesDiversity+Temperature	0.4914	0.3672	0.5709	100	2
SpeciesDiversity+Year	0.1784	0.0542	0.2579	100	2
Temperature+Year	0.3250	0.2008	0.4045	100	2
SpeciesDiversity+Temperature+Year	0.5133	0.3891	0.5928	100	1



方差分解的实际应用案例

```
f1m=lmer(DOY ~ MAT + Force5s + Chill_5_5 +SPEI1m_p+ (1|Species)+(1|site), data = E3)
```

```
var = c("MAT","Force5s","Chill_5_5", "SPEI1m_p")
```

```
PR2_all = var_decomp(f1m, var, E3) #####自定义函数，详见文章附件代码中注释
```

```
names(PR2_all) <- c( var, "Random", "Unexplained")
```

```
PR2_all
```

```
> PR2_all
```

MAT	Force5s	chill_5_5	SPEI1m_p	Random	Unexplained
53.4441051	23.7944966	6.3557217	0.5849362	10.1742082	5.6465322

此案例有数据，有代码（模型，作图）

Huang, J. G., Q. Ma, S. Rossi, F. Biondi, and E. Ziaco. 2020. Photoperiod and temperature as dominant environmental drivers triggering secondary growth resumption in Northern Hemisphere conifers. *Proceedings of the National Academy of Sciences* 117:202007058.

方差分解的实际应用案例

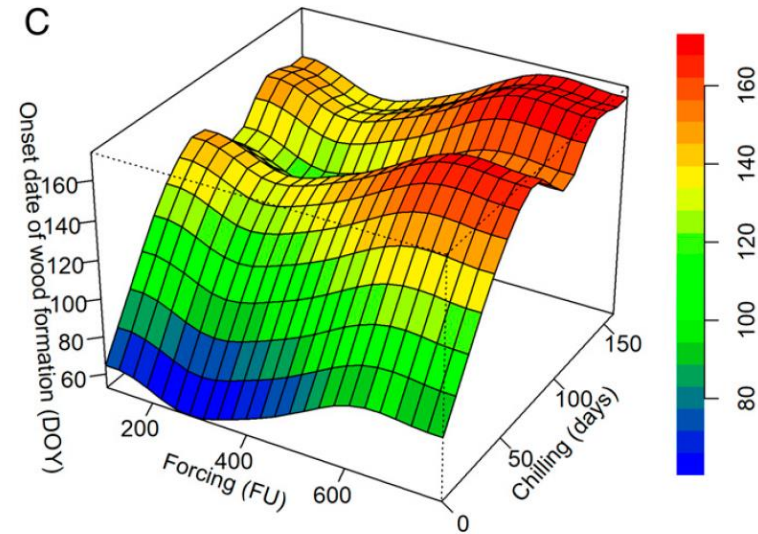
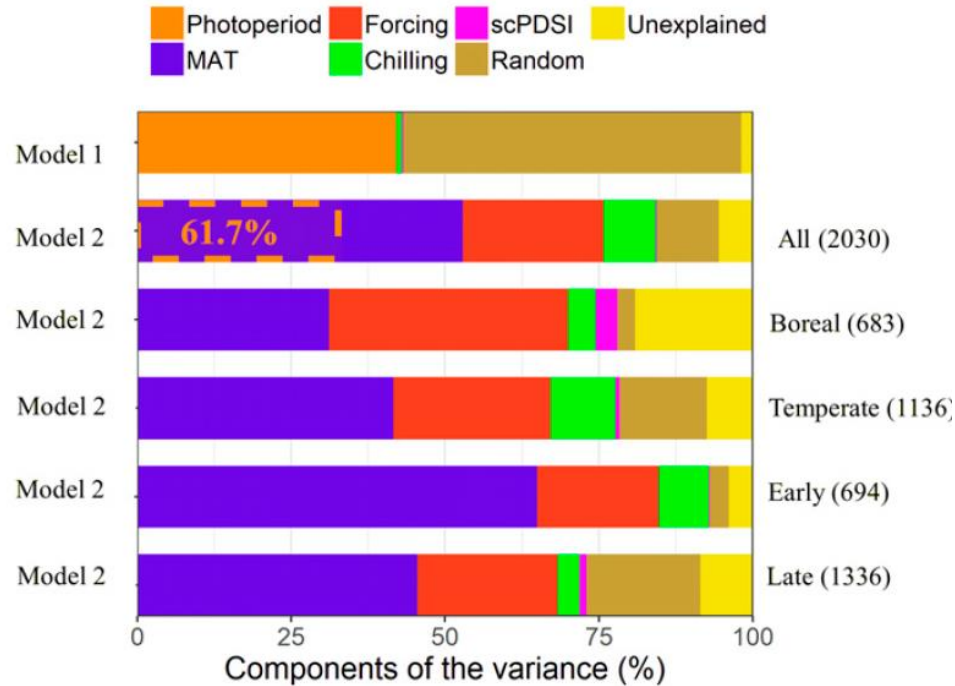


Fig. 2. Changes in the onset date of wood formation (DOY) in relation to mean annual temperature (MAT) (abbreviated as “temperature”) and forcing (A), MAT and chilling (B), and forcing (FU) and chilling (C) (fitted by a GAM).

一篇学习混合模型，方差分解和制图的佳作

Huang, J. G., Q. Ma, S. Rossi, F. Biondi, and E. Ziaco. 2020. Photoperiod and temperature as dominant environmental drivers triggering secondary growth resumption in Northern Hemisphere conifers. *Proceedings of the National Academy of Sciences* 117:202007058.

不同层级上X的交互作用

一个学生的成绩Y跟哪些因素有关？

1) 学生的个人智商IQ

2) 教师的教龄 texp

$\text{lmer}(Y \sim \text{IQ} + \text{texp} + (1 | \text{class}), \text{data} = d)$



现在，再思考一个更深入的问题：

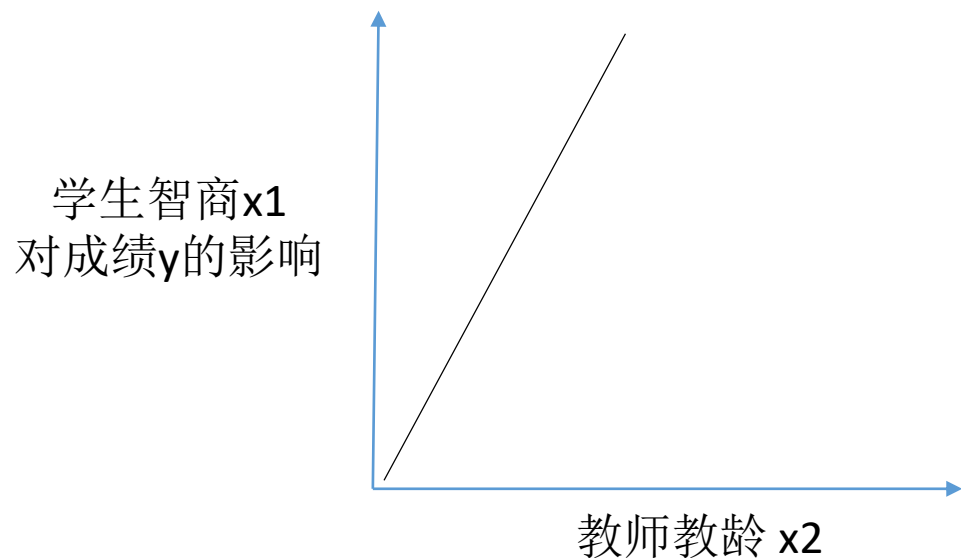
学员智商IQ对成绩的影响是否受到教师教龄 texp 的影响（比如，智商越高的学生，在教师很有经验的一个班级里，成绩是否会更好？）

$\text{lmer}(Y \sim \text{IQ} + \text{texp} + \text{IQ}:\text{texp} + (1 | \text{class}), \text{data} = d)$

不同层级上x的交互作用

$\text{lmer}(Y \sim IQ + \text{texp} + IQ:\text{texp} + (IQ | \text{class}), \text{data} = d)$

问题：学员智商（x1）对成绩的影响是否受到教师教龄的影响（比如，智商越高的学生，在教师很有经验的一个班级里，成绩是否会更好？）



模型中，IQ对成绩的影响：

- 1) 一个总体斜率 $Y \sim IQ$
- 2) 一个随class随机变化的随机斜率 $(IQ | \text{class})$
- 3) 一个随教师教龄而变化的斜率 $IQ:\text{year}$

三者相加，就是IQ对Y的总体影响，将此总体影响与x2作图，就能回答本问题

这是混合模型的一大优势

案例展示

`pdata<-read.csv("pdata.csv")` ####纯演示目的，案例无实际意义

####出于纯展示目的，
####将原数据中变量popular重新命名为score
####将原数据中变量extrav重新命名为IQ
`pdata$score<-pdata$popular`
`pdata$IQ<-pdata$extrav`



上层变量texp对score~IQ（下层变量）回归的截距beta0和斜率beta1都有影响

###IQ对score的影响随着不同的texp而变化

```
pmodel<- lmer(score ~ IQ + texp +IQ:texp +(IQ | class), data = pdata)
```

```
summary(pmodel)
```

###提取模型参数估计值作图

###获取随机斜率

```
r_slope<-as.data.frame(coef(pmodel)$class)
```

```
Random effects:
Groups      Name          Variance Std.Dev. Corr
class      (Intercept) 0.5289898 0.72732
            IQ          0.0003337 0.01827 -1.00
Residual    0.8898027 0.94329
Number of obs: 2000, groups:  class, 100

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept) -1.211e+00  3.210e-01 1.860e+02  -3.772 0.000217 ***
IQ           8.911e-01  4.495e-02 1.759e+03  19.822 < 2e-16 ***
texp        2.519e-01  1.975e-02 1.660e+02  12.758 < 2e-16 ***
IQ:texp     -2.746e-02  2.872e-03 1.725e+03  -9.563 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

上层变量 texp 对 $\text{score} \sim \text{IQ}$ （下层变量）回归的斜率 β_1 的影响

```
#### 获取各个班级教师的教龄信息
el<-pdata%>% group_by (class) %>% summarise
(texp=mean(texp))
### 获取随机斜率的SE, se.ranef
r_slope$se<-se.ranef(pmodel)$class[,2]
####将随机斜率与对应班级的老师教龄合并成一个数据集
r_slope$texp<-el$texp
### 求出考虑了老师教龄 $\text{texp}$ 之后, IQ与成绩的关系
r_slope$beta1<-r_slope$IQ +r_slope$`IQ:texp`*r_slope$texp
#### 计算斜率的置信区间
r_slope$lower<-r_slope$beta1-1.96*r_slope$se
r_slope$upper<-r_slope$beta1+1.96*r_slope$se
####把r_slope转为数据框格式
pd2<-as.data.frame(r_slope)
```

```
####对成绩与IQ关系的斜率 $\beta_1$ 与老师教龄作图
lm2<-lm(beta1~texp,data=pd2)
summary(lm2)

### 用predict命令求拟合线的置信区间
predlm = predict(lm2, interval = "confidence")

### 数据合并及作图
pd2 = cbind(pd2, predlm)
```

上层变量texp对score~IQ（下层变量）回归的斜率beta1的影响

Random effects:

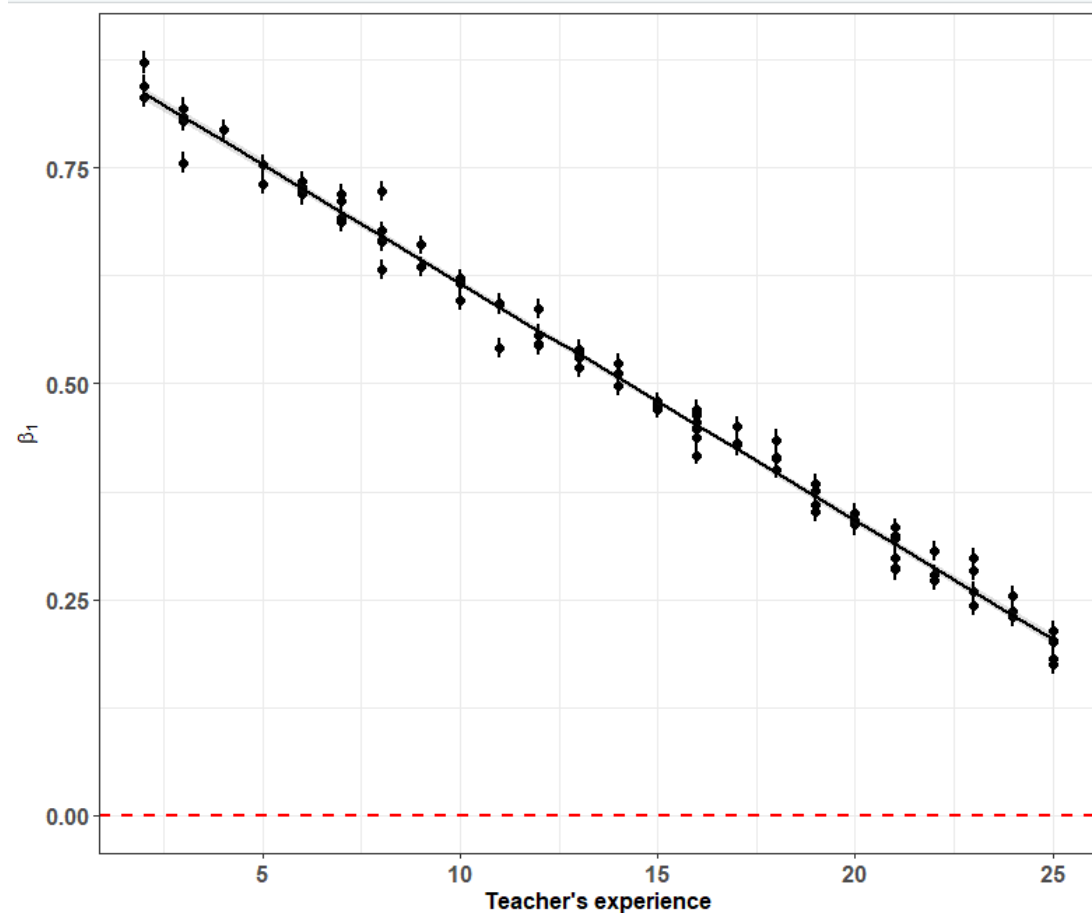
Groups	Name	Variance	Std.Dev.	Corr
class	(Intercept)	0.5289898	0.72732	
	IQ	0.0003337	0.01827	-1.00
	Residual	0.8898027	0.94329	

Number of obs: 2000, groups: class, 100

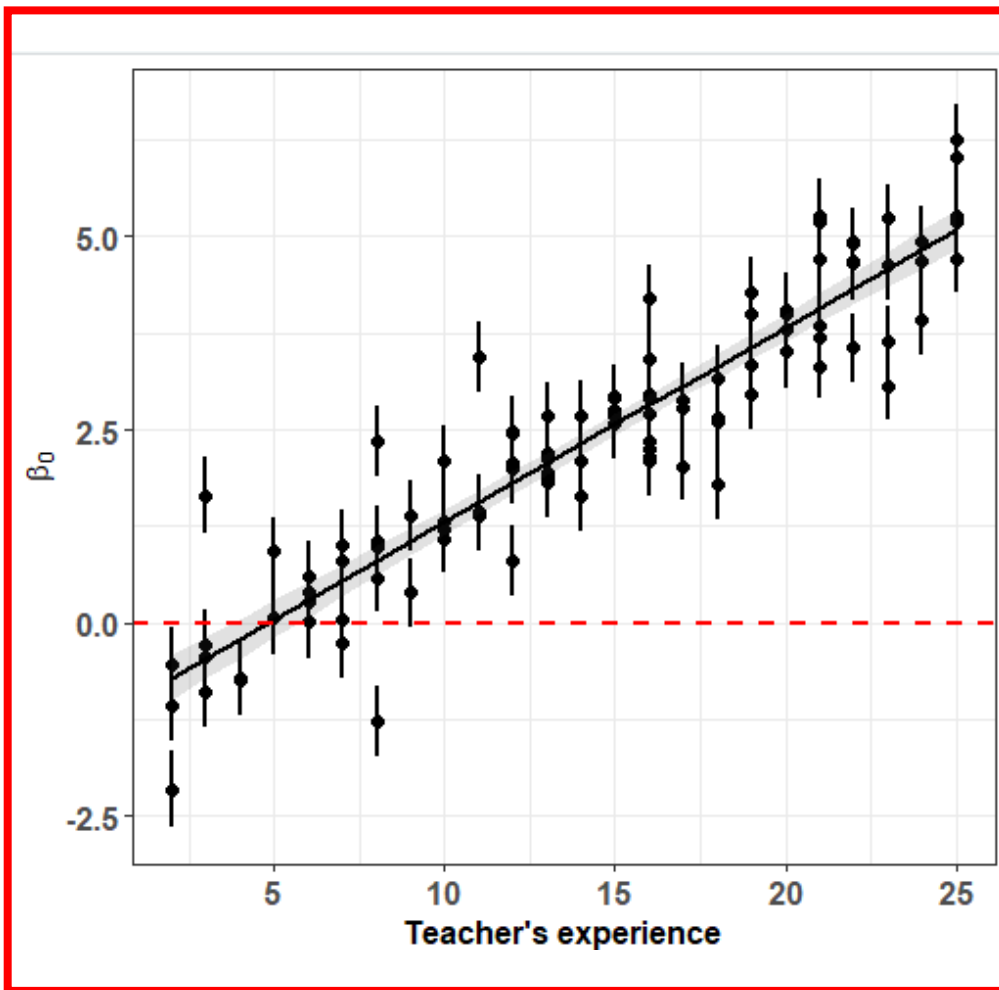
Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	-1.211e+00	3.210e-01	1.860e+02	-3.772	0.000217 ***
IQ	8.911e-01	4.495e-02	1.759e+03	19.822	< 2e-16 ***
texp	2.519e-01	1.975e-02	1.660e+02	12.758	< 2e-16 ***
IQ:texp	-2.746e-02	2.872e-03	1.725e+03	-9.563	< 2e-16 ***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



上层变量texp对score~IQ（下层变量）回归的截距beta0的影响



Random effects:

Groups	Name	Variance	Std.Dev.	Corr
class	(Intercept)	0.5289898	0.72732	
	IQ	0.0003337	0.01827	-1.00
	Residual	0.8898027	0.94329	

Number of obs: 2000, groups: class, 100

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	-1.211e+00	3.210e-01	1.860e+02	-3.772	0.000217	***
IQ	8.911e-01	4.495e-02	1.759e+03	19.822	< 2e-16	***
texp	2.519e-01	1.975e-02	1.660e+02	12.758	< 2e-16	***
IQ:texp	-2.746e-02	2.872e-03	1.725e+03	-9.563	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

生态学中的实际应用案例

Ecology, 91(2), 2010, pp. 355–361
© 2010 by the Ecological Society of America

On the application of multilevel modeling in environmental and ecological studies

SONG S. QIAN,^{1,3} THOMAS F. CUFFNEY,² IBRAHIM ALAMEDDINE,¹ GERARD MCMAHON,²
AND KENNETH H. RECKHOW¹

¹*Nicholas School of the Environment, Duke University, Durham, North Carolina 27708 USA*

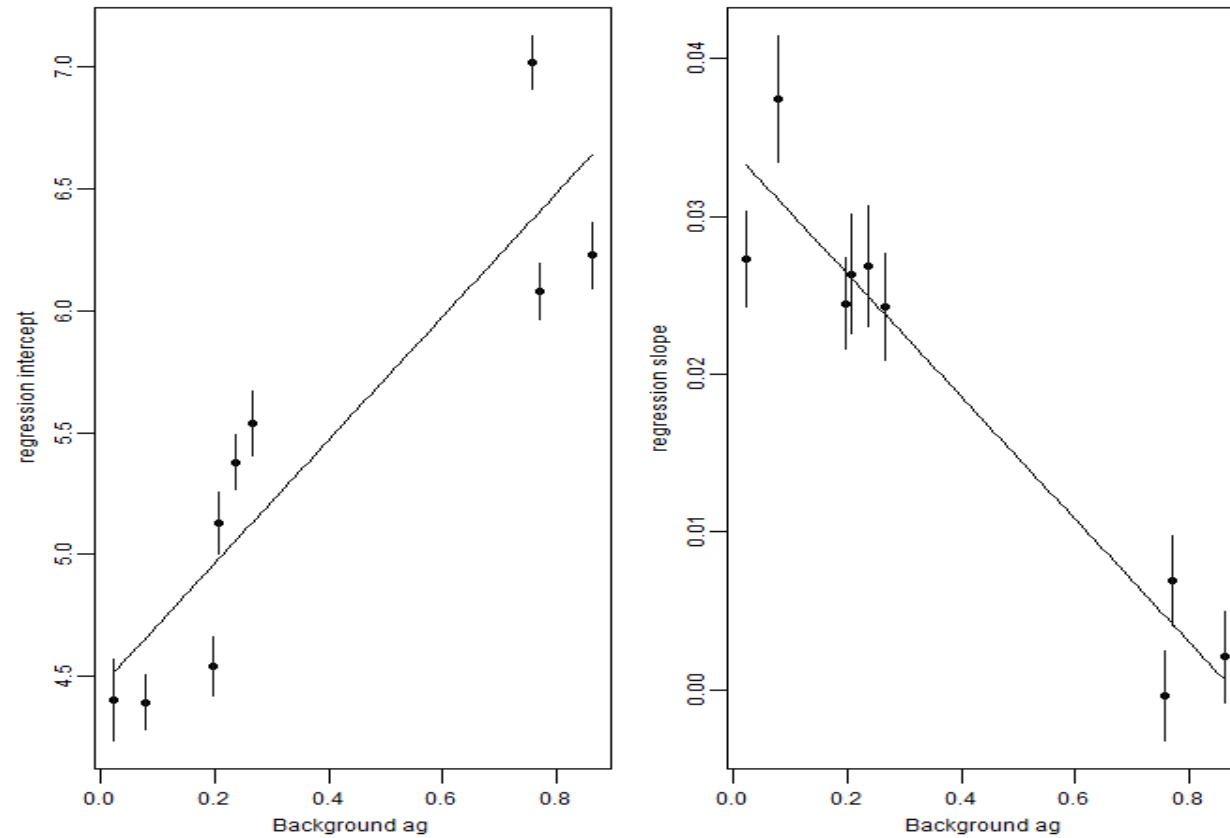
²*U.S. Geological Survey, 3916 Sunset Ridge Road, Raleigh, North Carolina 27607 USA*

Qian, S. S., Cuffney, T. F., Alameddine, I., McMahon, G. & Reckhow, K. H. On the application of multilevel modeling in environmental and ecological studies. *Ecology* **91**, 355–361, doi:<https://doi.org/10.1890/09-1043.1> (2010).

$Y \sim \text{nuii}$ (下层变量)的截距和斜率 都受上层变量ag.full的影响

随机截距和斜率都受site层上的变量ag.full的影响

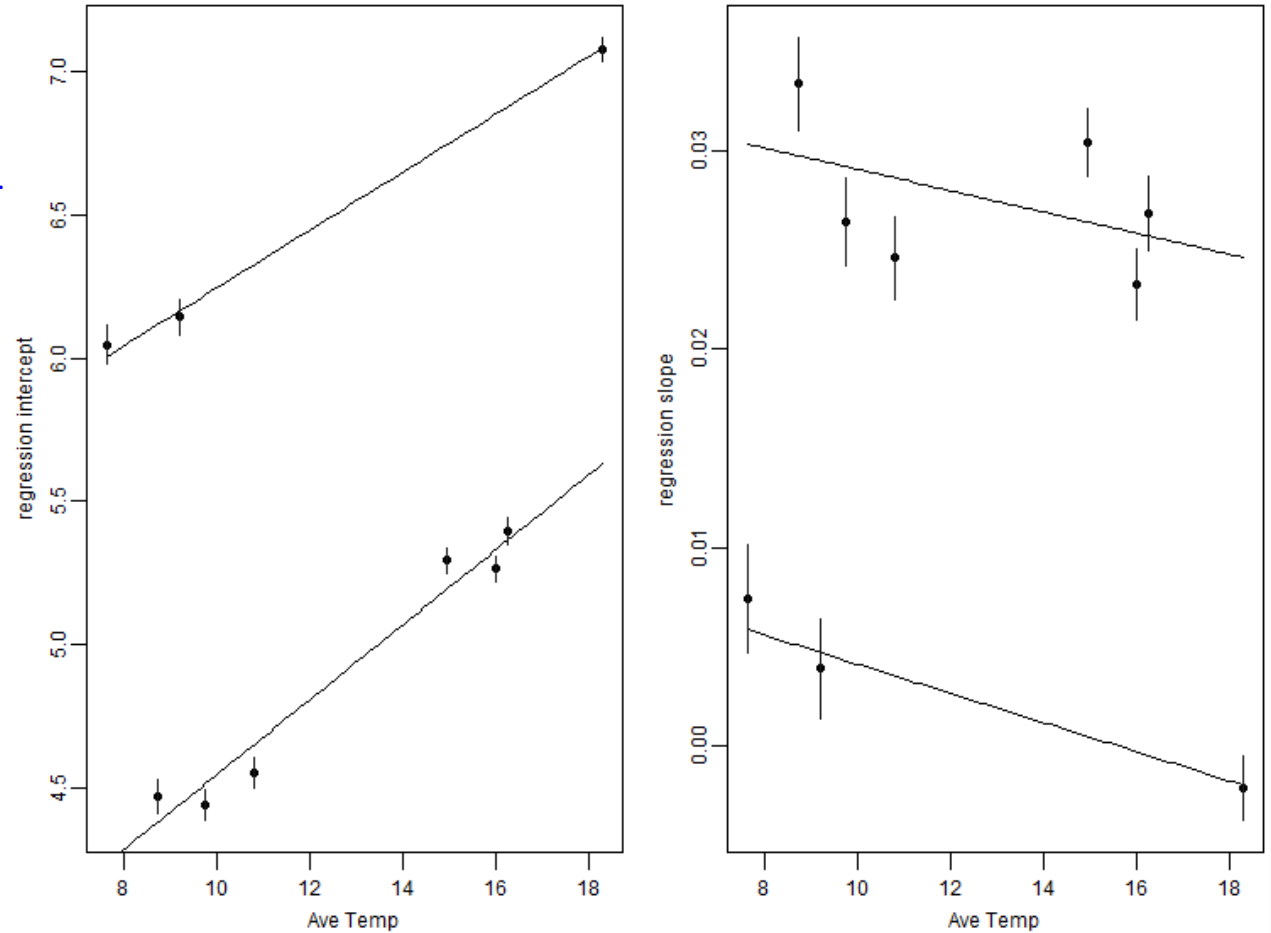
```
euse.lmer2 <- lmer(richtol ~ nuii + ag.full + nuii:ag.full + (1+nuii|site), data=rtol2)
```



Qian, S. S., Cuffney, T. F., Alameddine, I., McMahon, G. & Reckhow, K. H. On the application of multilevel modeling in environmental and ecological studies. *Ecology* **91**, 355-361, doi:<https://doi.org/10.1890/09-1043.1> (2010).

$Y \sim \text{nuii}$ (下层变量)的截距和斜率 都受上层变量 ag.full 的影响

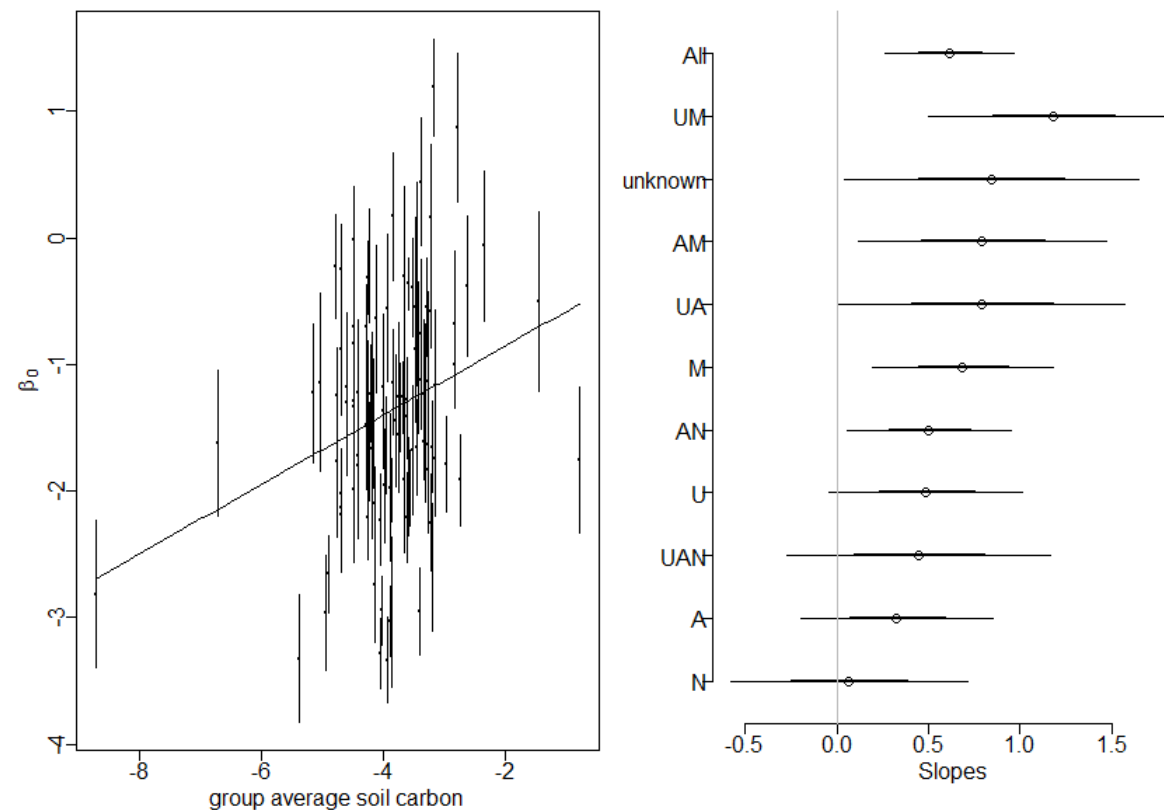
```
#### temp.full为site层上的变量，nuii为下层变量  
#### ag.cat为另一随机效应  
euse.lmer3.5 <- lmer(richtol ~ nuii+temp.full+nuii:temp.full+  
  (1+nuii|site)+  
  (1+nuii+temp.full+nuii:temp.full|ag.cat),  
  data=rtol2)
```



Qian, S. S., Cuffney, T. F., Alameddine, I., McMahon, G. & Reckhow, K. H. On the application of multilevel modeling in environmental and ecological studies. *Ecology* **91**, 355-361, doi:<https://doi.org/10.1890/09-1043.1> (2010).

$Y \sim \text{xx}$ (下层变量)的截距和斜率 都受上层(fertilizer)变量 $\text{logit}(\text{carbon.full})$ 的影响

```
n2o.M3 <- lmer(log(emission/x2) ~ xx + logit(carbon.full) +  
  (xx-1 | fertilizer) + (1 | group),  
  data=N2O.fertilized)
```



Qian, S. S., Cuffney, T. F., Alameddine, I., McMahon, G. & Reckhow, K. H. On the application of multilevel modeling in environmental and ecological studies. *Ecology* **91**, 355-361, doi:<https://doi.org/10.1890/09-1043.1> (2010).

广义线性混合效应模型中的应用案例— Two-level regression mixed effects model

FOREST ECOLOGY

Differential soil fungus accumulation and density dependence of trees in a subtropical forest

Lei Chen¹, Nathan G. Swenson², Niuniu Ji³, Xiangcheng Mi¹, Haibao Ren¹,
Liangdong Guo³, Keping Ma^{1*}

The mechanisms underlying interspecific variation in conspecific negative density dependence (CNDD) are poorly understood. Using a multilevel modeling approach, we combined long-term

Level 1

$$S_{(o)ijy} = \text{Bernoulli}(p_{(o)ijy}) \quad (1)$$

$$\text{logit}(p_{(o)ijy}) = \beta_{0j} + \beta_{1j} \times \text{Neigh}_{(o)ijy} + \beta_2 \times \log(\text{Height}_{(o)ijy}) + \Phi_1 + \Phi_y + v_t + v_{st} \quad (2)$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \times X_j + \pi_{0j} \quad (3)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} \times X_j + \pi_{1j} \quad (4)$$

混合效应模型一

Two-level regression mixed effects model

核心：让下层自变量x1对y的影响的截距和斜率随着上层自变量x2的值的而变化而变化

第一层

$$S_{(o)ijy} = \text{Bernoulli}(p_{(o)ijy}) \quad (1)$$

$$\text{logit}(p_{(o)ijy}) = \beta_{0j} + \beta_{1j} \times \text{Neigh}_{(o)ijy} + \beta_2 \times \log(\text{Height}_{(o)ijy}) + \underbrace{\Phi_i + \Phi_y + v_t + v_{st}}_{(1|\text{tree_id})+(1|\text{year})+(1|\text{station/plot})} \quad (2)$$

↑ ↑ 固定效应 固定效应 随机效应

At the second level, both species-level random intercept (β_{0j}) and slope (β_{1j}) from Eq. 2 was modeled as a function of each species-level predictor X_j

第二层

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \times X_j + \pi_{0j} \quad (3)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} \times X_j + \pi_{1j} \quad (4)$$

X_j 是一个物种水平上的因素

neighbor对y的影响的斜率和截距，都会受到另一个物种水平的变量，x的影响

Model < glmer(y ~ neighbor + X + neighbor : X + height + (1|tree_id) + (1|year) + (1|station/plot) + (neighbor | species), family = binomial)

x对y~neighbor截距的影响

x对y~neighbor斜率的影响

结果展示

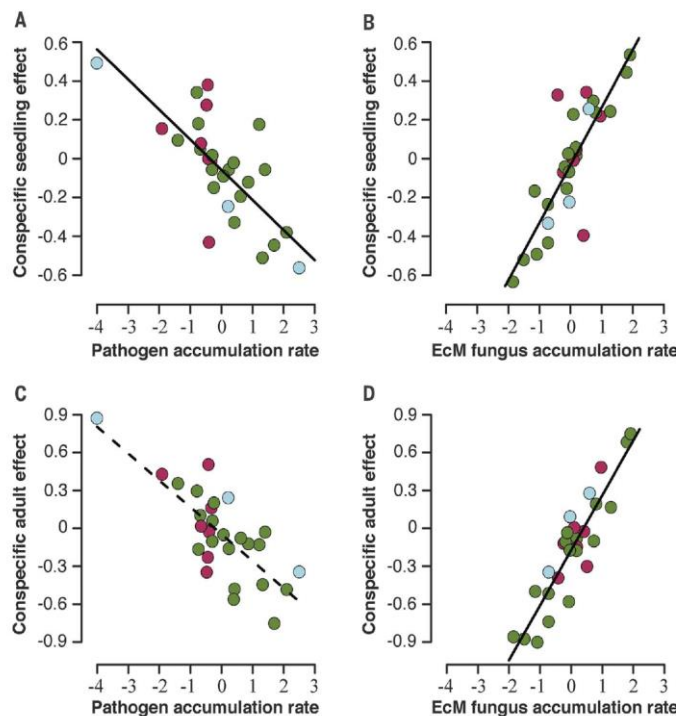


Fig. 3. Relationships between CNDD and pathogenic and EcM fungus accumulation rates over tree ontogeny. (A to D) The overall relationships fitted by two-level mixed models suggest that increasing pathogenic fungal density caused lower seedling survival, but increased EcM fungal density favored seedling survival by means of density-dependent effects. AM tree species are in green, EcM tree species are in red, and ErM tree species are in blue ($n = 28$ tree species). The black solid and dashed lines indicate that the fitted regressions were significant at $P < 0.05$ or not statistically significant, respectively. Parameter estimates are available in table S4.

##上层变量X

下层变量（物种层）

Y~Neighbor回归的截距和斜率都受到X的影响

```
Model<-glmer(y~neighbor+X+neighbor:X+height+(1|tree_id)+(1|year)+(1|station/plot)+(neighbor|species), family=binomial)
```

混合模型x来自不同层级的另一杰出案例

ECOLOGY LETTERS

Ecology Letters, (2019) 22: 1587–1598

doi: 10.1111/ele.13346

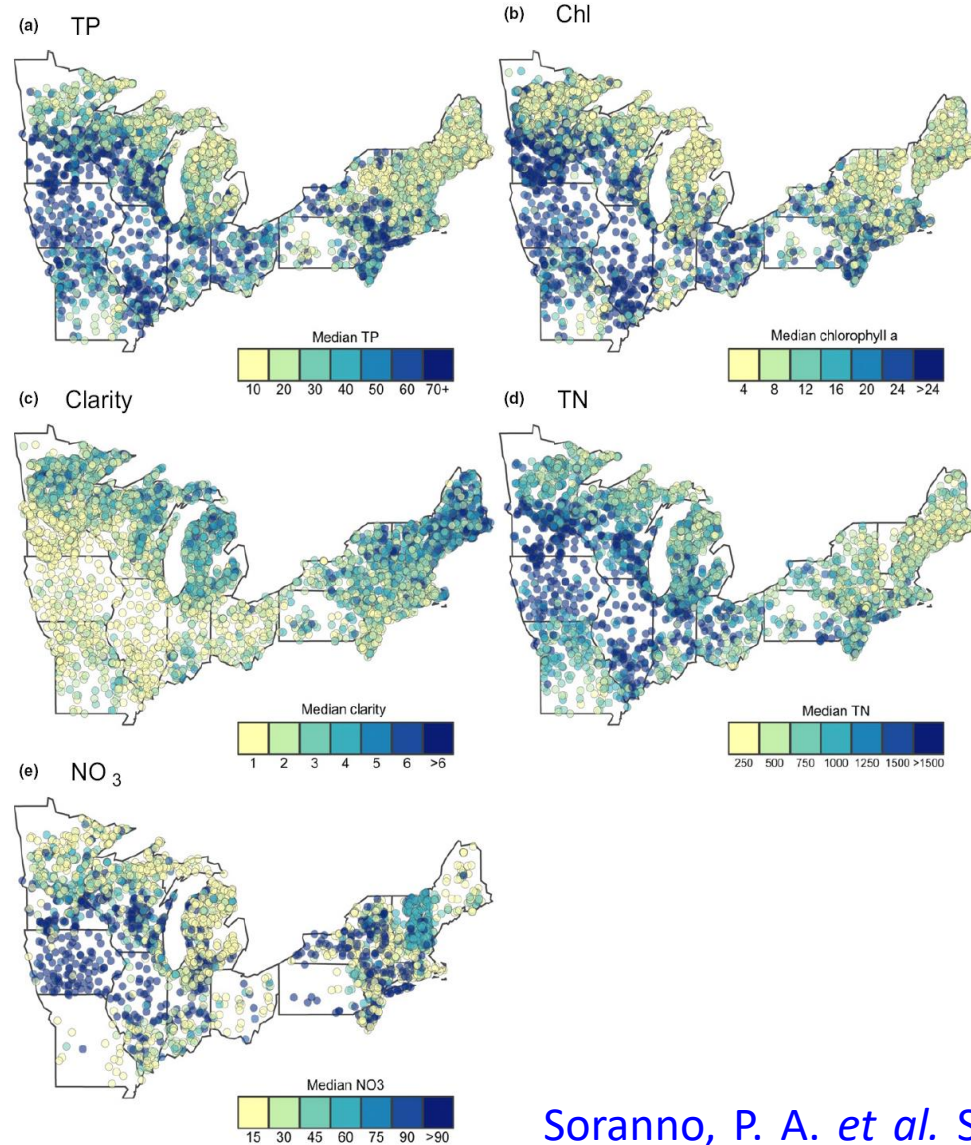
LETTER

Spatial and temporal variation of ecosystem properties at macroscales

We used a Bayesian hierarchical model to decompose the total variation in each of the five ecosystem properties into their four components, as well as a residual.

Soranno, P. A. *et al.* Spatial and temporal variation of ecosystem properties at macroscales. *Ecol. Lett.* **22**, 1587–1598, doi:<https://doi.org/10.1111/ele.13346> (2019).

分析对象(因变量)Y



The number of lakes with any measurement for one of these five properties range from 3560 (NO₃) to 7601 (clarity) lakes

Soranno, P. A. *et al.* Spatial and temporal variation of ecosystem properties at macroscales. *Ecol. Lett.* **22**, 1587-1598, doi:<https://doi.org/10.1111/ele.13346> (2019).

Figure 1 The median values by lake for all response variables, including TP (a), Chl (b), clarity (c), TN (d), and NO₃ (e).

自变量X的层级性

Table 2 Hypothesised ecological drivers of the four variance components

Variance component	Category	Predictor (units)
Regional层	Spatial, regional	Climate, normals
	Climate, normals	Precipitation, 30-year normal (mm)
	Atmospheric deposition	Temperature, 30-year normal (°C)
	Atmospheric deposition	N deposition in 1990, region (mean, kg ha ⁻¹)
	Atmospheric deposition	N deposition in 2010, region (mean, kg ha ⁻¹)
	Freshwater, hydrology	N deposition difference (1990–2010), region (kg ha ⁻¹)
	Freshwater, hydrology	Runoff, region (mean, mm year ⁻¹)
	Freshwater, hydrology	Baseflow, region (mean, %)
	Freshwater, hydrology	Lake area, region (mean, ha)
	Freshwater, hydrology	Lake area-isolated and headwater lakes, region (%)
	Freshwater, hydrology	Wetland area, region (%)
	Land use/cover	Urban area, region (%)
	Land use/cover	Agricultural area, region (%)
	Land use/cover	Forest area, region (%)
	Land use/cover	Lake depth, maximum (m)
Local层	Spatial, local	Lake area (ha)
	Freshwater, hydrology	Watershed area: lake area ratio
	Freshwater, hydrology	Stream density, watershed (km ha ⁻¹)
	Freshwater, hydrology	Wetland-woody, watershed (%)
	Freshwater, hydrology	Wetland-emergent, watershed (%)
	Freshwater, hydrology	Urban, watershed (%)
	Land use/cover	Agriculture-total, watershed (%)
	Land use/cover	Agriculture-row crop, watershed (%)
	Land use/cover	Agriculture-pasture, watershed (%)
	Land use/cover	Forest-total, watershed (%)
	Land use/cover	Forest-deciduous, watershed (%)
	Land use/cover	Forest-coniferous, watershed (%)
	Land use/cover	Forest-mixed, watershed (%)
	Land use/cover	Road density, watershed (km ha ⁻¹)
	Land use/cover	
Coherent层	Temporal, coherent	Climate, seasonal
	Climate, seasonal	Precipitation, winter and spring (mm)
	Climate, seasonal	Temperature, mean of summer months (°C)
	Climate, seasonal	Palmer hydrologic drought index, spring (unitless)
	Climate, monthly	Precipitation, previous November (mm)
	Climate, monthly	Precipitation, previous December (mm)
	Climate, monthly	Precipitation, previous January (mm)
	Climate, monthly	Precipitation, previous winter (mm)
	Climate, monthly	Precipitation, previous May (mm)
	Climate, monthly	Temperature, previous May (°C)
	Climate, monthly	Temperature, June (°C)
	Climate, monthly	Palmer hydrologic drought index, spring (unitless)
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
Local层	Temporal, local	Climate, monthly
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	
	Climate, monthly	

与空间相关的X

与时间相关的X

增加Letters案例

公式1

$$y_{ijkl} = \beta_0 + \mu_{1j} + \mu_{2k} + \mu_{3jk} + \mu_{4l} + \epsilon_{ijkl}$$

$$\epsilon_{ijkl} \sim N(0, \sigma_{\epsilon}^2)$$

$$\mu_{1j} \sim N(0, \sigma_{local}^2)$$

$$\mu_{2k} \sim N(0, \sigma_{coherent\ temporal}^2)$$

$$\mu_{3jk} \sim N(0, \sigma_{local\ temporal}^2)$$

$$\mu_{4l} \sim N(0, \sigma_{region}^2),$$

公式2

$$y_{ijkl} = \beta_0 + \mu_{1j} + \mu_{2k} + \mu_{3jk} + \mu_{4l} + \epsilon_{ijkl}$$

$$\epsilon_{ijkl} \sim N(0, \sigma_{\epsilon}^2)$$

$$\mu_{1j} \sim N(\beta_1 \cdot X_{1j} + \dots \beta_p \cdot X_{pj}, \sigma_{local}^2)$$

$$\mu_{2k} \sim N(\alpha_1 \cdot X_{1k} + \dots \alpha_n \cdot X_{nk}, \sigma_{coherent\ temporal}^2)$$

$$\mu_{3jk} \sim N(\gamma_1 \cdot X_{1jk} + \dots \gamma_m \cdot X_{mjk}, \sigma_{local\ temporal}^2)$$

$$\mu_{4l} \sim N(\psi_1 \cdot X_{1l} + \dots \psi_w \cdot X_{wl}, \sigma_{region}^2),$$

We used a Bayesian hierarchical model to decompose the total variation in each of the five ecosystem properties into their four components, as well as a residual.

$Y \sim 1 + (1 | local) + (1 | coh_temp) + (1 | local_temp) + (1 | region)$

$Y \sim 1 + X_1 + X_2 + \dots + X_n + (1 | local) + (1 | coh_temp) + (1 | local_temp) + (1 | region)$

各随机效应的R2计算

$$y_{ijkl} = \beta_0 + \mu_{1j} + \mu_{2k} + \mu_{3jk} + \mu_{4l} + \epsilon_{ijkl}$$

$$\epsilon_{ijkl} \sim N(0, \sigma_{\epsilon}^2)$$

$$\mu_{1j} \sim N(0, \sigma_{local}^2)$$

$$\mu_{2k} \sim N(0, \sigma_{coherent\ temporal}^2)$$

$$\mu_{3jk} \sim N(0, \sigma_{local\ temporal}^2)$$

$$\mu_{4l} \sim N(0, \sigma_{region}^2),$$

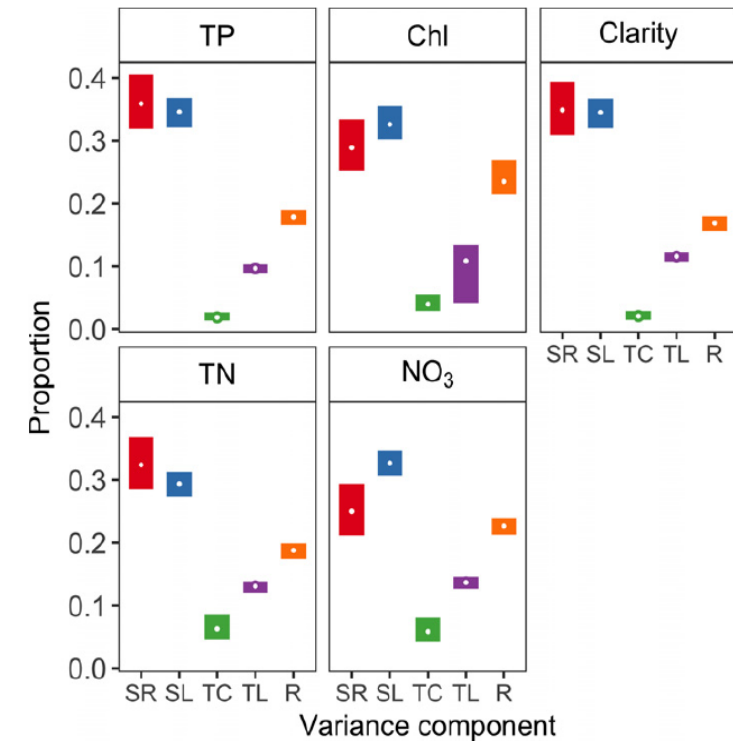


Figure 2 The proportion of total variation for each of the variance components of each of the ecological properties estimated from the unconditional variance component models. The white dots are posterior means and the rectangles are the 95% credible intervals. SR = spatial, regional; SL = spatial, local; TC = temporal, coherent; TL = temporal, local; R = residual. Colours are as for Fig. 3.

Soranno, P. A. *et al.* Spatial and temporal variation of ecosystem properties at macroscales. *Ecol. Lett.* **22**, 1587-1598, doi:<https://doi.org/10.1111/ele.13346> (2019).

不同的层级上的R2

$$R^2_{Spatial_local} = \frac{\sigma^2_{local(eqn_1)} - \sigma^2_{local(eqn_2)}}{\sigma^2_{local(eqn_1)}}$$

Table 3 The percentage of variation explained (approximate R^2) for each spatial and temporal component by the driver variables (posterior mean, followed by 95% credible intervals in parentheses). Negative % variance explained is possible with these models and they reflect poor predictive capability and are reported as zero. The estimated total variation is calculated by weighting the approximate R^2 values of each variance components by its proportion of total variance from Fig. 2, and assuming the negative approximate R^2 values are simply zero. For clarity, coherent temporal covariates were not included in the model due to convergence issues, so values of NA are reported

Variance component	Approximate R^2				
	TP	Chl	Clarity	TN	NO ₃
Spatial, regional, %	86 (77, 93)	81 (67, 90)	90 (81, 95)	94 (90, 97)	85 (68, 93)
Spatial, local, %	51 (44, 51)	38 (34, 42)	51 (48, 54)	48 (43, 52)	32 (26, 39)
Temporal, coherent, %	0	0	NA	20 (-115, 70)	0
Temporal, local, %	0	0	0	0	5 (-7, 16)
Estimated total variation, %:	49	36	49	39	33

Soranno, P. A. *et al.* Spatial and temporal variation of ecosystem properties at macroscales. *Ecol. Lett.* **22**, 1587-1598, doi:<https://doi.org/10.1111/ele.13346> (2019).

各个X对Y的影响显著性

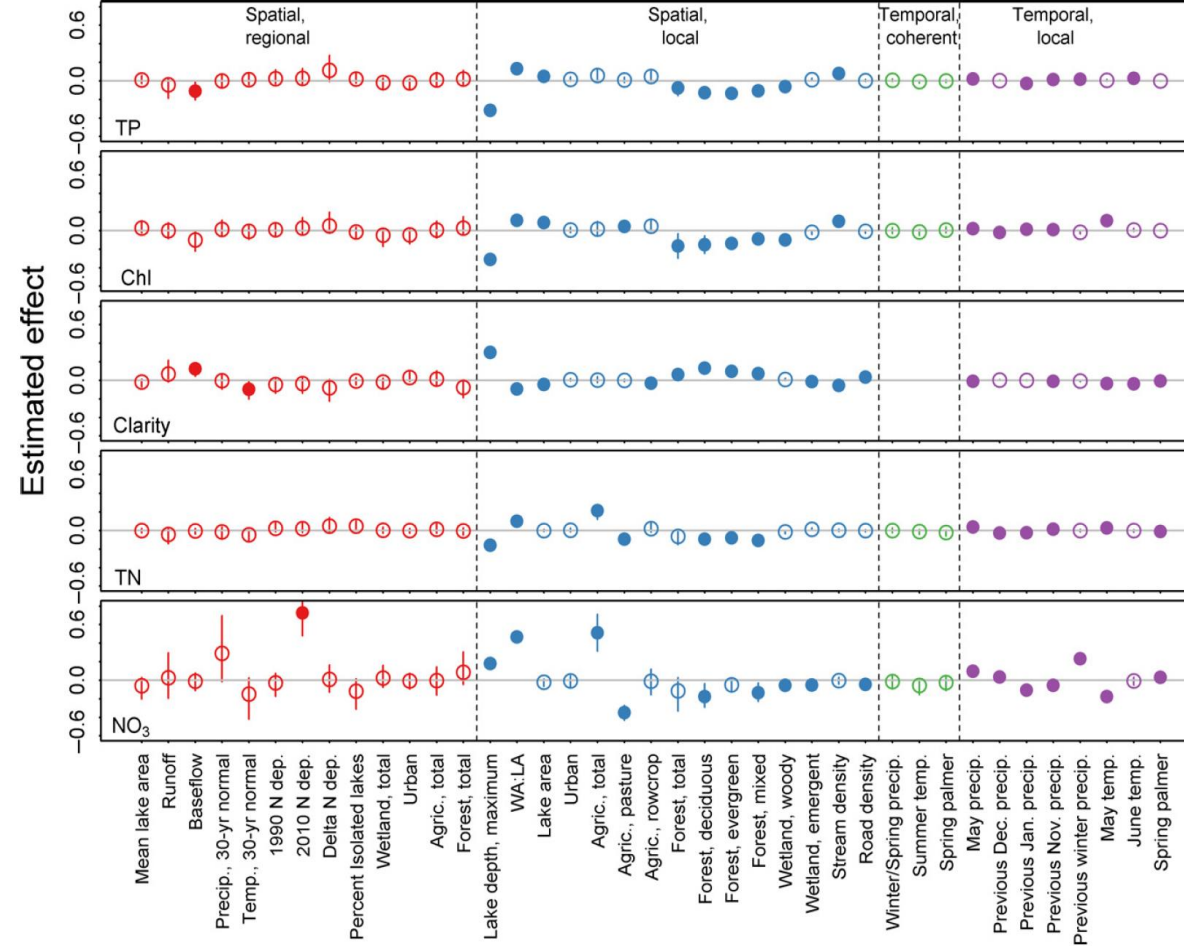


Figure 3 The estimated effects of driver variables on each of the four components of variation for five response variables. Filled symbols represent effect sizes whose 90% credible intervals (error bars) do not overlap zero. Symbols with very small effect sizes, but that are filled, do not overlap zero and only appear to do so due to the size of the symbol. Colours as are for Fig. 2.

Soranno, P. A. *et al.* Spatial and temporal variation of ecosystem properties at macroscales. *Ecol. Lett.* **22**, 1587-1598, doi:<https://doi.org/10.1111/ele.13346> (2019).

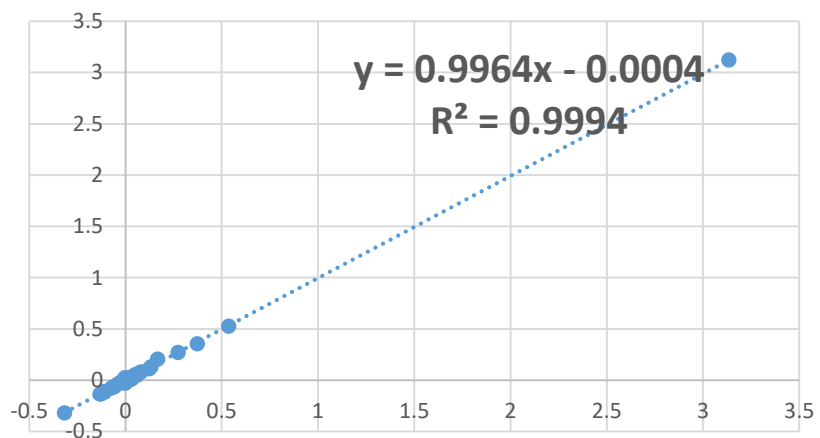
原模型采用JAGS(贝叶斯法)分析， 用stan, brms等最新贝叶斯工具同样轻松搞定

Three parallel Markov chains were run, each starting with a random value. Each was run for 80 000 iterations, discarding the first 60 000 samples. The remaining 60 000 samples were used to summarise posterior distributions. All models were fitted using the program JAGS (Plummer 2003) using the jags UI function (Kellner 2017) called from within the program R (R Core Team 2018), and are available at Wagner (2019). We assessed convergence using the Brooks–Gelman–Rubin statistic (Brooks & Gelman 1998) and trace plots.

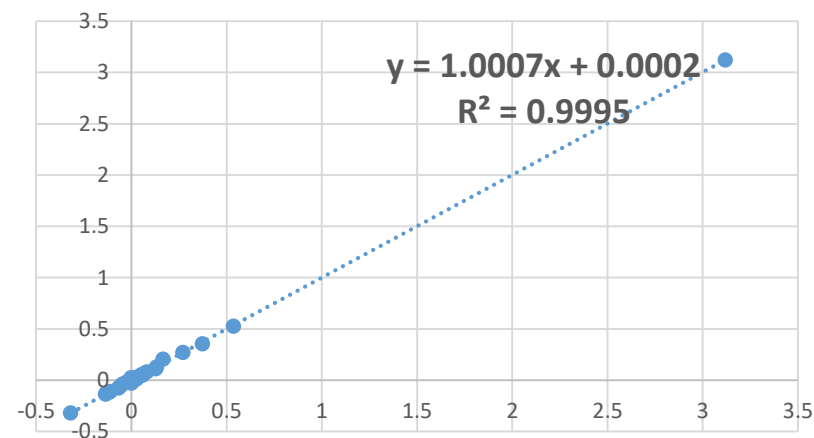
Soranno, P. A. *et al.* Spatial and temporal variation of ecosystem properties at macroscales. *Ecol. Lett.* **22**, 1587-1598, doi:<https://doi.org/10.1111/ele.13346> (2019).

不同贝叶斯分析工具结果对比

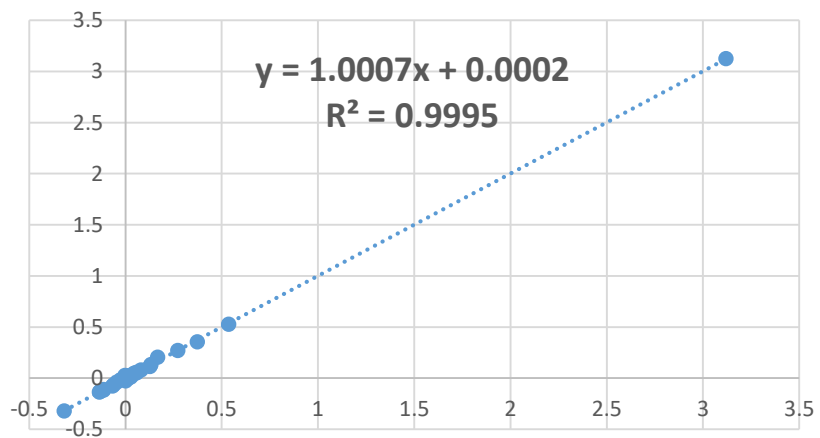
jags vs brms



jags vs stan



stan vs brms



混合效应模型--模型选择

Loglikelihood ratio test

- 多个因素 (x) 可能对因变量 y 有影响, 哪些因素最为重要?
- 首先排除共线性问题, 相关系数大于0.7的变量只保留一个
- 排除自变量之间的共线性问题之后, 可以先对 x 进行标准化 (scale命令), 然后可以直接比较回归参数的大小。
- 模型选择方法: LRT(x 个数较少时), AIC (x 个数较多时)

LMM模型对比-- Loglikelihood ratio test

```
m0 <- lm(Reaction~Days,sleepstudy)
```

```
m1 <- lmer(Reaction~Days+(1|Subject)+(0+Days|Subject),sleepstudy,REML=FALSE)
```

```
m2 <- update(m1,~Days+(1|Subject))
```

```
> anova(m2,m1,m0) ## two sequential tests
```

```
Data: sleepstudy
```

```
Models:
```

```
m0: Reaction ~ Days
```

```
m2: Reaction ~ Days + (1 | Subject)
```

```
m1: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
m0	3	1906.3	1915.9	-950.15	1900.3				
m2	4	1802.1	1814.8	-897.04	1794.1	106.214		1	< 2.2e-16 ***
m1	5	1762.0	1778.0	-876.00	1752.0	42.075		1	8.782e-11 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

混合效应模型选择案例（制图佳作）

ECOLOGY LETTERS

Ecology Letters, (2020) **23**: 79–87

doi: 10.1111/ele.13411

LETTER

Complementarity effects are strengthened by competition intensity and global environmental change in the central boreal forests of Canada

Eric B. Searle¹  and
Han Y. H. Chen^{1,2*} 

Abstract

Increases in niche complementarity have been hypothesised to reduce the intensity of interspecific competition within natural forests. In regions currently experiencing potentially enhanced growth under global environmental change, niche complementarity may become even more beneficial.

混合效应模型—基于AIC的模型选择

```
library(MuMIn)
```

```
##构建全模型
```

```
fullmodel<=lmer(lnG_BA~mod_logH12+mod_Year+mod_PSV12+mod_Fdis12+  
mod_Age+ (1|PLOT_ID/Phylo), grow)
```

```
summary(fullmodel)
```

```
##注意自变量x1,x2...不能包含空值
```

```
##设置自变量包含空值时，模型选择失败
```

```
options(na.action = na.fail)
```

案例数据来源 Searle, E. B. and H. Y. H. Chen. 2020. Complementarity effects are strengthened by competition intensity and global environmental change in the central boreal forests of Canada. Ecology Letters **23**:79-87.

混合效应模型—基于AIC的模型选择

##模型选择

```
model_selection <- dredge(fullmodel, evaluate = TRUE, rank = "AICc",REML=F)
```

```
model_selection
```

##与AICc值最低的模型差距在2以内的最优模型集合

```
subset(model_selection, delta<2)
```

##将最优模型集合内的参数估计按模型权重进行平均

```
best_est<-model.avg(model_selection,subset= delta <2, revised.var = TRUE)
```

```
summary(best_est)
```

##求出每个因素在最优模型中所占的重要性，以0.8为界限

```
importance(best_est)
```

混合效应模型—基于AIC的模型选择

```
Global model call: lmer(formula = lnG_BA ~ mod_logH12 + mod_Year + mod_PSV12 + mod_Fdis12 +
  mod_Age + (1 | PLOT_ID/Phylo), data = grow)
```

Model selection table

	(Int)	mod_Age	mod_F12	mod_lH1	mod_PSV	mod_Yer	df	logLik	AICc	delta
16	-0.5896000	-0.07818	1.373e-01	-1.466	-0.03716		8	-148471.8	296922.4	0.00
32	-0.5928000	-0.06551	1.370e-01	-1.466	-0.03692	-0.004495	9	-148475.8	296923.8	1.45
31	-0.6095000		1.371e-01	-1.465	-0.03583	-0.023170	8	-148478.8	296934.2	11.81
8	-0.6158000	-0.07792	1.247e-01	-1.465			7	-148475.7	296935.5	13.13
24	-0.6193000	-0.06298	1.245e-01	-1.465		-0.005310	8	-148479.5	296936.7	14.35
23	-0.6343000		1.249e-01	-1.465		-0.023250	7	-148482.1	296946.2	23.81
15	-0.6104000		1.438e-01	-1.462	-0.03650		7	-148512.6	297009.9	87.50
7	-0.6355000		1.315e-01	-1.461			6	-148516.2	297022.5	100.11
30	-0.5836000	-0.06406		-1.464	0.02564	-0.009073	8	-148609.9	297197.9	275.49
14	-0.5769000	-0.08953		-1.464	0.02533		7	-148606.9	297198.2	275.84
22	-0.5600000	-0.06616		-1.464		-0.008705	7	-148610.2	297204.1	281.74
6	-0.5537000	-0.09060		-1.464			6	-148607.1	297204.3	281.91
29	-0.6006000			-1.464	0.02687	-0.027320	7	-148612.8	297207.9	285.51
21	-0.5768000			-1.464		-0.027570	6	-148613.5	297214.9	292.57
13	-0.6015000			-1.459	0.02972		6	-148662.2	297314.7	392.38
5	-0.5752000			-1.459			5	-148663.9	297323.9	401.51
28	0.0143300	0.23830	2.546e-02		0.02506	0.023250	8	-199744.1	399471.2	102548.86
20	0.0339000	0.23580	3.423e-02			0.024050	7	-199742.3	399472.3	102549.92
26	0.0142100	0.23890			0.03753	0.022280	7	-199742.8	399473.6	102551.28
12	-0.0009446	0.30150	2.352e-02		0.02737		7	-199744.3	399477.4	102555.04
4	0.0199300	0.30120	3.306e-02				6	-199742.8	399479.1	102556.70
10	-0.0004989	0.29970			0.03884		6	-199742.7	399479.2	102556.81
18	0.0494700	0.23500				0.023110	6	-199743.5	399479.9	102557.56
2	0.0354900	0.29800					5	-199743.7	399486.1	102563.68
19	0.0889400		3.314e-02			0.091200	6	-199783.3	399557.9	102635.52
27	0.0754800		2.679e-02		0.01808	0.091120	7	-199785.8	399558.3	102635.93
25	0.0754900				0.03119	0.090320	6	-199784.8	399561.2	102638.82
17	0.1035000					0.090070	5	-199784.2	399564.9	102642.55
9	0.0835200				0.02297		5	-200054.0	400106.8	103184.43
1	0.1043000						4	-200053.1	400107.0	103185.54

模型总数= $2^k=2^5=32$

混合效应模型—基于AIC的模型选择

只显示最优模型的集合

```
> subset(model_selection, delta<2)
Global model call: lmer(formula = lnG_BA ~ mod_logH12 + mod_Year + mod_PSV12 + mod_Fdis12 +
  mod_Age + (1 | PLOT_ID/Phylo), data = grow)
---
Model selection table
      (Int)  mod_Age mod_F12 mod_lH1  mod_PSV   mod_Yer df    logLik      AICc delta weight
16 -0.5896 -0.07818  0.1373  -1.466 -0.03716      8 -148471.8 296922.4  0.00  0.674
32 -0.5928 -0.06551  0.1370  -1.466 -0.03692 -0.004495  9 -148475.8 296923.8  1.45  0.326
Models ranked by AICc(x, REML = FALSE)
Random terms (all models):
'1 | PLOT_ID/Phylo'
```

混合效应模型—基于AIC的模型选择

##基于最优模型集的参数估计

```
best_est<-model.avg(model_selection,subset= delta <2, revised.var = TRUE)
```

```
summary(best_est)
```

```
Call:
model.avg(object = model_selection, subset = delta < 2, r

Component model call:
lmer(formula = lnG_BA ~ <2 unique rhs>, data = grow)

Component models:
      df  logLik    AICc delta weight
1234   8 -148471.8 296922.4  0.00  0.67
12345  9 -148475.8 296923.8  1.45  0.33

Term codes:
      mod_Age mod_Fdis12 mod_logH12 mod_PSV12 mod_Year
           1           2           3           4           5
```

Model-averaged coefficients:
(full average)

	Estimate	Std. Error	Adjusted SE	z value	Pr(> z)	
(Intercept)	-0.590651	0.038958	0.038959	15.161	< 2e-16	***
mod_Age	-0.074046	0.013498	0.013498	5.486	< 2e-16	***
mod_Fdis12	0.137203	0.008124	0.008124	16.888	< 2e-16	***
mod_logH12	-1.465553	0.003708	0.003708	395.200	< 2e-16	***
mod_PSV12	-0.037081	0.009500	0.009500	3.903	9.49e-05	***
mod_Year	-0.001466	0.003862	0.003862	0.380	0.704	

(conditional average)

	Estimate	Std. Error	Adjusted SE	z value	Pr(> z)	
(Intercept)	-0.590651	0.038958	0.038959	15.161	< 2e-16	***
mod_Age	-0.074046	0.013498	0.013498	5.486	< 2e-16	***
mod_Fdis12	0.137203	0.008124	0.008124	16.888	< 2e-16	***
mod_logH12	-1.465553	0.003708	0.003708	395.200	< 2e-16	***
mod_PSV12	-0.037081	0.009500	0.009500	3.903	9.49e-05	***
mod_Year	-0.004495	0.005667	0.005667	0.793	0.428	

混合效应模型—基于AIC的模型选择

`importance(best_est)` ##最优模型中各个自变量的重要性（通常以0.8为阈值）

```
> importance(best_est)
```

	mod_Age	mod_Fdis12	mod_logH12	mod_PSV12	mod_Year
Importance:	1.00	1.00	1.00	1.00	0.33
N containing models:	2	2	2	2	1

Component models:

	df	logLik	AICc	delta	weight
1234	8	-148471.8	296922.4	0.00	0.67
12345	9	-148475.8	296923.8	1.45	0.33

The importance of mod_Year = $(0.33)/1 = 0.33$

Term codes:

mod_Age	mod_Fdis12	mod_logH12	mod_PSV12	mod_Year
1	2	3	4	5

模型选择方法的描述案例

For each model, a multimodel inference procedure was applied using the R MuMIn package. This method allowed us to perform model selection by creating a set of models with all possible combinations of the initial variables and sorting them according to the Akaike Information Criterion (AIC) fitted with Maximum Likelihood. We selected all models with $\Delta AIC < 2$ and used the model averaging approach using LMER to estimate parameters and associated P values, using the function `model.avg`.

C. Sirami *et al.*, Increasing crop heterogeneity enhances multitrophic diversity across agricultural regions. *Proc. Natl. Acad. Sci. USA*, 201906419 (2019).

多变量回归及作图问题

```
best_m<-lmer(lnG_BA~mod_logH12+mod_PSV12+mod_Age+mod_Fdis12+  
(1|PLOT_ID/Phylo), grow)
```

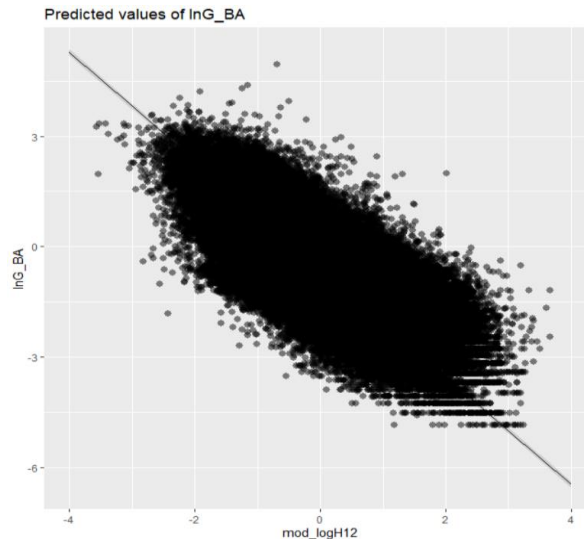
可调用专门的服务于模型作图的软件包来操作

```
library(sjPlot)
```

```
pd<-plot_model(best_m, type = "eff", terms = "mod_logH12", show.data=T)
```

###也可以把用来做拟合线的数据单独提取出来

```
dd<-get_model_data(best_m, type = "eff", terms = "mod_logH12")
```



也可以只现实与某单一因素作用有关的点 (控制其他因素后的数据点)

```
full_ran_mod<-lmer(lnG_BA~mod_logH12*mod_Year*(mod_PSV12+mod_Fdis12)+mod_Age+  
  (1|PLOT_ID/Phylo)+(1|T_N),grow)
```

###1, 先把模型信息转化为list格式

```
mod_list<-list(full_ran_mod)
```

#####2, 再依据模型对y值进行boot模拟

```
boot_list<-lply(mod_list,.fun = function(x)bootMer(x,fixef,10), .progress="win")
```

###3, 调用事先写好的useful_functions函数

```
source("useful_functions.R")##4, 运行此函数之后，即可产生拟合线数据和控制其他因素后的数据点
```

####拟合线数据

```
frame_logh<-cont_single_gg(data=grow, parm="logH12",coeff = "mod_logH12",  
  boot=boot_list[[1]],scale=T)
```

###背景数据点数据

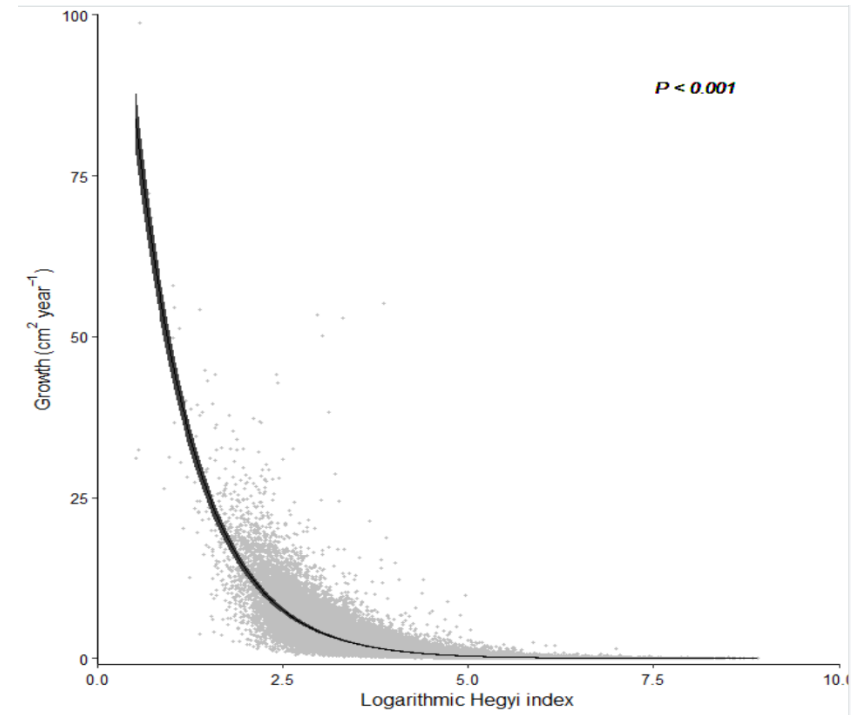
```
bg_logh<-cont_single_pts(data=grow, parm="logH12",coeff = "mod_logH12",  
  mod=mod_list[[1]])
```

(此案例代码来源 Searle and Chen 2020, Ecology Letters 23:79-87.)

也可以只现实与某单一因素作用有关的点 (控制其他因素后的数据点)

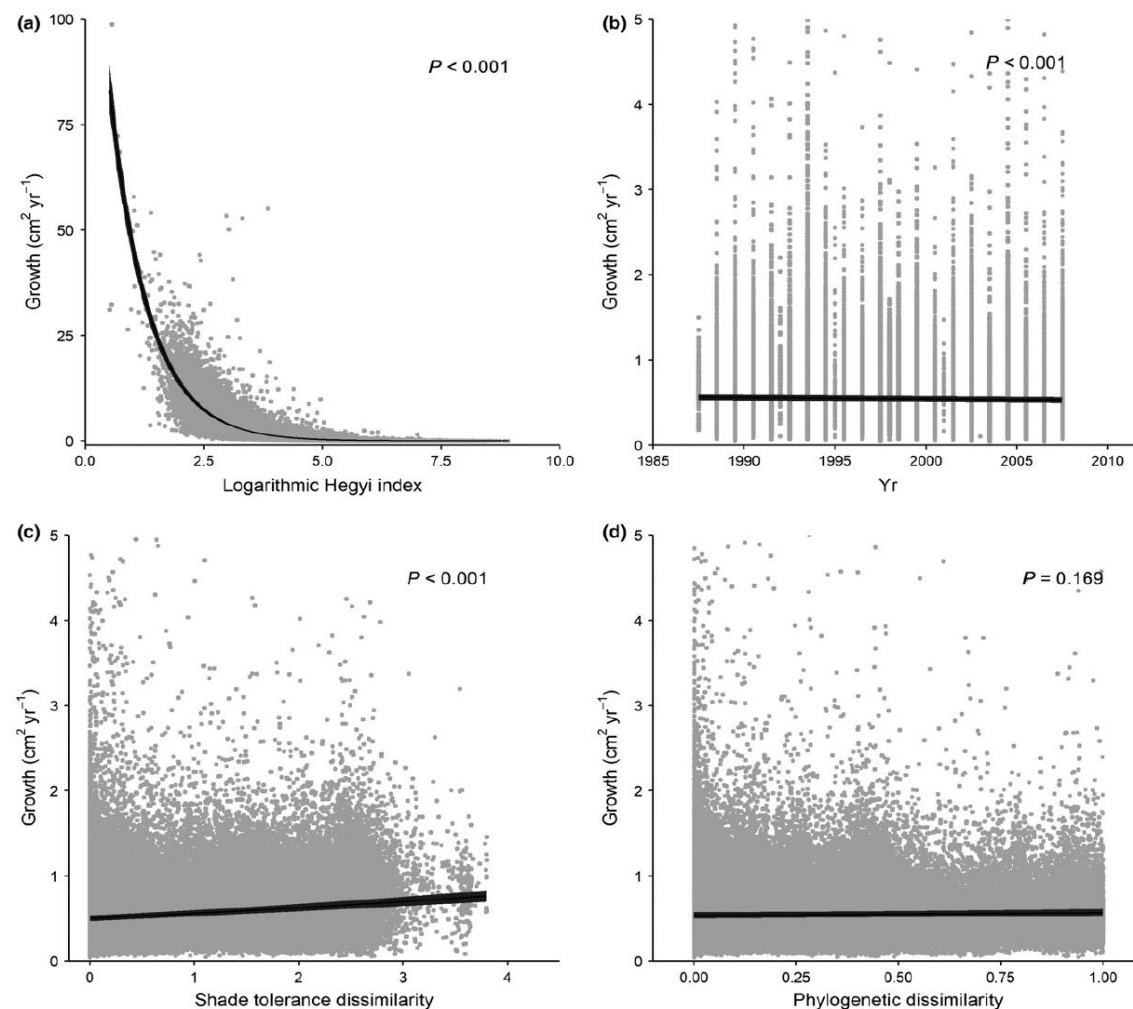
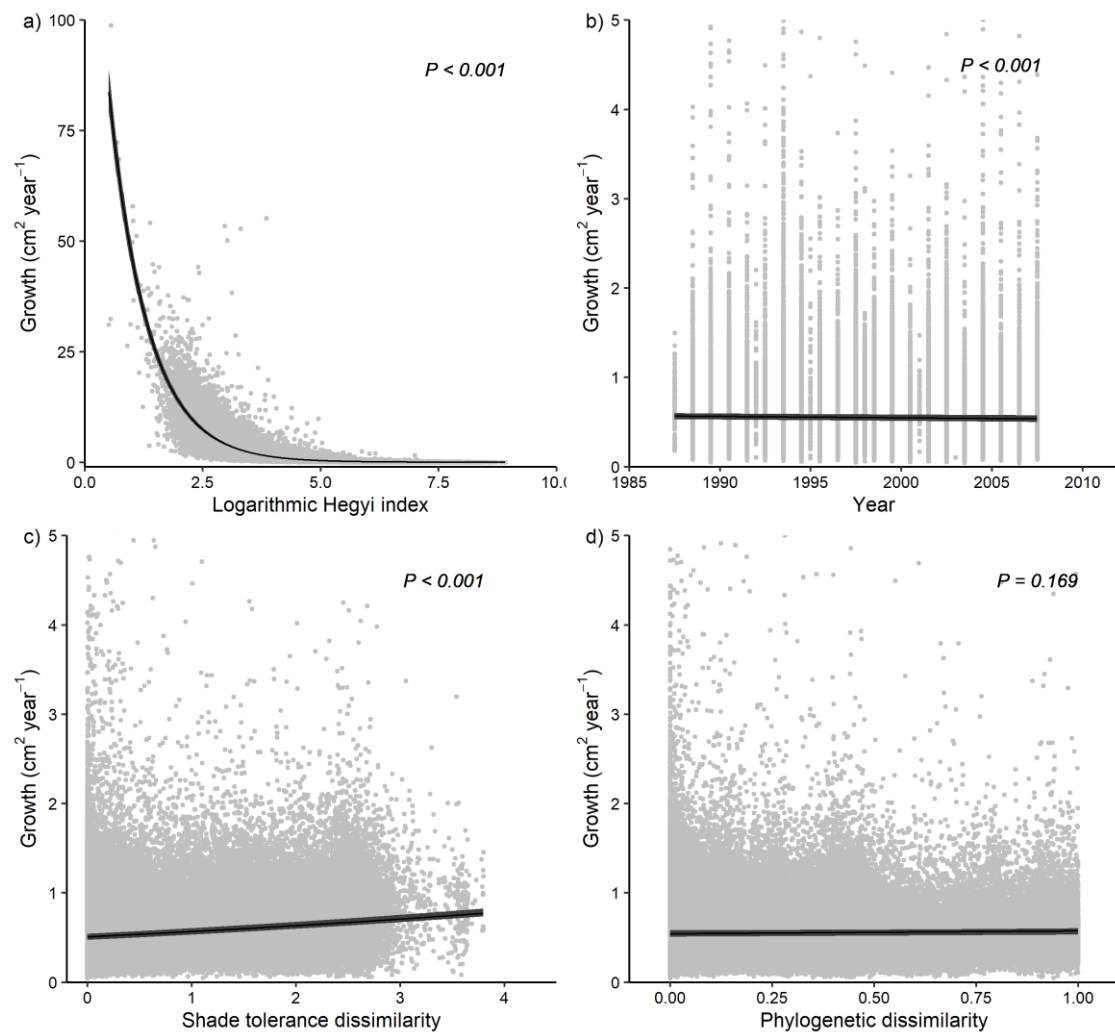
###最后依据前面产生的frame_logh和bg_logh这两个数据集，用ggplot2对结果进行做图

```
comp_gg<-ggplot()+  
  geom_point(data=bg_logh,aes(x=xaxis,y=exp(bg)),col='gray75',size=0.8)+  
  geom_ribbon(data=frame_logh,aes(x=xaxis,ymin=exp(low),ymax=exp(up)),  
    col=NA,fill="gray25")+  
  geom_line(data=frame_logh,aes(x=xaxis,y=exp(mean)),col="black")+  
  geom_text(data=bg_logh,aes(x=max(xaxis)-((max(xaxis)-min(xaxis))*0.10),  
    y=max(exp(bg))-((max(exp(bg))-min(exp(bg)))*0.10)),  
    label="P < 0.001",fontface='italic')+  
  theme_classic()+  
  xlab("Logarithmic Hegyi index")+  
  ylab(expression(Growth~('cm^2~year^{-1}')))+  
  theme(strip.background = element_blank(),  
    strip.placement="outside",  
    legend.position=c(0.9,0.5),  
    axis.text=element_text(colour="black"))+  
  coord_cartesian(ylim=c(-1,100),xlim=c(0,10),  
    expand=FALSE)
```



(此案例代码来源 Searle and Chen 2020, Ecology Letters **23**:79-87.)

对模型中的每一个自变量都重复上述过程即可得到各自变量独立对 y 的影响结果图



nlme模型选择案例

#Now run all combinations of models, including all interactions up to order 2.

#These models must initially be fit via maximum likelihood (ML) given different fixed effects.

#Fit using growing season values over 1901-1963

```
iWUE.m2 <- dredge(lme(iWUE.micromol.mol ~ Ca.cn + GRW.PPT.mm.cn + GRW.TMP.C.cn + GRW.VPD.kPa.cn +  
  Ca.cn:GRW.PPT.mm.cn + Ca.cn:GRW.TMP.C.cn + Ca.cn:GRW.VPD.kPa.cn +  
  GRW.PPT.mm.cn:GRW.TMP.C.cn + GRW.PPT.mm.cn:GRW.VPD.kPa.cn +  
  GRW.TMP.C.cn:GRW.VPD.kPa.cn,  
  data = meta.CRU.pre1963,  
  method = "ML",  
  random = ~1|Site/LatinBinomial,  
  correlation = corAR1()),  
  rank = "AICc")
```

注意，这里直接以AIC值最低的模型为最优模型

#Out of all resulting models, retrieve best models (lowest AICc)

```
iWUE.M2 <- model.avg(get.models(iWUE.m2, subset = delta < 2))
```

```
iWUE.M2a <- get.models(iWUE.m2, subset = delta == 0)
```

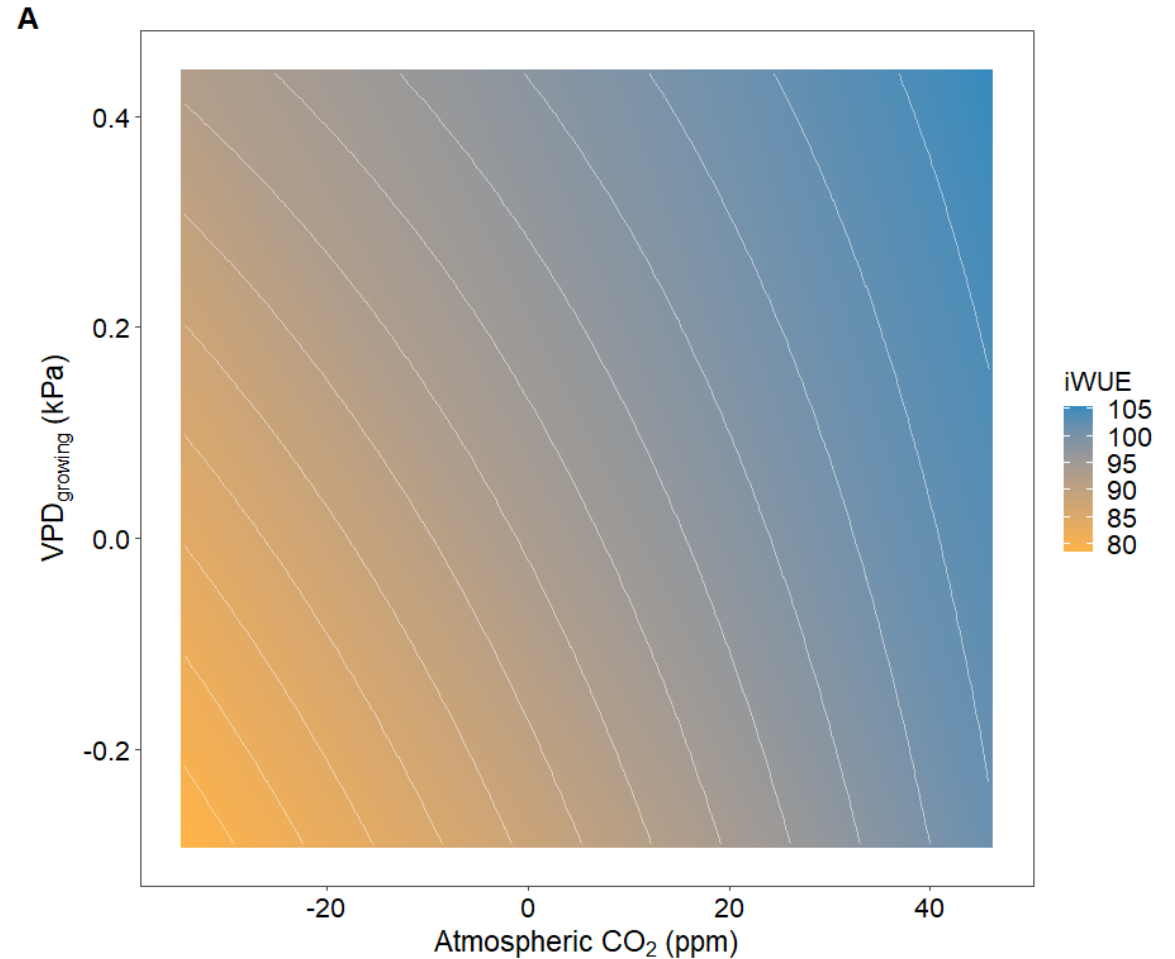
```
iWUE.M2a <- iWUE.M2a$`624`
```

```
summary(iWUE.M2a)
```

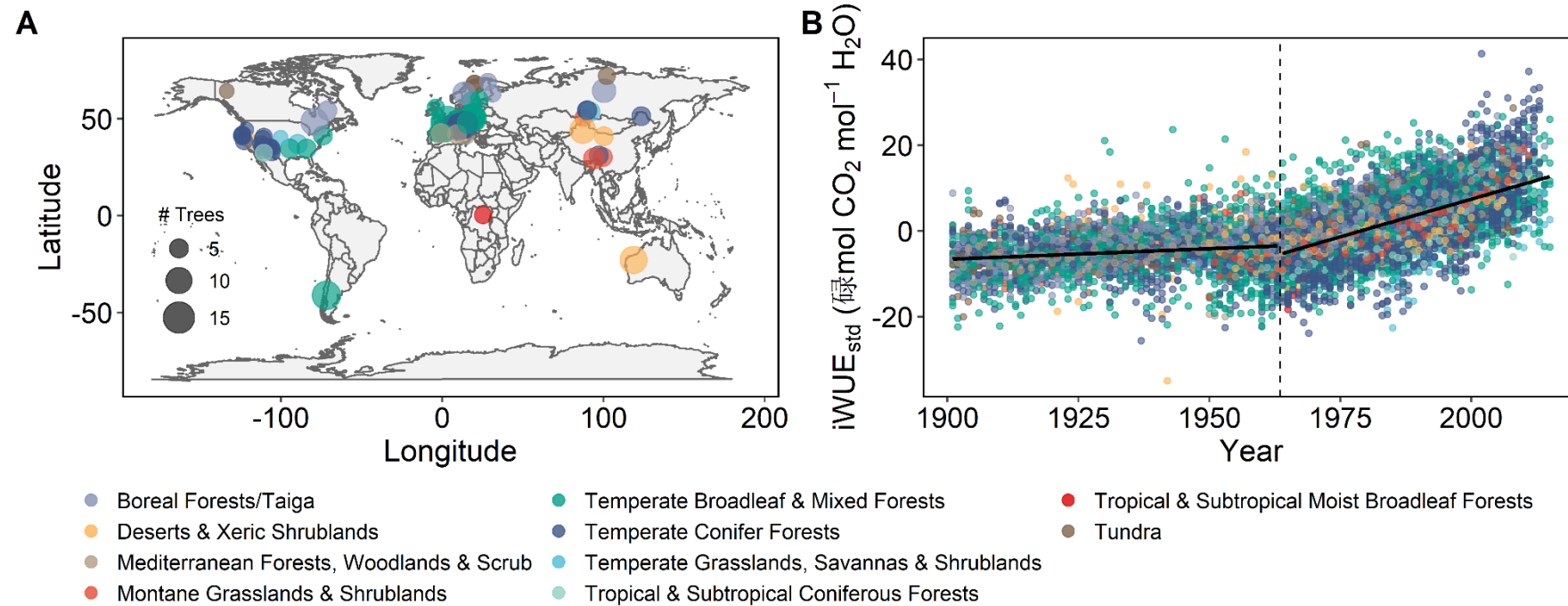
```
rsquared(iWUE.M2a)
```

nlme模型结果展示

```
iWUE.M2a <- get.models(iWUE.m2, subset = delta == 0)
iWUE.M2a <- iWUE.M2a$`624`
vis1 <- visreg2d(iWUE.M2a, xvar = "Ca.cn", yvar =
"GRW.VPD.kPa.cn")
#Extract contour data to plot with ggplot2
contour.data <- data.frame(vis1$z) #Extract values for surface
plot
colnames(contour.data) <- vis1$y #Extract Y axis values
(GRW.VPD.kPa) and name as column header
contour.data$Ca.cn <- vis1$x #Extract values for Ca.cn
```

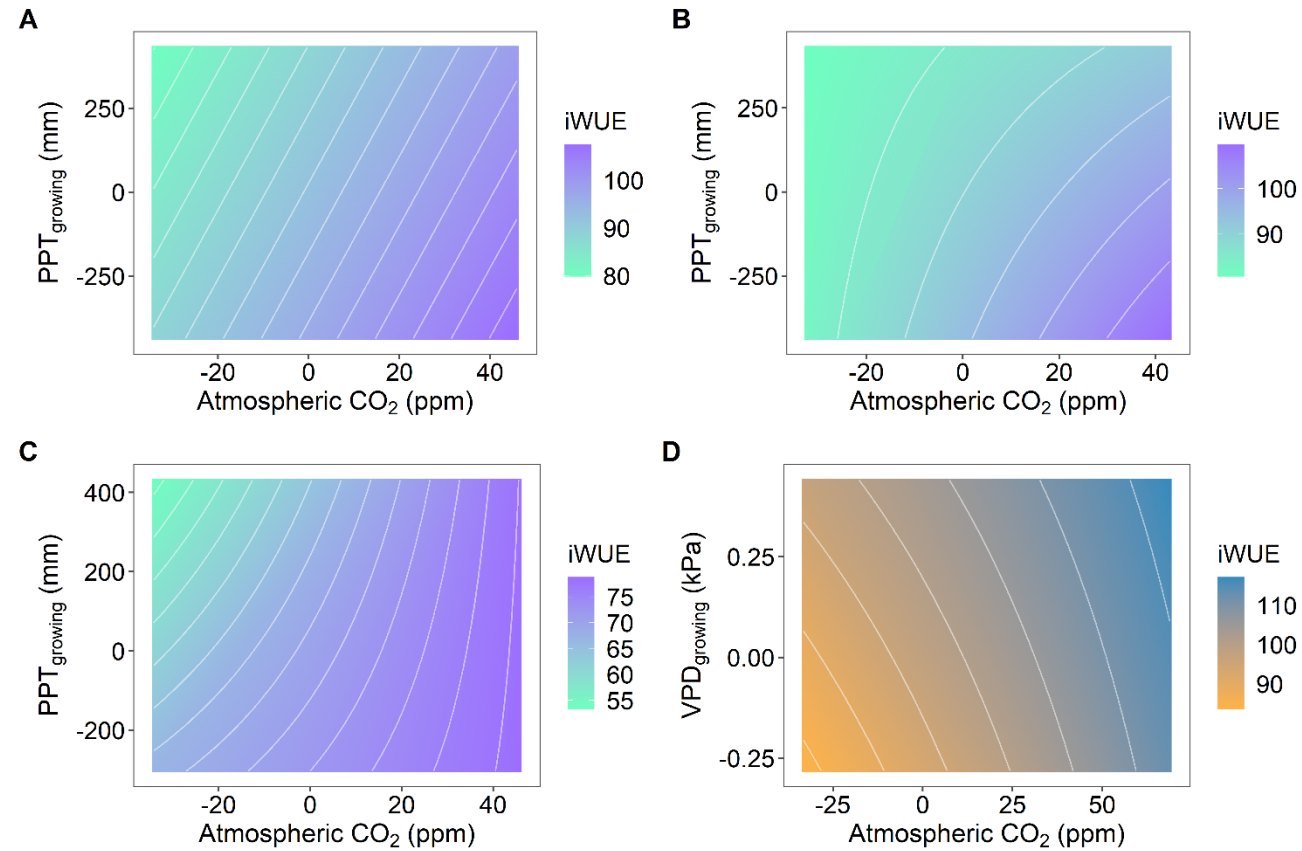


该案例也是学习作图和模型结果展示的佳作



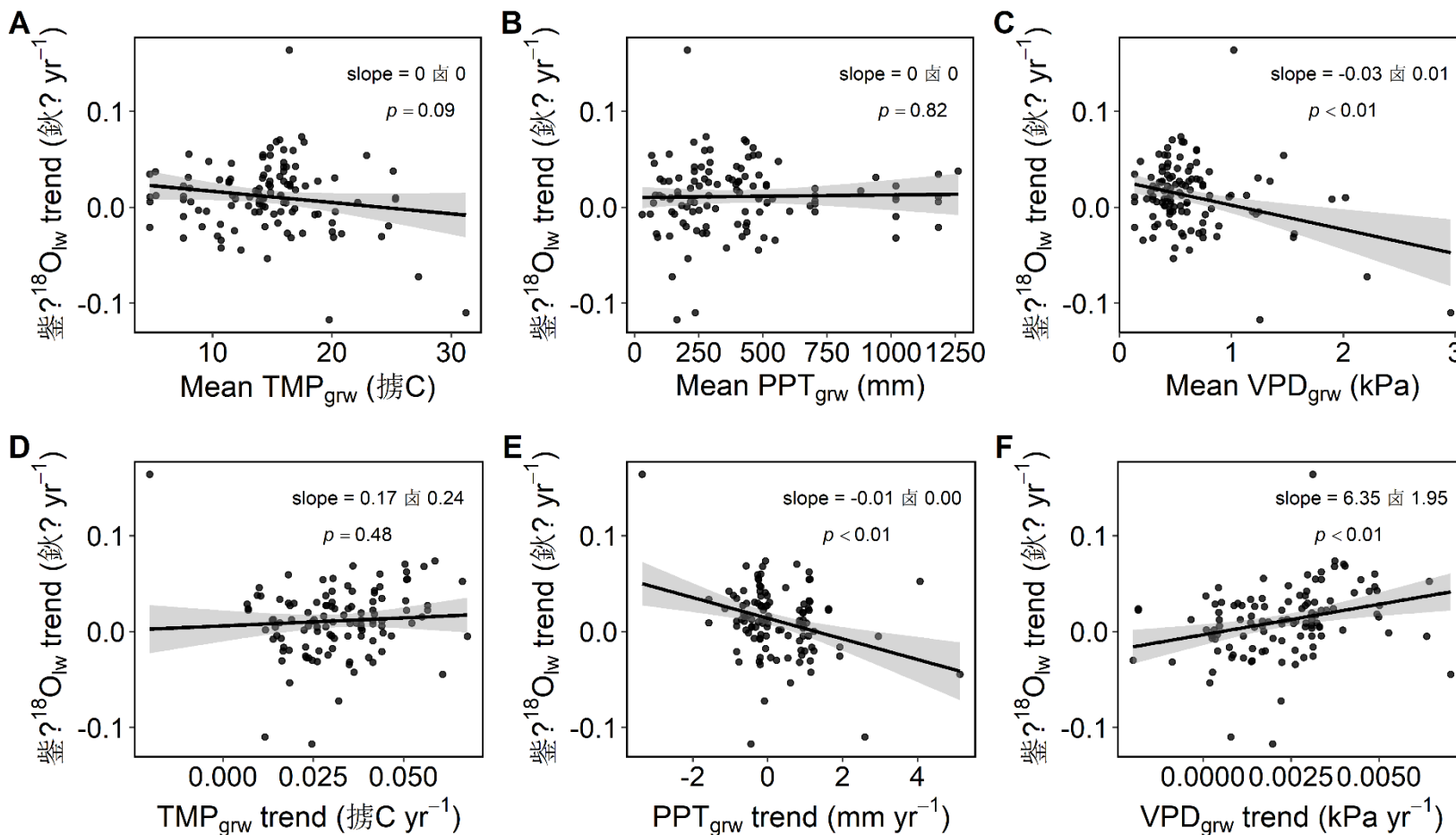
Mathias, J. M. & Thomas, R. B. *PNAS* **118**, e2014286118, doi:10.1073/pnas.2014286118 (2021).

该案例也是学习作图和模型结果展示的佳作

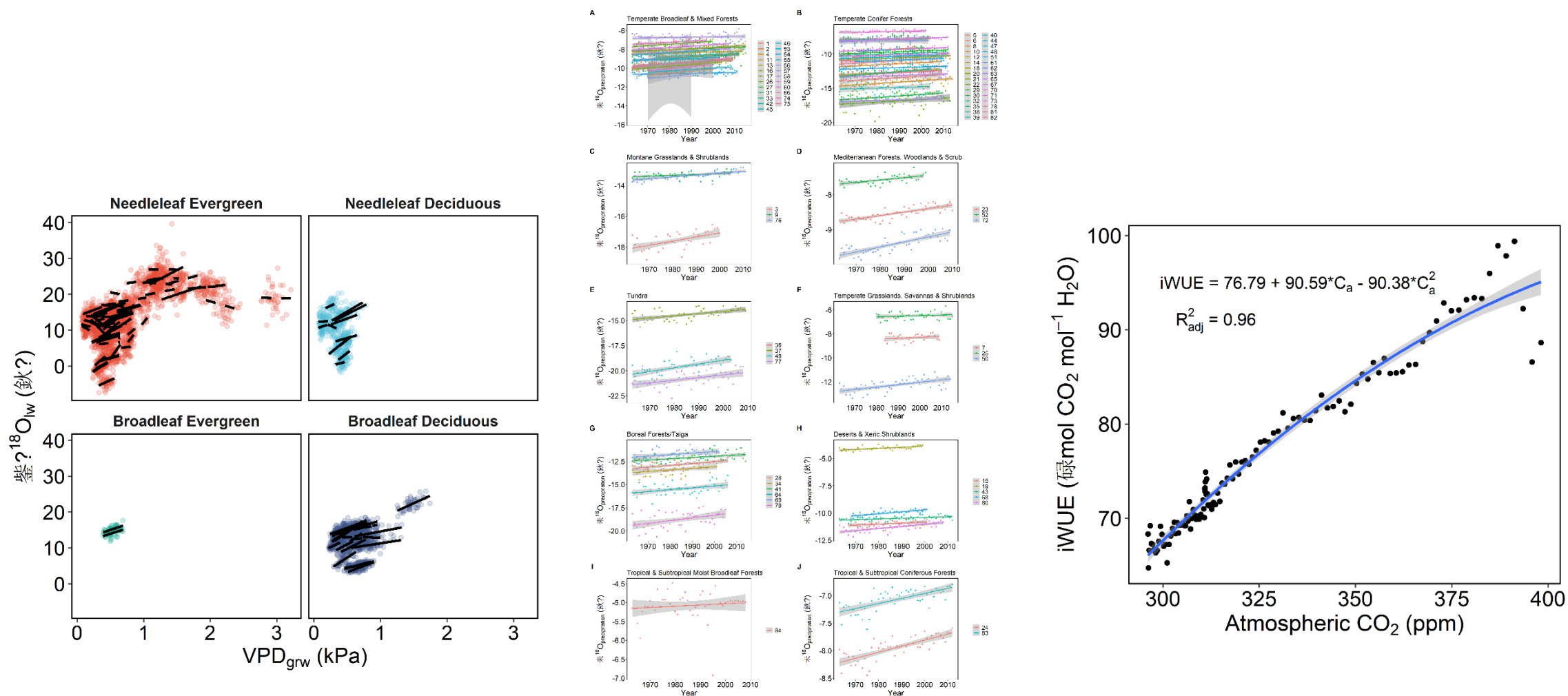


Mathias, J. M. & Thomas, R. B. *PNAS* **118**, e2014286118, doi:10.1073/pnas.2014286118 (2021).

该案例也是学习作图和模型结果展示的佳作



该案例也是学习作图和模型结果展示的佳作

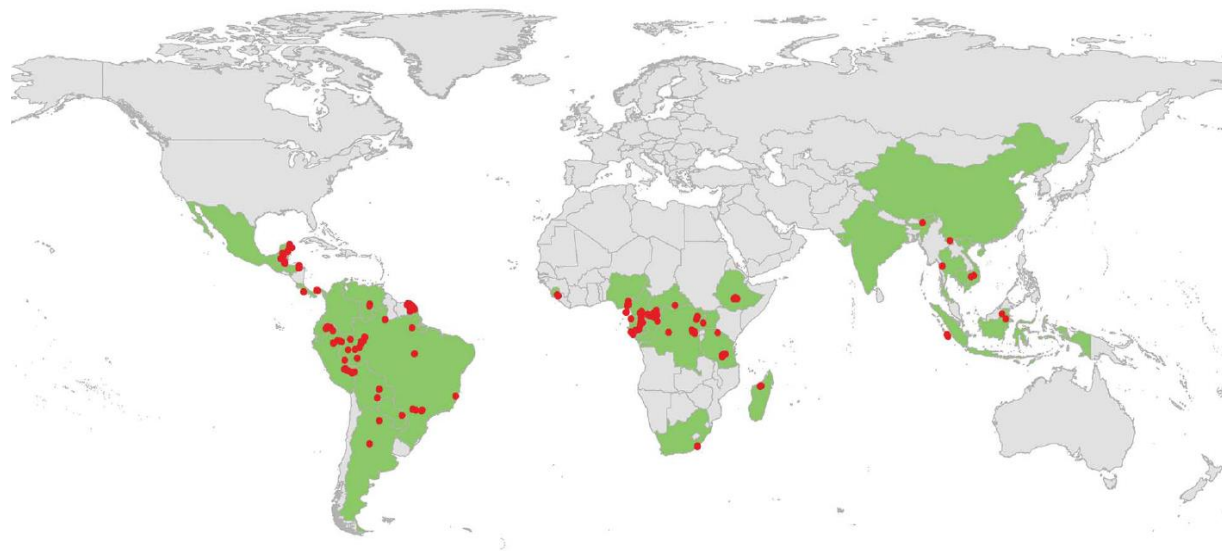
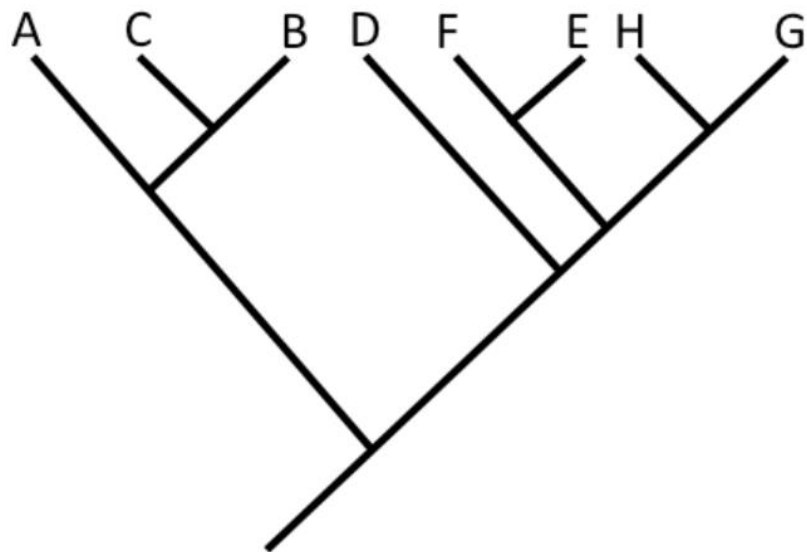


Mathias, J. M. & Thomas, R. B. *PNAS* **118**, e2014286118, doi:10.1073/pnas.2014286118 (2021).

混合效应模型

空间与系统发育信息的纳入

- 混合效应模型通常假定：不同水平的随机对 y 的影响的值，彼此独立，来自于从一个正态分布中的抽样。
- 但有些情况下，这一前提假设并不成立！



有时候，随机效应不同水平之间并不互相独立（时间，空间，亲缘关系）

问题：混合效应模型中，能否同时考虑多个随机效应的自相关结构？

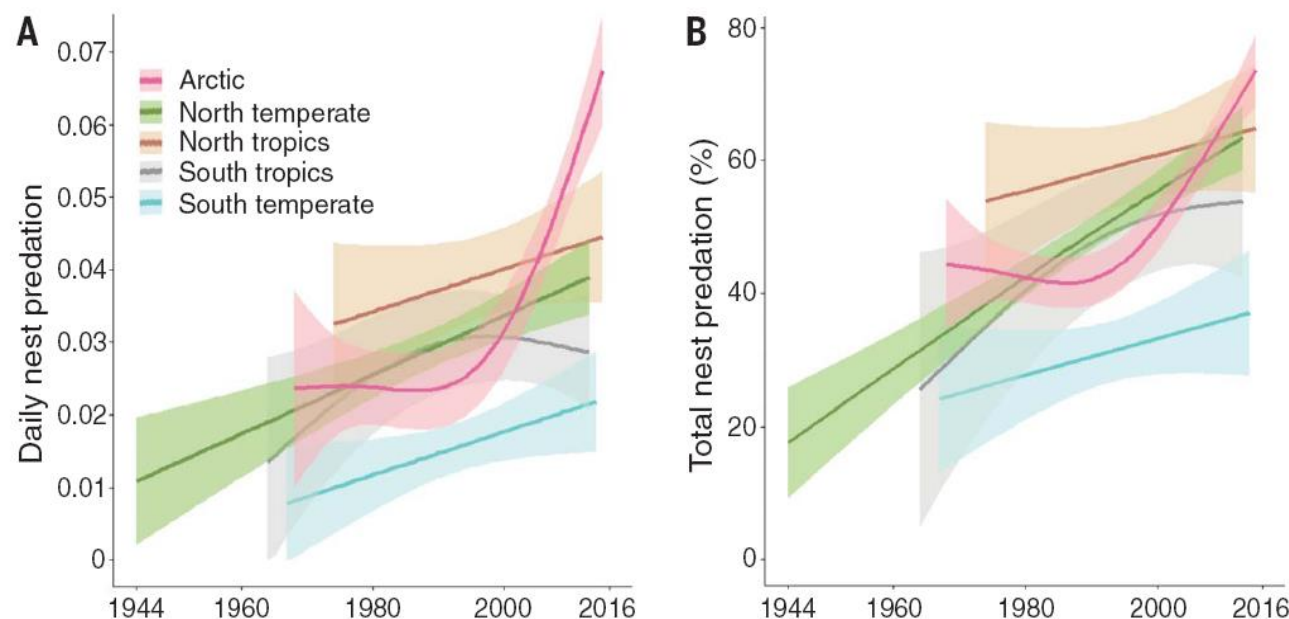
RESEARCH

PREDATION

Global pattern of nest predation is disrupted by climate change in shorebirds

Vojtěch Kubelka^{1,2*}, Miroslav Šálek³, Pavel Tomkovich⁴, Zsolt Végvári^{5,6}, Robert P. Freckleton⁷, Tamás Székely^{2,8,9,10*}

We collected data from both published and previously unpublished sources that included 38,191 nests in 237 populations of 111 shorebirds species from 149 locations, Encompassing all continents across a 70-year time span.



同时考虑多个随机效应的自相关结构的 混合效应模型

Using our comprehensive dataset **in a spatio-phylogenetic framework**
we show that...

We used phylogenetically and spatially controlled generalized linear models.

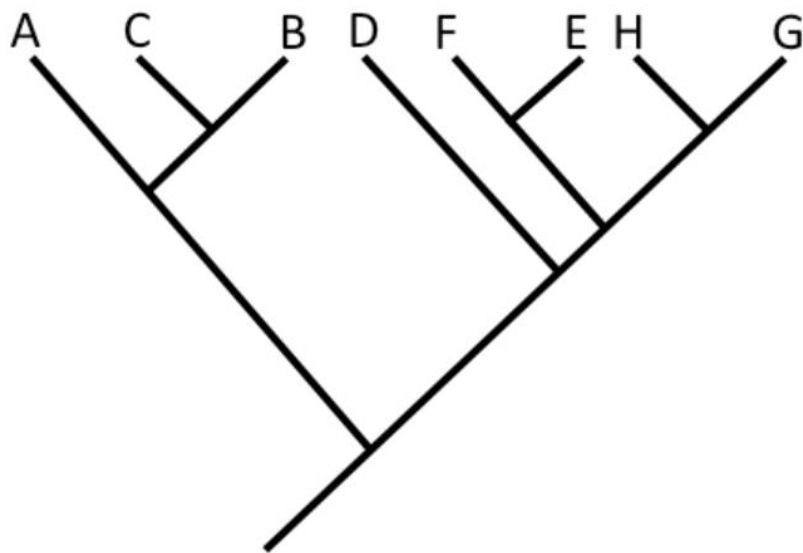
Specifically, we control for

- 1) phylogeny...
- 2) spatial autocorrelation,...
- 3) number of nests

其核心: **Linear mixed-effects kinship models** with control for phylogeny (species level of phylogeny + random effect of the species), spatial autocorrelation and number of nests per population.

同时考虑多个随机效应的自相关结构的混合效应模型—基本原理

- 告诉模型，随机效应的哪些水平彼此之间更为接近，哪些水平彼此差异较大



如，C, B之间的相似性，较A, B之间的相似性高，而A, B之间的相似性，又高于A, G之间的相似性

同时考虑多个随机效应的自相关结构的混合效应模型—基本原理

- 把相似性转化为一个相关性矩阵纳入到模型当中

sp1	1.00								
sp2	0.94	1.00							
sp3	0.94	0.97	1.00						
sp4	0.94	0.95	0.96	1.00					
sp5	0.92	0.90	0.91	0.93	1.00				
sp6	0.92	0.94	0.94	0.95	0.92	1.00			
sp7	0.89	0.91	0.91	0.91	0.88	0.91	1.00		
sp8	0.93	0.95	0.96	0.95	0.91	0.94	0.91	1.00	
sp9	0.88	0.89	0.90	0.89	0.87	0.89	0.86	0.93	1.00
	sp1	sp2	sp3	sp4	sp5	sp6	sp7	sp8	sp9

同时考虑多个随机效应的自相关结构的 混合效应模型—基本原理

对于含亲缘关系信息的数据一般可以把系统发育树转换为相关矩阵

```
library(readxl)
k1 <- read_excel("k_data.xls") ### 注意数据中，物种名小写，属名种名之间以下划线隔开，删掉数据库中没的物种
library(phyloIm)
library(brranching)
library(phytools)
sp_list_k <- unique(k1$Species) ### 提取物种名信息
phylo_tree <- phylomatic(taxa=sp_list_k, storedtree="slik2015", get="POST", lower.case=T) ### 构建系统发育树
plotTree(phylo_tree, main="phylo-tree", node.numbers=TRUE, fsize=0.2, type="fan") ### 作图查看树
phylo_tree$tip.label <- tolower(phylo_tree$tip.label) ### 保证数据的tip.label都小写，与数据匹配
library(phytools) ### 给树加上枝长
phylo_tree <- force.ultrametric(phylo_tree)
any(duplicated(phylo_tree$tip.label)) ### 检查tip.label有无重复
length(phylo_tree$tip) ### 查看树种包含了多少个物种
length(unique(phylo_tree$tip))
phyloMat <- vcv.phylo(phylo_tree) ### 用vcv.phylo函数把树转换为矩阵，并把矩阵中的值转化为0-1之间
phyloMat_k <- phyloMat / max(phyloMat)
View(phyloMat_k)
image(phyloMat_k)
```

得到的相关矩阵

	↑ aristotelia_chilensis ↓	↑ caldcluvia_nymanii ↓	↑ callicoma_serratifolia ↓	↑ ceratopetalum_gummiferum ↓	↑ connarus_suberosus ↓
aristotelia_chilensis	1.00000000	0.95650271	0.95650271	0.95650271	0.73008555
caldcluvia_nymanii	0.95650271	1.00000000	0.93282562	0.93282562	0.73008555
callicoma_serratifolia	0.95650271	0.93282562	1.00000000	0.93282562	0.73008555
ceratopetalum_gummiferum	0.95650271	0.93282562	0.93282562	1.00000000	0.73008555
connarus_suberosus	0.73008555	0.73008555	0.73008555	0.73008555	1.00000000
connarus_turczaninowii	0.73008555	0.73008555	0.73008555	0.73008555	0.85500519
agelaea_borneensis	0.73008555	0.73008555	0.73008555	0.73008555	0.85500519
byrsonima_crassa	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128
byrsonima_verbascifolia	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128
byrsonima_coccolobifolia	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128
dichapetalum_sp.	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128
caryocar_brasiliense	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128
carallia_brachiata	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128
ceriops_decandra	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128
ceriops_tagal_australis	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128
ceriops_tagal_tagal	0.66205128	0.66205128	0.66205128	0.66205128	0.66205128



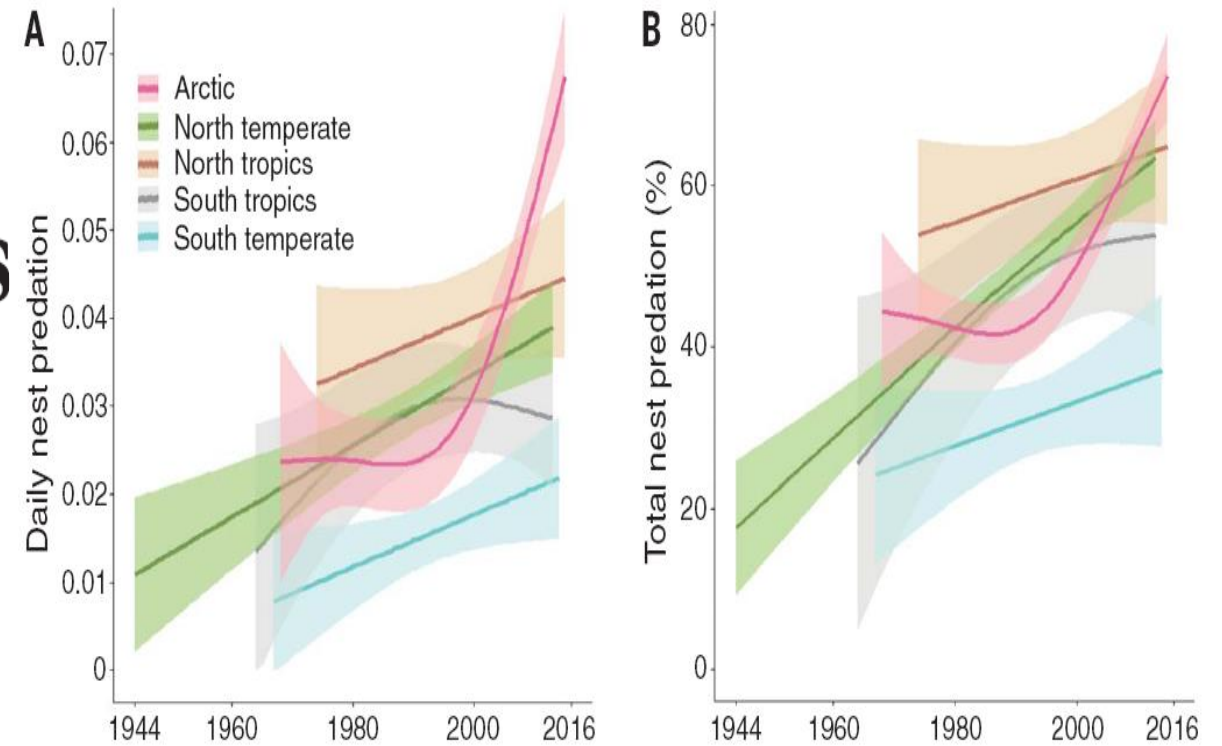
实际应用案例

RESEARCH

PREDATION

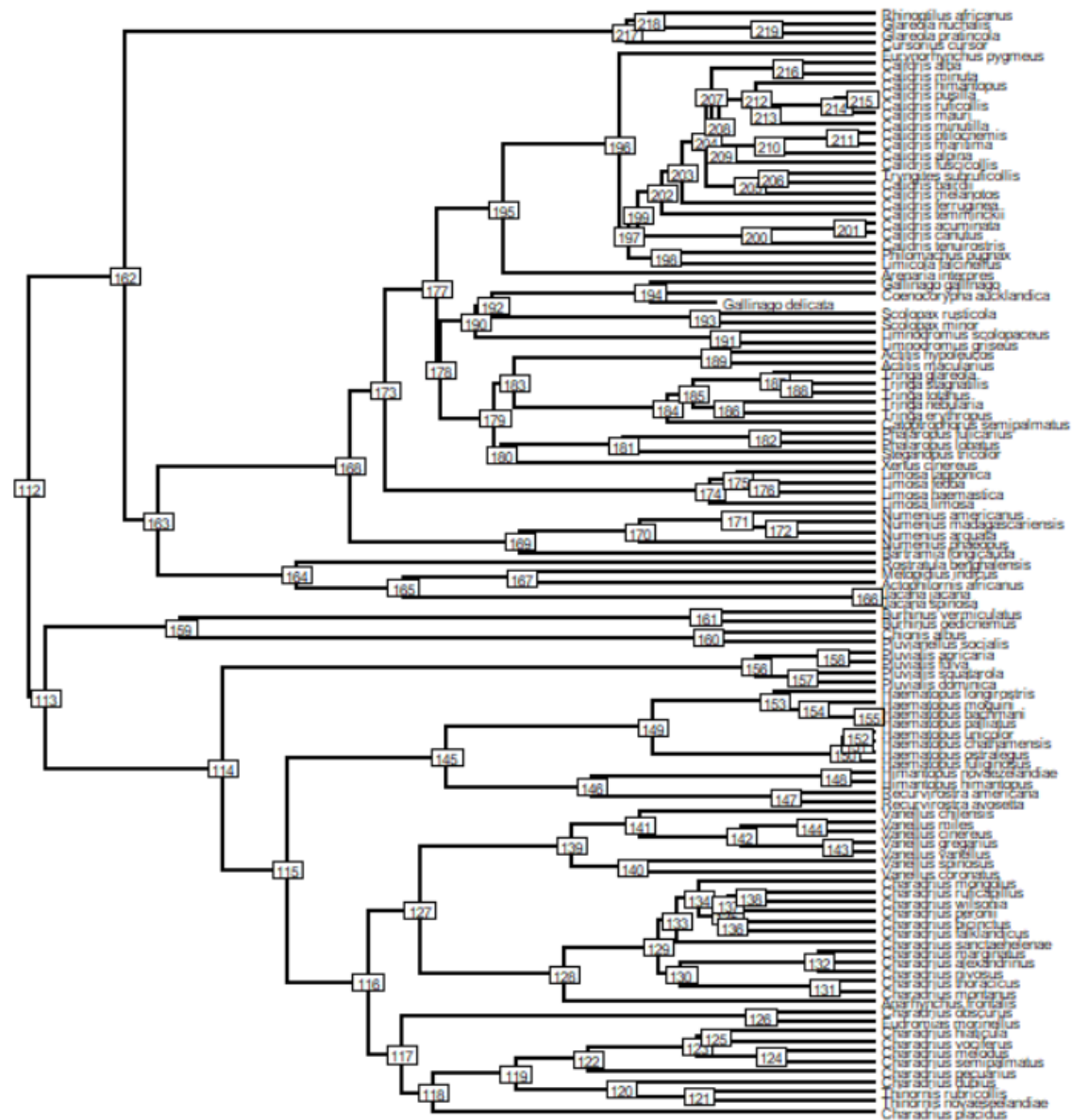
Global pattern of nest predation is disrupted by climate change in shorebirds

Vojtěch Kubelka^{1,2*}, Miroslav Šálek³, Pavel Tomkovich⁴, Zsolt Végvári^{5,6},
Robert P. Freckleton⁷, Tamás Székely^{2,8,9,10*}



1、亲缘关系矩阵(phyloMat)

```
tree2<-read.tree("tree2.phy")
plotTree(tree2, node.numbers = TRUE, fsize = 0.5)
#### 1, 构建系统发育矩阵
phyloMat <- vcv.phylo( tree2 )
phyloMat <- phyloMat / max( phyloMat )
```



2、空间距离矩阵(distanceMatrix)

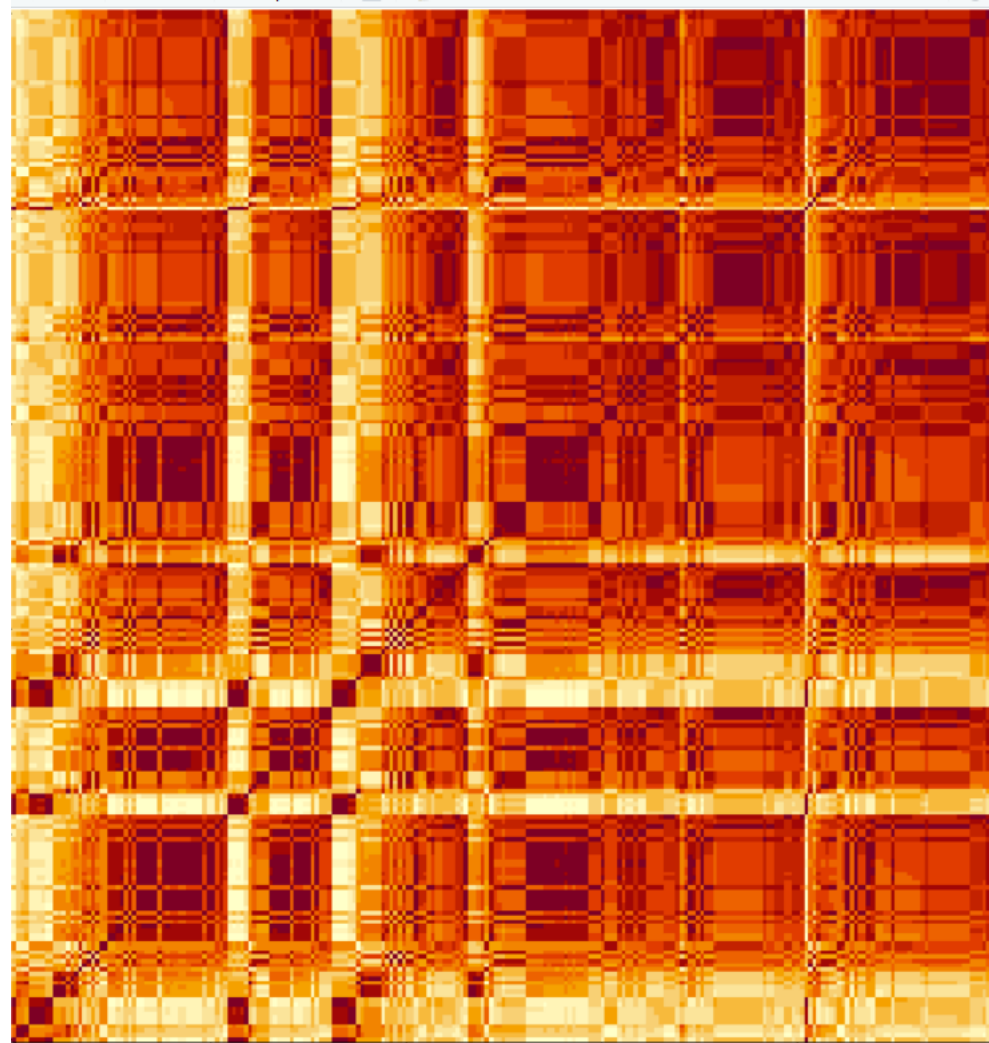
```
distanceMatrix <- dist.mat( data$Latitude, data$Longitude, data$species )
```

```
diag( distanceMatrix ) <- diag(distanceMatrix) + 0.01
```

```
distanceMatrix <- distanceMatrix / 1.01
```

```
rownames(distanceMatrix) <- colnames(distanceMatrix) <- data$species
```

	Himantopus_novaezelandiae	Chionis_albus	Pluvianellus_socialis	Charadrius_falklandicus
Himantopus_novaezelandiae	1.00000000	0.62218682	0.59575194	0.54891223
Chionis_albus	0.62218682	1.00000000	0.92186486	0.88120310
Pluvianellus_socialis	0.59575194	0.92186486	1.00000000	0.94045985
Charadrius_falklandicus	0.54891223	0.88120310	0.94045985	1.00000000
Haematopus_fuliginosus	0.89664904	0.57490932	0.53375817	0.48445371
Haematopus_longirostris	0.89664904	0.57490932	0.53375817	0.48445371
Thinornis_rubricollis	0.88201317	0.56371537	0.52098844	0.47155480
Thinornis_rubricollis.1	0.88008341	0.56285615	0.51984400	0.47038839
Charadrius_obscurus	0.94829512	0.59951426	0.58114476	0.53680913
Haematopus_unicolor	0.94667014	0.59277653	0.57384358	0.52939656
Haematopus_moquini	0.45642958	0.68704398	0.63889653	0.63531905
Haematopus_moquini.1	0.44955735	0.68961377	0.64328301	0.64131687
Burhinus_vermiculatus	0.45772753	0.64649253	0.59440679	0.58804432
Haematopus_palliatu	0.42417796	0.78088923	0.81851689	0.86444684
Vanellus_coronatus	0.39582870	0.58204224	0.53598407	0.53728485
Charadrius_sanctae	0.32965737	0.66501899	0.65301218	0.67746063
Charadrius_peronii	0.52324571	0.28750524	0.21979411	0.18249718
Charadrius_wilsonia	0.28323745	0.54838045	0.60768700	0.65556301
Charadrius_wilsonia.1	0.38863150	0.47470840	0.54293105	0.57674051
Burhinus	0.33383000	0.37551170	0.33660770	0.33451147



3、鸟巢数量相关性矩阵 (I)

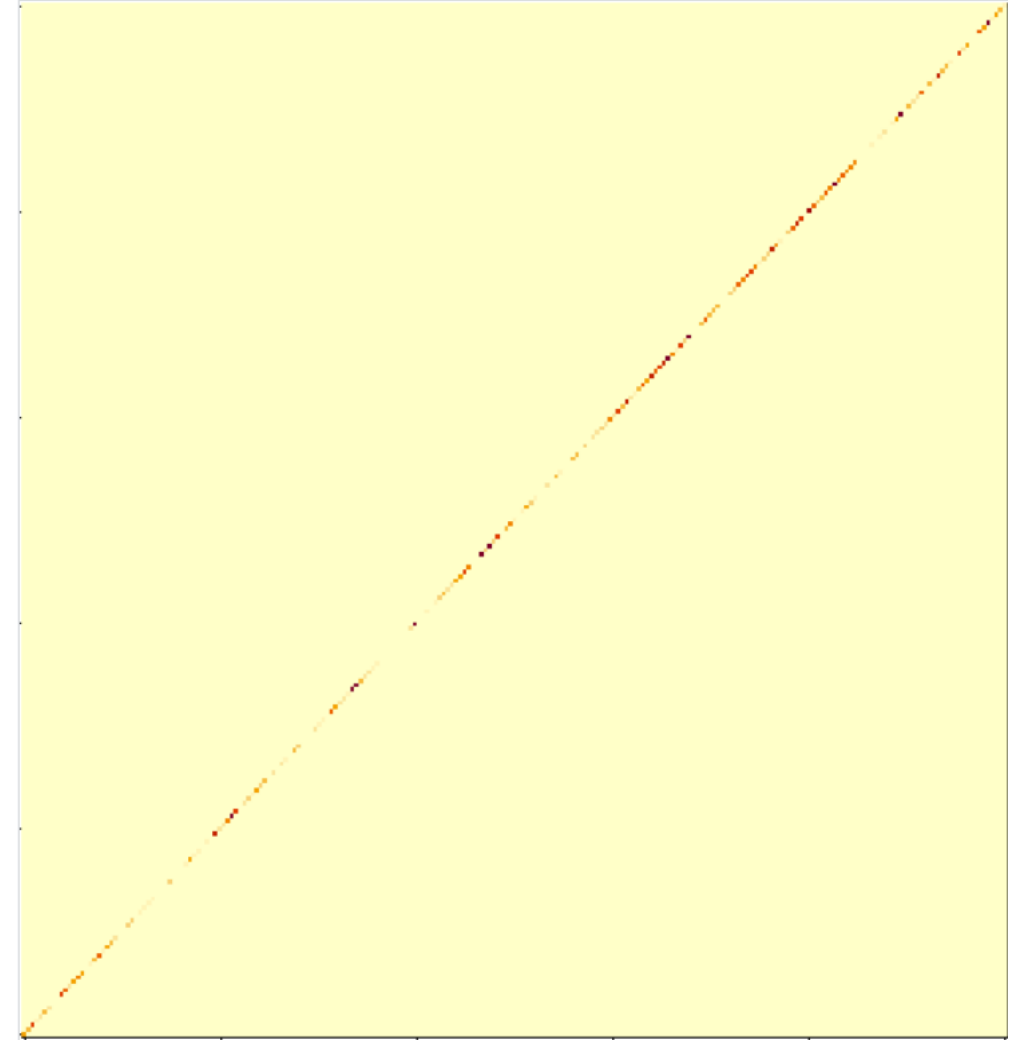
number of nests per population

```
I <- diag(1 / data$N_nests )
```

```
I <- I / max(I)
```

```
rownames(I) <- colnames(I) <- data$species
```

	Himantopus_novaezelandiae	Chionis_albus	Pluvianellus_socialis	Charadrius_falklandicus	Haematopus_fuliginosus
Himantopus_novaezelandiae	0.4444444	0.0000000	0.0000000	0.00	0.0000000
Chionis_albus	0.0000000	0.4137931	0.0000000	0.00	0.0000000
Pluvianellus_socialis	0.0000000	0.0000000	0.7058824	0.00	0.0000000
Charadrius_falklandicus	0.0000000	0.0000000	0.0000000	0.16	0.0000000
Haematopus_fuliginosus	0.0000000	0.0000000	0.0000000	0.00	0.1846154
Haematopus_longirostris	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Thinornis_rubricollis	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Thinornis_rubricollis.1	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Charadrius_obscurus	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Haematopus_unicolor	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Haematopus_moquini	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Haematopus_moquini.1	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Burhinus_vermiculatus	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Haematopus_palliatu	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Vanellus_coronatus	0.0000000	0.0000000	0.0000000	0.00	0.0000000
Charadrius_sanctae	0.0000000	0.0000000	0.0000000	0.00	0.0000000



把三个矩阵的信息加入到模型当中

Linear mixed-effects kinship models with control for **phylogeny** (species level of phylogeny + random effect of the species), **spatial autocorrelation** and **number of nests per population**

固定效应

随机效应

```
library(coxme)
```

```
m1_ML <- lmeKin( log(DPR) ~mean_year + log( N_nests)+ (1|species ),  
  varlist = list( I, phyloMat , distanceMatrix ),  
  data = data)
```

指定随机效应不同水平间的相关关系

结果

Linear mixed-effects kinship model fit by maximum likelihood

Data: data

Log-likelihood = -194.8911

n= 237

Model: $\log(\text{DPR}) \sim \text{mean_year} + \log(\text{N_nests}) + (1 \mid \text{species})$

Fixed coefficients

	value	std Error	z	p
(Intercept)	-33.065100787	4.971389299	-6.65	2.9e-11
mean_year	0.014901439	0.002496458	5.97	2.4e-09
log(N_nests)	-0.006628801	0.030362547	-0.22	8.3e-01

Random effects

Group	Variable	std Dev	Variance
species	vmat.1	3.030270e-01	9.182539e-02
	vmat.2	5.085568e-03	2.586300e-05
	vmat.3	1.138452e-03	1.296074e-06

Residual error= 0.5274976

`m1_ML <- lmeKin(log(DPR) ~ mean_year + log(N_nests) + (1|species), varlist = list(I, phyloMat , distanceMatrix),
data = data)`

所以，控制系统发育与空间距离的核心， 在于在模型中加入与随机效应对应的矩阵

- 一个随机效应可以对应一个矩阵，也可以对应多个矩阵

```
m1_ML1 <- lme4::lmer(log(DPR) ~ mean_year + log(N_nests) + (1|species),  
  varlist = list(phyloMat),  
  data = data)
```

```
m1_ML2 <- lme4::lmer(log(DPR) ~ mean_year + log(N_nests) + (1|species),  
  varlist = list(phyloMat, distanceMatrix),  
  data = data)
```

```
m1_ML3 <- lme4::lmer(log(DPR) ~ mean_year + log(N_nests) + (1|species),  
  varlist = list(I, phyloMat, distanceMatrix),  
  data = data)
```

对原文中模型的改进

是否可以设置地点为随机效应并加入之对应的矩阵？

```
###构造一个新变量，研究地点（location_id）
```

```
data$n1<-data$Latitude^5+data$Longitude^5
```

```
data$location_id<- as.numeric(factor(data$n1,  
levels=unique(data$n1)))
```

```
View(data)
```

```
###依据研究地点的经纬度坐标构建地点之间的距离矩阵distanceMatrix2
```

```
distanceMatrix2 <- dist.mat( data$Latitude, data$Longitude,  
                             data$location_id)
```

```
diag( distanceMatrix2 ) <- diag(distanceMatrix2) + 0.01
```

```
distanceMatrix2 <- distanceMatrix2 / 1.01
```

然后：在模型中 同时加入系统发育距离与地点空间距离矩阵

```
m2<- lme4::lmer(log(DPR) ~ mean_year + log(N_nests) +  
  (1|species) + (1|location_id),  
  varlist = list(covMlist(phyloMat),  
    distanceMatrix2),  
  data = data)
```

与species对应的系统发育距离矩阵

与location_id对应的系统发育距离矩阵

在模型中 同时加入系统发育距离与地点空间距离矩阵

```
> m2
```

```
Linear mixed-effects kinship model fit by maximum likelihood
```

```
Data: data
```

```
Log-likelihood = -189.8724
```

```
n= 237
```

```
Model: log(DPR) ~ mean_year + log(N_nests) + (1 | species) + (1 | location_id)
```

```
Fixed coefficients
```

	Value	Std Error	z	p
(Intercept)	-32.871974559	5.477488363	-6.00	2.0e-09
mean_year	0.014671294	0.002720404	5.39	6.9e-08
log(N_nests)	-0.007974759	0.031997833	-0.25	8.0e-01

```
Random effects
```

Group	Variable	Std Dev	Variance
species	Vmat.1	0.13531452	0.01831002
location_id	Vmat.1	1.11500704	1.24324069

```
Residual error= 0.4377888
```

###查看模型的AICc值

```
library(AICcmodavg)
```

```
AICc(m1_ml)
```

```
AICc(m2)
```

```
> AICc(m1_ML)
```

```
[1] 398.7445
```

```
> AICc(m2)
```

```
[1] 392.1101
```

其他实现方法1: lme4qtl包

####重新构造一个距离矩阵,

####数据中不能包括location重复的位置信息

```
d1 <- data[,c("Latitude","Longitude","location_id")]
```

```
d2 <- d1[!duplicated(d1$location_id),]
```

```
distanceMatrix3 <- dist.mat( d2$Latitude, d2$Longitude,  
                             d2$location_id)
```

```
diag( distanceMatrix3 ) <- diag(distanceMatrix3) + 0.01
```

```
distanceMatrix3 <- distanceMatrix3 / 1.01
```

```
colnames(distanceMatrix3) <- c(1:nrow(distanceMatrix3))
```

其他实现方法1: lme4qtl包

```
library(lme4qtl)
m3<-relmatLmer(log(DPR) ~mean_year + log( N_nests)+
               (1|species)+(1|location_id),
               data =data,
               relmat = list(species=phyloMat, location_id=distanceMatrix3))
summary(m3)
library(car)
Anova(m3, type="II", test.statistic ="Chisq")
Anova(m3, type="II", test.statistic ="F")
```

其他实现方法2: brms包(贝叶斯法)

The idea of **brms**: Fitting all kinds of regression models within one framework

Paul-Christian Bürkner , 2017

brms: An R Package for Bayesian Multilevel Models Using Stan

[Journal of Statistical Software](#)

其他实现方法2: brms包(贝叶斯法)

```
library(brms)
bm1<- brm(log(DPR) ~mean_year + log( N_nests)+
  (1|species)+(1|location_id),
  data =data,
  family = gaussian(),
  cov_ranef = list(species=phyloMat, location_id=distanceMatrix3),
  chains=1,iter=20000)
summary(bm1)
plot(bm1)
pp_check(bm1, iter=1000)
```

其他实现方法2: brms包(贝叶斯法)

Group-Level Effects:

~location_id (Number of levels: 152)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(Intercept)	1.18	0.37	0.42	1.89	685	1.00

~species (Number of levels: 111)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(Intercept)	0.15	0.10	0.01	0.37	1432	1.00

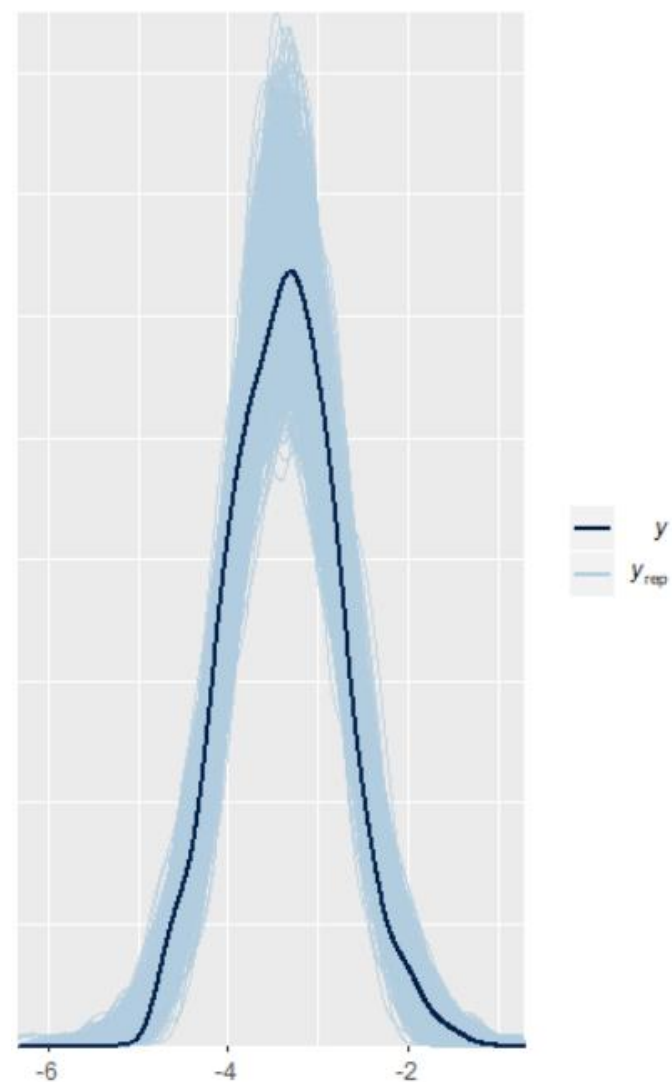
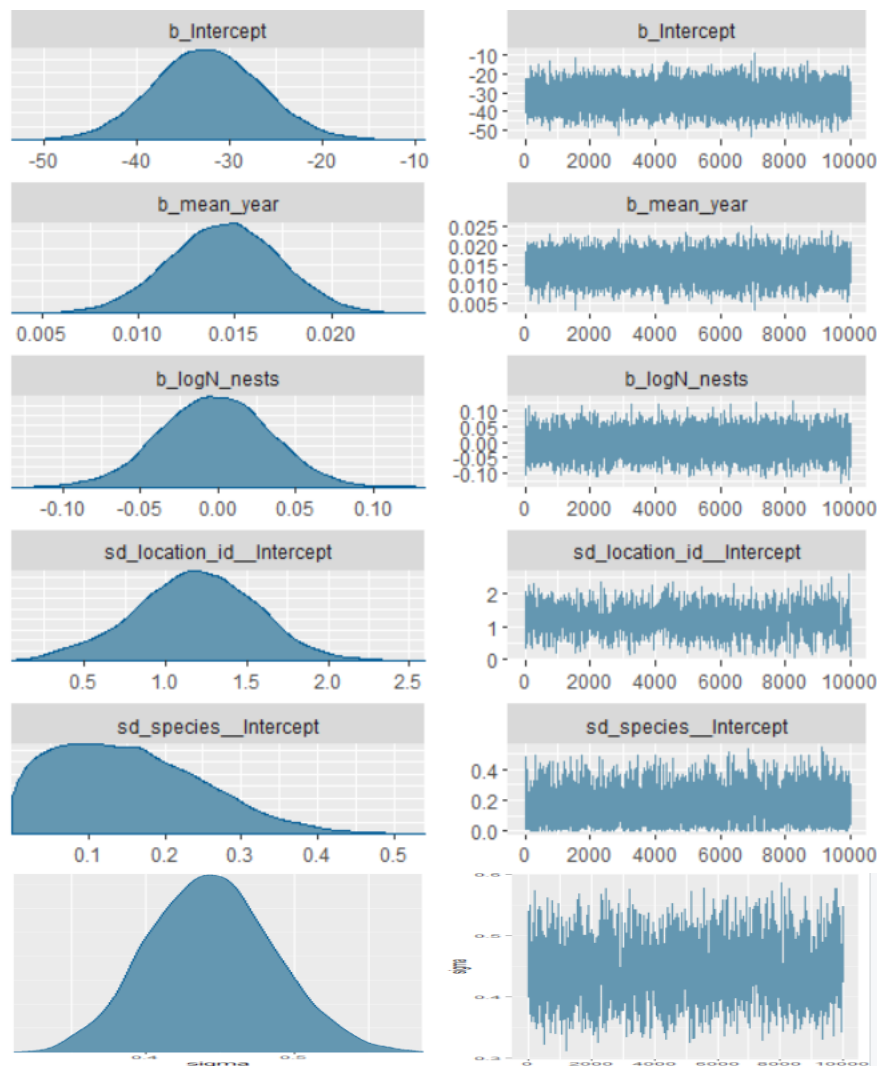
Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
Intercept	-32.37	5.76	-43.46	-21.12	8383	1.00
mean_year	0.01	0.00	0.01	0.02	8602	1.00
logN_nests	-0.00	0.03	-0.07	0.07	6280	1.00

Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sigma	0.44	0.04	0.36	0.53	846	1.00

其他实现方法2: brms包(贝叶斯法)



其他实现方法3： MCMCglmm包(贝叶斯法)

```
library(MCMCglmm)
inverse_matrix1<-solve(phyloMat)
inverse_phylo<-as(inverse_matrix1,"dgCMatrix")
inverse_dist_matrix1<-solve(distanceMatrix3)
inverse_dist_matrix2<-as(inverse_dist_matrix1,"dgCMatrix")
data<-as.data.frame(data)
data$species<-as.factor(data$species)
data$location_id<-as.factor(data$location_id)
data<-data[,-3]
bm2<-MCMCglmm(log(DPR)~mean_year + log(N_nests),
  random=~species+location_id, nitt=20000,
  ginverse=list(species=inverse_phylo,
    location_id=inverse_dist_matrix2),data=data)
summary(bm2)
plot(bm2)
```

MCMCglmm使用的是相关矩阵的逆矩阵

其他实现方法3: MCMCglmm包(贝叶斯法)

```
Iterations = 3001:199991
Thinning interval = 10
Sample size = 19700
```

```
DIC: 354.7387
```

```
G-structure: ~species
```

	post.mean	1-95% CI	u-95% CI	eff.samp
species	0.003288	1.489e-16	0.02101	299.5

```
~location_id
```

	post.mean	1-95% CI	u-95% CI	eff.samp
location_id	1.311	0.0007711	2.958	1724

```
R-structure: ~units
```

	post.mean	1-95% CI	u-95% CI	eff.samp
units	0.2179	0.1405	0.3009	1738

```
Location effects: log(DPR) ~ mean_year + log(N_nests)
```

	post.mean	1-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	-32.488070	-43.319183	-21.462747	18790	<5e-05 ***
mean_year	0.014490	0.009149	0.020010	18483	<5e-05 ***
log(N_nests)	0.000523	-0.064130	0.065425	12851	0.987

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

不同方法的参数估计结果对比

Variable	lme4	lmer	brms	MCMCglmm
Intercept	-32.72586909	-32.77	-32.37	-32.61133
mean_year	0.01473638	0.0146	0.01	0.014564
log(N_nests)	-0.01208028	0.0003	0.00	-0.001257

不同方法的对比

- `Lme4`, 只适用于正态分布数据, 不适用于泊松, 二项等分布数据与 `MuMIn` 能配套, 可进行自动模型选择, 但不能计算 R^2
- `Lme4qtl`, 适用于正态分布数据与泊松分布数据, 与 `MuMIn` 能配套, 可进行自动模型选择, R^2 计算等
- `brms`, 需要有贝叶斯统计基础, 运算速度慢, 但使用分布广语法与 `lme4` 高度相似, 易于掌握,
- `MCMCglmm`, 需要有贝叶斯基础, 相对于 `brms` 较快, 但语法不及 `brms` 简单

处理系统发育关系的另一种做法

```
dat<-read.csv("nature_example.csv")
```

```
terr1 <- lmer(tmax~tmax_acc + tmax_metric + (1|Phylum/Class/Order/Family/Genus), data=dat, subset=habitat=='terrestrial')
```

```
summary(terr1)
```

```
mari1 <- lmer(tmax~tmax_acc + tmax_metric + (1|Order/Family/Genus), data=dat, subset=habitat=='marine') #ignore phylum and class since invariant in this subset
```

```
summary(mari1)
```

in the lme4 version 1.1-18-1 package as nested random effects to help include duration, method (terrestrial or marine) and the all P values are two-sided. We calculate

```
Random effects:
Groups
Genus:(Family:(Order:(Class:Phylum))) (Intercept) 6.828e+00 2.61304
Family:(Order:(Class:Phylum)) (Intercept) 3.642e+00 1.90850
Order:(Class:Phylum) (Intercept) 7.014e+00 2.64848
Class:Phylum (Intercept) 5.568e-06 0.00236
Phylum (Intercept) 8.241e+00 2.87078
Residual 2.525e+00 1.58915
Number of obs: 319, groups:
Genus:(Family:(Order:(Class:Phylum))), 130; Family:(Order:(Class:Phylum)), 58; Order:(Class:Phylum), 15; Class:Phylum, 5; Phylum, 2

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  34.40581    2.30012    1.06464  14.958  0.0362 *
tmax_acc      0.24840    0.02158   278.39973  11.513  <2e-16 ***
tmax_metricleth -1.20897    1.70799   90.81468  -0.708  0.4809
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr) tmx_cc
tmax_acc      -0.201
tmx_mtrclth   -0.079 -0.037
```

Pinsky, M. L., A. M. Eikeset, D. J. McCauley, J. L. Payne, and J. M. Sunday. 2019. Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature* **569**:108-111.

混合效应模型在meta分析中的应用(lme4)

以单个案例研究的方差 v_i 作为权重进行分析

```
gi<- lmer(gi~ + (1 |id_estudo), data=subset(dat1,overall_effect_gi=="TRUE"), weight=1/var_gi)
```

效应值 y_i



案例内方差 v_i



Marino, N. d. A. C., G. Q. Romero, V. F. Farjalla, and M. Anderson. 2018. Geographical and experimental contexts modulate the effect of warming on top-down control: a meta-analysis. Ecology Letters 21:455-466.

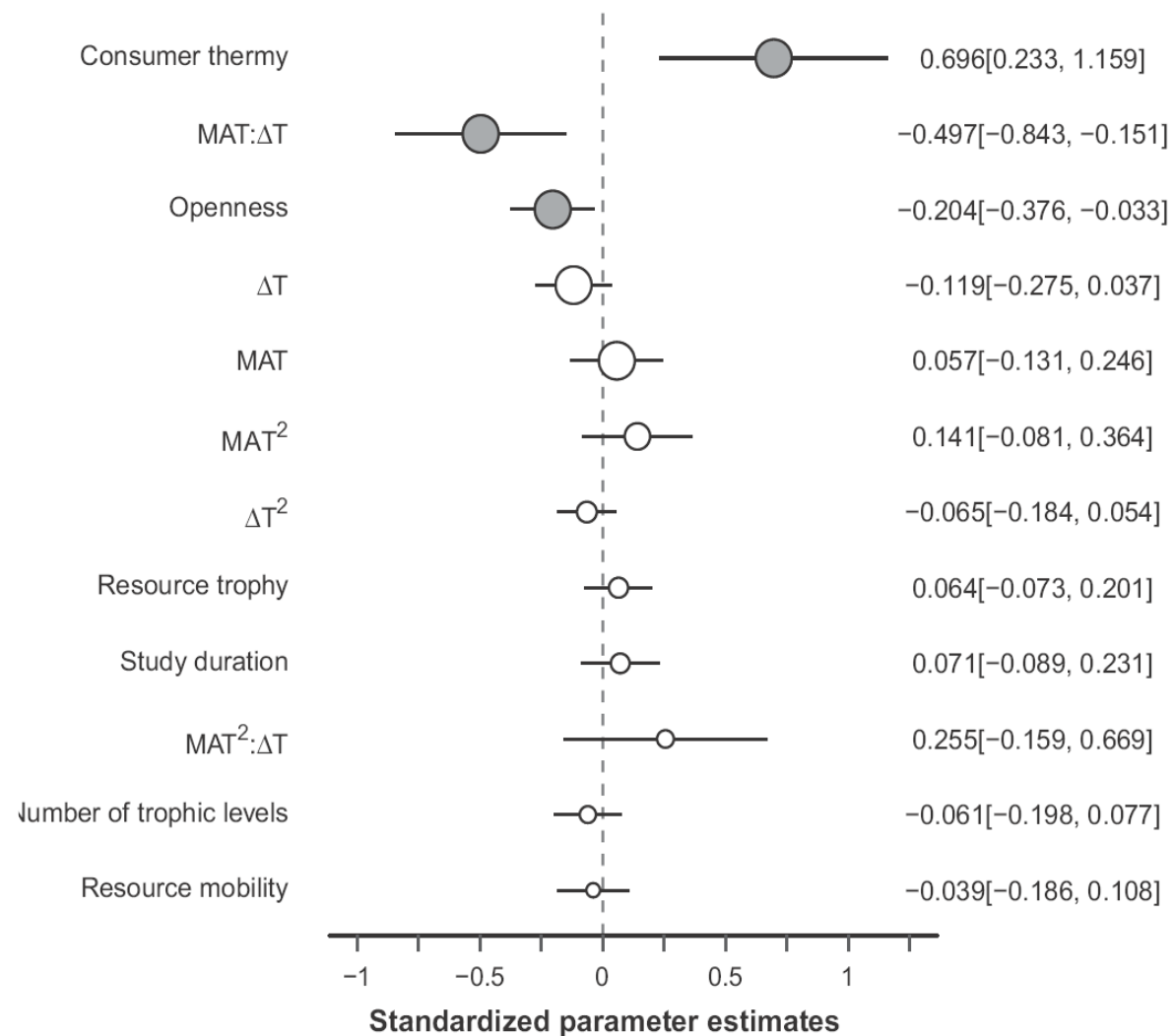
lme4 包meta分析模型选择

Table 1 Models that were better supported by the data according to the Akaike information criteria for (a) g_T , (b) g_W and (c) g_C

Model	Model structure	d.f.	logLik	AICc	Δ AICc	w_i
(a) Effect of warming on top-down control of resource biomass (g_T)						
#1	$g_T \sim \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \Delta T + \text{MAT}:\Delta T$	8	-105.48	227.83	0	0.17
#2	$g_T \sim \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \text{MAT}^2 + \Delta T + \text{MAT}:\Delta T$	9	-104.65	228.4	0.57	0.128
#3	$g_T \sim \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \Delta T + \Delta T^2 + \text{MAT}:\Delta T$	9	-104.86	228.82	0.99	0.104
#4	$g_T \sim \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \text{MAT}^2 + \Delta T + \text{MAT}:\Delta T + \text{MAT}^2:\Delta T$	10	-103.91	229.17	1.34	0.087
#5	$g_T \sim \text{Resource trophic} + \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \Delta T + \text{MAT}:\Delta T$	9	-105.07	229.23	1.41	0.084
#6	$g_T \sim \text{Consumer thermoregulation} + \text{Number of trophic levels} + \text{Openness} + \text{MAT} + \Delta T + \text{MAT}:\Delta T$	9	-105.1	229.29	1.46	0.082
#7	$g_T \sim \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \text{Study duration} + \Delta T + \text{MAT}:\Delta T$	9	-105.17	229.44	1.61	0.076
#8	$g_T \sim \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \text{MAT}^2 + \Delta T + \Delta T^2 + \text{MAT}:\Delta T$	10	-104.12	229.57	1.74	0.071
#9	$g_T \sim \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \text{MAT}^2 + \text{Study duration} + \Delta T + \text{MAT}:\Delta T$	10	-104.16	229.66	1.83	0.068
#10	$g_T \sim \text{Resource trophic} + \text{Consumer thermoregulation} + \text{Openness} + \text{MAT} + \text{MAT}^2 + \Delta T + \text{MAT}:\Delta T$	10	-104.2	229.75	1.92	0.065
#11	$g_T \sim \text{Consumer thermoregulation} + \text{Openness} + \text{Resource mobility} + \text{MAT} + \Delta T + \text{MAT}:\Delta T$	9	-105.34	229.77	1.95	0.064
(b) Effect of warming on resource biomass (g_W)						
#1	$g_W \sim \text{Ecosystem type} + \text{Resource trophic} + \text{Resource generation time} + \Delta T + \Delta T^2$	8	-94.9	206.67	0	0.265
#2	$g_W \sim \text{Ecosystem type} + \text{Resource generation time} + \Delta T + \Delta T^2$	7	-96.15	206.97	0.3	0.228
#3	$g_W \sim \text{Ecosystem type} + \text{Resource trophic} + \text{Resource generation time} + \text{MAT} + \Delta T + \Delta T^2$	9	-94.4	207.89	1.22	0.144
#4	$g_W \sim \text{Ecosystem type} + \text{Resource generation time} + \text{MAT} + \Delta T + \Delta T^2$	8	-95.59	208.05	1.38	0.133
#5	$g_W \sim \text{Ecosystem type} + \text{Resource trophic} + \text{Resource generation time} + \Delta T + \Delta T^2 + \text{Experimental venue}$	9	-94.56	208.22	1.55	0.122
#6	$g_W \sim \text{Ecosystem type} + \text{Resource generation time} + \Delta T + \Delta T^2 + \text{Experimental venue}$	8	-95.78	208.43	1.76	0.11
(c) Effect of top-down control on resource biomass (g_C)						
#1	$g_C \sim \text{Ecosystem type} + \text{MAT} + \text{Number of trophic levels}$	6	-94.75	202.01	0	0.246
#2	$g_C \sim \text{Ecosystem type} + \text{Number of trophic levels}$	5	-96.14	202.63	0.62	0.181
#3	$g_C \sim \text{Ecosystem type} + \text{MAT} + \text{Number of trophic levels} + \text{Experimental venue}$	7	-94.14	202.95	0.94	0.154
#4	$g_C \sim \text{Ecosystem type} + \text{Number of trophic levels} + \text{Experimental venue}$	6	-95.49	203.49	1.48	0.117
#5	$g_C \sim \text{Consumer foraging} + \text{Ecosystem type} + \text{MAT} + \text{Number of trophic levels}$	7	-94.52	203.72	1.71	0.105
#6	$g_C \sim \text{Ecosystem type} + \text{MAT} + \text{Number of trophic levels} + \text{Resource mobility}$	7	-94.52	203.72	1.71	0.105
#7	$g_C \sim \text{Ecosystem type} + \text{Resource generation time} + \text{Number of trophic levels}$	6	-95.73	203.96	1.95	0.093

We fitted a weighted mixed-effects model for each response variable and conducted a model selection, whereby we ranked models according to their AICc values and selected the models within $\Delta\text{AICc} \leq 2$ of the model with the lowest AICc value for inference. The Akaike weight for a given model, w_i , represents the probability that a given model is the best model in the set. We used this confidence set in the model averaging. MAT: mean annual temperature (in degrees centigrade). ΔT : intensity of the experimental warming (ln-transformed, in degrees centigrade). Study duration: ln-transformed, in days.

lme4 包meta分析标准化参数估计



混合效应模型--统计效力（power）的计算

- 统计效力：当原假设 H_0 错误，且模型能拒绝该假设的概率
- 一般认为， $\text{power} > 0.8$ 表示模型具有较好的统计效力
- 也可以计算模型能检测到某一定量差异的统计效力
- 比如模型对某运动能使体重降低10kg时的统计效力

混合效应模型--统计效力 (power) 的计算 simr包

计算统计效力，指定的作用影响程度下进行

```
mt2<-lmer(y~x2+(1|species),data=multi)
```

```
summary(mt2)
```

```
library(simr)
```

```
p1<-powerSim(mt2,nsim=100)
```

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	8.935e-01	2.721e-02	1.970e+02	32.83	<2e-16	***
x2	<u>2.824e-02</u>	1.603e-03	1.970e+02	17.61	<2e-16	***

```
Power for predictor 'x2', (95% confidence interval):  
100.0% (96.38, 100.0)
```

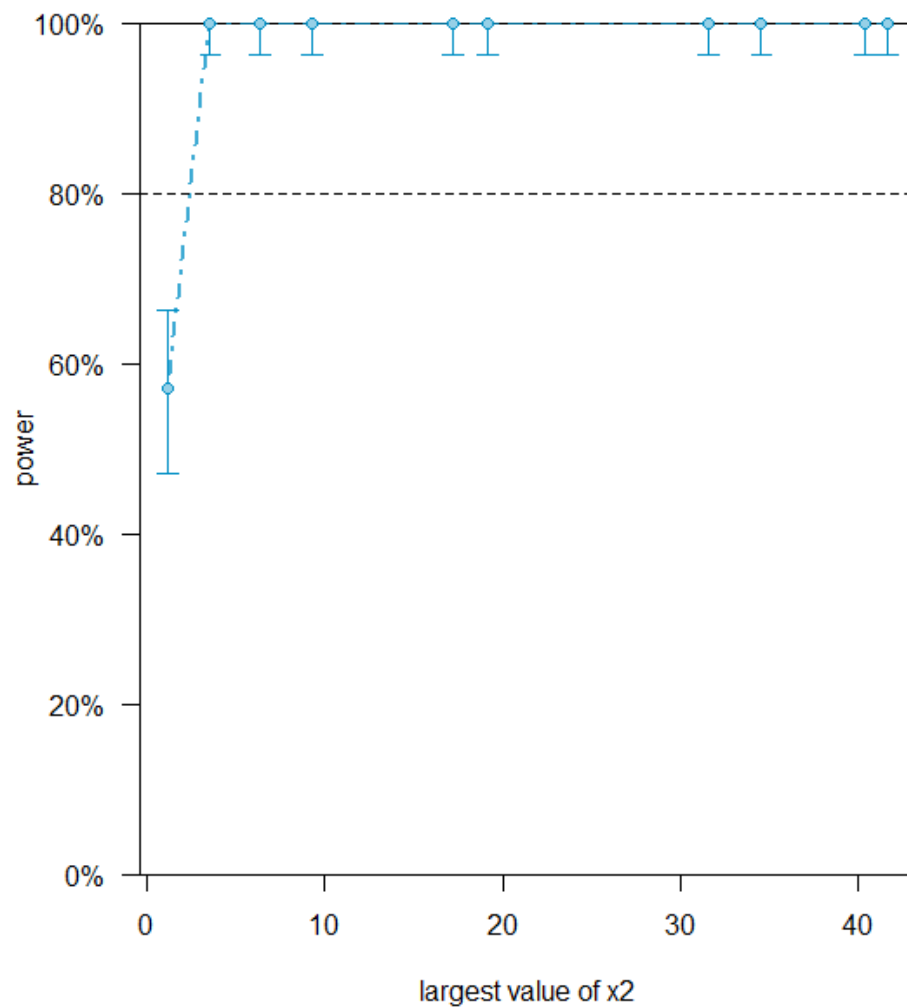
```
Test: unknown test  
Effect size for x2 is 0.028
```

```
Based on 100 simulations, (0 warnings, 0 errors)  
alpha = 0.05, nrow = 199
```

```
Time elapsed: 0 h 0 m 15 s
```

```
nb: result might be an observed power calculation
```

混合效应模型--统计效力（power）的计算



扩大自变量的取样范围是
提高统计效力的有效方法

混合效应模型--统计效力（power）的计算

- 某因素对因变量的影响越强，则模型越易于检测到这种差异，越容易拒绝原假设，即power越强。Power越强，较小的影响也能被检测到到概率越大。



混合效应模型--统计效力 (power)

检验一个模型模型对指定大小的效应的检出能力

###对指定差异大小的模型的效力计算

```
fixef(mt2)["x2"] <-0.0001
```

```
p2<-powerSim(mt2,nsim=100)
```

```
p2
```

```
pc3<-powerCurve(mt2,nsim=100)
```

```
plot(pc3)
```

```
fixef(mt2)["x2"] <-0.3
```

```
p3<-powerSim(mt2,nsim=100)
```

```
p3
```

```
> fixef(mt2)["x2"] <-0.0001
> p2<-powerSim(mt2,nsim=100)
> p2lating: |=====
Power for predictor 'x2', (95% confidence interval):
      1.00% ( 0.03,  5.45)

Test: unknown test
      Effect size for x2 is 0.00010

Based on 100 simulations, (0 warnings, 0 errors)
alpha = 0.05, nrow = 199

Time elapsed: 0 h 0 m 15 s
```

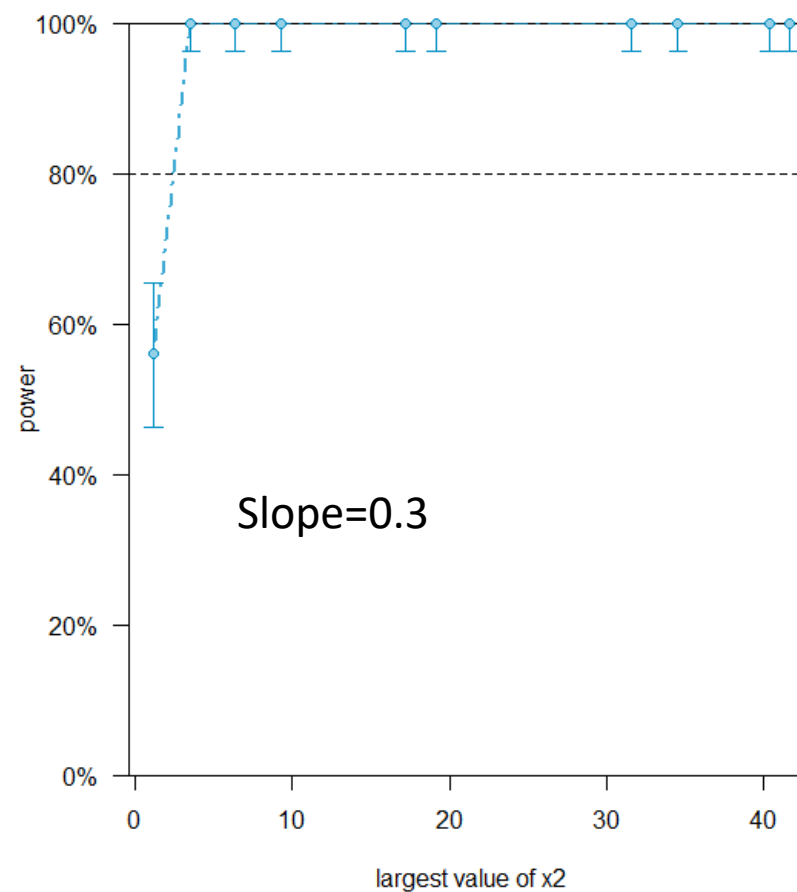
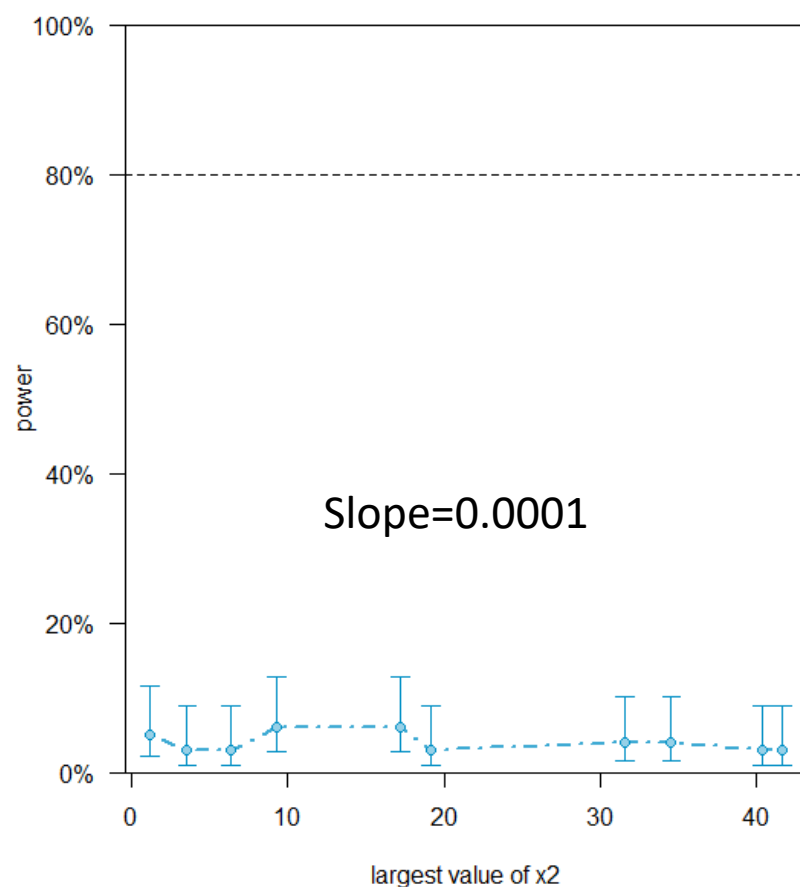
```
> fixef(mt2)["x2"] <-0.3
> p3<-powerSim(mt2,nsim=100)
> p3lating: |=====
Power for predictor 'x2', (95% confidence interval):
     100.0% (96.38, 100.0)

Test: unknown test
      Effect size for x2 is 0.30

Based on 100 simulations, (0 warnings, 0 errors)
alpha = 0.05, nrow = 199

Time elapsed: 0 h 0 m 16 s
```

混合效应模型--统计效力（power）的计算



`mt2<-lmer(y~x2+(1|species),data=multi)`，根据现有数据，当 x_2 对 y 的影响的实际斜率为0.0001时，模型mt2很难检出此效应，但是如果该影响的斜率为0.3，那么mt2模型则很容易检出这一效应。

技术问题已非第一难题！
认识你的数据及数据结构，至关重要！
统计，只是为了回答你的科学问题！

不忘初心， 保持学习， 不断进步！！！！



我的个人公众号

谢谢大家，以后多交流！！！！