

广义线性混合模型

(generalized linear mixed effect model, GLMM)

张 霜

中国科学院生态环境研究中心

2021-05-09



北京凌云翼数据科技有限公司

事实上，对违反正态分布假设的数据 线性混合模型的结果仍具有一定的可靠性

- Model estimates were usually robust to violations of assumptions, **with the exception** of slight upward biases in estimates of random effect variance **if the generating distribution was bimodal but was modelled by Gaussian error distributions.**
- Further, estimates for (random effect) components that violated distributional assumptions became less precise but remained unbiased. However, this particular problem did not affect other parameters of the model.
- The same pattern was found for strongly correlated fixed effects, which led to imprecise, but unbiased estimates, with uncertainty estimates reflecting imprecision.

Schielzeth, H. *et al.* Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution* **n/a**, doi:10.1111/2041-210x.13434 (2020).

当然，对于不满足正态性假设的数据，
还是广义线性混合模型比较安全，让你远
离被批判或撤稿风险！

glmm-依据数据类型选取合适的回归模型

数据类型	案例	理论分布	对应模型
连续变量	如身高、体重、血压、作物产量等，可以对数据进行log, logit等装换，以促进其正态性	正态分布	一般线性模型
计数数据	物种个数，病人个数等，可能存在过度离散或零膨胀问题	泊松分布（均值=方差）	泊松回归 伪泊松回归 负二项回归 零膨胀泊松回归 零膨胀负二项回归
0,1数据	成功/失败，男/女，是否出现	二项分布	二项回归
由计数数据转化而来的比率数据	成功率，康复率（可能存在过度离散或零膨胀）	二项分布	二项回归 伪二项回归（过度离散） 零膨胀二项回归
非负数据，但0较多	每个人逛街的花费		Hurdle模型
连续数据型比率数据	土壤含水率，空气湿度	Beta分布	Beta回归（也可logit转换后线性回归）
连续变量[0，正无穷)	生物量	Gamma分布（均值越大，方差越大）	Gamma回归

广义线性模型

广义线性混合模型（GLMM）

View(RIKZ)

Sample	Richness	Exposure	NAP	Beach	fBeach
1	11	10	0.045	1	1
2	10	10	-1.036	1	1
3	13	10	-1.336	1	1
4	11	10	0.616	1	1
5	10	10	-0.684	1	1
6	8	8	1.190	2	2
7	9	8	0.820	2	2
8	8	8	0.635	2	2
9	19	8	0.061	2	2
10	17	8	-1.334	2	2
11	6	11	-0.976	3	3
12	1	11	1.494	3	3
13	4	11	-0.201	3	3

Richness 为计数数据，所以应采用与计数数据对应的广义线性混合模型来分析，

1) GLMM with poisson zero structure

2) GLMM with negative binomial zero structure

并考虑是否存在过度离散或零膨胀问题

GLMM的参数估计方法

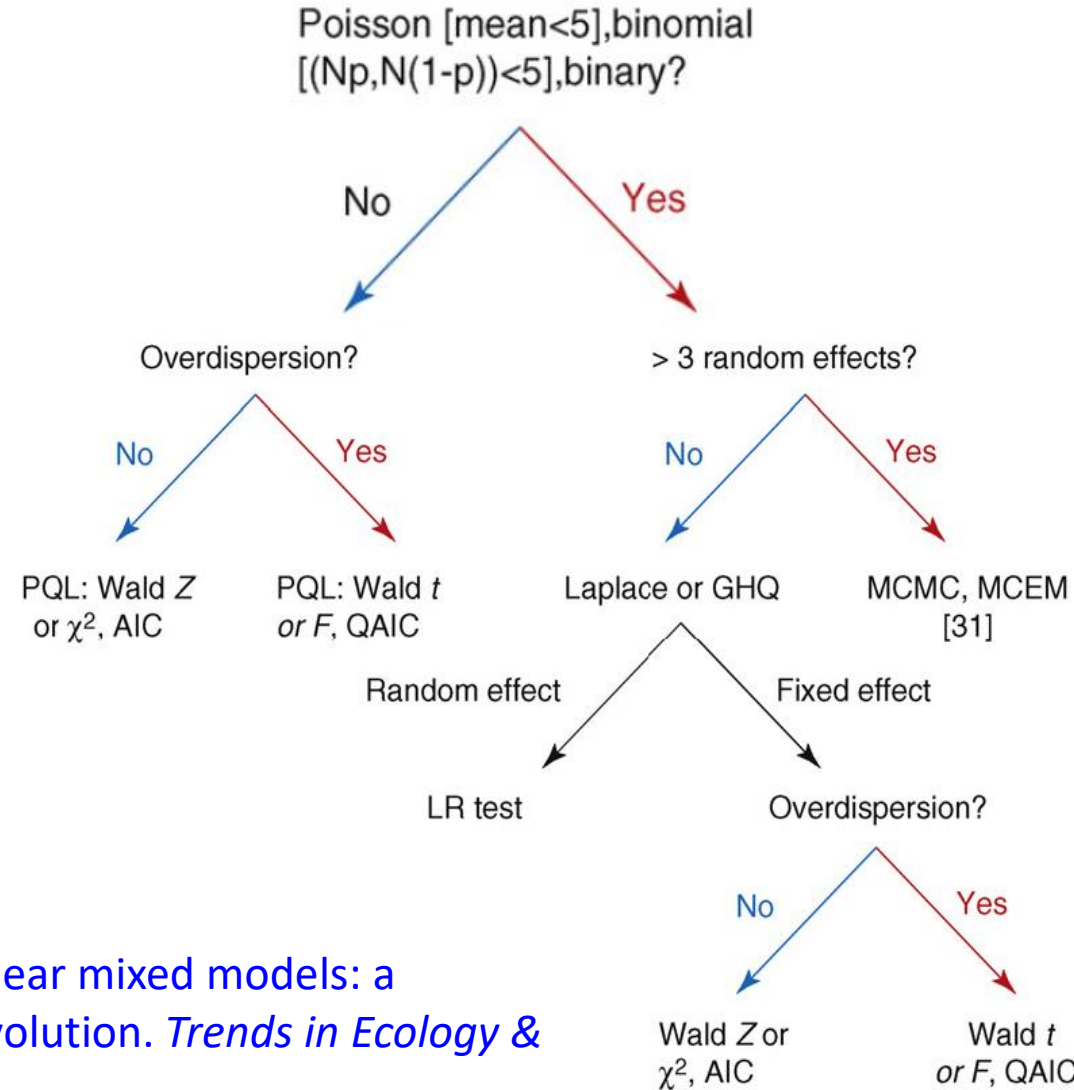
What methods are available to fit (estimate) GLMMs?

(adapted from Bolker et al TREE 2009)

Method	Advantages	Disadvantages	Packages
Penalized quasi-likelihood	Flexible, widely implemented	Likelihood inference may be inappropriate; biased for large variance or small means	PROC GLIMMIX (SAS), GLMM (GenStat), glmmPQL (R:MASS), ASREML-R
Laplace approximation	More accurate than PQL	Slower and less flexible than PQL	glmer (R:lme4,lme4a), glmm.admb (R:glmmADMB), INLA, glmmTMB, AD Model Builder, HLM
Gauss-Hermite quadrature	More accurate than Laplace	Slower than Laplace; limited to 2-3 random effects	PROC NLMIXED (SAS), glmer (R:lme4, lme4a), glmmML (R:glmmML), xtlogit (Stata)
Markov chain Monte Carlo	Highly flexible, arbitrary number of random effects; accurate	Slow, technically challenging, Bayesian framework	MCMCglmm (R:MCMCglmm), rstanarm (R), brms (R), MCMCpack (R), WinBUGS/OpenBUGS (R interface: BRugs/R2WinBUGS), JAGS (R interface: rjags/R2jags), AD Model Builder (R interface: R2admb), glmm.admb (post hoc MCMC after Laplace fit) (R:glmmADMB)

截图来源: <https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>

glmm模型参数估计方法的选择



Bolker, B. M. *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* **24**, 127-135 (2009).

广义线性混合模型GLMM— 依据数据特征选取参数估计方法

- 如果数据类型不符合正态分布，那么最大似然（ML）与限制性最大似然法(REML)对参数的估计都不再适用，因为二者都是在数据残差符合正态分布的前提下才可应用。
- 广义线性混合模型的参数估计统计量为Z值(即判断参数是否和零有显著差异的依据)，而线性混合模型的参数估计统计量为t值

广义线性混合模型GLMM— 依据数据特征选取参数估计方法

- 方法1， 采用penalized quasilikelihood (PQL)法， PQL法是一种弹性技术， 可以处理非正态， 不平衡的数据结构， 随机效应有交叉的情况(mass::glmmPQL)。
- 但是， 如果响应变量为计数型的离散分布（如泊松）， 均值小于5， 这时， PQL法会带来具有偏爱性的估计。所以泊松分布均值小于5时， 不宜采用PQL法。

所以计数数据， 且均值 >5 ， 可选用PQL法

广义线性混合模型GLMM— 依据数据特征选取参数估计方法

- 方法二：计数数据，或0/1数据，采用the Laplace approximation法（glmer in lme4, glmmTMB），但此方法仅能处理不多于3个随机变量的模型。
- 方法三：Markov chain Monte Carlo algorithms (MCMC)，在数据足够支持的前提下，能处理任意个随机变量的模型。

随机变量不多于3个， the Laplace approximation （ glmer in lme4 ）

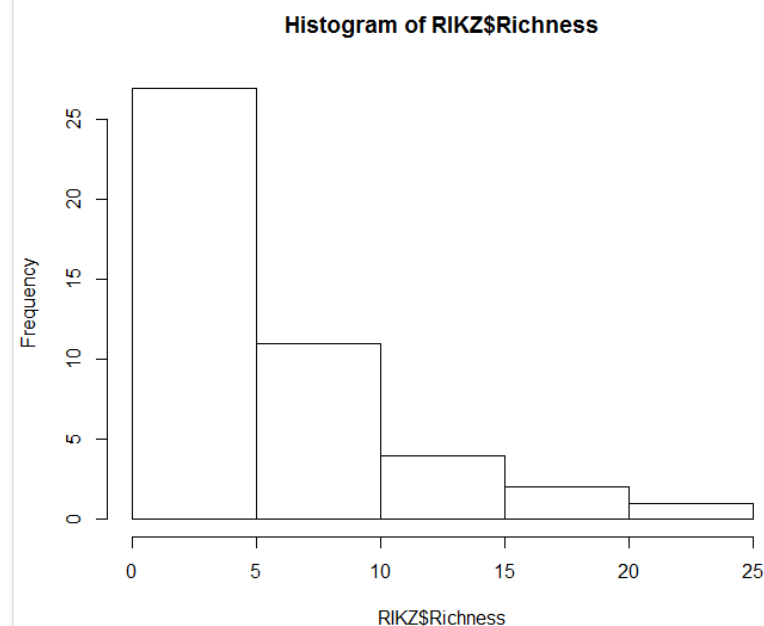
任意多个随机变量， MCMC （ 贝叶斯法 ）

广义线性混合模型GLMM— 依据数据特征选取参数估计方法

- 泊松分布，均值 >5 ，PQL
- the Laplace approximation, 随机变量不超过3个（lme4）
- Markov chain Monte Carlo algorithms (MCMC)，任意多个随机变量

示例数据RIKZ中，物种丰富度Richness为计数数据，且均值大于5

```
> mean(RIKZ$Richness)
[1] 5.688889
```



广义线性混合模型

GLMM-with poisson error structure-- penalized quasiliikelihood (PQL) 法

用此法时，对于泊松分布，其均值大于5时

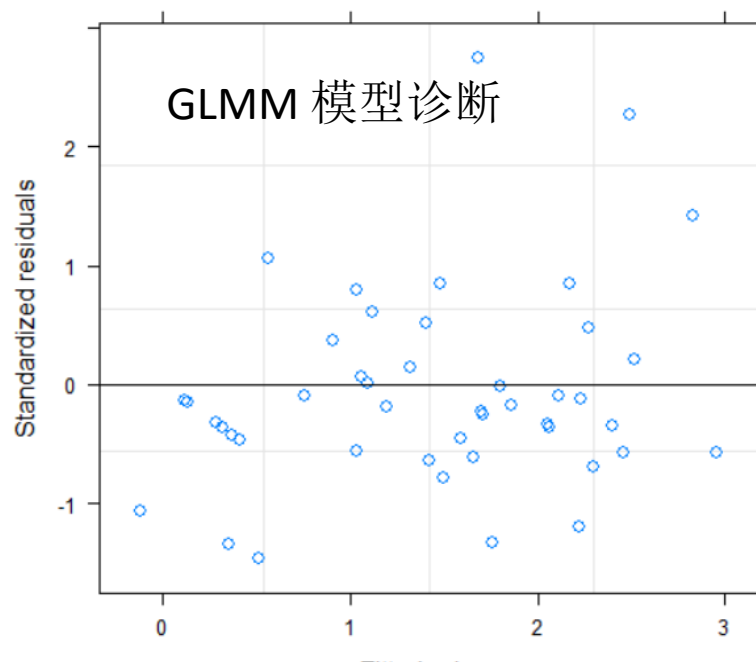
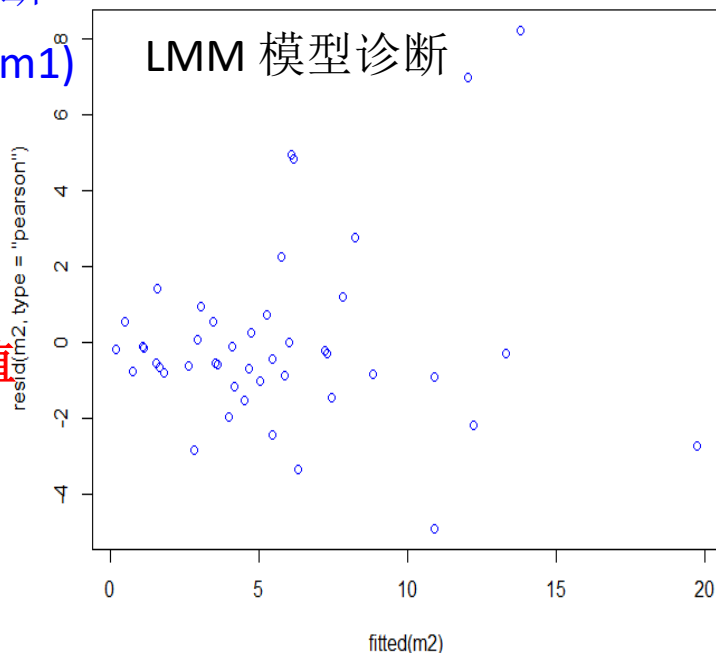
```
library(MASS)
```

```
PQL_m1 <- glmmPQL(Richness~NAP, ~NAP | fBeach, family = poisson(link = "log"), data = RIKZ, verbose = FALSE)
```

```
summary(PQL_m1)
```

##模型诊断

```
plot(PQL_m1)
```



```
> library(r2glmm)
> r2beta(PQL_m1)
```

	Effect	Rsqr	upper.CL	lower.CL
1	Model	0.628	0.794	0.423
2	NAP	0.628	0.794	0.423

缺陷：目前PQL法
无法获得模型AIC值

GLMM-with poisson error structure— the Laplace approximation

##两种意义相同的写法

```
gm1<-glmer(Richness~NAP+(NAP|fBeach), data =RIKZ, family = poisson(link = "log"))
```

```
gm1_1<-glmer(Richness~NAP+(NAP|fBeach), data =RIKZ, family = "poisson")
```

其他分布类型与link函数

glmer命令可处理的
分布类型

```
binomial(link = "logit")  
gaussian(link = "identity")  
Gamma(link = "inverse")  
inverse.gaussian(link = "1/mu^2")  
poisson(link = "log")  
quasi(link = "identity", variance = "constant")  
quasibinomial(link = "logit")  
quasipoisson(link = "log")
```

```
> gm1<-glmer(Richness~NAP+(NAP|fBeach), data =RIKZ, family = "quasipoisson")  
Error in lme4::glFormula(formula = Richness ~ NAP + (NAP | fBeach), data = RIKZ, :  
  "quasi" families cannot be used in glmer
```

GLMM-with poisson error structure— the Laplace approximation 法

```
gm1<-glmer(Richness~NAP+(NAP|fBeach), data =RIKZ, family = poisson(link = "log"))
```

```
> summary(gm1)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: poisson ( log )
Formula: Richness ~ NAP + (NAP | fBeach)
Data: RIKZ
```

```
      AIC      BIC    logLik deviance df.resid
  218.7    227.8   -104.4    208.7      40
```

Scaled residuals:

```
      Min       1Q   Median       3Q      Max
-1.35846 -0.51129 -0.21846  0.09802  2.45384
```

Random effects:

```
Groups Name      Variance Std.Dev. Corr
fBeach (Intercept) 0.2630   0.5128
      NAP         0.0891   0.2985  0.18
```

Number of obs: 45, groups: fBeach, 9

Fixed effects:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.6942     0.1868    9.071 < 2e-16 ***
NAP          -0.6074     0.1374   -4.421 9.81e-06 ***
```

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

Correlation of Fixed Effects:

```
(Intr)
```

```
NAP 0.121
```

```
tab_model(gm1, show.icc=T)
```

Richness			
Predictors	Incidence Rate Ratios	CI	p
(Intercept)	5.44	3.77 – 7.85	<0.001
NAP	0.54	0.42 – 0.71	<0.001
Random Effects			
σ^2	0.17		
τ_{00} fBeach	0.26		
τ_{11} fBeach.NAP	0.09		
ρ_{01} fBeach	0.18		
ICC	0.69		
N _{fBeach}	9		
Observations	45		
Marginal R2 / Conditional R2	0.400 / 0.815		

随机效应部分

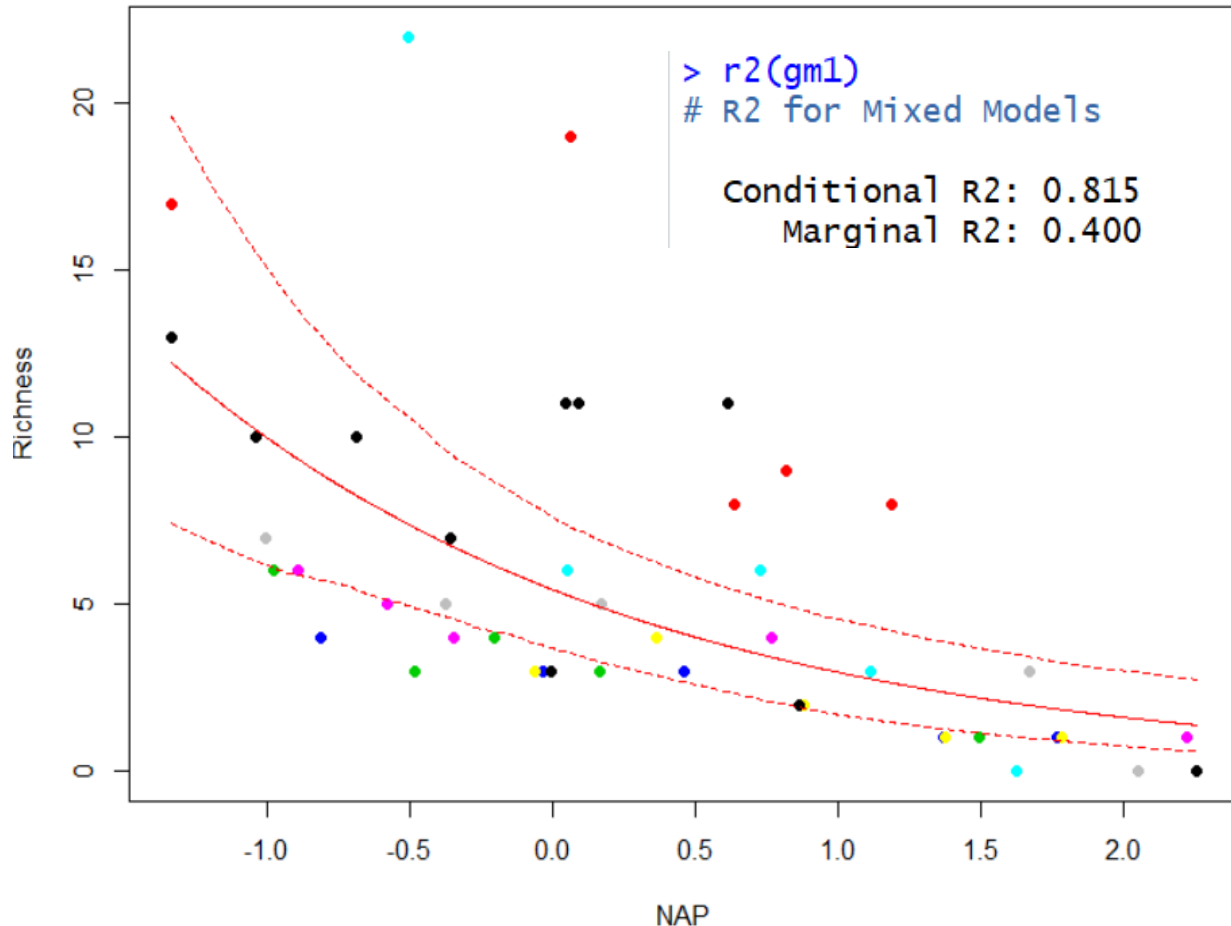
固定效应部分

$$\text{Log(Richness)} = 1.6942 - 0.6074 * \text{NAP}$$

GLMM-with poisson error structure– the Laplace approximation 法—结果制图

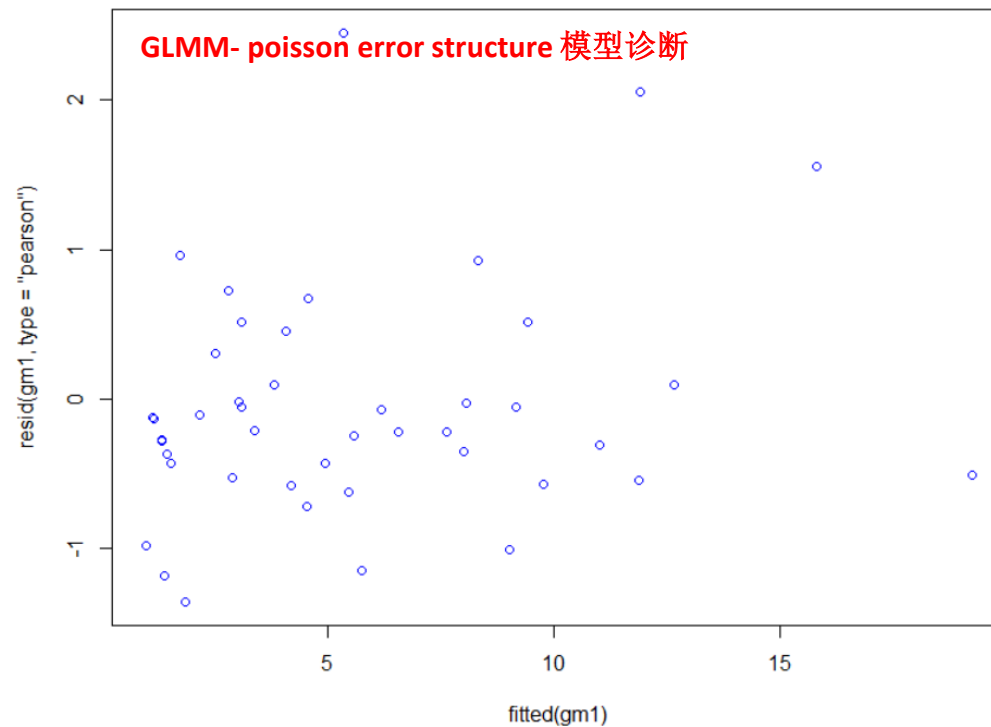
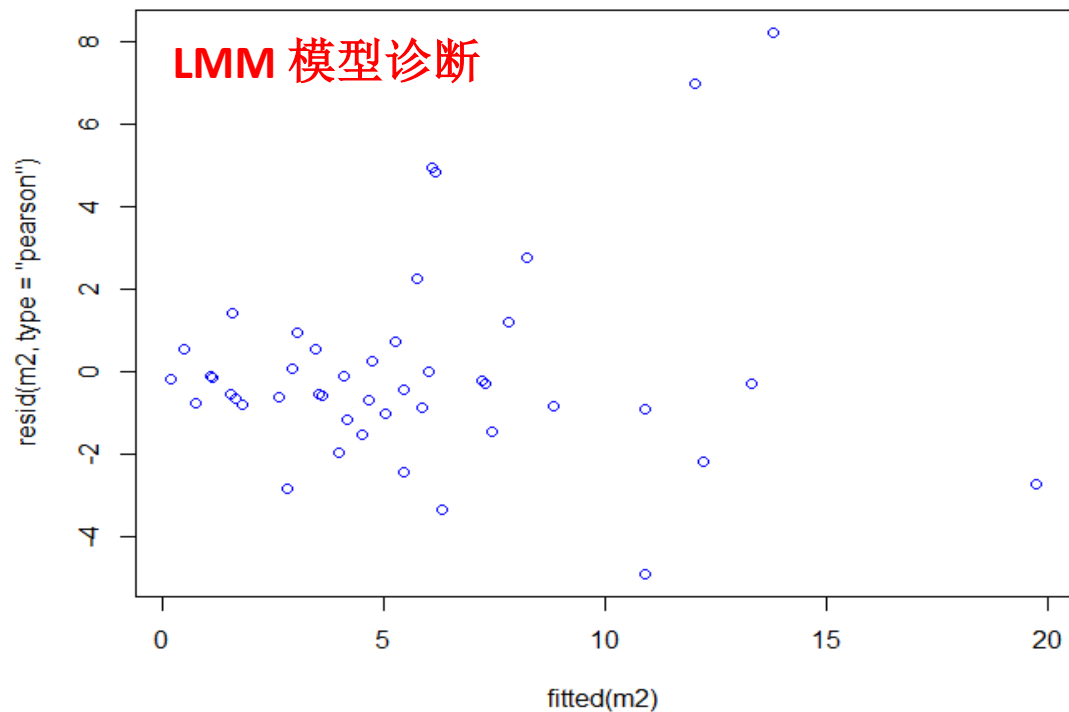
```
new_NAP<-data.frame(NAP=seq(-1.336,2.2550,length=1000))
pred_R3<-exp(predict(gm1,newdata=new_NAP,re.form=~0))
ci_line3<-bootMer(gm1,FUN=function(.)
predict(.,newdata=new_NAP,
          re.form=~0),nsim=1000)
ci_R3<-apply(ci_line3$t,2,function(x) x[order(x)][c(25,975)])
lb_R3<-exp(ci_R3[1,])
ub_R3<-exp(ci_R3[2,])

plot(Richness~NAP,data=RIKZ,col=fBeach, pch=19)
lines(new_NAP$NAP,pred_R3,lwd=1,col="red")
lines(new_NAP$NAP,lb_R3,lty=2,col="red")
lines(new_NAP$NAP,ub_R3,lty=2,col="red")
```



GLMM-with poisson error structure— the Laplace approximation 法—模型诊断

```
plot(fitted(gm1),resid(gm1,type="pearson"),col="blue")
```



GLMM-with poisson error structure— 过度离散 (overdispersion)

➤如果模型存在过度离散，则可以加入一个水平数量与样本量相等的随机变量

```
> check_overdispersion(gm1)
# Overdispersion test
```

```
dispersion ratio = 0.660
Pearson's Chi-Squared = 26.402
p-value = 0.952
```

No overdispersion detected.

<https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>

```
gm2<-glmer(Richness~NAP+(NAP|fBeach)+(1|Sample),
data =RIKZ, family = poisson(link = "log"))
```

Richness			
Predictors	Incidence Rate Ratios	CI	p
(Intercept)	5.40	3.75 – 7.80	<0.001
NAP	0.54	0.42 – 0.71	<0.001
Random Effects			
σ^2	0.18		
τ_{00} Sample	0.01		
τ_{00} fBeach	0.26		
τ_{11} fBeach.NAP	0.08		
ρ_{01} fBeach	0.21		
ICC	0.67		
N fBeach	9		
N Sample	45		
Observations	45		
Marginal R ² / Conditional R ²	0.401 / 0.801		

```
> AIC(gm1, gm2)
      df      AIC
gm1    5 218.7413
gm2    6 220.6535
```

实际应用案例 Poisson models included an observation-specific random effect to account for potential overdispersion.

Blowes, S. *et al. Science* **366**, 339-345, (2019).

GLMM— Gauss-Hermite quadrature (GHQ)法

- 严格来讲，Laplace approximation(lme4:glmer的默认值)是参数估计方法GHQ里的一种特殊情况，只进行一次迭代。
- GHQ,由于可以进行重复的迭代，比Laplace更精确。
- 但GHQ法的可塑性较差，只适用于仅含一个随机效应的情况。

```
GHQ_m1<-glmer(Richness~NAP+(1|fBeach),nAGQ = 30, data =RIKZ, family = poisson(link = "log"))
```

```
summary(GHQ_m1)
```

```
tab_model(GHQ_m1)
```

Richness			
Predictors	Incidence Rate Ratios	CI	p
(Intercept)	5.27	3.75 – 7.42	<0.001
NAP	0.60	0.52 – 0.70	<0.001

Random Effects

σ^2	0.19
τ^2_{00} fBeach	0.23
ICC	0.55
N fBeach	9

Observations	45
Marginal R2 / Conditional R2	0.378 / 0.718

```
> r2(GHQ_m1)
```

```
# R2 for Mixed Models
```

```
Conditional R2: 0.718  
Marginal R2: 0.378
```

glmer-poisson, binomial回归应用案例

Because the number of shoots produced (clonal propagation) consists of **count data**, we used **a generalized linear mixed-effects model in order to specify a Poisson error distribution**. This model also included a fixed effect of temperature treatment as well as a random effect of mesocosm.

```
m_SH <- glmer (Shoot_Count ~ SM2Treatment + SM1Treatment + (1|SM2_Bin), data=T2, na.action = na.exclude,  
family="poisson")
```

We modeled species-specific colonization and extinction rates **by fitting generalized linear mixed-effects models with binomial errors and logit link functions**. Patch and year were treated as random factors.

```
mMelCc <- glmer(MelCc ~ logPLM2+VS+logNH+Road+logMayPrec+(1|Patch)+(1|Year), family="binomial",  
data=MelCc_dat)
```

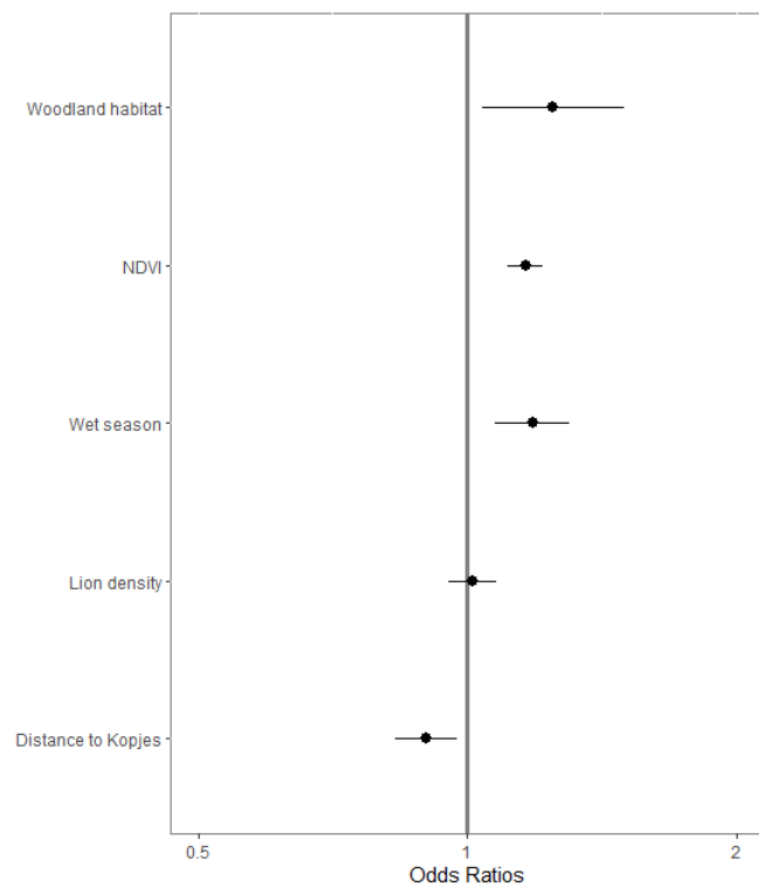
DuBois, K., Williams, S. L. & Stachowicz, J. J. Previous exposure mediates the response of eelgrass to future warming via clonal transgenerational plasticity. *Ecology* **101**, e03169, (2020).

Opedal, Ø. H., Ovaskainen, O., Saastamoinen, M., Laine, A.-L. & van Nouhuys, S. Host-plant availability drives the spatiotemporal dynamics of interacting metapopulations across a fragmented landscape. *Ecology* **101**, e03186, (2020).

广义线性混合模型应用案例

```
model <- glmer(assoc ~ season + BINHAB + scale(encount_risk) + scale(KOP.DIST.M) + scale(NDVI) + (1|SiteID),  
family=binomial, data=grazers, control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
```

```
summary(model)
```



可促进模型收敛

Beaudrot, L., M. S. Palmer, T. M. Anderson, and C. Packer. 2020. Mixed-species groups of Serengeti grazers: a test of the stress gradient hypothesis. *Ecology*.

glmm- negative binomial model

```
m_nb<-glmer.nb(y~x+(1|group), data=d)
```

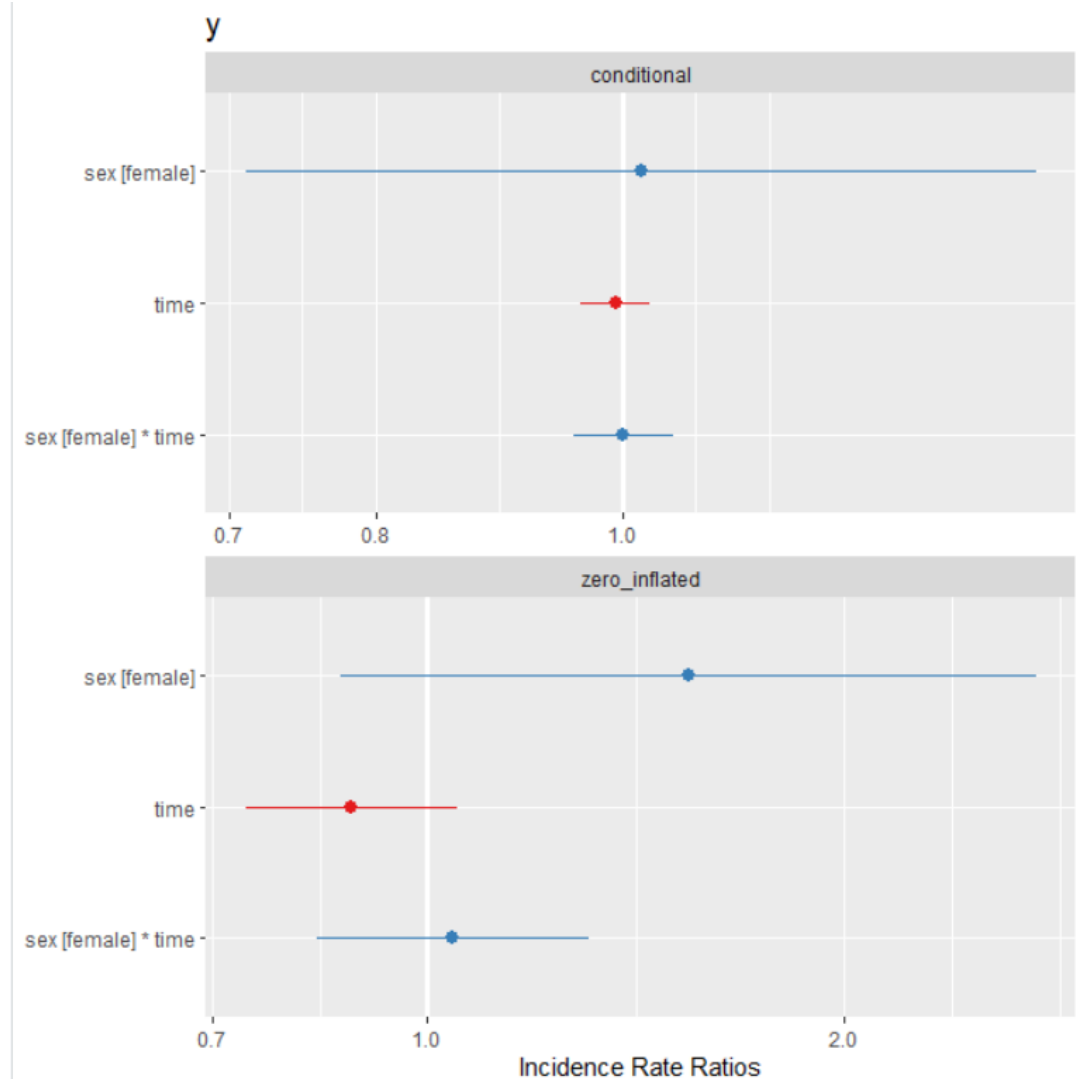
充分了解你的数据是选择合适模型的前提

glmm-zero inflated poisson model

➤ glmm- zero-inflated poisson model

`library(GLMMadaptive)`###要加载此包

```
glmm_zero_inflated_p <- mixed_model(y ~ sex  
* time, random = ~ 1 | id, data = DF, family =  
zi.poisson(), zi_fixed = ~ sex* time)
```

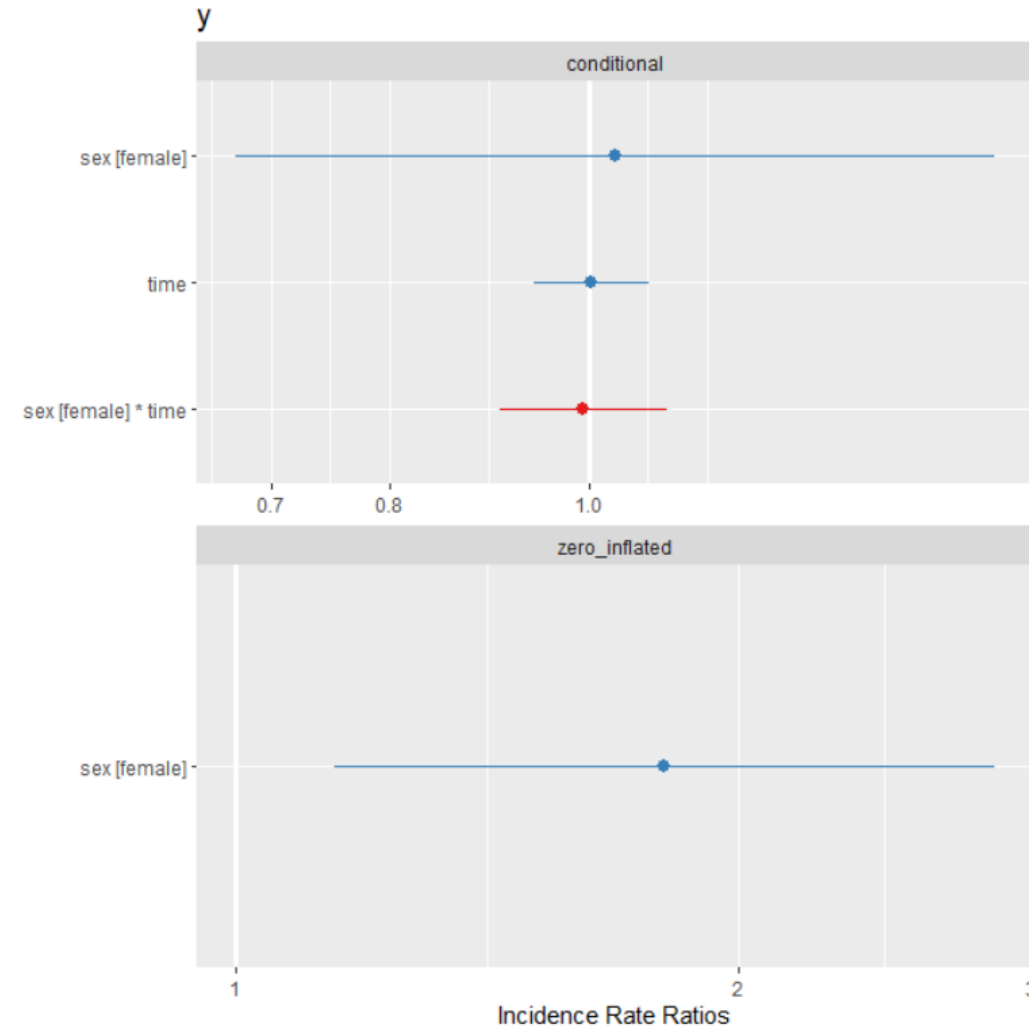


glmm-zero inflated negative binomial model

➤ glmm- zero-inflated negative binomial model

```
glmm_zero_inflated_nb <- mixed_model(y ~ sex * time,  
random = ~ 1 | id, data = DF,
```

```
family = zi.negative.binomial(), zi_fixed = ~ sex)
```

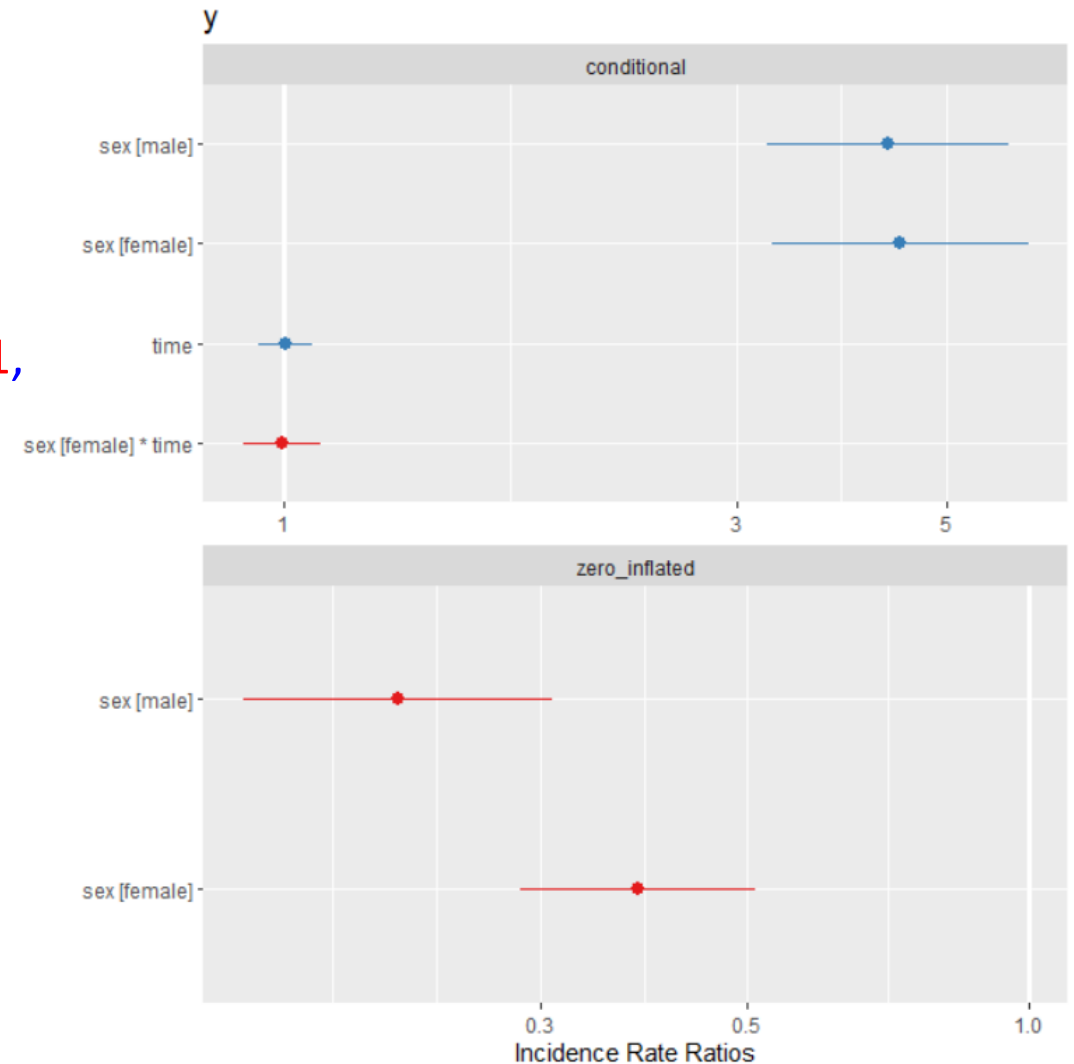


glmm- zero-inflated negative binomial model

去掉截距， 直接查看模型对各因素不同水平的参数估计值

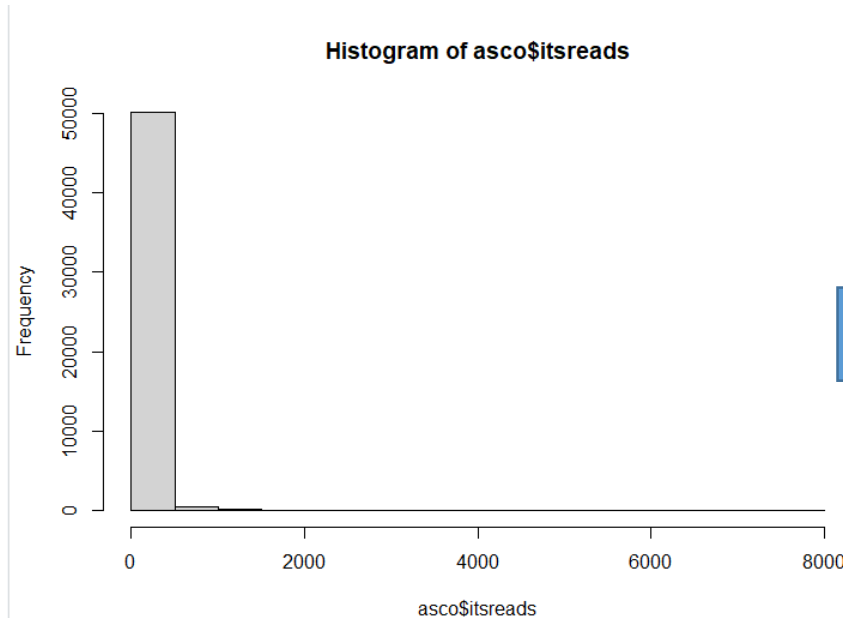
➤ glmm- zero-inflated negative binomial model

```
glmm_zero_inflated_nb2 <- mixed_model(y ~ sex * time-1,  
random = ~ 1 | id, data = DF,  
family = zi.negative.binomial(),  
zi_fixed = ~ sex-1)  
plot_model(glmm_zero_inflated_nb2)
```

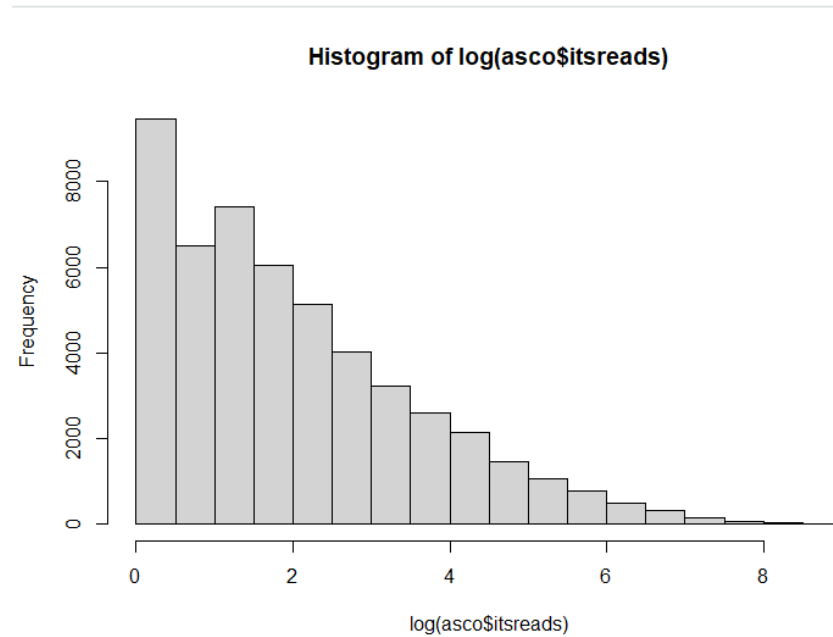


glmm-gamma regresion

We used linear mixed effects models (for normally distributed data) or **generalized linear models with a Gamma distribution** in the “lmer” and “glmer” functions in the *lme4* package in R. **Read abundances (logged, zeroes removed)** of microbial phyla were the response variable, vegetation type (**woody/herbaceous**) was a **fixed effect** and **site was included as a random effect**.



`hist(asco$itsreads)# not normal`



`hist(log(asco$itsreads))#not normal`

```
asco_veg<-glmer((log(itsreads)+1)~ veg + (1|site), asco, family = Gamma)
```

Collins, C. G. *et al.* Belowground impacts of alpine woody encroachment are determined by plant traits, local climate, and soil conditions. *Global Change Biol.* **26**, 7112-7127, doi:<https://doi.org/10.1111/gcb.15340> (2020).

glmm-gamma regression model

```
> asco_veg<-glmer((log(itsreads)+1)~veg + (1|site), asco, family = Gamma)
> summary(asco_veg)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: Gamma (inverse)
Formula: (log(itsreads) + 1) ~ veg + (1 | site)
Data: asco

            AIC      BIC   logLik deviance df.resid
182295.3 182330.7 -91143.7 182287.3    50885

Scaled residuals:
    Min       1Q   Median       3Q      Max
-1.3551 -0.7780 -0.1783  0.6090  4.6695

Random effects:
 Groups   Name      Variance Std.Dev.
 site    (Intercept) 0.000358 0.01892
 Residual                    0.295742 0.54382
Number of obs: 50889, groups: site, 13

Fixed effects:
              Estimate Std. Error t value Pr(>|z|)
(Intercept)  0.333774   0.009536   35.00  <2e-16 ***
vegS         -0.002776   0.001624   -1.71   0.0873 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr)
vegS -0.091
```

Collins, C. G. *et al.* Belowground impacts of alpine woody encroachment are determined by plant traits, local climate, and soil conditions. *Global Change Biol.* **26**, 7112-7127, doi:<https://doi.org/10.1111/gcb.15340> (2020).

其他可完成GLMM模型的R包

The most commonly used functions for mixed modeling in R are

- *linear mixed models*: `aov()`, `nlme::lme`¹, `lme4::lmer`; `brms::brm`
- *generalized linear mixed models (GLMMs)*
 - frequentist: `MASS::glmmPQL`, `lme4::glmer`; `glmmTMB`
 - Bayesian: `MCMCglmm::MCMCglmm`; `brms::brm`
- *nonlinear mixed models*: `nlme::nlme`, `lme4::nlmer`; `brms::brm`
- *GNLMMs*: `brms::brm`

推荐网页：GLMM常见问题

<https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>

glmmTMB包，处理零膨胀有一套

Feature	glmmTMB	glmmADMB	MCMCglmm	brms	INLA	mgcv	gamlss
hurdle models ¹	✓		✓	✓	✓		✓
predictors of zero-inflation	✓		✓	✓			✓
predictors of dispersion	✓			✓			✓
zero-truncated distributions	✓	✓	✓	✓	✓		✓
nbinom2 distribution	✓	✓		✓	✓	✓ ²	✓
nbinom1 distribution	✓	✓					✓ ²
compois distribution	✓ ³						
Delaporte distribution							✓ ²
Sichel distribution							✓
geometric distribution			✓	✓			✓ ²
PIG distribution							✓
weights	✓ ⁴	✓		✓	✓	✓	✓
offsets	✓	✓	5	✓	✓	✓	
various RE structures	✓ ⁶		✓	✓	✓	✓	
RE cor across components ⁷			✓	✓			
MLE	✓	✓				✓	✓
MCMC samples ⁸	✓ ⁹	✓	✓	✓	✓ ¹⁰	✓	
multivariate responses ¹¹			✓	✓			
GAM ¹²				✓		✓	✓

glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling

[\[PDF\] ethz.ch](#)

[ME Brooks, K Kristensen, KJ Van Benthem... - The R ..., 2017 - research-collection.ethz.ch](#)

Count data can be analyzed using generalized linear mixed models when observations are correlated in ways that require random effects. However, count data are often zero-inflated, containing more zeros than would be expected from the typical error distributions. We ...

☆ 被引用次数: 1505 相关文章 所有 11 个版本

glmmTMB案例1：一篇疯狂使用广义线性混合模型的Science论文

SHARE

REPORT



The human dimension of biodiversity changes on islands

Sandra Nogué^{1,†,*}, Ana M. C. Santos^{2,3,4,5}, H. John B. Birks^{6,7}, Svante Björck⁸, Alvaro Castilla-Beltrán¹, ...

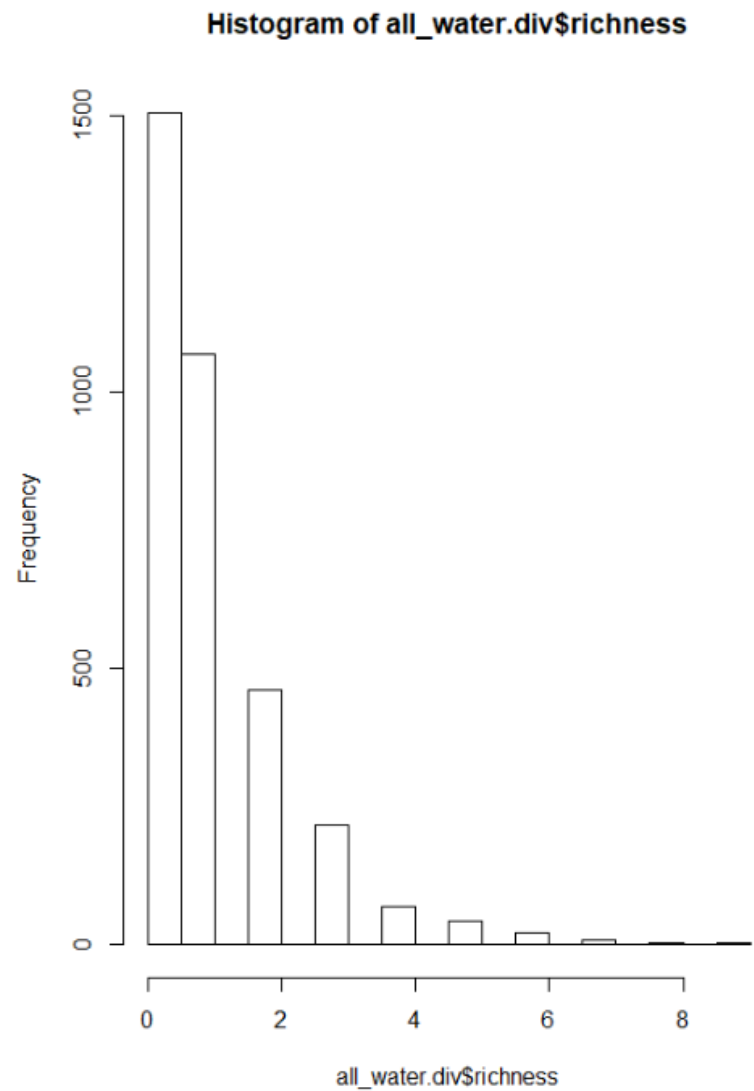
+ See all authors and affiliations

Science 30 Apr 2021:
Vol. 372, Issue 6541, pp. 488-491
DOI: 10.1126/science.abd6706

2021年4月30号, Science

We then used **generalized linear mixed effect models** in the R package ‘**glmmTMB**’ v1.0.1 to analyze how **resource types** (i.e. equid wells, background waters, and dry controls) and **environmental variables** influenced **vertebrate activity**. We **chose distributions appropriate for each response variable** and **nested date within site as a random effect**. We employed multi-model inference techniques to eliminate spurious and correlated environmental variables. Across all analyses, **omnibus tests** were conducted with the ‘Anova’ function in R package ‘car’ v3.0-8(35), while **post-hoc tests** were conducted using the ‘emmeans’ package v1.4.7 (36).

Y (richness) 的分布 (计数数据)



glmmTMB 包--GLMM-零膨胀泊松模型

```
richness.cels.glmmad.poisson <- glmmTMB(richness ~  
  water +  
  tmax +  
  last_3_ppt +  
  (water:tmax)+  
  (water:last_3_ppt)+  
  #last_3_tmax +  
  #last_7_ppt +  
  #last_7_tmax + # insignificant  
  #daysSinceRainInclusive +  
  # daysSinceRain +  
  #last_14_ppt +  
  # last_14_tmax + (1|numeric_day/site),  
  data=all_water.div,  
  ziformula = ~1, family=poisson)
```

Getting started with the glmmTMB package

Ben Bolker

July 1, 2020

across groups specified by block. A model of nested random effects (block within site) could be 1|site/block if block labels are reused across multiple sites, or (1|site)+ (1|block) if the nesting structure is explicit in the data and each level of block only occurs within one site. A model of crossed random effects (block and year) would be (1|block)+(1|year).

“nested date within site as a random effect “
遗憾的是其所有模型都嵌套错误

```

> summary(richness.cels.glmmad.poisson)
Family: poisson ( log )
Formula: richness ~ water + tmax + last_3_ppt + (water:tmax) + (water:last_3_ppt) +      (1 | nu
meric_day/site)
Zero inflation: ~1
Data: all_water.div

      AIC      BIC   logLik deviance df.resid
 5313.3   5379.7 -2644.6   5289.3     1864

Random effects:

Conditional model:
Groups          Name      Variance Std.Dev.
site:numeric_day (Intercept) 1.168e-10 1.081e-05
numeric_day      (Intercept) 1.700e-10 1.304e-05
Number of obs: 1876, groups: site:numeric_day, 1081; numeric_day, 705

Conditional model:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.2166906  0.1262707   1.716 0.086147 .
waterEquid-wells -0.9011003  0.2335292  -3.859 0.000114 ***
waterNatural    -0.3440486  0.1875676  -1.834 0.066615 .
tmax            0.0026582  0.0038556   0.689 0.490538
last_3_ppt     -0.0002931  0.0055047  -0.053 0.957535
waterEquid-wells:tmax 0.0405989  0.0069229   5.864 4.51e-09 ***
waterNatural:tmax 0.0262829  0.0057512   4.570 4.88e-06 ***
waterEquid-wells:last_3_ppt 0.0064125  0.0099445   0.645 0.519036
waterNatural:last_3_ppt -0.0132994  0.0104754  -1.270 0.204231
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

glmmTMB 包--GLMM-零膨胀泊松模型

```
> summary(richness.cels.glmmad.poisson)
Family: poisson ( log )
Formula: richness ~ water + tmax + last_3_ppt + (water:tmax) + (water:last_3_ppt) +
  (1 | numeric_day/site)
Zero inflation: ~1
Data: all_water.div
```

AIC	BIC	logLik	deviance	df.resid
5313.3	5379.7	-2644.6	5289.3	1864

Random effects:

Conditional model:

Groups	Name	Variance	Std.Dev.
site:numeric_day	(Intercept)	1.317e-10	1.148e-05
numeric_day	(Intercept)	1.882e-10	1.372e-05

Number of obs: 1876, groups: site:numeric_day, 1081; numeric_day, 705

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.2166906	0.1262707	1.716	0.086147 .
waterEquid-wells	-0.9011003	0.2335292	-3.859	0.000114 ***
waterNatural	-0.3440486	0.1875676	-1.834	0.066615 .
tmax	0.0026582	0.0038556	0.689	0.490539
last_3_ppt	-0.0002931	0.0055047	-0.053	0.957535
waterEquid-wells:tmax	0.0405989	0.0069229	5.864	4.51e-09 ***
waterNatural:tmax	0.0262829	0.0057512	4.570	4.88e-06 ***
waterEquid-wells:last_3_ppt	0.0064125	0.0099445	0.645	0.519037
waterNatural:last_3_ppt	-0.0132994	0.0104754	-1.270	0.204231

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

Zero-inflation model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-23.68	3197.36	-0.007	0.994

```
> # Use the car library's Anova function to produce omnibus p-values:
> Anova(richness.cels.glmmad.poisson)
Analysis of Deviance Table (Type II Wald chisquare tests)
```

Response: richness

	Chisq	Df	Pr(>Chisq)
water	153.4210	2	< 2.2e-16 ***
tmax	62.0938	1	3.275e-15 ***
last_3_ppt	0.1394	1	0.7089
water:tmax	40.9762	2	1.265e-09 ***
water:last_3_ppt	2.7354	2	0.2547

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

omnibus tests were conducted with the 'Anova' function in R package 'car'

glmmTMB 包--GLMM-零膨胀泊松模型

```
> emms1 <- emmeans(richness.cels.glmmad.poisson, ~(water*tmax))
```

```
> emms1
```

water	tmax	emmean	SE	df	lower.CL	upper.CL
Dry	31.1	0.299	0.0294	1864	0.241	0.356
Equid-wells	31.1	0.672	0.0357	1864	0.602	0.742
Natural	31.1	0.745	0.0295	1864	0.687	0.802

Results are given on the log (not the response) scale.
Confidence level used: 0.95

```
> con1 <- contrast(regrid(emms1))
```

```
> pairs(con1, by = NULL)
```

contrast	estimate	SE	df	t.ratio	p.value
Dry,31.0645127914099 effect - Equid-wells,31.0645127914099 effect	-0.609	0.0804	1864	-7.581	<.0001
Dry,31.0645127914099 effect - Natural,31.0645127914099 effect	-0.758	0.0736	1864	-10.287	<.0001
Equid-wells,31.0645127914099 effect - Natural,31.0645127914099 effect	-0.148	0.0934	1864	-1.588	0.2510

P value adjustment: tukey method for comparing a family of 3 estimates

post-hoc tests were conducted using the 'emmeans' package

对其模型嵌套进行修正后结果对比

```
richness.cels.glmmad.poisson2 <- glmmTMB(richness ~
```

```
  water +
```

```
  tmax +
```

```
  last_3_ppt +
```

```
  (water:tmax)+
```

```
  (water:last_3_ppt)+
```

```
  (1|site/numeric_day),
```

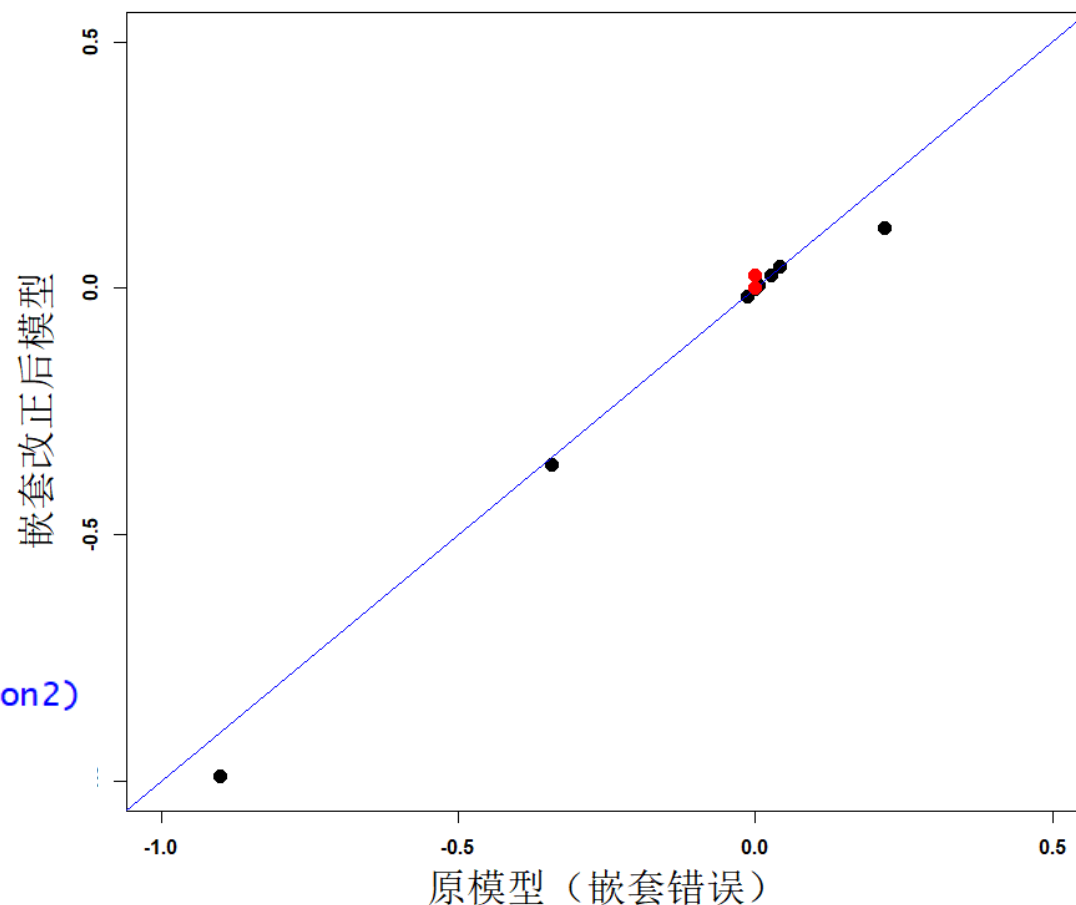
```
  data=all_water.div,
```

```
  ziformula = ~1,
```

```
  family=poisson)
```

```
> AIC(richness.cels.glmmad.poisson, richness.cels.glmmad.poisson2)
```

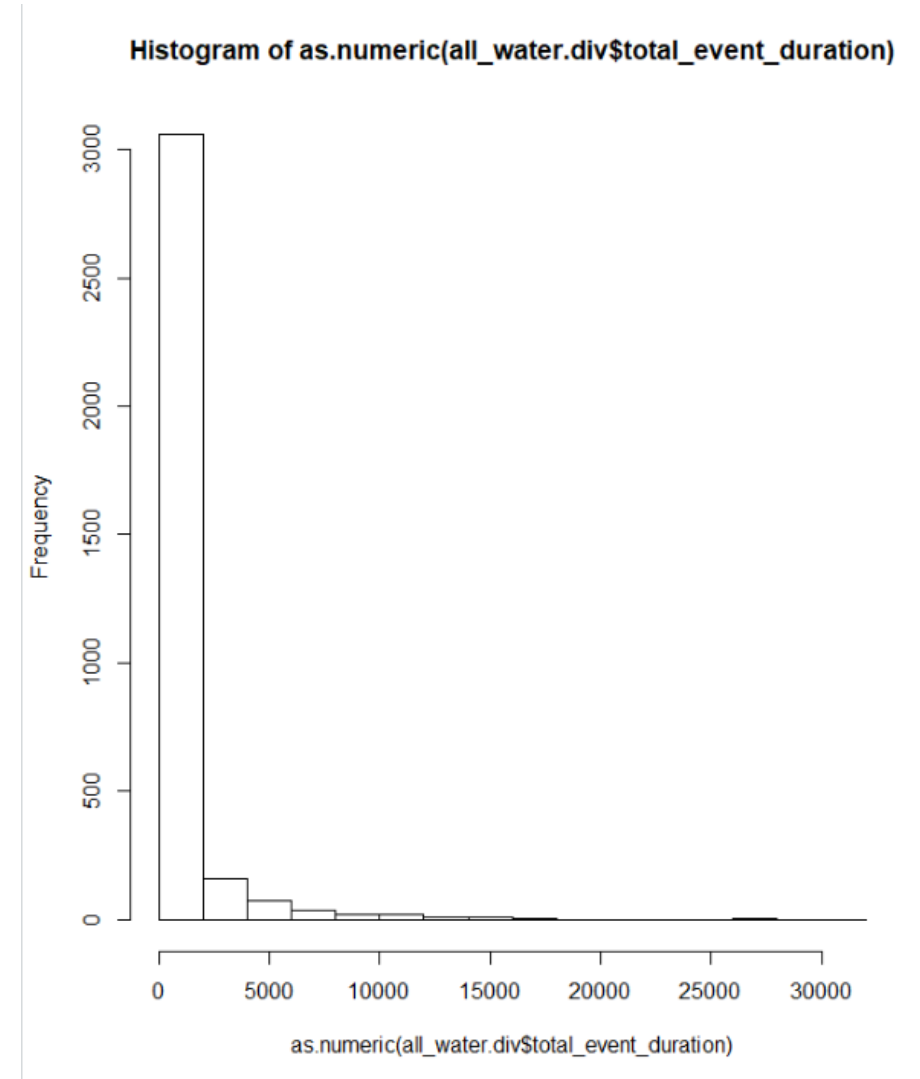
	df	AIC
richness.cels.glmmad.poisson	12	5313.282
richness.cels.glmmad.poisson2	12	5262.870



GLMM-零膨胀负二项nbinom2分布

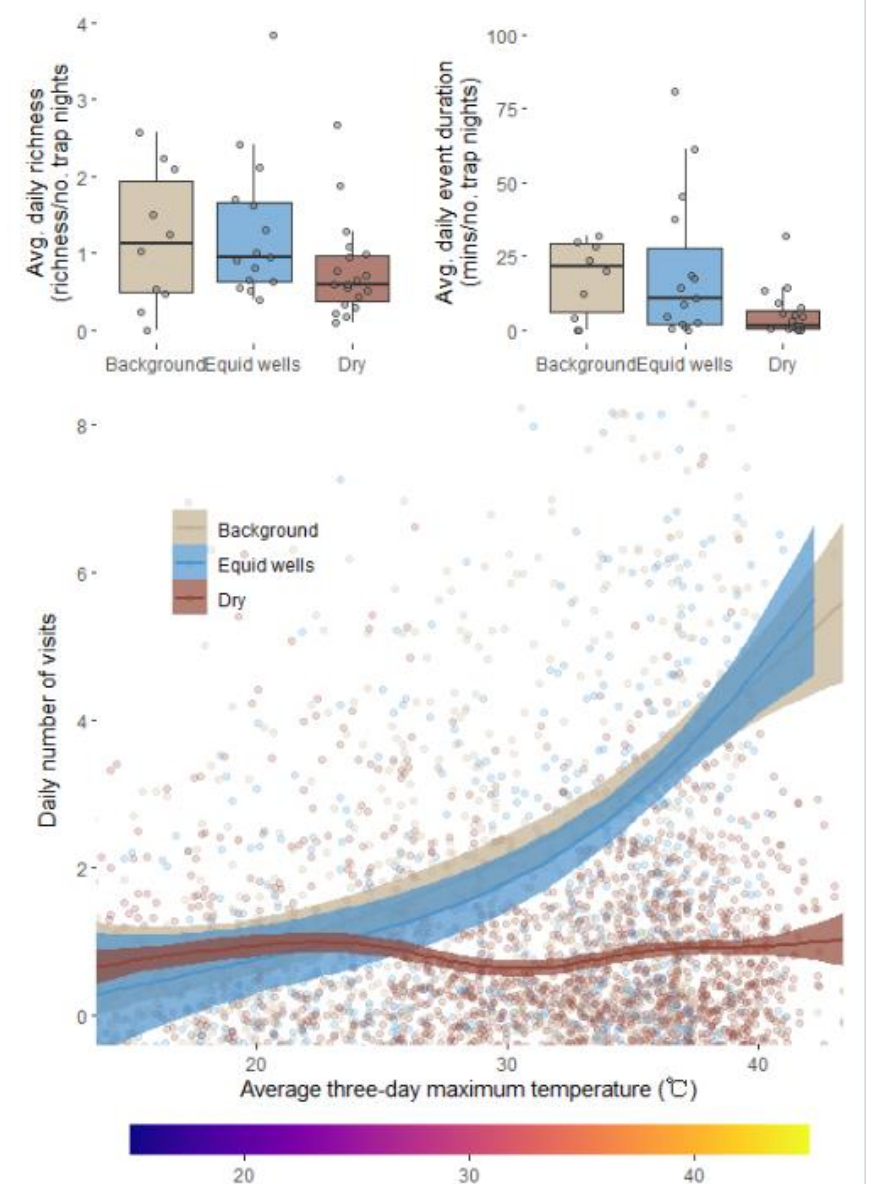
```
# negbin2 is the best fitting distributional family.  
duration.cels.glmmTMB_nbinom2 <-  
glmmTMB(as.numeric(total_event_duration) ~  
  water +  
  #tmax +  
  last_3_tmax +  
  #last_7_tmax +  
  #last_14_tmax +  
  #ppt +  
  last_3_ppt +  
  (water:last_3_tmax)+  
  (water:last_3_ppt)+  
  #last_7_ppt +  
  #last_14_ppt+  
  #daysSinceRain_mm +  
  (1|numeric_day/site),  
data=all_water.div,  
ziformula = ~0,  
family=nbinom2)
```

同样模型嵌套错误



用glmm说明x对Y的影响， 用gam模型作图

```
ggplot(data = all_water.div,  
       aes(x = last_3_tmax, y = (no_events),  
           fill = water, color = water,  
           group = water, text = last_3_tmax))+  
geom_jitter(alpha = .2, height = .5, width = 0)+  
geom_smooth(alpha = .7, method = "gam")+  
scale_fill_manual(name = NULL, values = pal, labels = labs)+  
scale_color_manual(name = NULL, values = pal, labels = labs)+  
xlab("Average three-day maximum temperature (°C)")+  
ylab("Daily number of visits")+  
coord_cartesian(ylim = c(0, 11), xlim = c(15, 40))+  
theme_bw()+  
theme(panel.grid = element_blank(),  
      panel.border = element_blank(),  
      strip.background = element_blank(),  
      strip.text = element_text(size = 12),  
      text = element_text(color = "black", family = "Helvetica"))
```



GLMM-零膨胀负二项nbinom1分布

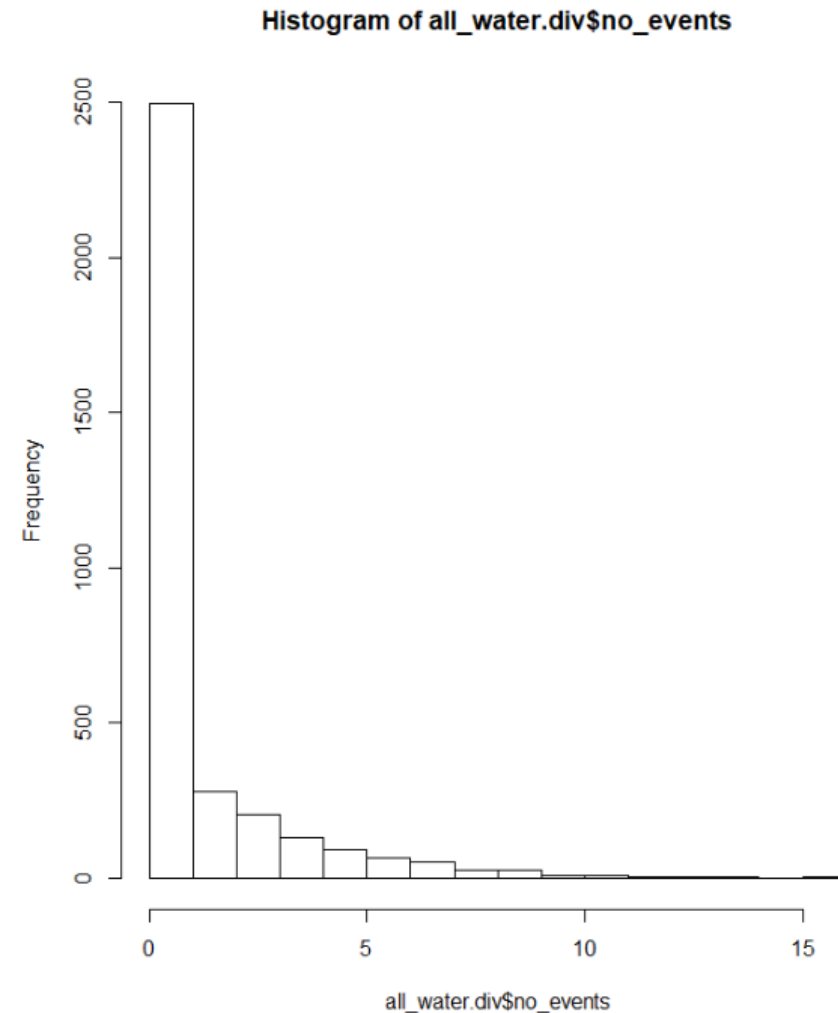
```
events.cels.glmmTMB.negbin <- glmmTMB(no_events ~ water +  
  #tmax +  
  last_3_tmax +  
  #last_7_tmax +  
  #last_14_tmax +  
  #ppt +  
  last_3_ppt +  
  (water:last_3_tmax)+  
  (water:last_3_ppt)+  
  #last_7_ppt +  
  #last_14_ppt+  
  #daysSinceRain_mm +  
  (1|numeric_day/site),  
  data=all_water.div,  
  ziformula = ~1, # this tells it it needs to be zero-
```

inflated

family=nbinom1)

```
events.cels.glmmTMB.negbin
```

```
Anova(events.cels.glmmTMB.negbin)
```



GLMM-零膨胀gamma模型

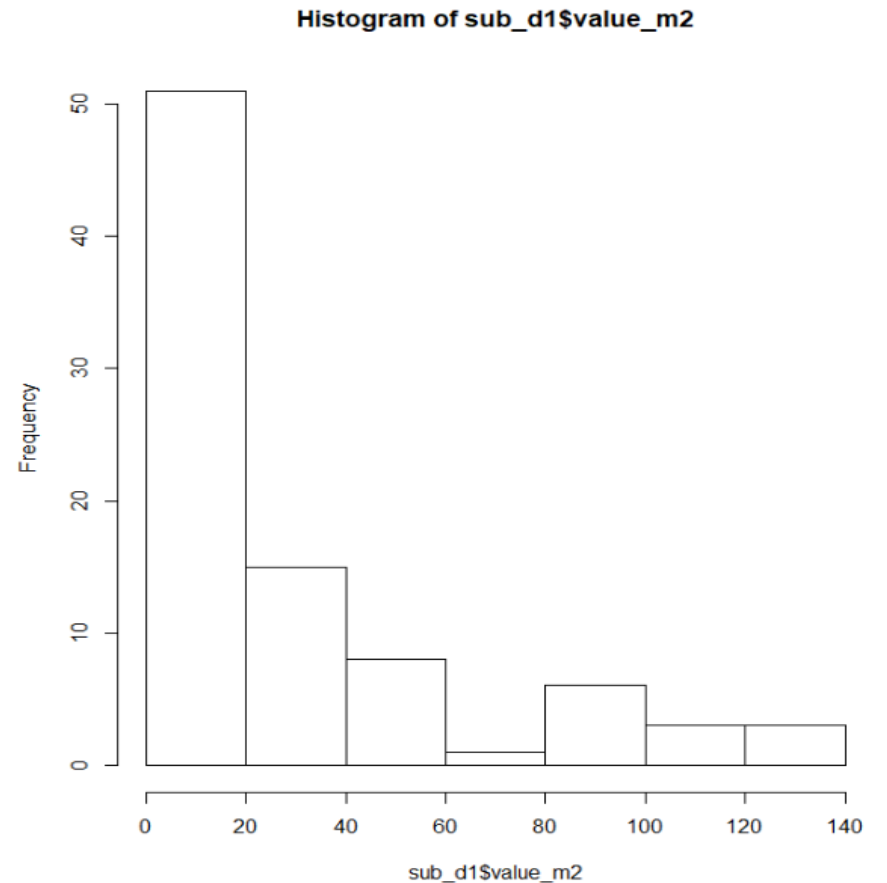
```
> sub_d1$value_m2
[1] 3.00 6.05 38.05 40.00 5.05 1.00 5.05 135.00 30.00 60.00 25.05
[12] 31.00 16.00 93.00 41.00 86.00 130.00 115.00 6.20 0.05 0.05 10.05
[23] 0.10 3.00 5.00 41.00 85.00 48.00 12.00 10.05 111.00 32.00 36.00
[34] 7.00 0.05 0.10 0.05 5.00 6.05 0.10 7.05 22.00 25.05 8.05
[45] 1.00 2.00 35.05 32.00 30.00 6.05 6.05 55.00 93.00 81.00 2.05
[56] 115.00 135.00 2.05 0.05 5.00 25.00 7.00 0.05 18.00 50.00 42.00
[67] 1.00 95.00 30.00 13.00 5.00 5.00 35.00 45.00 70.00 8.05 9.00
[78] 5.00 17.00 15.00 0.05 6.00 0.05 0.05 1.00 0.05 1.00
```

```
out <- glmmTMB(value_m2 ~ plot_type + season +
  (1|Site_Code),
  data = veg.sum[veg_type == "herb"],
  ziformula = ~1,
  family=Gamma)
```

```
> Anova(out)
Analysis of Deviance Table (Type II Wald chisquare tests)
```

```
Response: value_m2
      Chisq Df Pr(>Chisq)
plot_type 14.0404 2 0.0008937 ***
season      0.0904 1 0.7636816
```

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



glmmTMB包应用案例2 （无数据）

We used a negative **binomial generalized linear mixed model** [glmmTMB function in R package glmmTMB to test whether **commercial success (i.e., number of sellers by species)** was linked to five ecological characteristics associated with invasiveness in ants (polygyny, worker body size, habitat generalism, nesting generalism, and total range size), while accounting for species' geographical origin (i.e., in which ecozone species occur) and **controlling for phylogenetic effects by using ants' superfamily and genus as nested random effects**.

We determined the best-fitting model using stepwise model selection by AIC (Akaike information criteria) (stepAIC function in R package MASS).

Gippet, J. M. W. & Bertelsmeier, C. Invasiveness is linked to greater commercial success in the global pet trade. *PNAS*, **118**, e2016337118, doi:10.1073/pnas.2016337118 (2021). 2021年4月6号

glmmTMB应用案例-负二项回归

全模型

```
NB <- glmmTMB(NbSellers ~ 1 +  
  Gyny +  
  MeanSize +  
  HabitatGeneralism +  
  NestingGeneralism +  
  TotalRangeSize +  
  Af +  
  Au +  
  In +  
  Nea +  
  Neo +  
  EPa +  
  WPa +  
  (1|subFamily/genus),  
  data= Traded_ants_v4,  
  family=nbinom2)  
summary(NB)  
Anova(NB, type="III")
```

模型精简得到最优模型



best model selection based on AIC
stepAIC function in R package MASS

```
bestModel <- glmmTMB(NbSellers ~ 1 +  
  #Gyny +  
  #Workers +  
  MeanSize +  
  HabitatGeneralism +  
  #NestingGeneralism +  
  TotalRangeSize +  
  Af +  
  #Au +  
  #In +  
  #Nea +  
  #Neo +  
  #EPa +  
  WPa +  
  (1|subFamily/genus),  
  data= Traded_ants_v4,  
  family=nbinom2)
```

glmmTMB-负二项模型--查看结果及作图

We calculated the coefficient of determination of the best-fitting model using Nakagawa's pseudo-R-squared [r2_nakagawa function in R package performance].

```
Anova(bestModel, type="III")  
plot(simulateResiduals(bestModel))  
summary(bestModel)  
performance::r2_nakagawa(bestModel)  
# marginal effects 结果制图  
effects2 <- plot_model(bestModel2,  
                        show.values = TRUE,  
                        value.offset = .3,  
                        type="eff",  
                        transform=NULL)
```



Gippet, J. M. W. & Bertelsmeier, C. Invasiveness is linked to greater commercial success in the global pet trade. *PNAS*, **118**, e2016337118, doi:10.1073/pnas.2016337118 (2021). 2021年4月6号

glmmTMB案例3-零膨胀负二项模型 (zero inflated negative binomial model)

Ecology, 101(3), 2020, e02971
© 2020 by the Ecological Society of America

Trait–environment interactions affect plant establishment success
during restoration

CHAD R. ZIRBEL ^{1,2,3} AND LARS A. BRUDVIG¹

¹*Department of Plant Biology and Program in Ecology, Evolutionary Biology and Behavior, Michigan State University, East Lansing,
Michigan 48824 USA*

²*Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, Minnesota 55108 USA*

**We included random effects of field, plot, and species to account for non-independence of samples.
We used plot-level counts of each species sown into each field as our response variable and fit a
zero-inflated negative binomial error distribution.**

glmmTMB案例3-零膨胀负二项模型 (zero inflated negative binomial model)

####Build GLMM -----

```
mod.tmb.nb<-glmmTMB(count~seeds_m2+total_biomass*moist.avg+seed_mass*light_available+  
  NITROGEN.*herbivory+root.mf*moist.avg+sla*moist.avg+sla*light_available+  
  (1|transect_id/plot)+(1|species),  
  zi=~1,  
  family=nbinom2(link = "log"),  
  data=data.scale,  
  control=glmmTMBControl(optCtrl=list(iter.max=1e3,eval.max=1e3)))###促进模型收敛  
summary(mod.tmb.nb)  
Anova(mod.tmb.nb, type=3)
```

We included random effects of field, plot, and species to account for non-independence of samples. We used plot-level counts of each species sown into each field as our response variable and fit a zero-inflated negative binomial error distribution.

```
> summary(mod.tmb.nb)
Family: nbinom2 ( log )
Formula: count ~ seeds_m2 + total_biomass * moist.avg + seed_mass * light_available +
  NITROGEN. * herbivory + root.mf * moist.avg + sla * moist.avg +
  sla * light_available + (1 | transect_id/plot) + (1 | species)
Zero inflation: ~1
Data: data.scale
```

AIC	BIC	logLik	deviance	df.resid
3850.2	3955.8	-1904.1	3808.2	1109

Random effects:

Conditional model:

Groups	Name	Variance	Std.Dev.
plot:transect_id	(Intercept)	0.2310	0.4806
transect_id	(Intercept)	1.1384	1.0670
species	(Intercept)	0.7588	0.8711

Number of obs: 1130, groups: plot:transect_id, 95; transect_id, 19; species, 41

Overdispersion parameter for nbinom2 family (): 1.41

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.02043	0.32568	0.063	0.949980
seeds_m2	-0.12007	0.15715	-0.764	0.444833
total_biomass	0.05628	0.24763	0.227	0.820218
moist.avg	0.14273	0.08717	1.637	0.101553
seed_mass	-0.13568	0.23262	-0.583	0.559701
light_available	0.07198	0.08522	0.845	0.398311
NITROGEN.	-0.70869	0.25192	-2.813	0.004906 **
herbivory	-0.08418	0.08614	-0.977	0.328424
root.mf	-0.29777	0.19042	-1.564	0.117879
sla	0.12436	0.20760	0.599	0.549166
total_biomass:moist.avg	0.07315	0.05087	1.438	0.150452
seed_mass:light_available	0.08845	0.08695	1.017	0.309019
NITROGEN.:herbivory	-0.11385	0.04970	-2.291	0.021983 *
moist.avg:root.mf	0.17629	0.04645	3.795	0.000148 ***
moist.avg:sla	-0.01417	0.05672	-0.250	0.802688
light_available:sla	-0.10010	0.05799	-1.726	0.084307 .

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

Zero-inflation model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.135	1.937	-2.135	0.0328 *

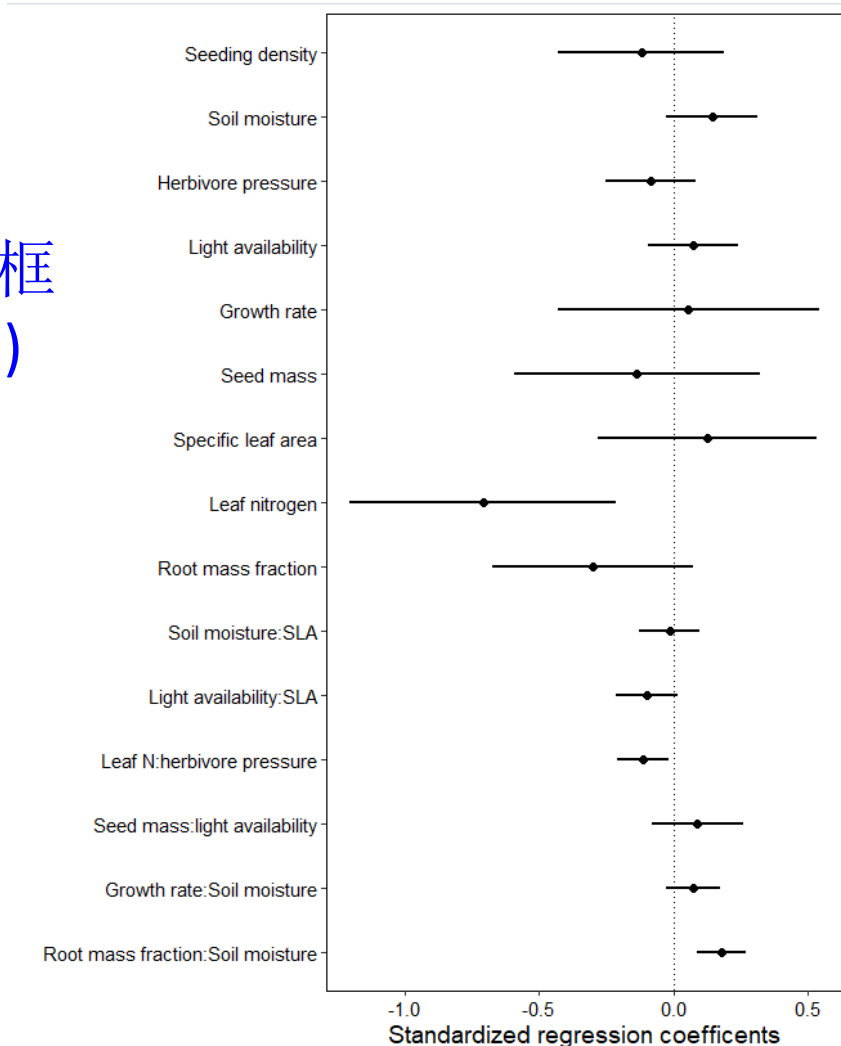
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Zirbel, C. R. & Brudvig, L. A. Trait–environment interactions affect plant establishment success during restoration. *Ecology* **101**, e02971, doi:<https://doi.org/10.1002/ecy.2971> (2020).

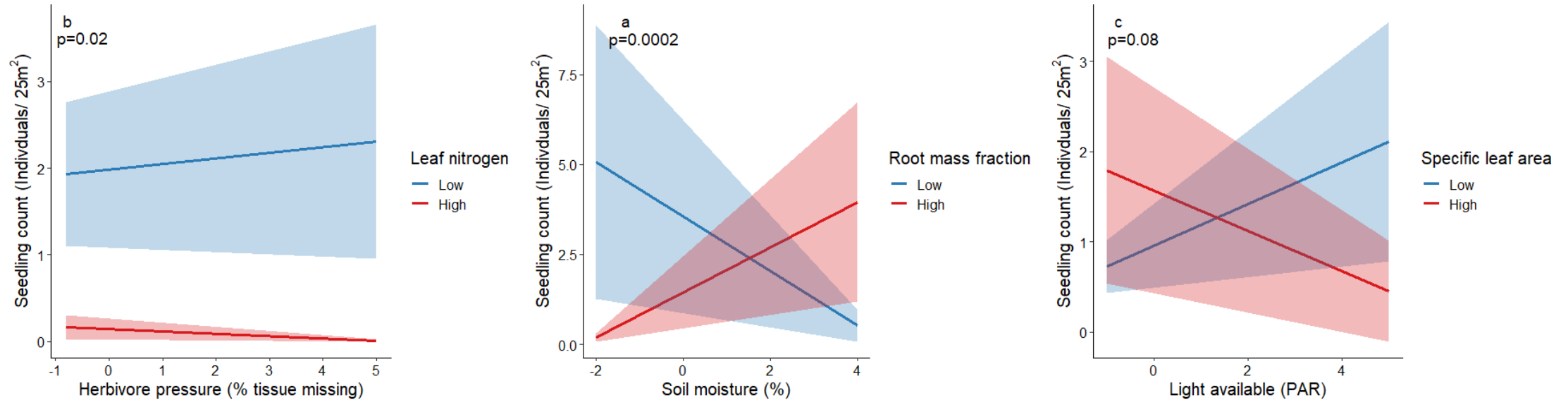
glmmTMB-零膨胀负二项模型 (zero inflated negative binomial model)

核心命令提取模型参数估计值，并转化为数据框
`results.table<-as.data.frame(confint(mod.tmb.nb))`

Zirbel, C. R. & Brudvig, L. A. Trait–environment interactions affect plant establishment success during restoration. *Ecology* **101**, e02971, doi:<https://doi.org/10.1002/ecy.2971> (2020).



glmmTMB-零膨胀负二项模型： 交互作用作图

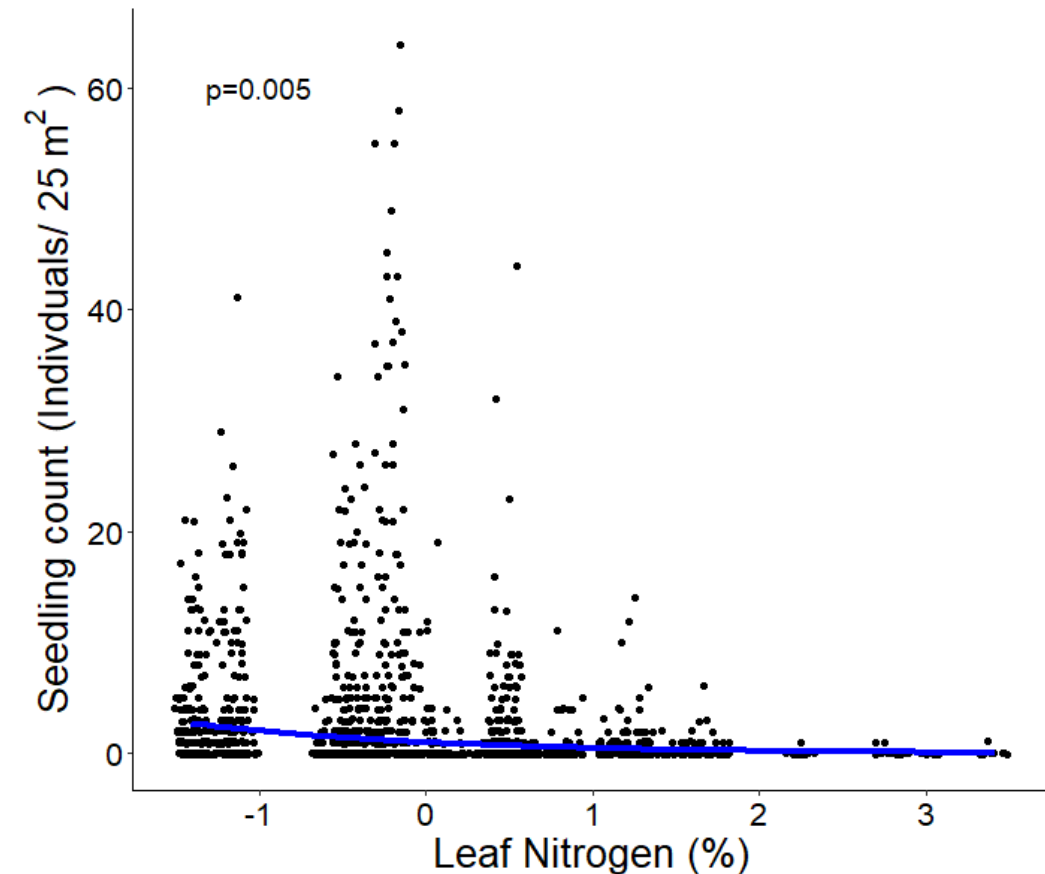


用effect命令提取交互作用的参数估计值，作图
`eff.herb<-effect("NITROGEN.:herbivory",mod.nb.2,xlevels=2)`
`eff_df.herb <- data.frame(eff.herb)`

Zirbel, C. R. & Brudvig, L. A. Trait–environment interactions affect plant establishment success during restoration. *Ecology* **101**, e02971, doi:<https://doi.org/10.1002/ecy.2971> (2020).

glmmTMB-零膨胀负二项模型- 对某因子的效应单独作图

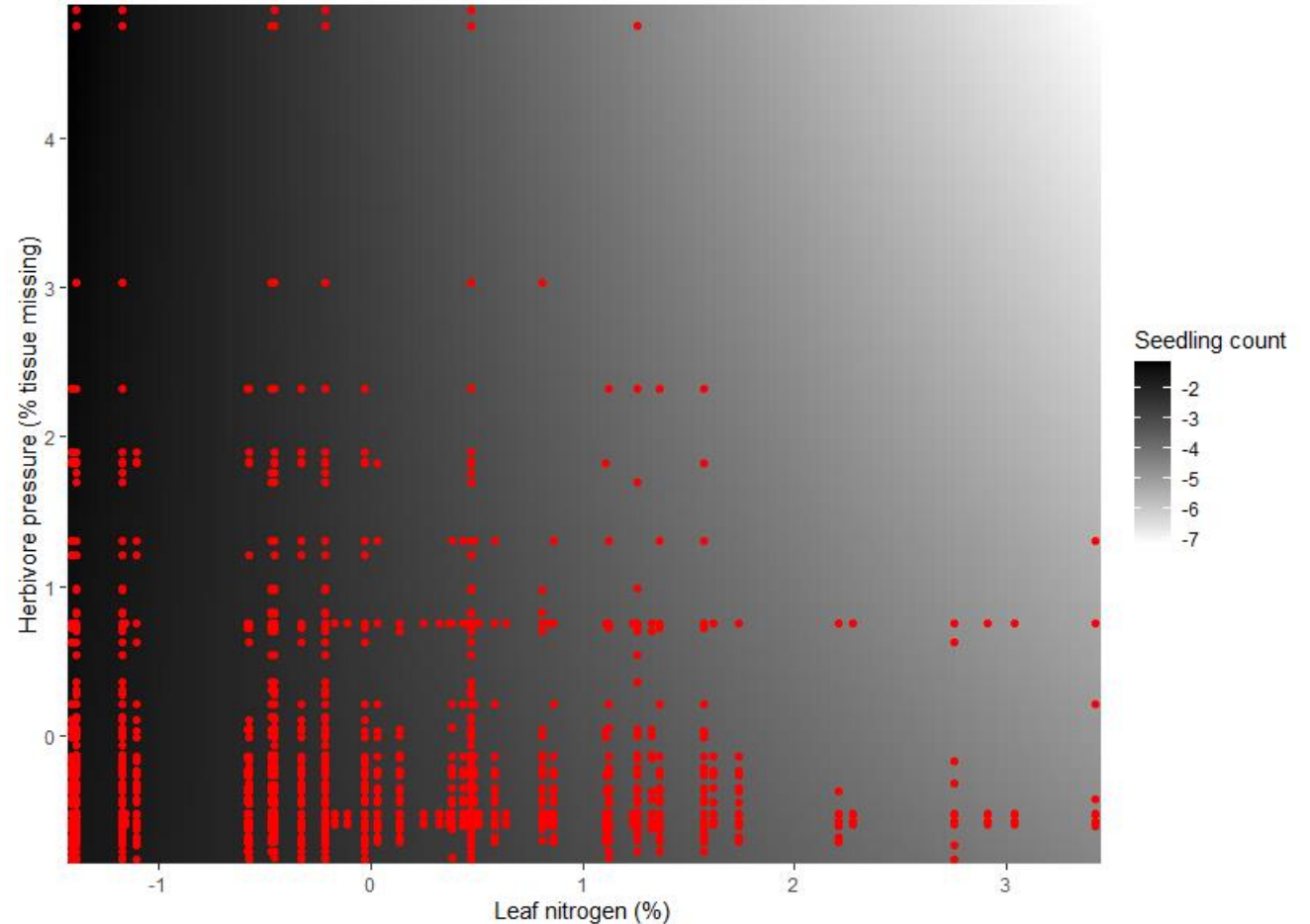
```
ggplot(data.scale)+  
  geom_point(aes(NITROGEN.,count),color="black",size=1.5,position =  
  position_jitter(width = 0.1, height = 0.1))+  
  stat_function(fun=function(x)exp(fixef(mod.tmb.nb)[[1]][1] +  
  fixef(mod.tmb.nb)[[1]][7]*x),size=1.5,color="blue")+  
  annotate("text",label="p=0.005",x=-1,y=60,size=5)+  
  labs(x = "Leaf Nitrogen (%)", y =bquote('Seedling count (Indivduals/  
  25~m^2~)'))+  
  theme(text =  
  element_text(size=20),axis.text=element_text(color="black"),panel.back  
  ground=element_blank(),  
  
  panel.grid.major=element_blank(),panel.grid.minor=element_blank(),  
  axis.line = element_line(size=.7, color="black"),legend.position =  
  "none")
```



Zirbel, C. R. & Brudvig, L. A. Trait–environment interactions affect plant establishment success during restoration. *Ecology* **101**, e02971, doi:<https://doi.org/10.1002/ecy.2971> (2020).

glmmTMB-零膨胀负二项模型： 两数值型变量对Y的影响作图（visreg包）

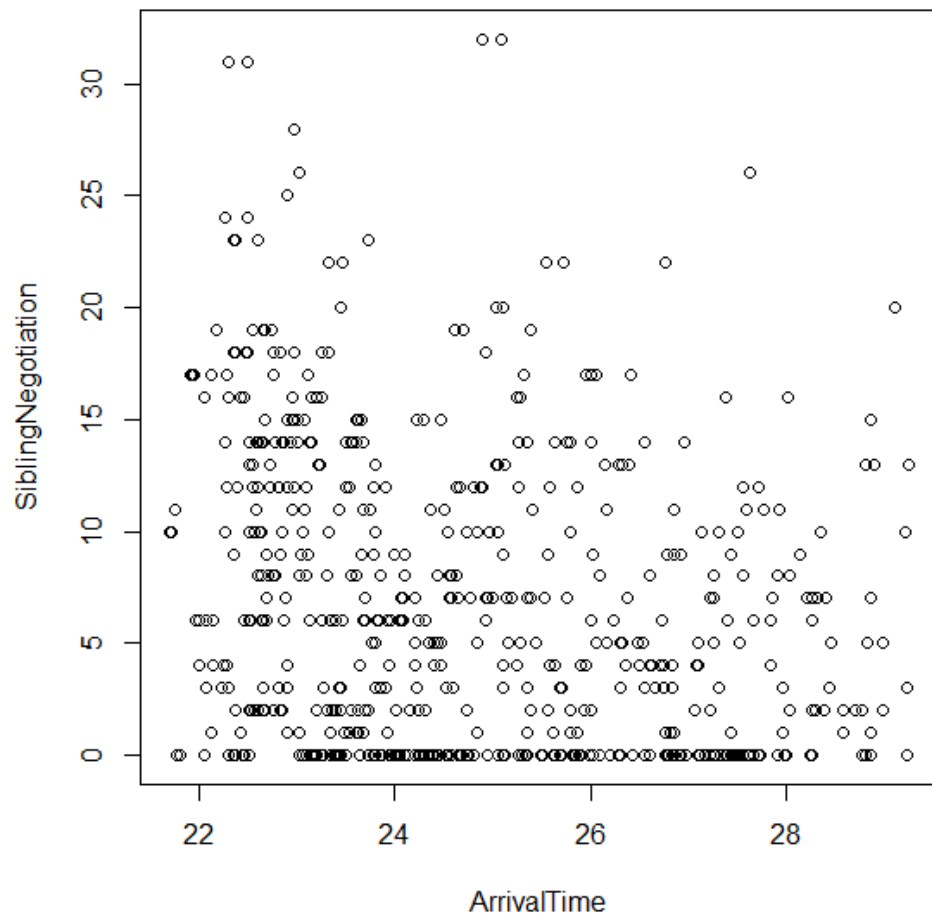
```
visreg2d(mod.tmb.nb, x="NITROGEN.", y="herbivory",  
plot.type="gg",  
xlab="Leaf nitrogen (%)", ylab="Herbivore pressure  
(% tissue missing)", main="", cex.lab=1.7, cex.axis = 2,  
col=colorRampPalette(c("white", "black"))(30))+  
geom_point(aes(NITROGEN., herbivory),  
data=data.scale, col='red')+  
scale_fill_gradient(name = "Seedling  
count", low="white", high="black")
```



Zirbel, C. R. & Brudvig, L. A. Trait–environment interactions affect plant establishment success during restoration. *Ecology* **101**, e02971, doi:<https://doi.org/10.1002/ecy.2971> (2020).

非线性混合效应模型： generalized additive mixed model

```
library(mgcv)
owls<- read.delim("Owls.txt")
owls$NCalls<-owls$SiblingNegotiation
owls$LBroodSize<-log(owls$BroodSize)
owls$fNest<-factor(owls$Nest)
plot(SiblingNegotiation~ArrivalTime,data=owls)
```



非线性混合效应模型： generalized additive mixed model

```
gamm1<-gamm(SiblingNegotiation~ +s(ArrivalTime), random=list(fNest=~1),data=owls,  
family=poisson(link=log))
```

随机因素 与因变量有非线性关系的因素

```
summary(gamm1$gam)
```

```
Family: poisson  
Link function: log  
  
Formula:  
SiblingNegotiation ~ +s(ArrivalTime)  
  
Parametric coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)   1.6857     0.0901   18.7   <2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Approximate significance of smooth terms:  
              edf Ref.df    F p-value  
s(ArrivalTime) 8.39   8.39 50.5  <2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
R-sq.(adj) =  0.0999  
scale est. = 1          n = 599
```

非线性混合效应模型： generalized additive mixed model

##对结果作图predict命令

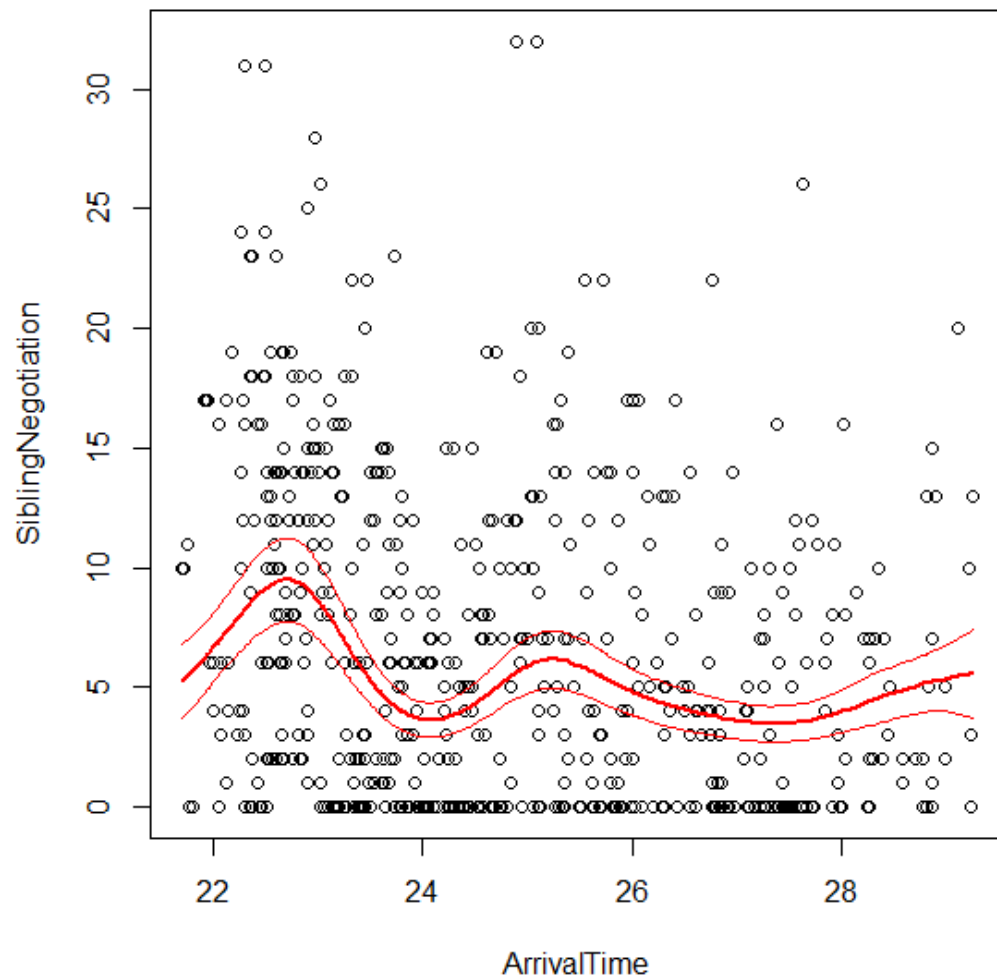
```
M <- predict(gamm1$gam,type="response",se.fit=T)
```

```
plot(SiblingNegotiation~ArrivalTime,data=owls)
```

```
lines(owls$ArrivalTime,M$fit,lwd=2,col="red")
```

```
lines(owls$ArrivalTime,M$fit+1.96*M$se.fit,lwd=1,col="red")
```

```
lines(owls$ArrivalTime,M$fit-1.96*M$se.fit,lwd=1,col="red")
```



非线性混合效应模型： generalized additive mixed model

```
gamm2<-gamm(SiblingNegotiation~+FoodTreatment+s(ArrivalTime),  
random=list(fNest=~1),data=owls, family=poisson(link=log))
```

```
summary(gamm2$gam)
```

```
> summary(gamm2$gam)
```

```
Family: poisson  
Link function: log
```

```
Formula:  
SiblingNegotiation ~ +FoodTreatment + s(ArrivalTime)
```

```
Parametric coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.9360	0.0927	20.9	<2e-16	***
FoodTreatmentsatiated	-0.5967	0.0361	-16.5	<2e-16	***

```
---
```

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

```
Approximate significance of smooth terms:
```

	edf	Ref.df	F	p-value	
s(ArrivalTime)	8.4	8.4	51.1	<2e-16	***

```
---
```

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

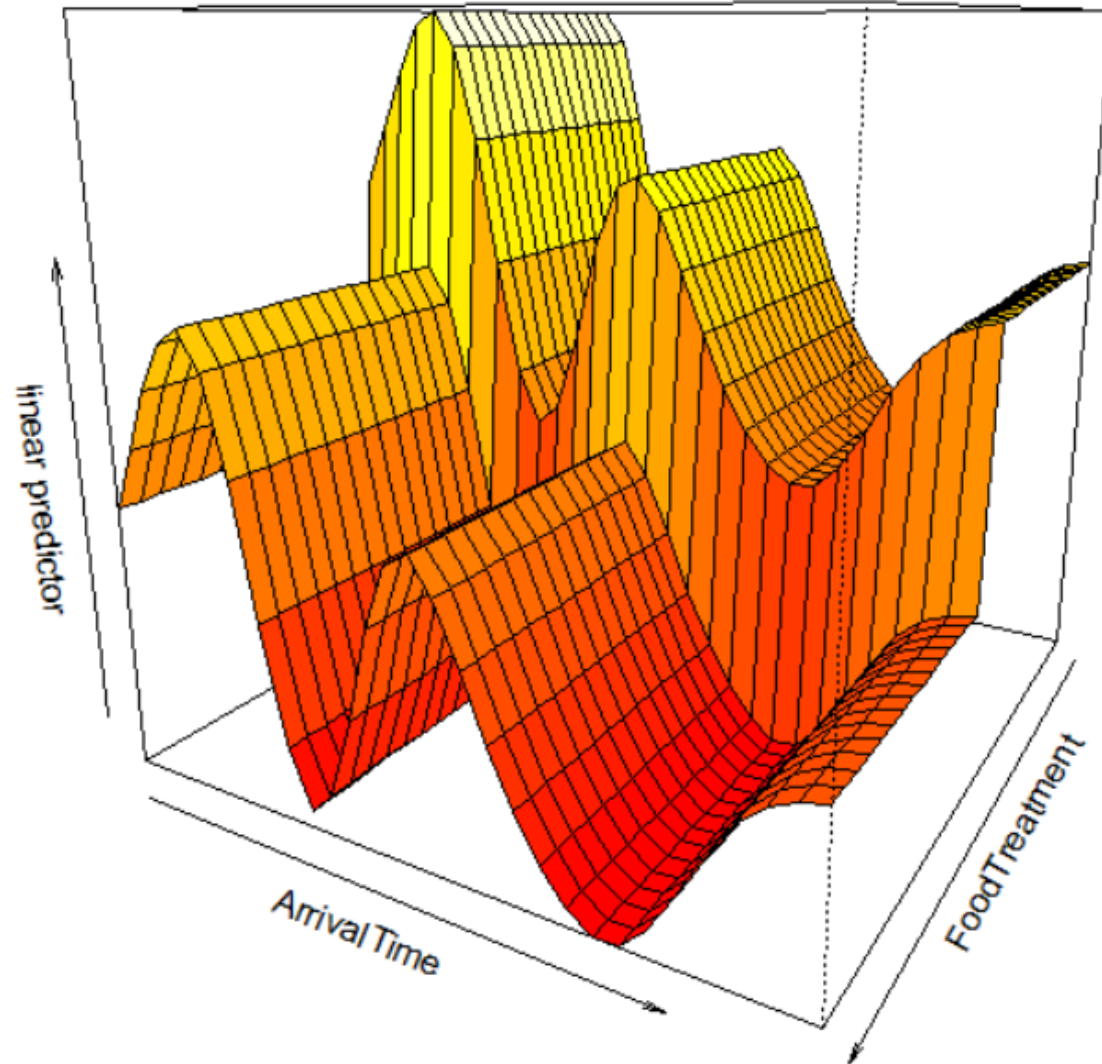
```
R-sq.(adj) =  0.162
```

```
Scale est. = 1
```

```
n = 599
```

非线性混合效应模型

- `vis.gam(gamm2$gam,thet=120)`



Greater vulnerability to warming of marine versus terrestrial ectotherms

Malin L. Pinsky¹

We used a GAMM to calculate latitudinal gradients of $T_{\max}$ ' (smoothed interaction term between latitude and habitat) while accounting for fixed differences among habitats (ocean versus land), a fixed effect of experiment type (critical versus lethal $T_{\max}$ measurement) and taxonomic non-independence (phylum, class, order, family and genus as nested random effects). We fit the model using the mgcv package version 1.8-24⁴⁸ in R version 3.4.2. Reported P values are two-sided. We report r^2 calculations from the summary.gam() function in the mgcv package. We fit all possible models with subsets of these variables and ranked them with Akaike's information criterion with the finite sample size correction (AICc)⁴⁹. AICc

非线性混合模型应用案例

uGamm（而不用gamm）可以进行模型自动选择

模型结构

```
modTmax <- uGamm(tmax.accsum.elev ~ Realm + tmax_metric + s(lat, by=Realm),  
random=list(Phylum=~1, Class=~1, Order=~1, Family=~1, Genus=~1), data=modTmaxdat)
```

#依据原始数据得到拟合线数据

```
temp <- predict(modTmax$gam, newdata=newdat, se.fit=TRUE)
```

```
newdat$tmax <- temp$fit
```

```
newdat$tmaxse <- temp$se.fit
```

Pinsky, M. L., Eikeset, A. M., McCauley, D. J., Payne, J. L. & Sunday, J. M. Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature* **569**, 108-111, doi:10.1038/s41586-019-1132-4 (2019).

非线性混合模型应用案例—结果制图

```
plot(tmax ~ lat, data=newdat[newdat$Realm=='Terrestrial'], type='l', col='green',  
ylim=c(10,50), xlim=c(-80,80))  
lines(tmax + tmaxse ~ lat, data=newdat[newdat$Realm=='Terrestrial'], col='green', lty=2)  
lines(tmax - tmaxse ~ lat, data=newdat[newdat$Realm=='Terrestrial'], col='green', lty=2)  
lines(tmax ~ lat, data=newdat[newdat$Realm=='Marine'], col='blue')  
lines(tmax - tmaxse ~ lat, data=newdat[newdat$Realm=='Marine'], col='blue', lty=2)  
lines(tmax + tmaxse ~ lat, data=newdat[newdat$Realm=='Marine'], col='blue', lty=2)  
points(tmax.accsum.elev ~ lat, data=dat[dat$Realm=='Terrestrial'], col='green', cex=0.5)  
points(tmax.accsum.elev ~ lat, data=dat[dat$Realm=='Marine'], col='blue', cex=0.5)
```

Pinsky, M. L., Eikeset, A. M., McCauley, D. J., Payne, J. L. & Sunday, J. M. Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature* **569**, 108-111, doi:10.1038/s41586-019-1132-4 (2019).

Generalized linear mixed models: a practical guide for ecology and evolution

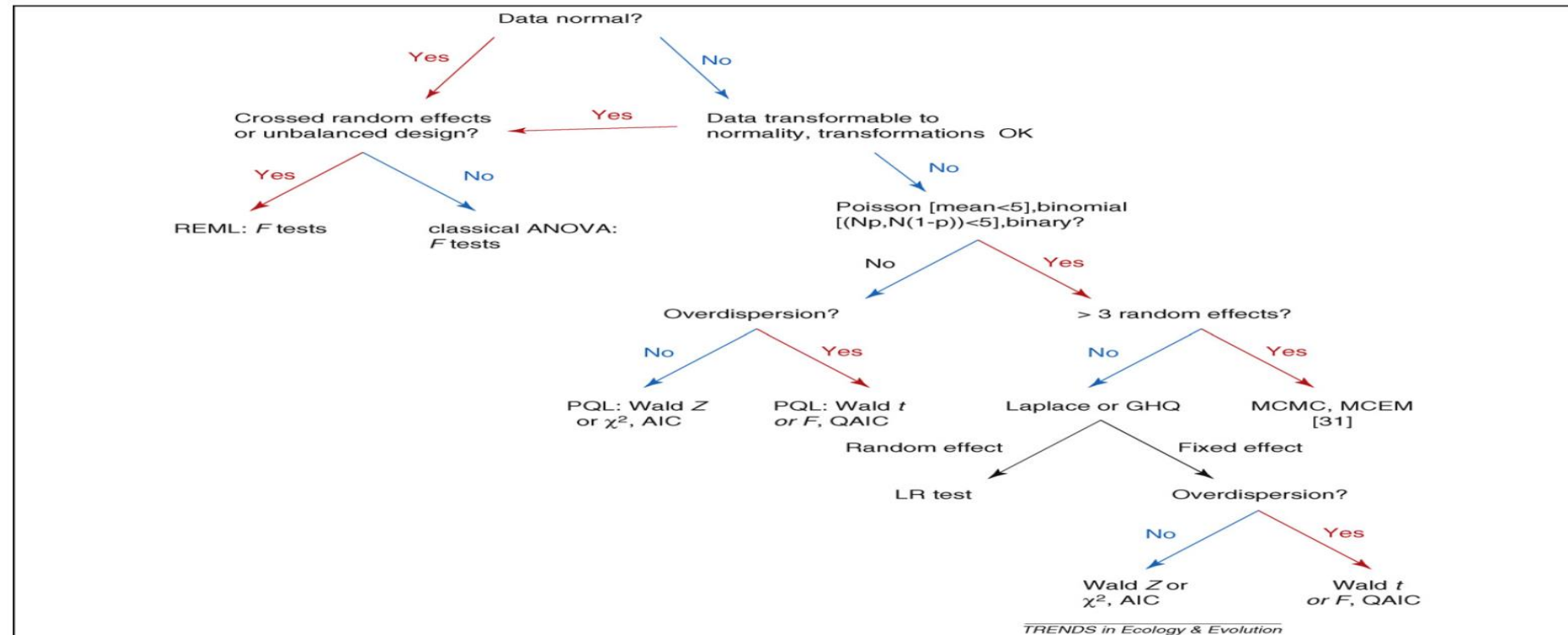


Figure 1. Decision tree for GLMM fitting and inference. Conditions on the Poisson and binomial distributions along the right branch refer to penalized quasiliikelihood (PQL) rules of thumb [30]: to use PQL, Poisson distributions should have mean > 5 and binomial distributions should have the minimum of the number of successes and failures > 5 . MCEM = Monte Carlo expectation-maximization [40].

截止2021年5月5日，已被引用7022次（数据来源，[googlescholar](#)）

清醒一下，再战江湖！

Mixed effect model 涉及的其他问题

线性混合模型--注意事项

对于参数估计

- 1) 如果数据正态，或者数据结构平衡，则可用最大似然法（ML）
- 2) 如果随机效应是嵌套的（不存在交叉），或者只有一个随机效应,可采用 (ML),否则一律REML
- 3)如果模型不收敛，那么说明 “That means your model has too many factors and not a big enough sample size, and cannot be fit” 就要丢掉一些固定效应或随机效应，如保持随机效应不变，每次只减掉一个固定效应。

<http://ase.tufts.edu/gsc/gradresources/guidetomixedmodelsinr/mixed%20model%20guide.html>

模型不收敛

```
control=glmerControl(optimizer="bobyqa",  
  optCtrl=list(maxfun = 2e5))
```

```
model <- glmer(assoc ~ season + BINHAB + scale(encount_risk) + scale(KOP.DIST.M) +  
  scale(NDVI) + (1 | SiteID), family=binomial, data=grazers,  
  control=glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))  
summary(model)
```

Beaudrot, L., Palmer, M. S., Anderson, T. M. & Packer, C. Mixed-species groups of Serengeti grazers: a test of the stress gradient hypothesis. *Ecology* **101**, e03163, doi:<https://doi.org/10.1002/ecy.3163> (2020).

Singular问题： 对（某些）随机效应方差的参数估计值接近零 或相关性估计为+/-1

```
> gm3<-glmer(GHQgroup~tree.canopy+timegroup;
  =mydata, family = poisson)
boundary (singular) fit: see ?isSingular
```

Singular models: random effect variances estimated as zero, or correlations estimated as +/- 1

It is very common for overfitted mixed models to result in singular fits. Technically, singularity means that some of the θ (variance-covariance Cholesky decomposition) parameters corresponding to diagonal elements of the Cholesky factor are exactly zero, which is the edge of the feasible space, or equivalently that the variance-covariance matrix has some zero eigenvalues (i.e. is positive semidefinite rather than positive definite), or (*almost* equivalently) that some of the variances are estimated as zero or some of the correlations are estimated as +/-1. This commonly occurs in two scenarios:

- small numbers of random-effect levels (e.g. <5), as illustrated in [these simulations](#) and discussed (in a somewhat different, Bayesian context) by Gelman (2006).
- complex random-effects models, e.g. models of the form $(f|g)$ where f is a categorical variable with a relatively large number of levels, or models with several different random-slopes terms.
- When using `lme4`, singularity is most obviously detectable in the output of `summary.merMod()` or `VarCorr.merMod()` when a variance is estimated as 0 (or very small, i.e. orders of magnitude smaller than other variance components) or when a correlation is estimated as exactly ± 1 . However, as pointed out by D. Bates, Kliegl, et al. (2015), singularities in larger variance-covariance matrices can be hard to detect: checking for small values among the diagonal elements of the Cholesky factor is a good start.

- 1) 随机效应的水平数过少导致，将其作为固定效应
- 2) 随机效应过于复杂，如 $(f|g)$ 其中 f 为分类变量，且水平数较多
- 3) 可以尝试贝叶斯多水平模型

Mixed effect model—嵌套型随机效应 (nested random effect structure)

- 模型至少含有两个随机效应
- 其中一个嵌套于另一个之中
- 比如班级嵌套于学校中，物种嵌套于属内，属嵌套于科内...

例如：

```
n4<-lmer(mathgain ~ mathknow  +(1 | schoolid/classid),data=nested)
```

```
(1 | cityid/districtid/schoolid/classid)
```

Mixed effect model—嵌套型随机效应 (nested random effect structure)

school	class
1	1
1	1
1	1
1	2
1	2
1	2
1	2
1	2
2	3
2	3
2	3
2	1
2	1
2	1
2	1
2	2
2	2
3	2
3	3
3	3
3	3
3	3
3	2
3	2

如果不同school中，class的命名有重复那么嵌套必须写成 $(1 | \text{school}/\text{class})$

而不能写成 $(1 | \text{school}) + (1 | \text{class})$
因为这时名称相同的class会在多个school出现，写成 $(1 | \text{school}) + (1 | \text{class})$ ，school和class会被认为是两个交叉的变量，即我中有你，你中有我，不存在嵌套。

建议：每层随机效应的每个水平都进行唯一化命名

混合效应模型--嵌套型随机效应结构

写法格式示例，前提schoolid和classid都唯一命名

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

```
View(nested)
```

```
n3<-lmer(mathgain ~ mathknow +(1 | schoolid) + (1 | classid), data=nested)
```

```
n4<-lmer(mathgain ~ mathknow +(1 | schoolid/classid), data=nested)
```

tab_model(n3)

mathgain			
Predictors	Estimates	CI	p
(Intercept)	57.78	54.77 – 60.80	<0.001
mathknow	2.45	-0.11 – 5.01	0.062
Random Effects			
σ^2	1018.83		
τ_{00} classid	104.65		
τ_{00} schoolid	90.17		
ICC classid	0.09		
ICC schoolid	0.07		
Observations	1081		
Marginal R2 / Conditional R2	0.005 / 0.165		

tab_model(n4)

mathgain			
Predictors	Estimates	CI	p
(Intercept)	57.78	54.77 – 60.80	<0.001
mathknow	2.45	-0.11 – 5.01	0.062
Random Effects			
σ^2	1018.83		
τ_{00} classid:schoolid	104.65		
τ_{00} schoolid	90.17		
ICC classid:schoolid	0.09		
ICC schoolid	0.07		
Observations	1081		
Marginal R2 / Conditional R2	0.005 / 0.165		

Mixed effect model—嵌套型随机结构的重要应用— 方差分解

用varcomp命令直接提取方差项

```
varcomp.LMA<-varcomp(lme(log(LMA)~1, random=1|Site/Plot/Species/Tree/Strata, data = d, na.action = na.omit))
```

Table 2 Variance partitioning of the full nested linear models on Leaf Mass Area (LMA) and Leaf Dry Matter Content (LDMC) across six nested ecological scales. All data were $\log_{10}$ transformed prior to analysis. $n = 1910$ leaves. Square brackets represent the 95% confidence intervals, which were calculated by bootstrapping (500 runs with 1910 randomly sampled data points with replacement)

Ecological scale	% variance of trait [95% C.I.]	
	log LMA	log LDMC
Leaf and error	10 [7–9]*	15 [9–15]*
Strata	16 [16–21]	16 [12–26]
Tree	22 [19–28]	17 [12–30]
Species	21 [16–26]	35 [25–43]
Plot	0 [0–0]	0 [0–0]
Site	30 [27–32]	16 [13–19]

Messier, J., B. J. McGill, and M. J. Lechowicz. 2010. How do traits vary across ecological scales? A case for trait-based ecology. *Ecology Letters* **13**:838–848.

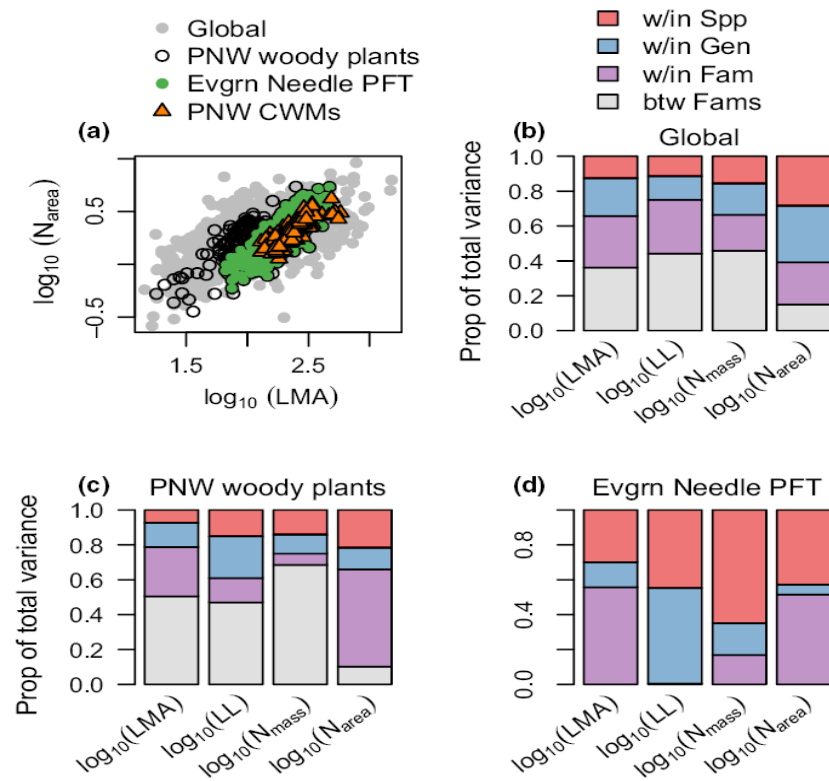
Mixed effect model—嵌套型随机结构的重要应用— 方差分解

##同样可以用VarCorr命令提取方差项

```
logLMAvar <- lmer(log.LMA ~ 1 + (1|Project) + (1|Family) + (1|Genus) + (1|Species), data.all)
```

```
LMAvariance <- data.frame(VarCorr(logLMAvar))
```

Anderegg, L. D. L., L. T. Berner, G. Badgley, M. L. Sethi, B. E. Law, J. HilleRisLambers, and J. Penuelas. 2018. Within-species patterns challenge our understanding of the leaf economics spectrum. *Ecology Letters* 21:734-744.



Mixed effect model—嵌套型随机结构的重要应用— 方差分解案例

```
library(ape)
```

```
library(nlme)
```

```
data(carnivora)
```

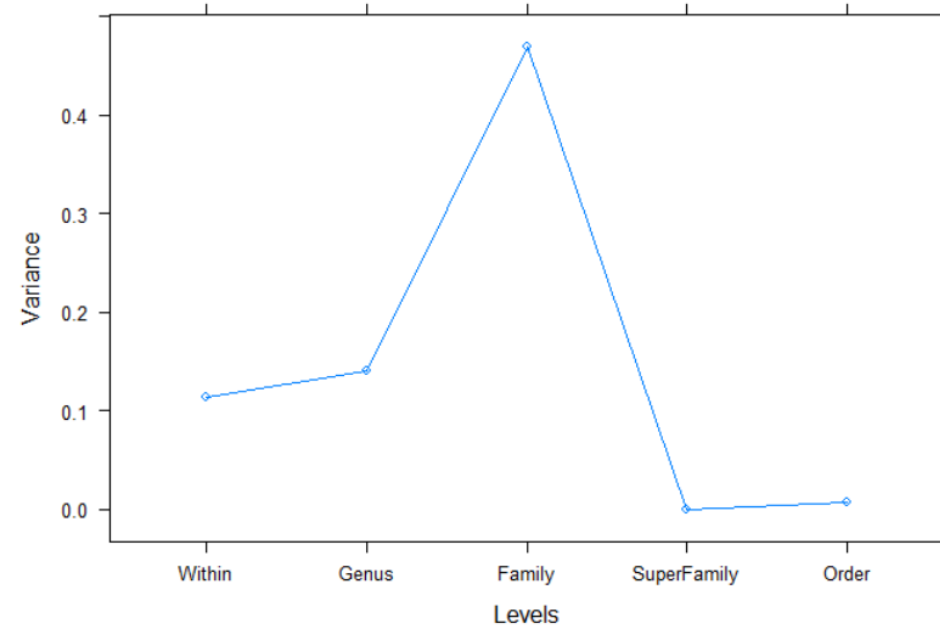
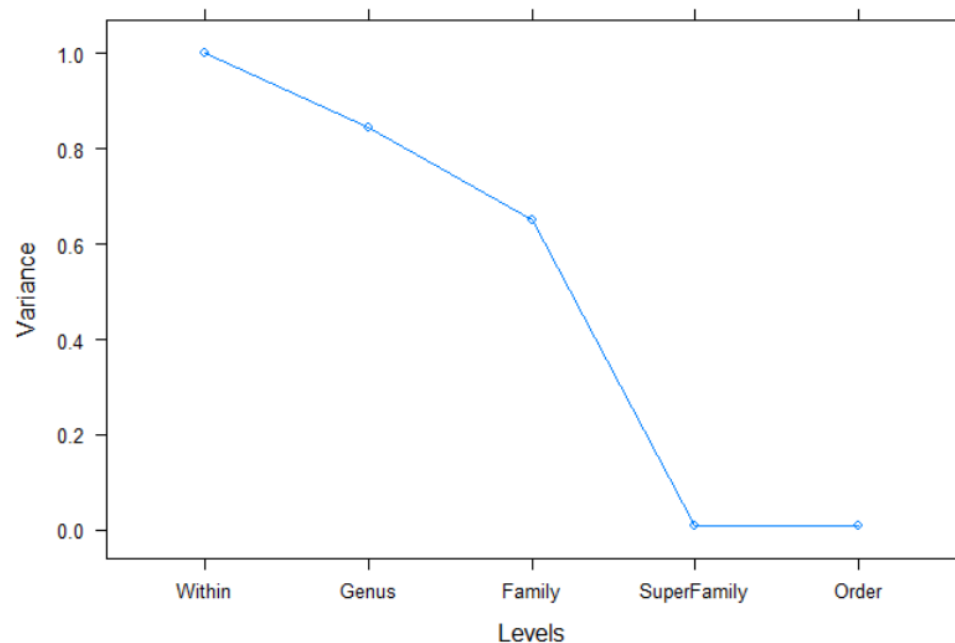
```
m <- lme(log10(SW) ~ 1, random = ~ 1 | Order/SuperFamily/Family/Genus, data=carnivora)
```

```
v1 <- varcomp(m, TRUE, TRUE)
```

```
v2 <- varcomp(m, scale = FALSE, cum = FALSE)
```

```
plot(v1)
```

```
plot(v2)
```



多重嵌套结构的应用案例

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

We tested whether marine or terrestrial species had a higher proportion of local extirpations with a GLMM with binomial errors and a logit link (function `glmer()` in the `lme4` version 1.1-18-1 package⁸⁷ for R v.3.2.4). We included taxonomic levels as nested random effects to help to account for taxonomic non-independence, and

Letter | Published: 24 April 2019

Greater vulnerability to warming of marine versus terrestrial ectotherms

Malin L. Pinsky , Anne Maria Eikeset, Douglas J. McCauley, Jonathan L. Payne & Jennifer M. Sunday

Nature **569**, 108–111(2019) | [Cite this article](#)

13k Accesses | **117** Citations | **565** Altmetric | [Metrics](#)

Mixed effect model—交叉型随机效应结构

```
library(readxl)
```

```
k1 <- read_excel("Kozlov data.xls")
```

```
k1$n1 <- k1$Latitude^5 + k1$Longitude^5
```

```
k1$Site <- as.numeric(factor(k1$n1, levels=unique(k1$n1)))
```

```
k1$Site <- as.factor(k1$Site)
```

```
mk1 <- lmer(log(Herbivory/100 + 0.1) ~ abs(Latitude) + (1 | Species) + (1 | Site), data=k1)
```

一个地点可以有多个物种，一个物种也可以在多个地点出现

```
> summary(mk1)
```

Linear mixed model fit by REML ['lmerMod']

Formula: log(Herbivory/100 + 0.1) ~ scale(Latitude) + (1 | Species) + (1 | Site)

Data: k1

REML criterion at convergence: 616

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.3385	-0.5381	-0.1275	0.4316	3.5619

Random effects:

Groups	Name	Variance	Std.Dev.
Species	(Intercept)	0.02484	0.1576
Site	(Intercept)	0.04201	0.2050
Residual		0.05342	0.2311

Number of obs: 1355, groups: Species, 739; Site, 423

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-1.79549	0.01774	-101.222
scale(Latitude)	-0.06267	0.01673	-3.747

Correlation of Fixed Effects:

	(Intr)
scale(Ltttd)	0.125

不同地点造成的方差大于物种造成的方差

log(Herbivory/100 + 0.1)			
Predictors	Estimates	CI	p
(Intercept)	-1.80	-1.83 – -1.76	<0.001
Latitude	-0.06	-0.10 – -0.03	<0.001
Random Effects			
σ^2	0.05		
τ_{00} Species	0.02		
τ_{00} Site	0.04		
ICC	0.56		
N Species	739		
N Site	423		
Observations	1355		
Marginal R ² / Conditional R ²	0.032 / 0.570		

用nlme设置交叉型随机结构模型

```
library(readxl)
k1 <- read_excel("Kozlov data.xls")
k1$n1<-k1$Latitude^5+k1$Longitude^5
k1$Site<- as.numeric(factor(k1$n1, levels=unique(k1$n1)))
k1$Site<-as.factor(k1$Site)
k1$l=1
k1$SL<-scale(abs(k1$Latitude))
k1$y<-log(k1$Herbivory/100+0.1)
k2<-k1[1:300,]
nlm1<- nlme(model=y~b0+b1*SL,
```

```
  fixed=b0+b1~1,
  random=list(l=pdIdent(b0~factor(Species)-1),
             l=pdIdent(b0~factor(Site)-1)),
  start=c(-2, -0.1), data=k2)
```

```
Fixed effects: b0 + b1 ~ 1
              Value Std.Error DF   t-value p-value
b0.(Intercept) -1.7954538  0.01770836 1353  -101.39020  0e+00
b1              -0.0626649  0.01669973 1353   -3.75245  2e-04
Correlation:
  b0.(I)
b1 0.125
```

写法复杂，运算缓慢，结果凌乱，
简直无可救药！！！！

但另一方面，nlme在够造自定义模型方面
无以伦比

###如模型 $y=c0+c1*(x) +\sin(c2*(x))$

```
diy_d<-read.csv("DIY.csv")
```

```
fit <- nlme(
```

```
  y ~ c0 + c1 + sin(c2),
```

```
  fixed = list(c0 ~ 1, c1 ~ x - 1, c2 ~ x - 1),
```

```
  random = list(u = pdBlocked(list(pdIdent(c0 ~ 1), pdIdent(c1 ~ x - 1))))
```

```
  start = c(0, 0.5, 1),
```

```
  data=diy_d)
```

```
summary(fit)
```

```
Fixed effects: list(c0 ~ 1, c1 ~ x - 1, c2 ~ x - 1)
              Value Std.Error DF t-value p-value
c0.(Intercept) 0.1747719 0.07592855 897 2.30179 0.0216
c1.x            0.9949405 0.02222992 897 44.75682 0.0000
c2.x            2.0289522 0.03208897 897 63.22896 0.0000
```

方差异质性处理—应用案例

```
mod21=lme(data=data1, aboveC ~ age + traj.20,  
           weights=varComb(varExp(form=~age),  
                             varIdent(form=~1|traj.20)),  
           random=~1|project/fire,  
           na.action=na.omit, method="ML")  
summary(mod21)
```

For the LMMs of aboveground C, C:N, and aboveground C to belowground C we included a variance structure (**varIdent**) to account for differences in variance among trajectories that were apparent in the residuals. For the LMM of aboveground C we also included a variance structure (**varExp**) to account for increasing variance with age that was apparent in the residuals.

Mack, M. C. *et al.* Carbon loss from boreal forest wildfires offset by increased dominance of deciduous trees. *Science* 372, 280-283, doi:10.1126/science.abf3903 (2021).

混合效应模型—时间(重复测量)与空间相关性的处理

常用做法：将测量时间或者采样地点作为随机效应

- We examined neighborhood effects on y by adopting a mixed modelling approach. ...To control for temporal, spatial and interspecific variation, we included census year, species identity and 1-m² seedling plot nested in each census station as random factors.

(Chen et al. 2019, Science 366:124-128.)

```
Model<-lmer(y~x+(1|year)+(1|species)+(1|station/plot),data=data)
```

混合效应模型—时间(重复测量)与空间相关性的处理

- 常用做法：将测量时间或者采样地点作为随机效应

We fitted a linear mixed model with restricted maximum likelihood using the landscape-level multidiversity index (T60.landscape) as the response variable.... To reflect the large-scale spatial and temporal structure of our dataset, we added sampling year (Year), nested within study region (Region), as a random effect. To reflect the spatial structure of our dataset within each region, we included the longitude and latitude of the center of each landscape (Lat, Lon) as covariates.

```
Model<-lmer(y~x+Lat+Lon+(1|region/year)+(1|species)+(1|station/plot), data=data)
```

(Sirami et al. 2019, PNAS)

混合效应模型--重复测量 也可以指定时间自相关结构

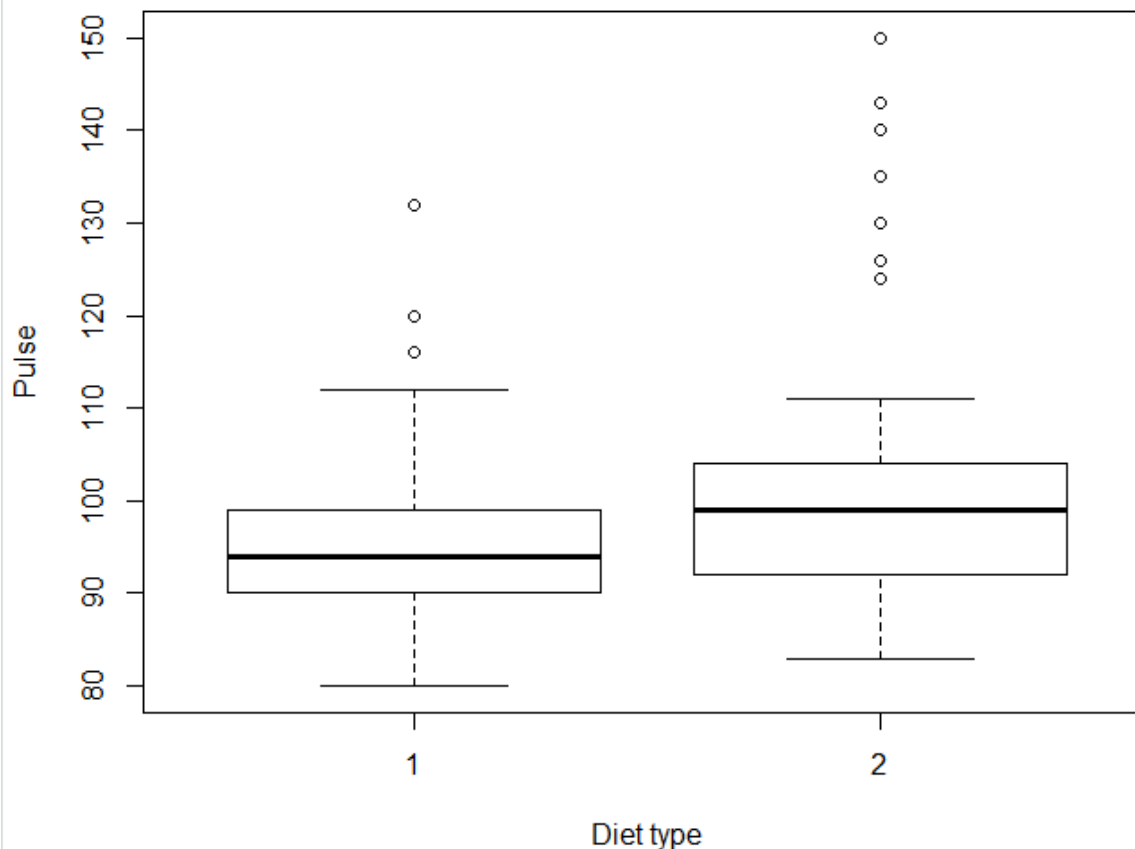
对同一对象进行多次测量

##

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

```
View(longg)
```

```
boxplot(pulse~diet,data=longg,  
xlab="Diet type", ylab="Pulse")
```



混合效应模型--重复测量（nlme包） （repeated measures）

```
rm1<-lme(pulse~diet, random=~1|id, data=longg, correlation=corCAR1(form = ~time|id))
```

测量对象（随机效应）

指定时间自相关的类型

测量次数

###其他相关矩阵类型

- `rm2<-lme(pulse~diet,random=~1|id,data=longg, correlation=corAR1(form = ~time|id))`
- `rm3<-lme(pulse~diet,random=~1|id,data=longg, correlation=corCompSymm(form = ~time|id))`
- `rm4<-lme(pulse~diet,random=~1|id,data=longg, correlation=corSymm(form = ~time|id))`

##查看更过相关矩阵结构类型

?lme

注意： 同一观测对象一次观测只能有一个数据

混合效应模型--重复测量 (nlme包)

--指定相关矩阵结构, 同时可设置方差不齐次

##设置不同的diet或id, 方差不一致

```
rm5<-lme(pulse~diet,random=~1|id,data=longg,correlation=corSymm(form = ~time|id),  
weights=varFunc(~diet))
```

##设置不同的id,方差不一致

```
rm6<-lme(pulse~diet,random=~1|id,data=longg, correlation=corSymm(form = ~time|id),  
weights=varFunc(~id))
```

混合效应模型--重复测量 (repeated measures)

选取合适的相关矩阵结构

```
> AIC(rm1,rm2,rm3,rm4,rm5,rm6)
```

	df	AIC
rm1	5	709.2112
rm2	5	709.2112
rm3	5	722.0902
rm4	7	703.5242
rm5	7	693.2538
rm6	7	657.4748

```
> summary(rm6)
```

Linear mixed-effects model fit by REML

Data: longg

	AIC	BIC	logLik
	657.4748	674.8161	-321.7374

Random effects:

Formula: ~1 | id

(Intercept) Residual

StdDev: 2.520716 3.381619

Correlation Structure: General

Formula: ~time | id

Parameter estimate(s):

Correlation:

1 2

2 0.648

3 0.318 0.779

Variance function:

Structure: fixed weights

Formula: ~id

Fixed effects: pulse ~ diet

	Value	Std.Error	DF	t-value	p-value
(Intercept)	84.56055	4.617486	60	18.313116	0.0000
diet	6.29858	3.280366	28	1.920086	0.0651

Correlation:

(Intr)

diet -0.943

Standardized within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-1.5045876	-0.2388306	0.1804497	0.6969361	2.8049339

Number of observations: 90

Number of Groups: 30

时间自相关应用案例

Each LME model examined included a temporal autocorrelation structure of order as well as species nested within site as a random.

```
ar1<-lme(iWUE.micromol.mol ~ Year, data = meta.CRU.pre1963, random = ~ 1|Site/LatinBinomial,  
correlation = corAR1())
```

```
rm1<-lme(pulse~diet, random=~1|id, data=longg, correlation=corCAR1(form = ~time|id))
```

测量对象（随机效应）

指定时间自相关的类型

测量次数

Mathias, J. M. & Thomas, R. B. Global tree intrinsic water use efficiency is enhanced by increased atmospheric CO₂ and modulated by climate and plant functional types. *PNAS*, **118**, e2014286118(2021).

Generalised estimation equation— 非正态分布数据重复测量

- lme4的glmer包的不足：无法指定重复测量的相关矩阵结构

```
library(geepack)
```

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

```
Birds$fField <- factor(Birds$fField)
```

```
View(Birds)
```

```
##数据为泊松分布时
```

```
M.gee1<-geeglm(Richness~offset(LA)+fSptreat+DEPTH+I(DEPTH^2),  
               data = Birds, family="poisson", id=fField, corstr = "ar1")
```

```
summary(M.gee1)
```

数据分布类型

测量对象

时间自相关结构

```
anova(M.gee1)
```

Generalised estimation equation— 非正态分布数据重复测量

```
> summary(M.gee1)
```

```
call:
geeglm(formula = Richness ~ offset(LA) + fSptreat + DEPTH + I(DEPTH^2),
       family = "poisson", data = Birds, id = fField, corstr = "ar1")
```

Coefficients:

	Estimate	Std.err	wald	Pr(> w)
(Intercept)	-0.678203	0.261009	6.75	0.0094 **
fSptreatrlfld	-0.522314	0.228764	5.21	0.0224 *
DEPTH	0.049824	0.019315	6.65	0.0099 **
I(DEPTH^2)	-0.001141	0.000491	5.41	0.0201 *

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

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	2.33	0.293

Correlation: Structure = ar1 Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha	0.422	0.0903

Number of clusters: 11 Maximum cluster size: 10

Richness			
<i>Predictors</i>	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.51	0.30 – 0.85	0.009
fSptreat [rlfld]	0.59	0.38 – 0.93	0.022
DEPTH	1.05	1.01 – 1.09	0.010
DEPTH^2	1.00	1.00 – 1.00	0.020
N _{fField}	11		
Observations	110		

混合效应模型—空间自相关的控制（nlme包）

缺陷：

- 空间距离不能为零，即同一地点不能有多个值，同时不能设置交叉型随机结构，只能设置嵌套型
- 解决方案，分别在经纬度上加上一个很小的随机数，让其距离不为零。

```
nrow(k1)
```

```
k1$r<-rnorm(1355, 0, 0.05)
```

```
k1$lat<-k1$Latitude+k1$r
```

```
k1$lon<-k1$Longitude+k1$r
```

```
lmm_sp_auto <- lme(fixed = log(Herbivory/100+0.1) ~ abs(Latitude), data = k1,  
  correlation = corHaversine(1,~lat+lon | Species/Site),  
  random = ~ 1 | Species, method = "ML")
```

##查看更过相关矩阵结构类型，corHaversine为自定义函数，能够吧经纬度坐标转换为距离，使用前要先运行此函数。

混合效应模型--重复测量 (repeated measures)

选取合适的相关矩阵结构

时间自相关矩阵结构类型

Available standard classes:

<code>corAR1</code>	autoregressive process of order 1.
<code>corARMA</code>	autoregressive moving average process, with arbitrary orders for the autoregressive and moving average components.
<code>corCAR1</code>	continuous autoregressive process (AR(1) process for a continuous time covariate).
<code>corCompSymm</code>	compound symmetry structure corresponding to a constant correlation.

空间自相关矩阵结构类型

选择方法: AIC值

<code>corExp</code>	exponential spatial correlation.
<code>corGaus</code>	Gaussian spatial correlation.
<code>corLin</code>	linear spatial correlation.
<code>corRatio</code>	Rational quadratics spatial correlation.
<code>corSpher</code>	spherical spatial correlation.
<code>corSymm</code>	general correlation matrix, with no additional structure.

nlme包内置的空间自相关结构

Available standard classes:

corAR1	autoregressive process of order 1.		
corARMA	autoregressive moving average process, with arbitrary orders for the autoregressive and moving average components.		
corCAR1	continuous autoregressive process (AR(1) process for a continuous time covariate).		
corCompSymm	compound symmetry structure corresponding to a constant correlation.		
corExp	exponential spatial correlation.	corRatio	Rational quadratics spatial correlation.
corGaus	Gaussian spatial correlation.	corSpher	spherical spatial correlation.
corLin	linear spatial correlation.	corSymm	general correlation matrix, with no additional structure.

广义线性混合模型 (glmm)

处理空间自相关

```
library(MASS)
```

```
pql_sp_auto<- glmmPQL(log(Herbivory/100+0.1) ~ abs(Latitude),  
random= ~ 1|Species, correlation = corExp(form = ~ lat+lon),  
family= gaussian, data = k1)
```

```
binomial(link = "logit")  
gaussian(link = "identity")  
Gamma(link = "inverse")  
inverse.gaussian(link = "1/mu^2")  
poisson(link = "log")  
quasi(link = "identity", variance = "constant")  
quasibinomial(link = "logit")  
quasipoisson(link = "log")
```

- 可以通过**family**函数指定不同的数据分布类型
- 缺陷：随机效应的一个水平在同一个位置不有有多个数据（空间距离为0）
- 同样可以通过给经纬度坐标加上一个很小的数字来解决

对比nlme和glmmPQL空间自相关回归的结果

```
> summary(lmm_sp_auto1)
Linear mixed-effects model fit by maximum likelihood
Data: k1
      AIC      BIC logLik
807.5 833.5 -398.7

Random effects:
Formula: ~1 | Species
      (Intercept) Residual
StdDev:      0.186   0.2852

Correlation Structure: corHaversine
Formula: ~lat + lon | Species
Parameter estimate(s):
range
0.0366
Fixed effects: log(Herbivory/100 + 0.1) ~ abs(Latitude)
              Value Std.Error   DF t-value p-value
(Intercept) -1.7535  0.019552  738  -89.68  0.0000
abs(Latitude) -0.0014  0.000648  615   -2.23  0.0262
Correlation:
      (Intr)
abs(Latitude) -0.81

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
-2.2566 -0.6398 -0.1766  0.4964  3.4646

Number of Observations: 1355
Number of Groups: 739
```

```
> summary(pql_sp_auto)
Linear mixed-effects model fit by maximum likelihood
Data: k1
      AIC      BIC logLik
      NA      NA      NA

Random effects:
Formula: ~1 | Species
      (Intercept) Residual
StdDev:      0.186   0.2851

Correlation Structure: Exponential spatial correlation
Formula: ~lat + lon | Species
Parameter estimate(s):
range
0.000309
Variance function:
Structure: fixed weights
Formula: ~invwt
Fixed effects: log(Herbivory/100 + 0.1) ~ abs(Latitude)
              Value Std.Error   DF t-value p-value
(Intercept) -1.7535  0.019552  738  -89.69  0.0000
abs(Latitude) -0.0014  0.000648  615   -2.23  0.0262
Correlation:
      (Intr)
abs(Latitude) -0.81

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
-2.2516 -0.6398 -0.1766  0.4964  3.4647

Number of Observations: 1355
Number of Groups: 739
```

- 一定要学会依据自己数据的特征寻找合适的统计模型！！！！
- 要了解原理，不要迷信于各种技术！！！！

混合效应模型的使用要切忌任意发挥

Trends in Ecology & Evolution

CellPress
REVIEWS

Opinion

Mixed Models Offer No Freedom from Degrees of Freedom

Göran Arnqvist^{1,*}

混合模型几大忠告

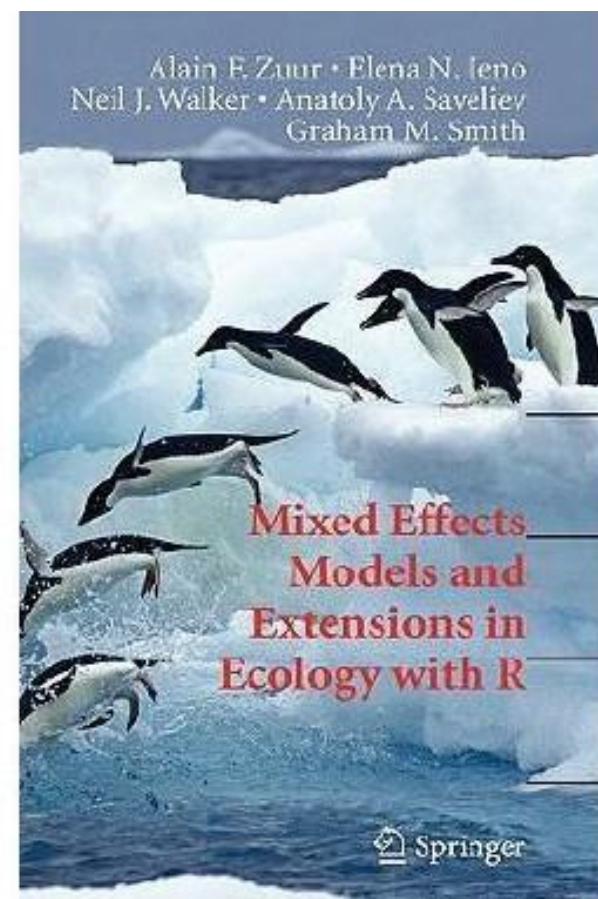
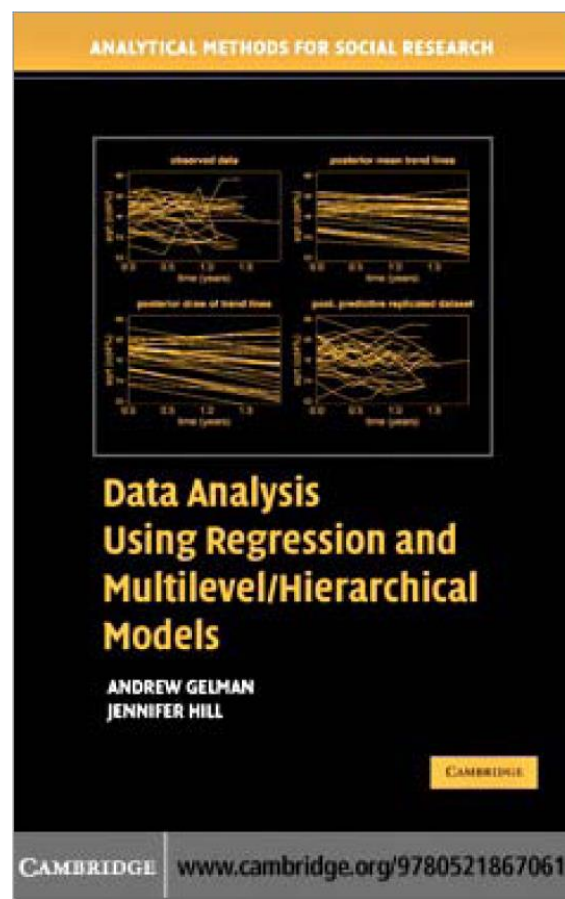
1. 要在文章中详尽介绍模型的结构（如family, link, 随机效应结构等），最好提供R代码在文章附件中。
2. 判定参数的显著性，最好能伴以F检验。
3. 如果混合模型的结果比一般模型（不包含随机效应）更显著，那么检查随机效应结构设置是否正确。
4. 参数估计最好用REML。
5. 数值存在重复测量时，一定要采用重复测量混合模型，而不是将同一观测对象的值进行平均。
6. 单个随机效应的水平数不可太少，至少应大于5个，模型含有多个随机效应时，更是如此。

Arnqvist, G. 2020. Mixed models offer no freedom from degrees of freedom. Trends in Ecology & Evolution 35:329-335.

混合效应模型的应用

不同的叫法:

- Mixed effect models
- Multi-level models
- Hierarchical models
- Nested models



休息一下，继续前进！