



结构方程模型在生态学中的应用

硕 士 导 师：曹文侠 教 授

博 士 导 师：董全民 研究员

汇 报 人：刘玉祯

2022年12月4日



目录

01

SEM基本概念

02

以单个因子为目标的变量筛选

03

以群落组成为目标的变量筛选

04

piecewiseSEM的多种应用场景



目录

01

SEM基本概念

结构方程模型 (structural equation modeling: SEM)



基于先验知识和我们的理解，首先能够给出一个或几个有因果逻辑的**结构**，反映我们感兴趣的科学问题。

要求：专业知识和SEM的基础概念



我们需要将这些结构转换为
我们熟悉的数学**方程**，即统计回归模型。

要求：回归，因子分析、广义线性模型、混合效应模型等统计知识



基于我们收集的数据开展**模型**拟合与检验，判断我们的结构是否合适，给出最终模型的参数。

要求：软件帮我们做了，我们只要理解背后原理，读懂这些结果并图形展示就可以了！

思路与我们传统的统计模型没有太大差别，主要区别就是**SEM从一开始就要有图形结构**(初始模型、先验模型、概念模型、元模型或叫meta-模型)。它是一种多元统计方法，同时处理**多个变量之间的因果关系**。

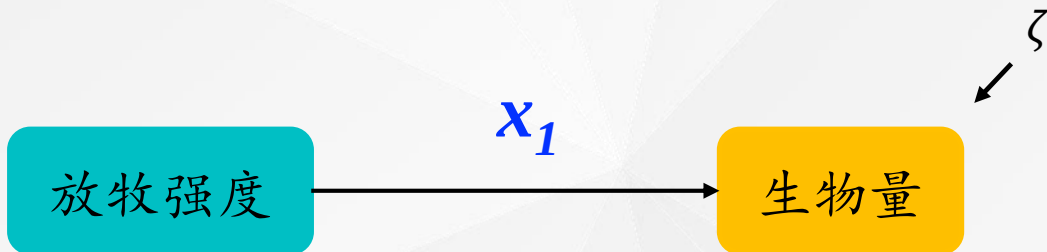
● 什么是SEM

- SEM假定了因果关系；
- SEM通过结合多个结构方程（SE）进行推断，获得一个复杂系统中变量间的直接和间接效应；
- 可以对无法直接测量的变量开展分析。

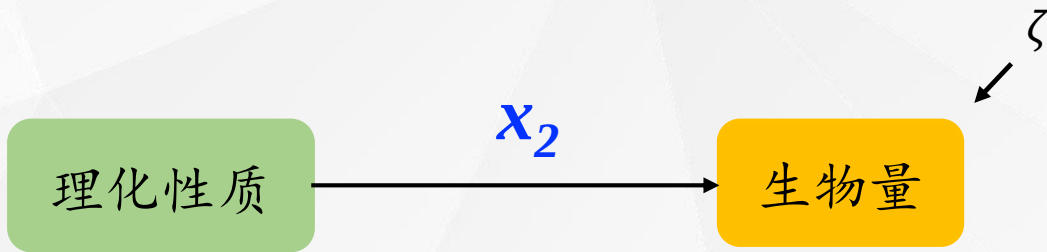
SEM广义上是一种构建、评估和量化因果关系网络中直接和间接效应的框架模型。包含建立因果关系网、起始模型，检验模型结构和估计模型参数，图形表达、解释、推广模型的结果等一系列的过程。

● 为什么使用SEM?

$$y = x_1 * \zeta$$

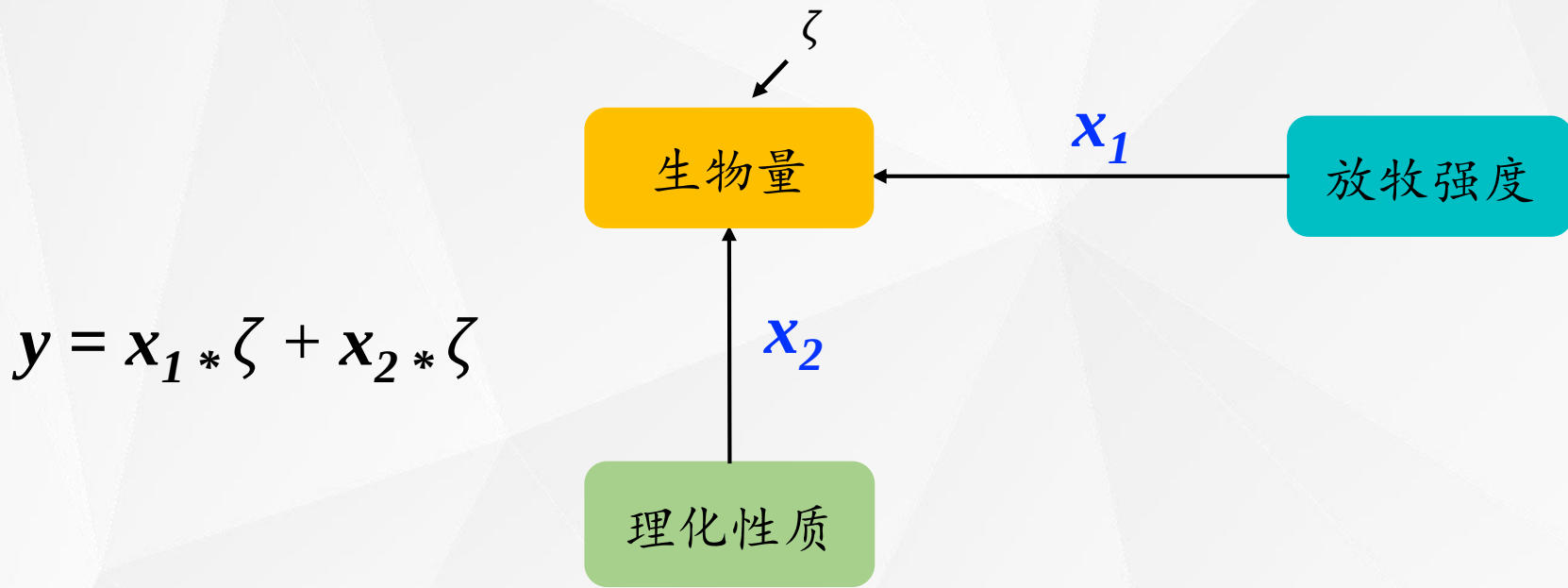


$$y = x_2 * \zeta$$



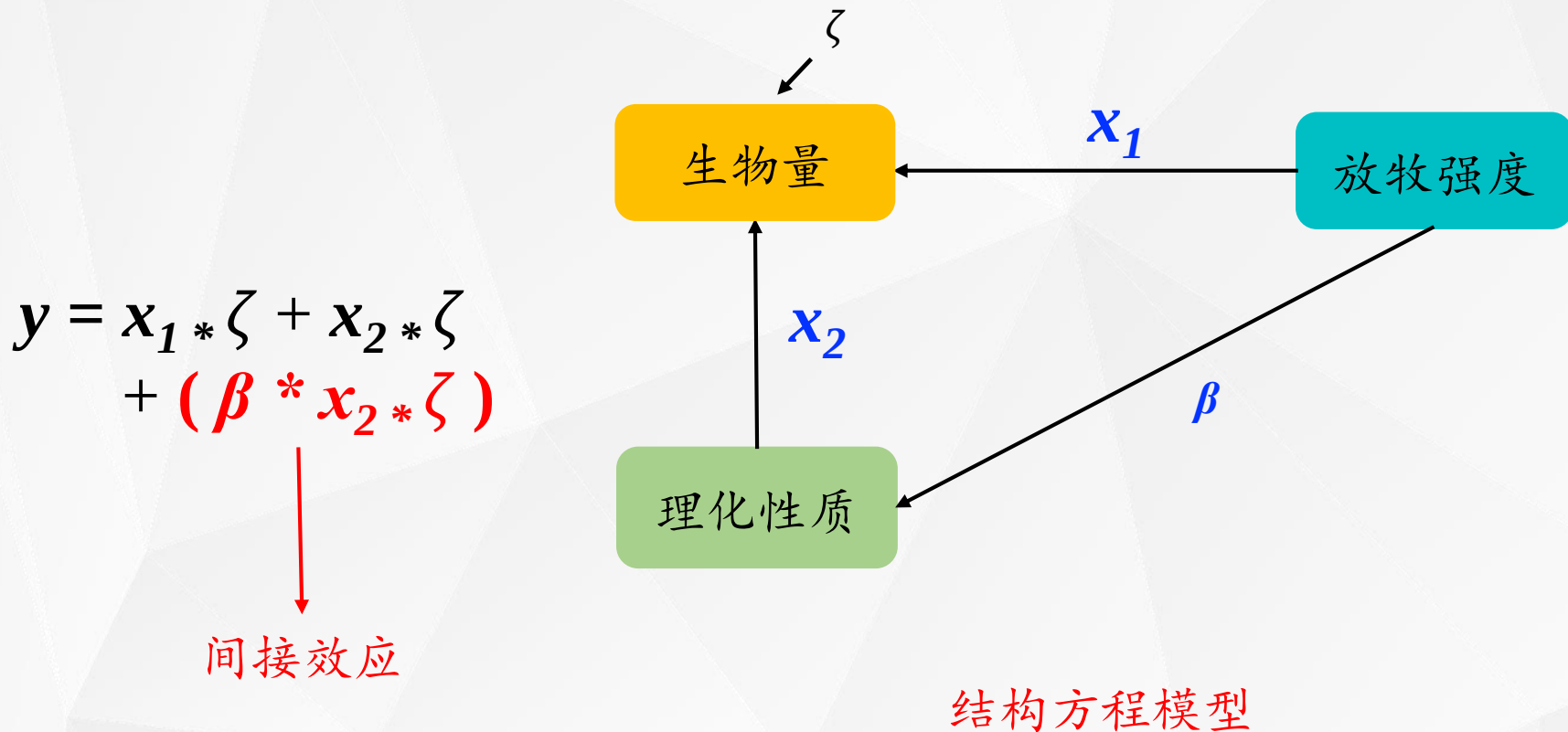
一元线性回归模型

● 为什么使用SEM?

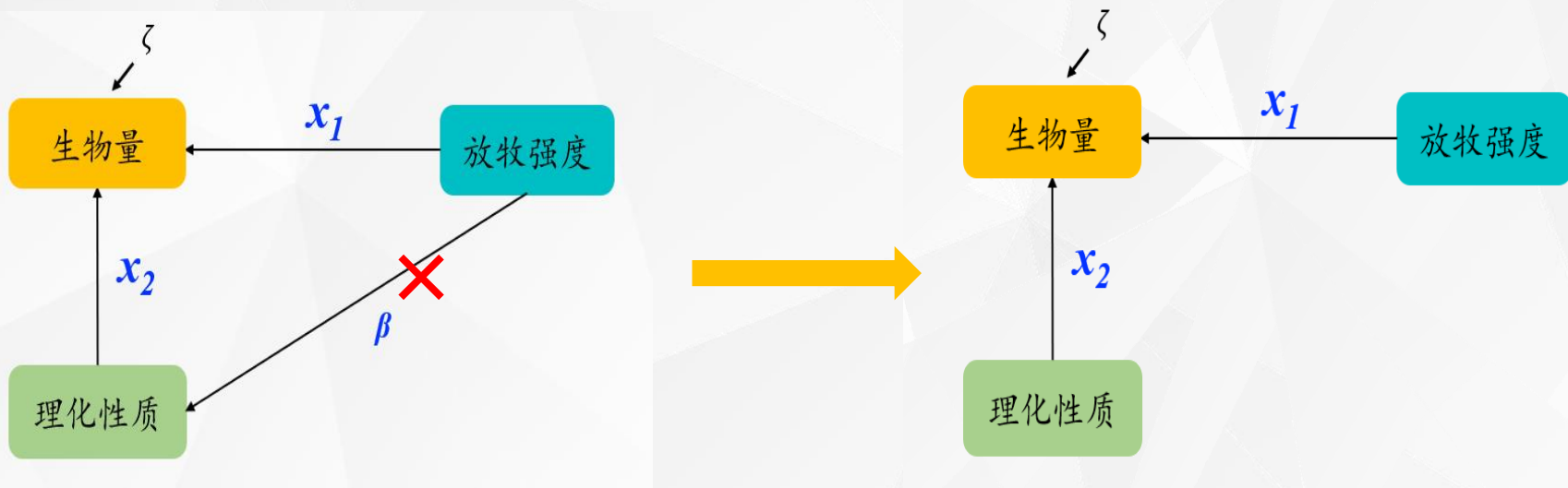


多元线性回归模型

● 为什么使用SEM?



为什么使用SEM?



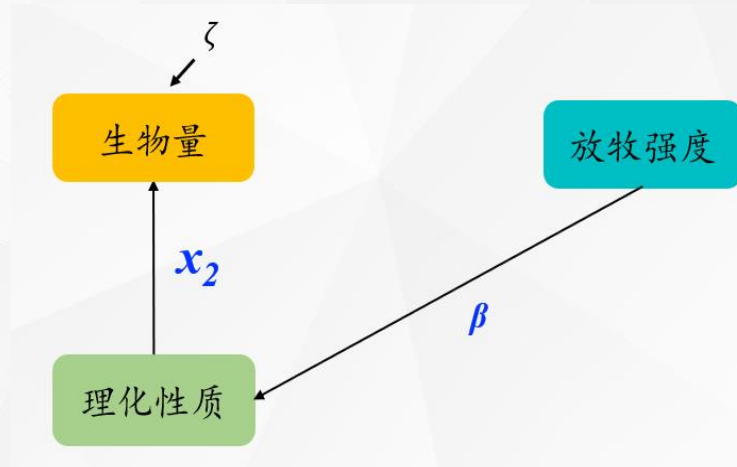
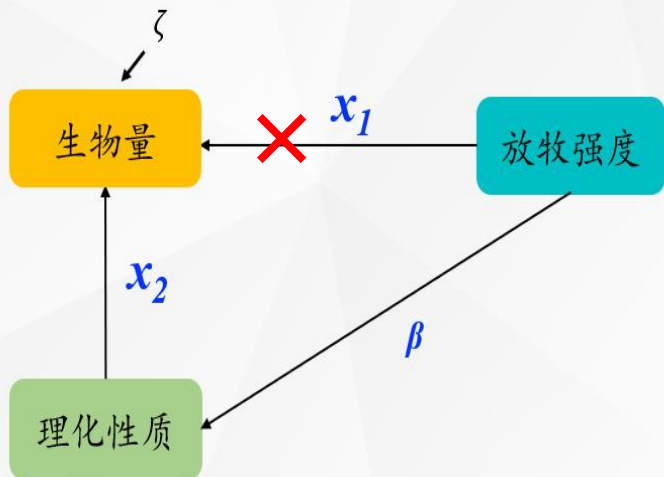
$$y = x_1 * \zeta + x_2 * \zeta + (\beta * x_2 * \zeta)$$

结构方程模型

$$y = x_1 * \zeta + x_2 * \zeta$$

多元线性回归模型

为什么使用SEM?



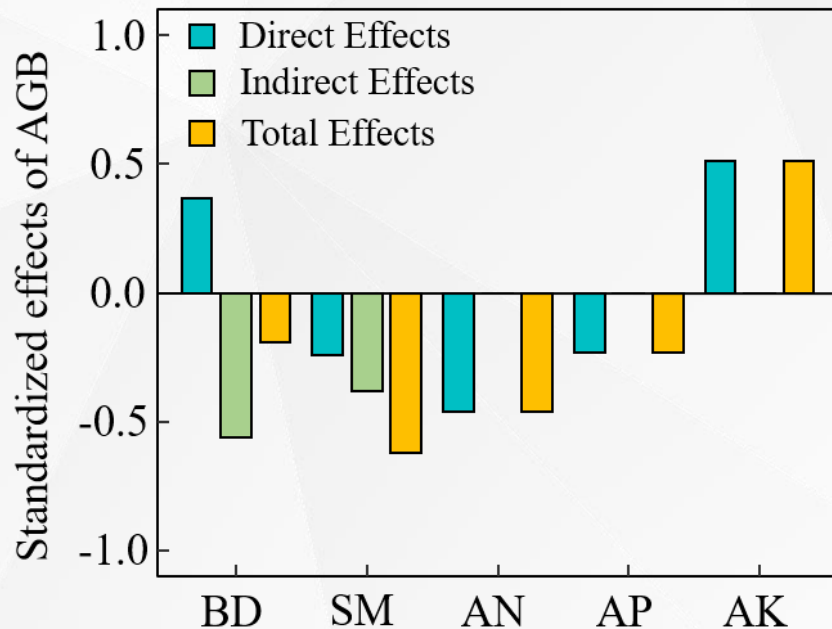
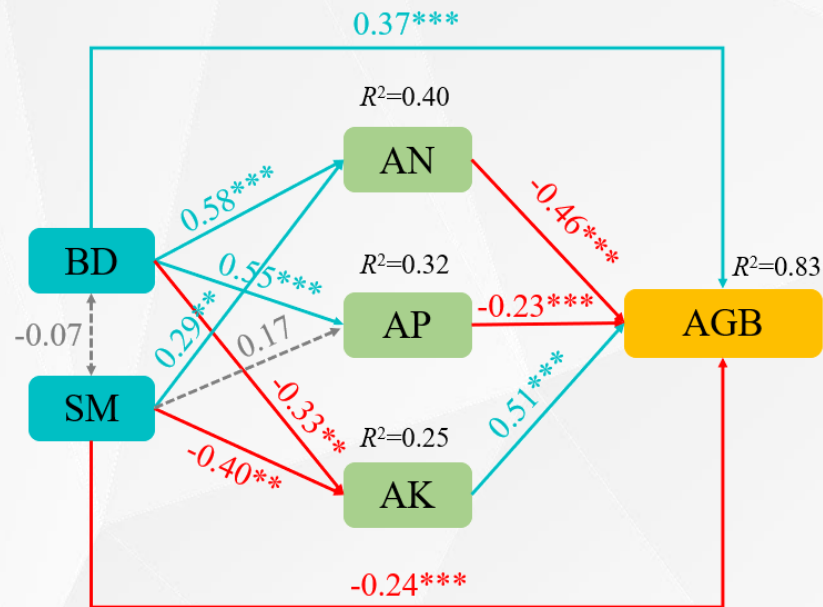
$$y = x_1 * \zeta + x_2 * \zeta + (\beta * x_2 * \zeta)$$

$$y = x_2 * \zeta + (\beta * x_2 * \zeta)$$

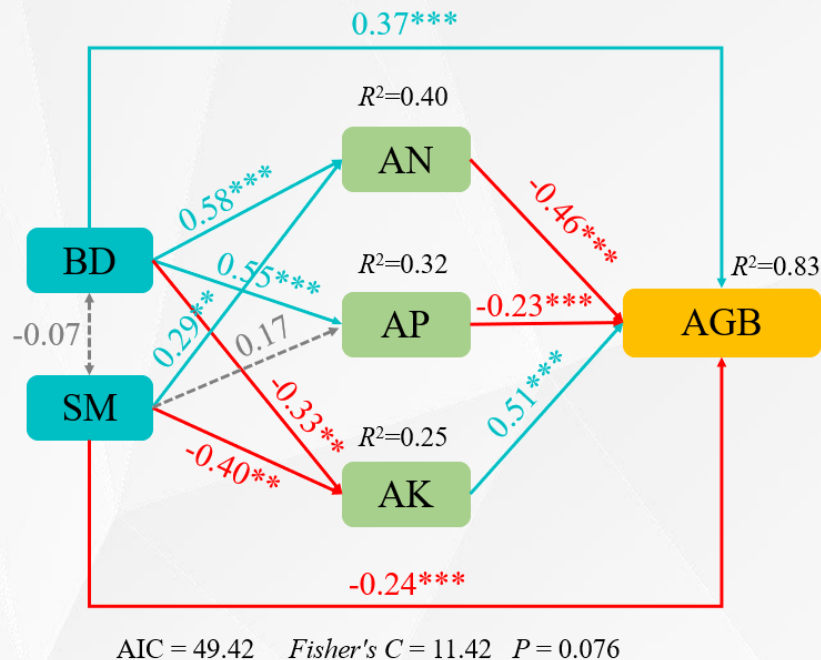
结构方程模型

结构方程模型

● 解读一个SEM



● 计算各种效应量



对AGB有**直接效应**的路径一共有5条，分别是：

BD → AGB 0.37

SM → AGB -0.24

AN → AGB -0.46

AP → AGB -0.23

AK → AGB 0.51

D=样本量/路径数

对AGB有**间接效应**的路径一共有6条，分别是：

BD → AN → AGB $0.58 \times (-0.46) = -0.2668$

BD → AP → AGB $0.55 \times (-0.23) = -0.1265$

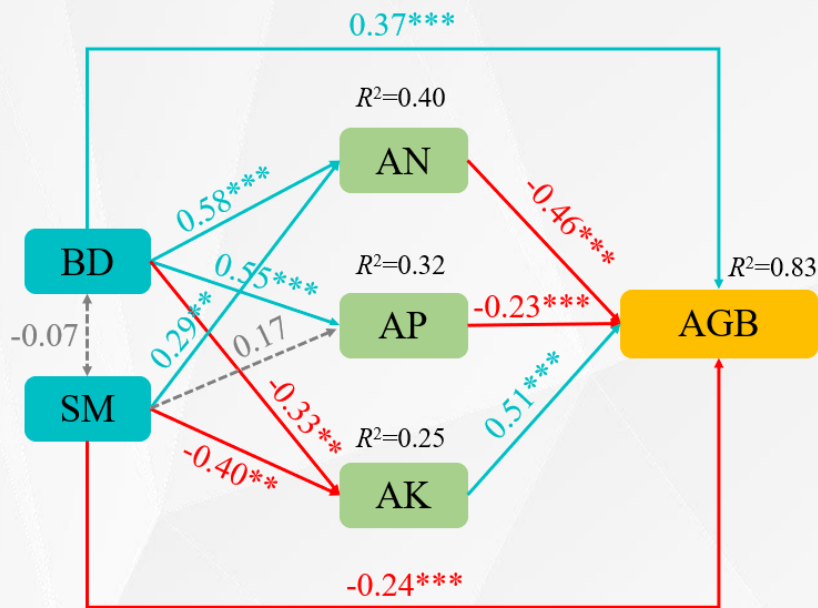
BD → AK → AGB $(-0.33) \times 0.51 = -0.1683$

SM → AN → AGB $0.29 \times (-0.46) = -0.1334$

SM → AP → AGB $0.17 \times (-0.23) = -0.0391$

SM → AK → AGB $(-0.40) \times 0.51 = -0.204$

● 计算各种效应量



那么我们的总效应其实就是
直接效应加上间接效应：

BD对AGB的总效应就是：

$$0.37 + (-0.2668) + (-0.1265) + (-0.1683) = -0.1916$$

SM对AGB的总效应就是：

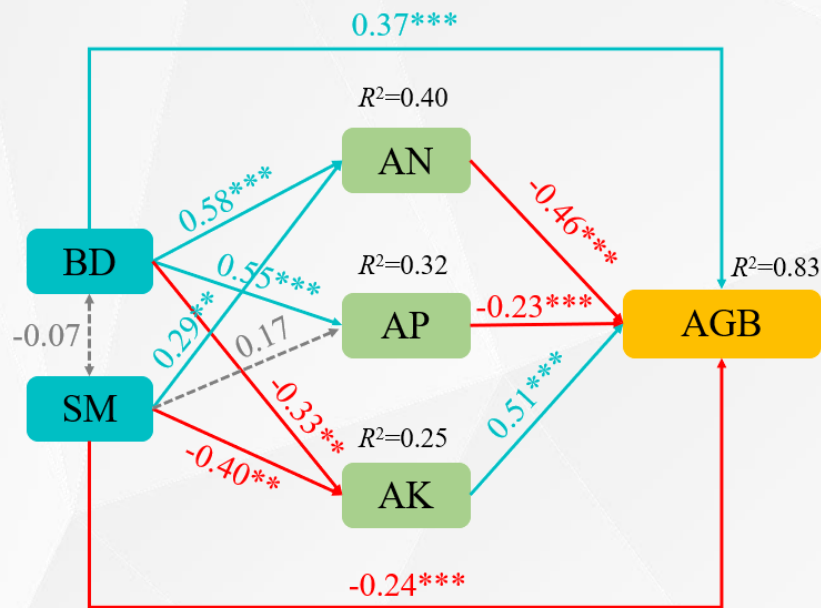
$$(-0.24) + (-0.1334) + (-0.0391) + (-0.204) = -0.6165$$

$$y = x_1 * \zeta + x_2 * \zeta + (\beta * x_2 * \zeta)$$

● 什么是SEM

- 回归只是SEM的特例，只考虑了直接影响；
- 可以同时考虑多个因果关系，量化直接效应、间接效应和总效应；
- 通过SEM可以探究当前获得的因果关系是否由另外因素共同导致，或通过其他因素介导；
- 常常和mixed model, meta-analysis 等配合使用，增加结果的系统性和可靠性；
- 让你的研究看起来高大上，但请不要过分复杂和花哨！

● 思考一下



AIC = 49.42 Fisher's C = 11.42 P = 0.076

我们的**目标因子**很好确定，
但是我们的**解释因子**是如何选择的？

我们**解释因子**间的相互关系是如何确定的？

什么情况下我们才**适合**做
SEM？

● 进入实际研究案例

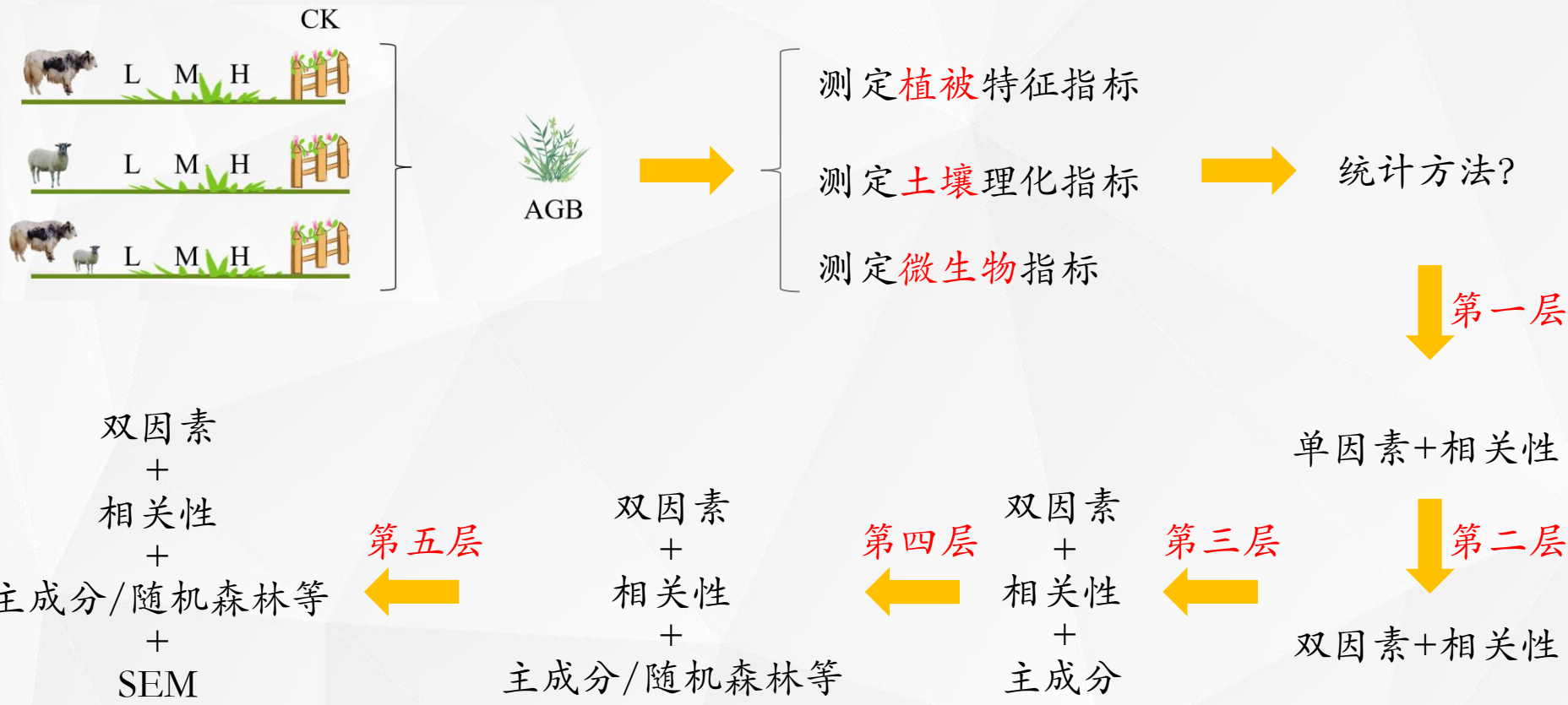
CK



AGB

● 怎样展开研究说明问题

试验设计



● 第二层 双因素+相关性

双因素

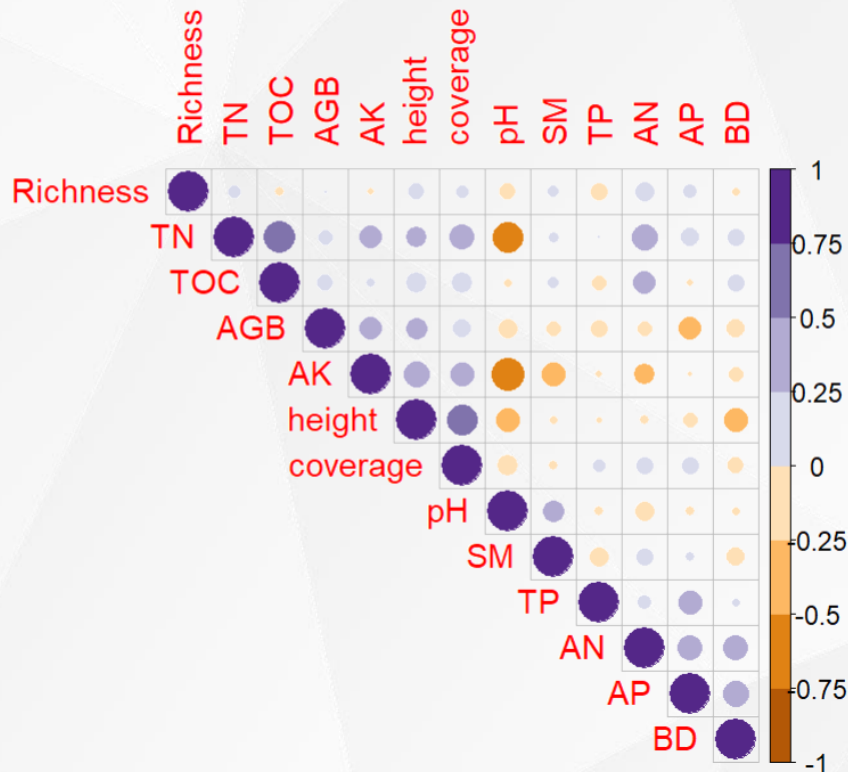
```
##-----双因素-----##  
mydata<- read.csv(file.choose())  
table(mydata$species, mydata$intensity)  
fit <- aov(AGB ~ species * intensity,data=mydata)  
twoway <- summary(fit)  
twoway  
""
```

```
> twoway
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
species	1	11567	11567	2.528	0.1190
intensity	1	29830	29830	6.520	0.0142 *
species:intensity	1	4103	4103	0.897	0.3488
Residuals	44	201296	4575		

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

相关性

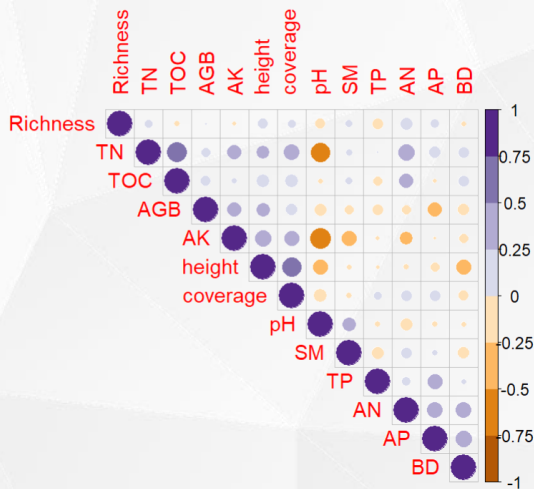


第三层 双因素+相关性+PCA

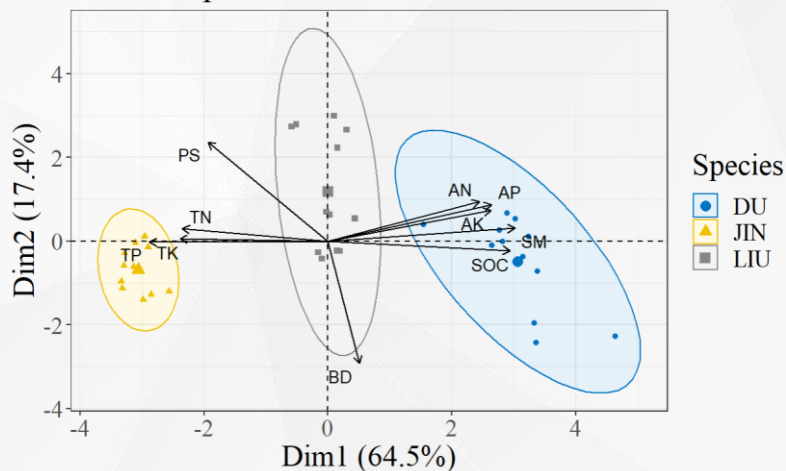
```
> twoway
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
species	1	11567	11567	2.528	0.1190
intensity	1	29830	29830	6.520	0.0142 *
species:intensity	1	4103	4103	0.897	0.3488
Residuals	44	201296	4575		

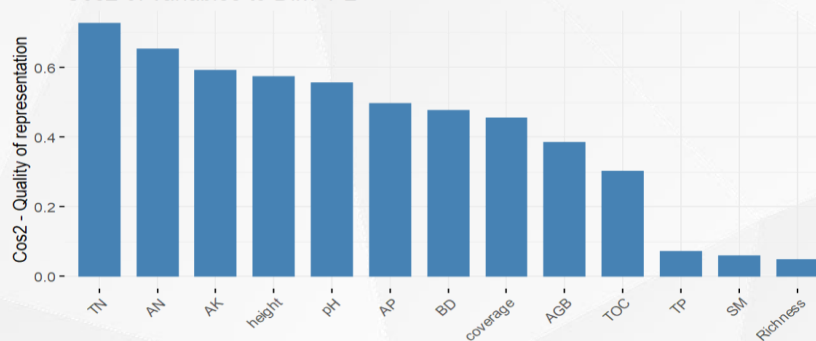
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



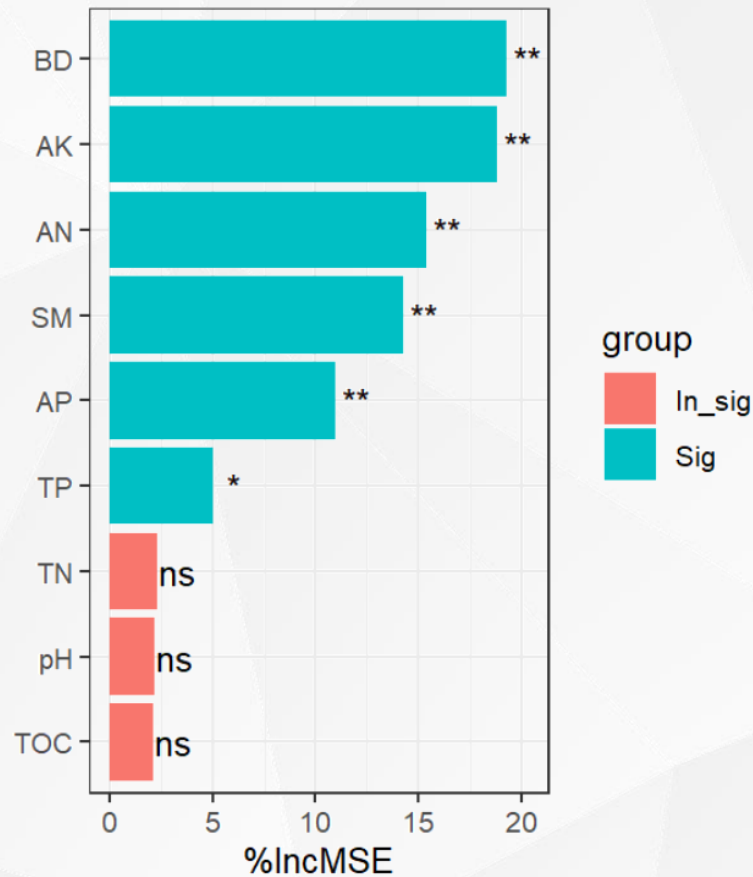
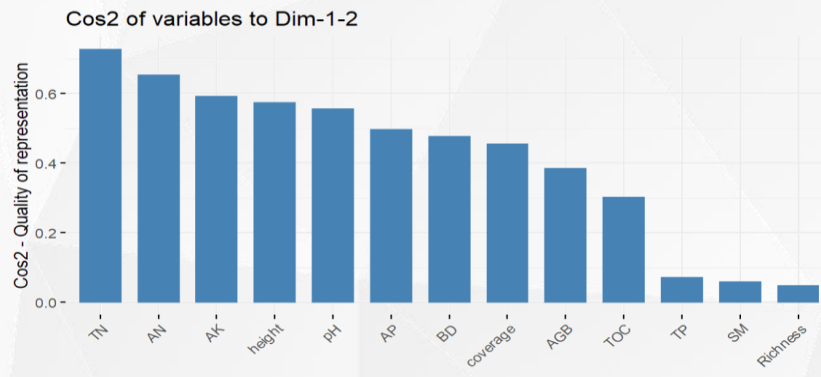
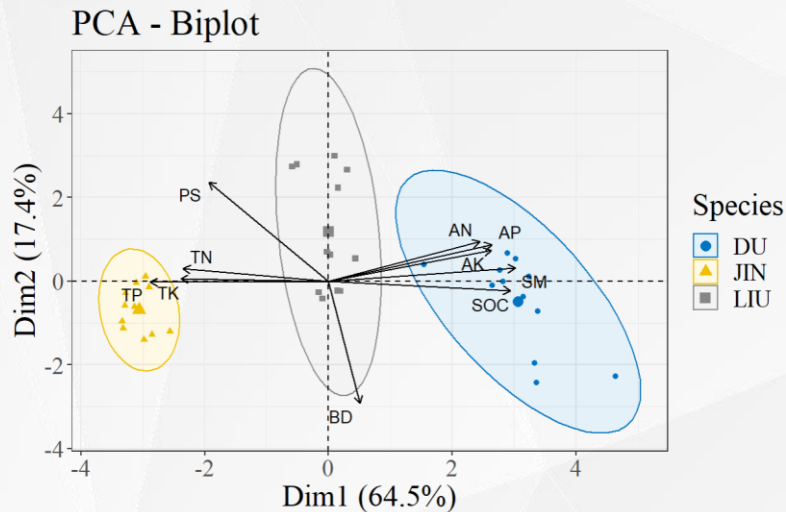
PCA - Biplot



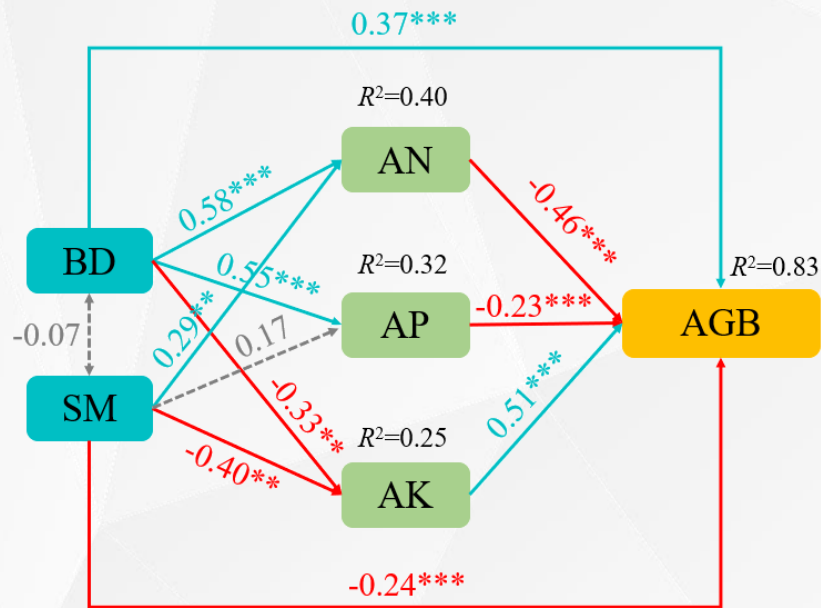
Cos2 of variables to Dim-1-2



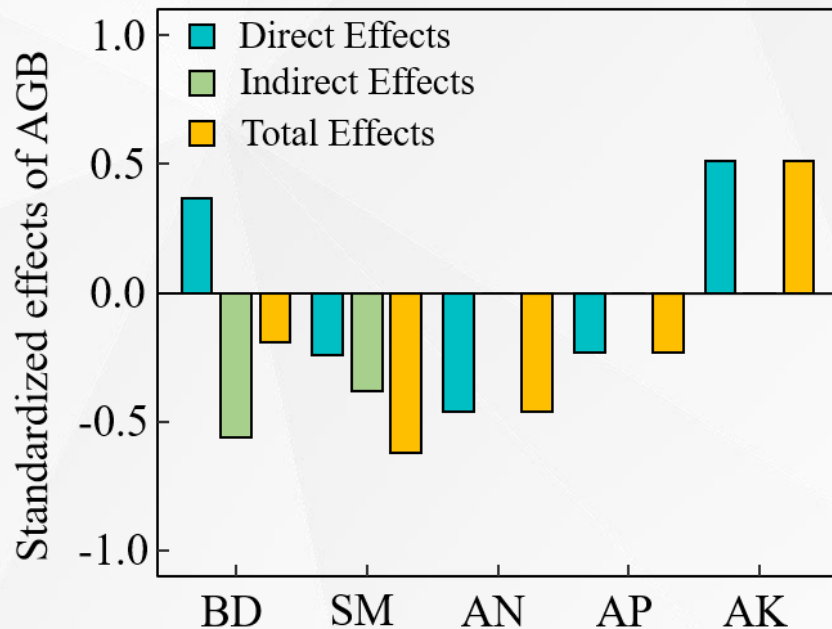
● 第四层 双因素+相关性+PCA+随机森林等



第五层 双因素+相关性+PCA+随机森林等+SEM



AIC = 49.42 Fisher's C = 11.42 P = 0.076



● SEM最常用的包

lavaan: An R package for structural equation modeling

[Y Rossee](#) - Journal of statistical software, 2012 - jstatsoft.org

... **lavaan** and how my initial objectives impacted the software design. Next, I illustrate the most characteristic feature of **lavaan**: the '**lavaan** ... SEM programs to use **lavaan**, there must be ...

☆ 保存 ↯ 引用 被引用次数: 16708 相关文章 所有 23 个版本 ↯

各种领域

PIECEWISESEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics

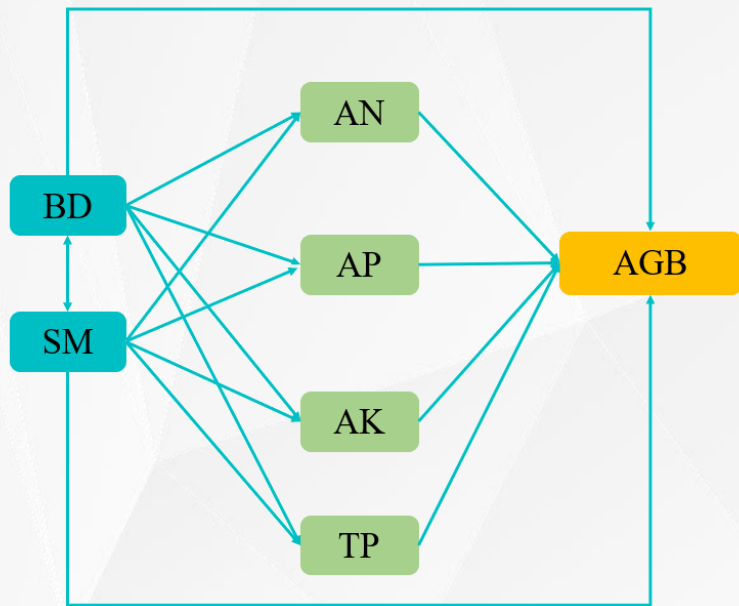
[JS Lefcheck](#) - Methods in Ecology and Evolution, 2016 - Wiley Online Library

... Here, I present a fully documented, open-source package **piecewiseSEM**, a practical implementation of confirmatory path analysis for the R programming language. The package ...

☆ 保存 ↯ 引用 被引用次数: 2209 相关文章 所有 8 个版本

绝大多数为生态领域

● piecewiseSEM



$BD \rightarrow AGB$

$SM \rightarrow AGB$

$AN \rightarrow AGB$

$AP \rightarrow AGB$

$AK \rightarrow AGB$

$TP \rightarrow AGB$

$BD \rightarrow AN \rightarrow AGB$

$BD \rightarrow AP \rightarrow AGB$

$BD \rightarrow AK \rightarrow AGB$

$BD \rightarrow TP \rightarrow AGB$

.....

```
model <- psem(  
  lm(AGB ~ BD+AK+AN+SM+AP+TP, litter ),  
  lm(AN ~ BD+SM, litter ),  
  lm(AP ~ BD+SM, litter ),  
  lm(AK ~ BD+SM, litter ),  
  lm(TP ~ BD+SM, litter ),  
  (SM%~~%BD)  
)
```

● 那么问题来了

1. 筛选变量重要性的方法只有相关性、主成分、随机森林吗？
2. 筛选时我们的目标因子只是一个指标，如果目标因子是一个矩阵（例如群落数据），有对应的方法筛选变量重要性吗？
3. SEM就只有这一种形式吗？



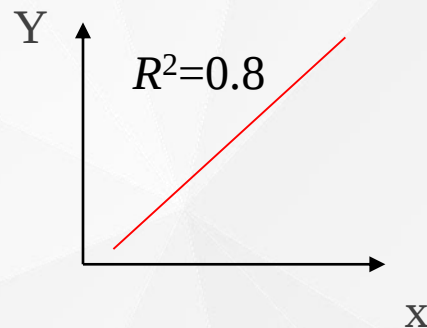
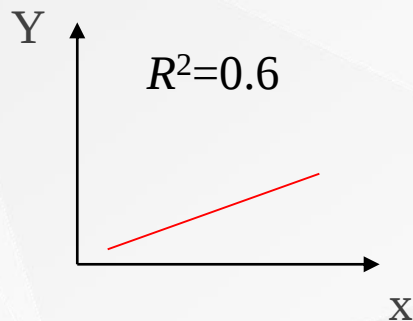
目录

02

以单个因子为目标的变量筛选

● 目标是单一指标

一元回归



多元回归

Received: 18 October 2021 | Accepted: 10 December 2021

DOI: 10.1111/2041-210X.13800

APPLICATION

Methods in Ecology and Evolution 

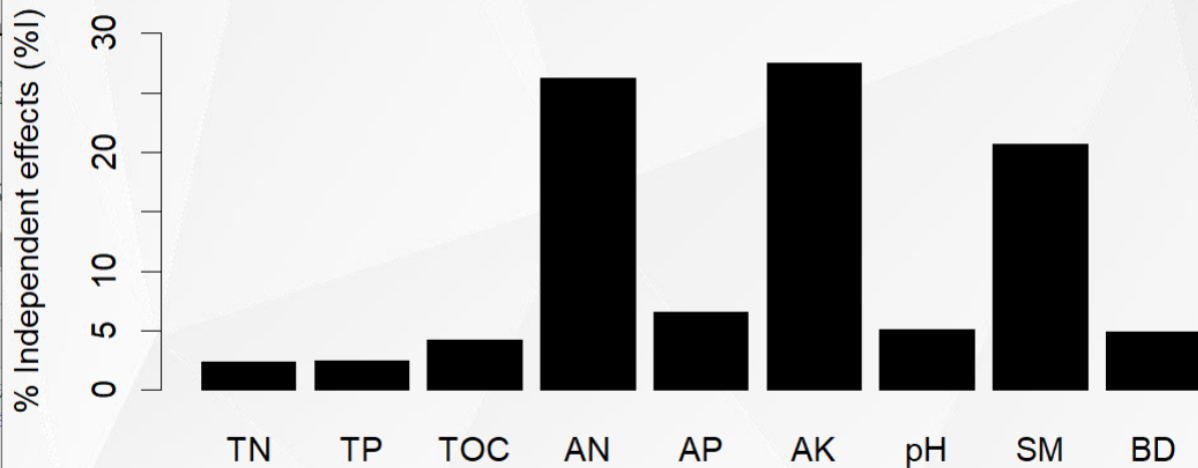
Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the `rdacca.hp` R package

Jiangshan Lai^{1,2}  | Yi Zou³  | Jinlong Zhang⁴  | Pedro R. Peres-Neto⁵ 

● 目标是单一指标

多元回归

hier.part 分解广义线性模型中变量的重要性

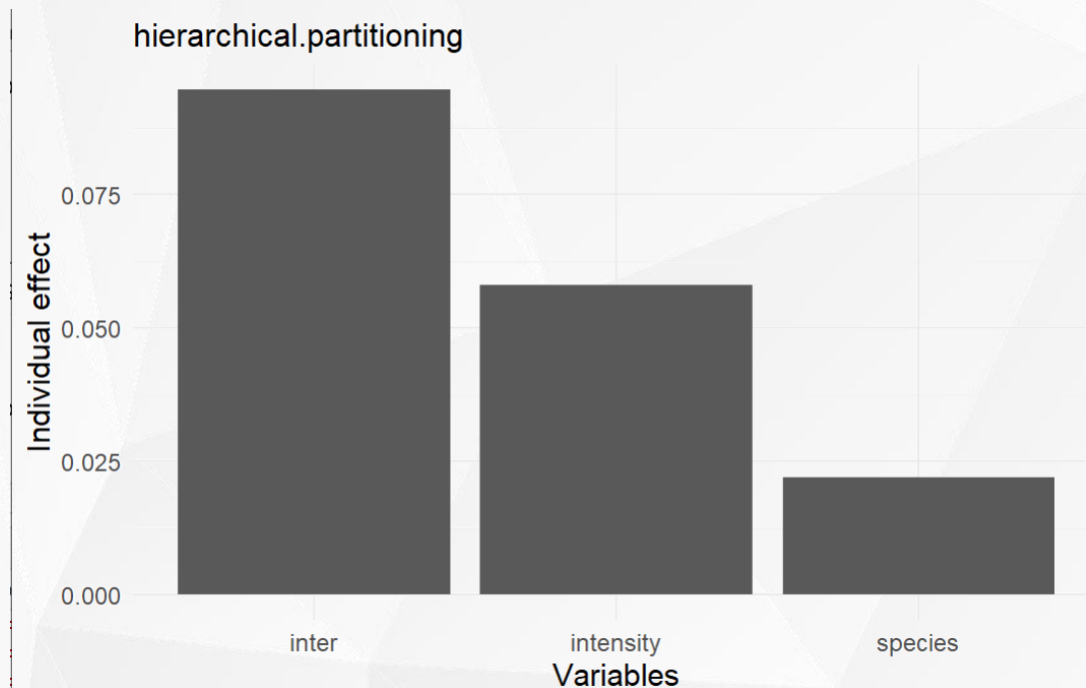


```
> #展示每个解释变量占点
> hier_part$I.perc
      ind.exp.var
TN      2.344309
TP      2.497052
TOC     4.250247
AN     26.258416
AP      6.568495
AK     27.459742
pH      5.081951
SM     20.690535
BD      4.849253
```

● 目标是单一指标

有交互的时候呢？

glmm.hp 分解混合效应模型中变量的重要性



```
$r.squaredGLMM
              R2m          R2c
[1,] 0.1746512 0.1746512

$hierarchical.partitioning
              Individual I.perc(%)
species          0.0220         12.59
intensity        0.0581         33.26
inter            0.0946         54.15

attr(,"class")
[1] "glmmhp"
> |
```



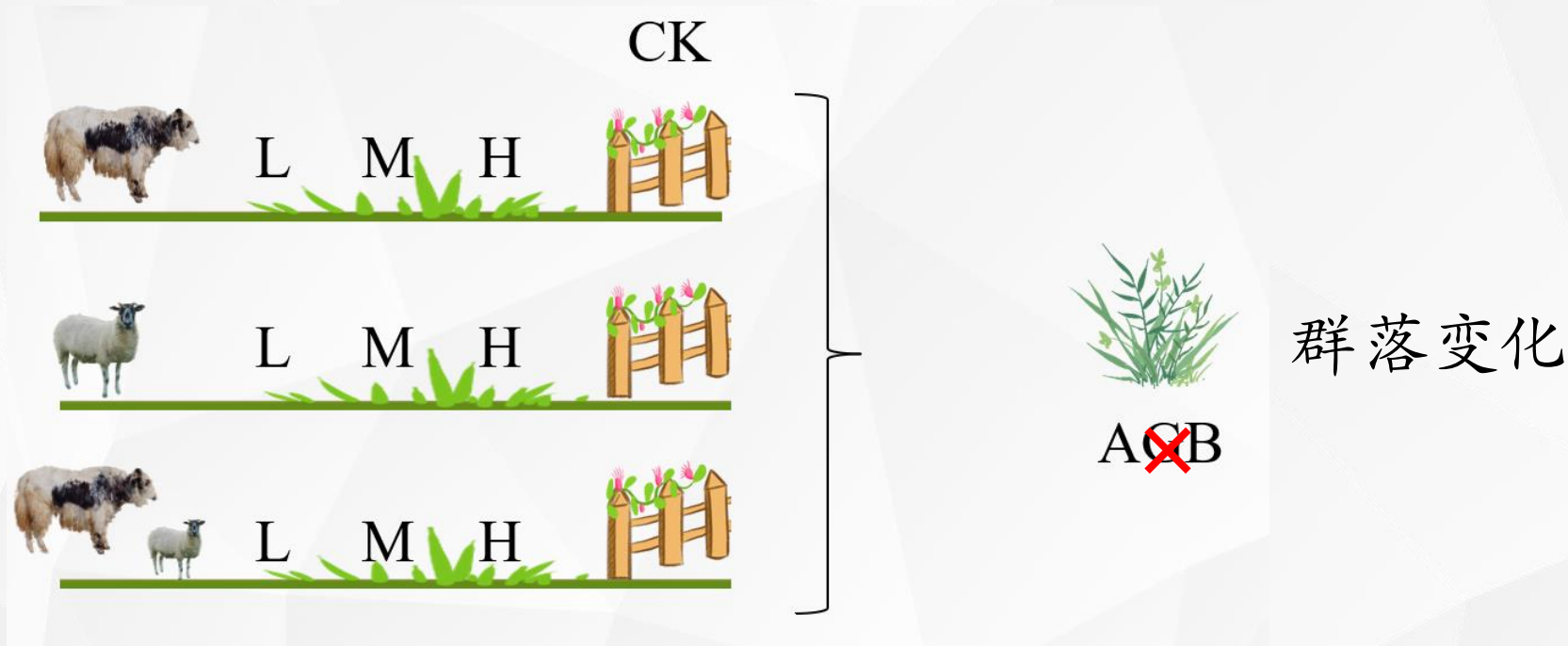
目录

03

以群落组成为目标的变量筛选

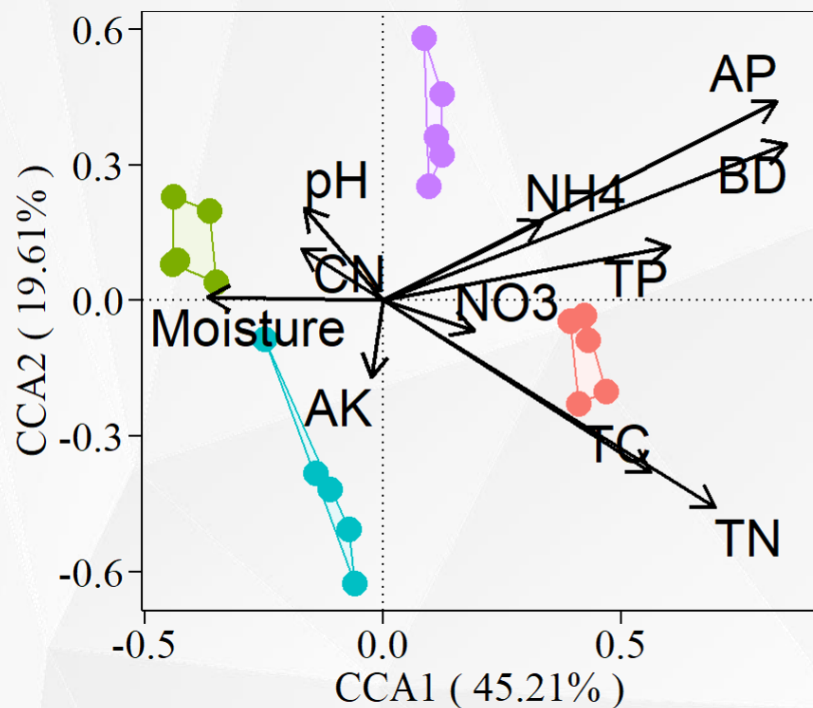
● 目标是矩阵

RDA/CCA/Mantel



● 目标是矩阵

RDA/CCA



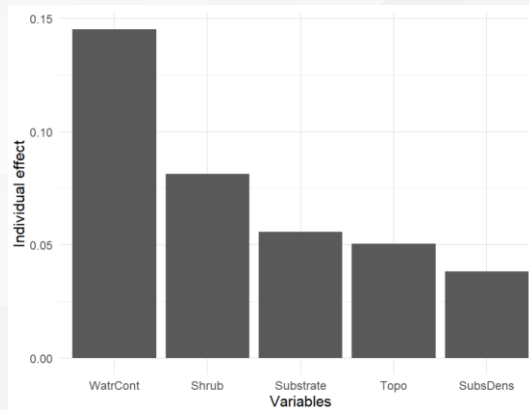
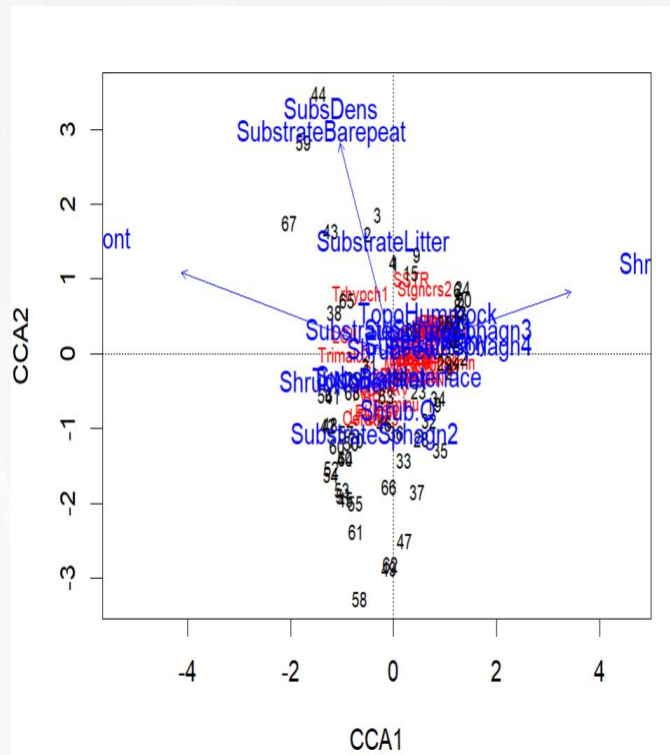
狼堡
青青草原
西域刀羊
有座桥

> K

	r2	p-value
TC	0.40524048	0.010
TN	0.61575203	0.001
CN	0.03735918	0.709
TP	0.36340831	0.017
NO3	0.03708327	0.715
NH4	0.13981037	0.278
AP	0.85959860	0.001
AK	0.02638697	0.802
pH	0.05957245	0.606
Moisture	0.12862063	0.316
BD	0.81756787	0.001

● 目标是矩阵

RDA/CCA 的层次分割



```
> mite.cca.hp
$Method_Type
[1] "CCA" "adjR2"

$Total_explained_variation
[1] 0.371

$Hier.part
      Unique Average.share Individual I.perc(%)
SubsDens 0.0299      0.0084      0.0383      10.32
WatrCont 0.0858      0.0594      0.1452      39.14
Substrate 0.0519      0.0038      0.0557      15.01
Shrub     0.0212      0.0602      0.0814      21.94
Topo      0.0270      0.0235      0.0505      13.61

attr(,"class")
[1] "rdaccahp"
> plot(mite.cca.hp)
```

目标是矩阵

Mantel检验

Mantel's p

— < 0.01

— 0.01 - 0.05

— >= 0.05

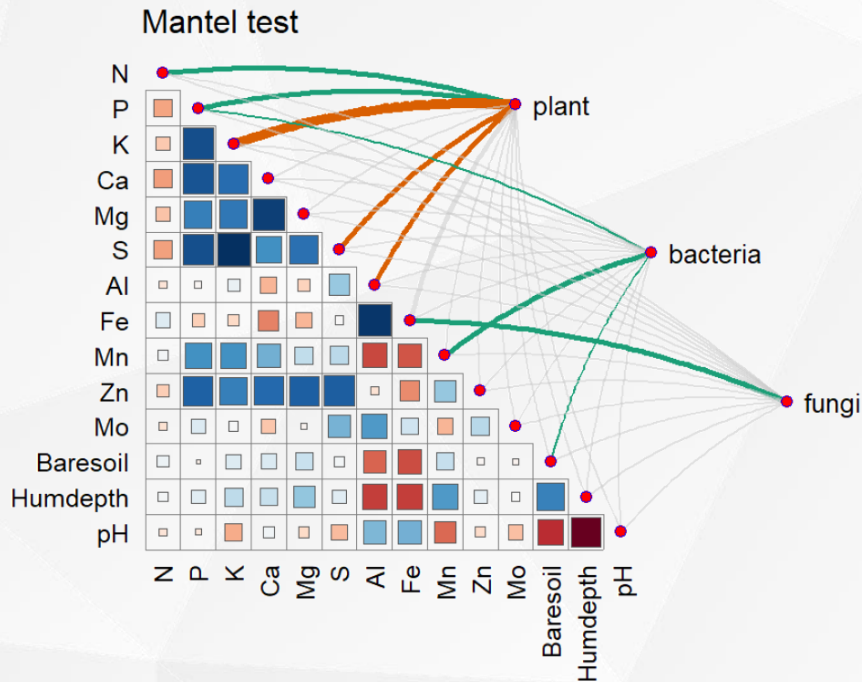
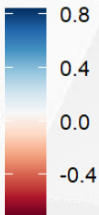
Mantel's r

— < 0.2

— 0.2 - 0.4

— >= 0.4

Pearson's r



```
> mantel01
```

```
# A tibble: 42 x 6
```

	spec	env	r	p rd	pd
	<chr>	<chr>	<dbl>	<dbl>	<fct>
1	plant	N	0.238	0.013	0.2 - 0.4 0.01 - 0.05
2	plant	P	0.226	0.015	0.2 - 0.4 0.01 - 0.05
3	plant	K	0.440	0.001	>= 0.4 < 0.01
4	plant	Ca	0.171	0.069	< 0.2 >= 0.05
5	plant	Mg	0.136	0.14	< 0.2 >= 0.05
6	plant	S	0.295	0.003	0.2 - 0.4 < 0.01
7	plant	Al	0.387	0.001	0.2 - 0.4 < 0.01
8	plant	Fe	0.221	0.054	0.2 - 0.4 >= 0.05
9	plant	Mn	0.117	0.169	< 0.2 >= 0.05
10	plant	Zn	0.0785	0.223	< 0.2 >= 0.05

```
# ... with 32 more rows
```

```
# i Use `print(n = ...)` to see more rows
```



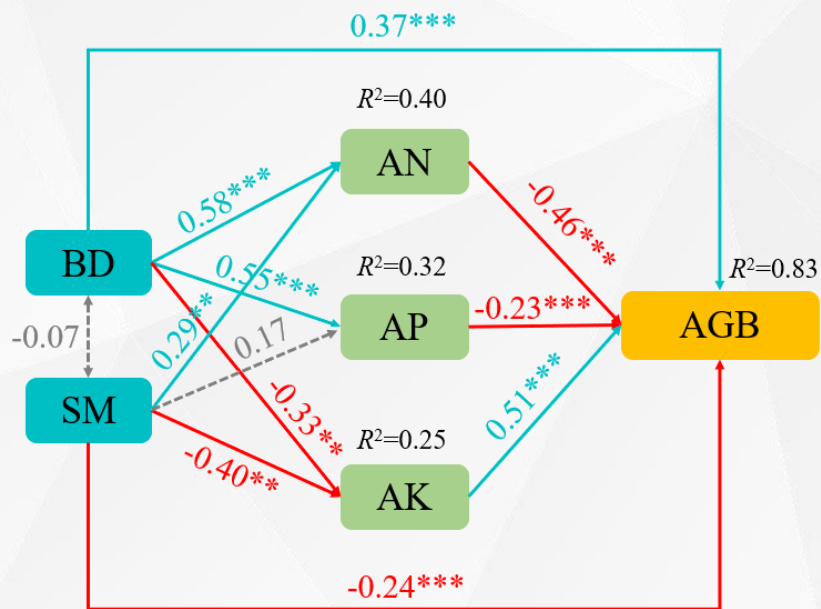
目录

04

piecewiseSEM的多种应用场景

● 应用场景一

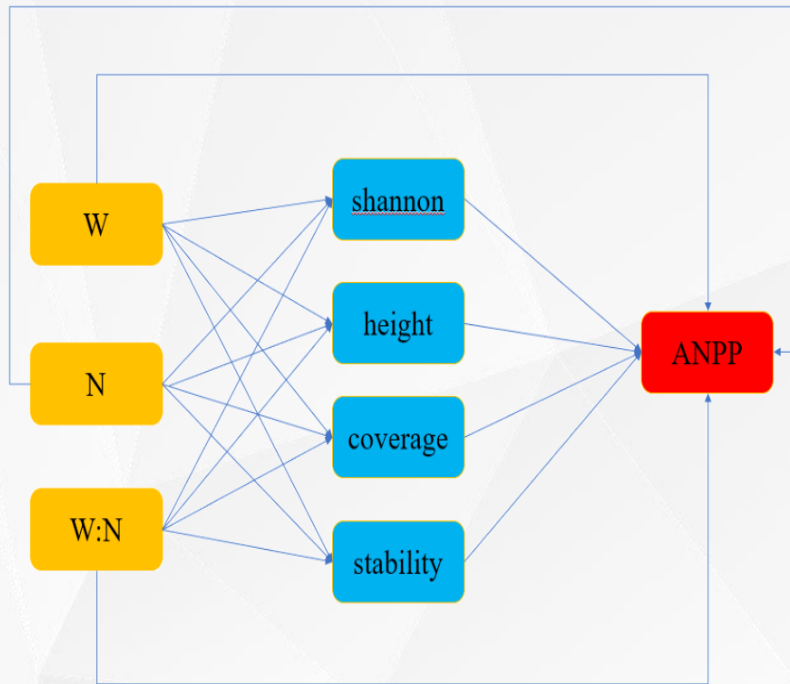
普通的SEM



AIC = 49.42 Fisher's C = 11.42 P = 0.076

● 应用场景二

嵌套双因素的SEM



#进行分析

```
model <- psem(  
  lm(ANPP ~ W*N+shannon+coverage+stability, mydata ),  
  lm(shannon ~ W*N, mydata ),  
  lm(coverage ~ W*N, mydata ),  
  lm(stability ~ W*N, mydata )  
)
```

#查看模型结果

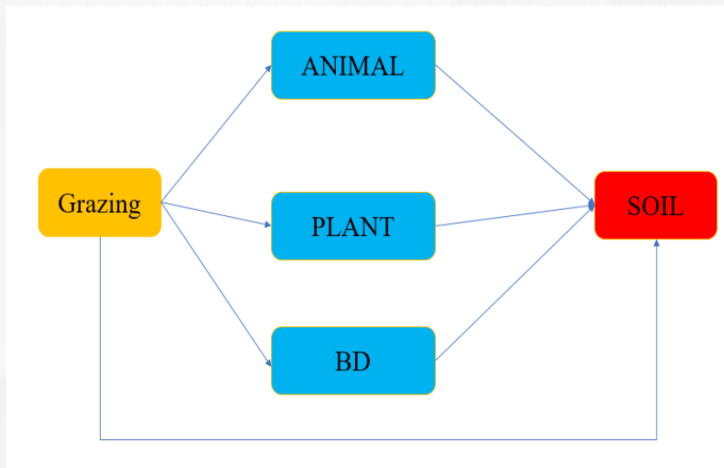
```
summary(model, .progressBar = F)
```

#简单绘图

```
plot(model)
```

● 应用场景三

嵌套混合效应模型的SEM

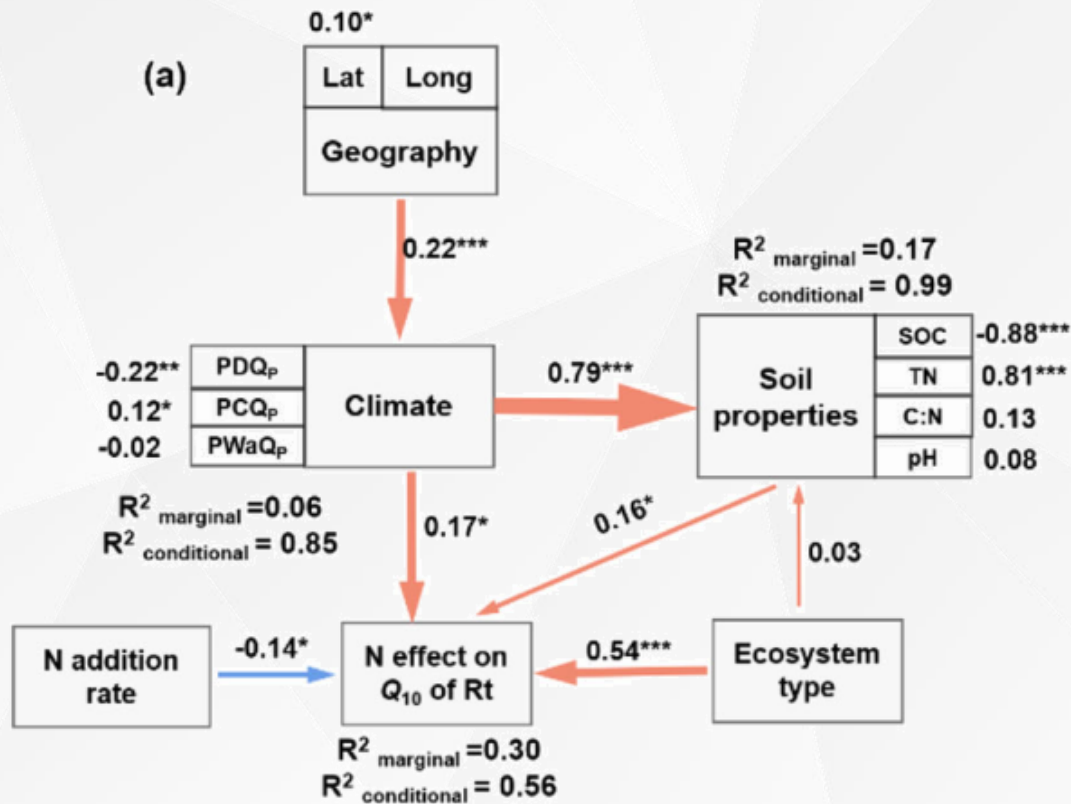


#进行分析

```
model <- psem(  
  lmer(SOIL ~ Grazing+ANIMAL + PLANT + BD + (1|Block), mydata ),  
  lmer(ANIMAL ~ Grazing+(1|Block), mydata ),  
  lmer(PLANT ~ Grazing+(1|Block), mydata ),  
  lmer(BD ~ Grazing+(1|Block), mydata )  
)
```

● 应用场景五

复合变量的SEM



Fisher's C = 9.99; $P = 0.44$; $df = 10$

AIC = 41.99; BIC = 95.77

● 总结一下

平时我们写论文时，大概就是这个思路



常规分析（单因素、双因素、多因素）



变量重要性排序（相关、PCA、RDA、
CCA、随机森林、层次分割、mantel检验）



SEM



谢谢
请大家批评指正