

结构方程模型在土壤微生物生态学中的应用

Applications of Structure Equation Modeling (SEM) in Soil Microbial Ecology

杨腾¹, 褚海燕^{1,2,*}

作者单位: 1. 中国科学院南京土壤研究所, 土壤与农业可持续发展国家重点实验室, 江苏, 南京; 2. 中国科学院大学, 北京

*通讯作者邮箱: hychu@issas.ac.cn

引用格式: 杨腾, 褚海燕. (2021). 结构方程模型在土壤微生物生态学中的应用. *Bio-101* e2003678. Doi: 10.21769/BioProtoc. 2003678.

How to cite: Yang, T. and Chu, H. Y. (2021). Applications of Structure Equation Modeling (SEM) in Soil Microbial Ecology. *Bio-101* e2003678. Doi: 10.21769/BioProtoc.2003678. (in Chinese)

摘要: 结构方程模型(SEM)结合了验证性因子分析和路径分析的思想, 可同时考虑多个自变量和因变量间的复杂因果关系, 通过建立、估计、检验和比较模型达到确定最佳模型, 解析变量间可能的因果关系的目标, 因而是一种高效的多元数据统计方法。经过逾百年的发展, SEM已广泛运用在生态学研究各领域。多种软件系统可执行SEM分析。本文以青藏高原高寒草地土壤真菌多样性分布研究为例, 结合气候、土壤属性、植物多样性和植物生产力等指标, 采用AMOS 21.0构建、验证并解析其结构方程模型, 阐述土壤真菌多样性在青藏高原高寒草原的分布规律, 给出各预测因子对土壤真菌多样性的直接与间接效应, 为结构方程模型在土壤微生物生态学中的进一步应用提供科学案例与参考。

关键词: 结构方程模型; 土壤真菌多样性; 植物多样性; 间接效应; 直接效应

研究背景: SEM的原理来自验证性因子分析和路径分析统计思想的结合。其中, 验证性因子分析可视为探索性因子分析的扩展, 而路径分析则为多重回归的扩展。根据研究者的先验知识预先设定理论模型, 而后通过估计、检验、比较和修缮(重设)得到最佳模型是SEM操作的基本步骤。SEM不仅可以处理测量变量, 也可执行潜变量的分析, 因而是一种高效的多元数据统计方法。

早期的 SEM 可追溯到 Wright 在上世纪 20 年代初有关路径分析的工作。通过路径分析获得路径图和路径系数,研究者可以对系统内各因子的关系进行分解并揭示这些关系背后的潜在机理,确定不同因子对同一过程的直接和间接作用,并判别它们的相对重要性等(王茜石和储诚进, 2011)。传统的生物统计学、计量经济学、心理测验学和社会计量学等均对当今 SEM 的发展起到推动作用。

土壤微生物生态学是生态学领域的“后起之秀”,其研究成果有助于我们更好地应对气候变化、防治土地污染,建立绿色农业生态系统等。得益于生物信息和测序技术的发展,土壤微生物生态学已越来越受到关注和青睐,而高效先进的统计学方法对土壤微生物生态学的发展同样必不可少。如 SEM,作为重要的多元统计方法,必将为土壤微生物生态学研究增添新的活力。

仪器设备

1. 普通 Windows 系统个人电脑,内存 8G,硬盘空间(含软件)2G 以上。

软件

1. IBM SPSS Statistics Version 20, IBM SPSS Amos 21.0.0 (Build 1178), Excel 2016, PowerPoint 2016。

注:本教程是基于已经在个人电脑上安装好的相关软件。安装问题不在本文范围内,请读者自行上网搜索。

实验步骤

1. 测试数据与理论模型

本文的测试数据来源先前对青藏高原高寒草地土壤真菌多样性分布规律的研究(Yang *et al.*, 2017),涉及 60 个 100 × 100 m 样地的地上植物多样性、土壤真菌多样性(以丰富度代表)、地上植物生产力(以生长季 NDVI 平均值代表)、土壤属性和气候因子。首先,根据前人研究和先验知识构建 SEM 理论模型(图 1)。其中每条路径均可标注对应的参考文献以及假设的正负影响,此处省略;需要的小伙伴可参考 Lange *et al.* (2015)的文章。具体的数据操作中,我们以年降雨(MAP)代表气候,以土壤总氮(TN)代表土壤养分,以土壤碳氮比(C:N ratio)代表土壤化学计量,以生长季

NDVI 的均值(NDVI mean)代表地上植物生产力,以真菌和植物的丰富度指数(F.richness 和 P.richness)分别代表土壤真菌和地上植物的多样性。数据表以 excel 形式保存(图 2)。

注:测试数据可通过百度网盘下载(链接: <https://pan.baidu.com/s/11z0-QJAU2LQ-JDC61NvqLA> 提取码: 8688)。

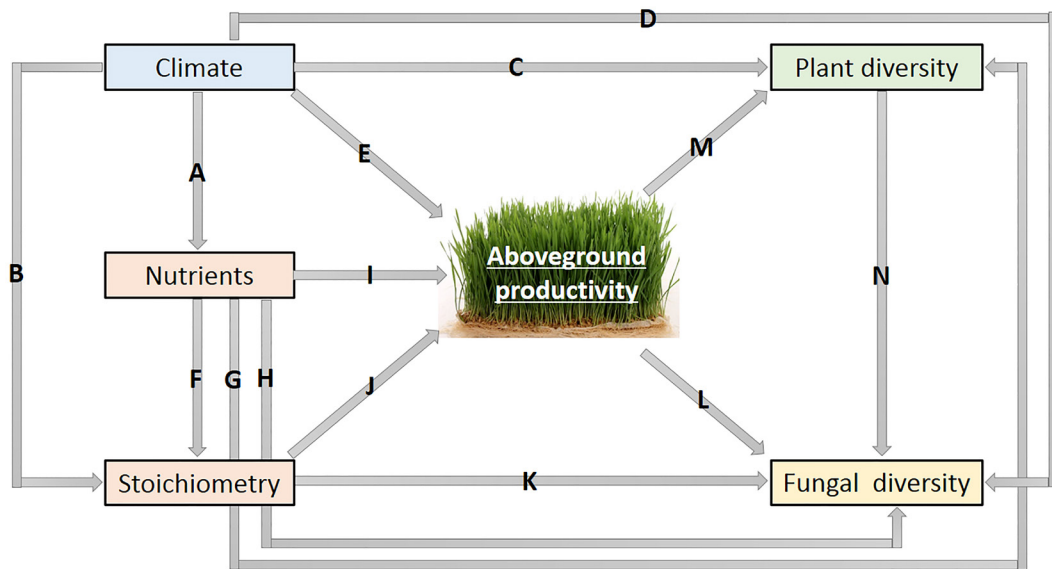


图 1. SEM 的理论模型

	A	B	C	D	E	F	G
1	Sample. Code	MAP	TN	CN. ratio	NDVI. mean	P. richness	F. richness
2	TP1	321	0.34	9.3	0.50	10	4002
3	TP2	333	0.07	22.0	0.26	7	3323
4	TP3	343	0.14	15.2	0.28	9	3787
5	TP4	405	0.51	11.1	0.80	28	4030
6	TP5	432	0.16	14.3	0.56	15	3710
7	TP6	494	0.92	11.9	0.65	21	3972
8	TP7	523	0.62	12.6	0.63	18	4402
9	TP8	503	0.63	12.4	0.61	27	3824
10	TP9	377	1.09	13.3	0.27	8	3636
11	TP10	335	0.10	26.9	0.27	6	3412
12	TP11	341	0.09	29.2	0.29	18	3571
13	TP12	478	0.47	11.9	0.62	20	3500
14	TP13	489	1.45	11.6	0.69	24	3317
15	TP14	511	0.91	12.6	0.69	22	3629
16	TP15	560	0.78	12.5	0.68	13	3690
17	TP16	485	0.64	13.3	0.48	12	3244
18	TP17	458	0.86	12.5	0.66	16	3383

图 2. 测试数据表格(SEM60.xlsx)

2. SPSS 操作—数据导入和转换

打开 SPSS 20 软件，点击“文件(F)-打开(O)-数据(A)”，进入“打开数据”对话框，将“文件类型”调为“Excel (*xls, *xlsx, *xlsm)”，导航到相应的目标路径下，点击“打开”按钮导入图 2 Excel 表格(图 3)。

点击“文件(F)-另存为(A)”，进入“将数据保存为”对话框，将“保存类型”调为“SPSS Statistics (*.sav)”，点击“保存”为 SEM60.sav(图 4)，以备 Amos 21.0.0 使用。

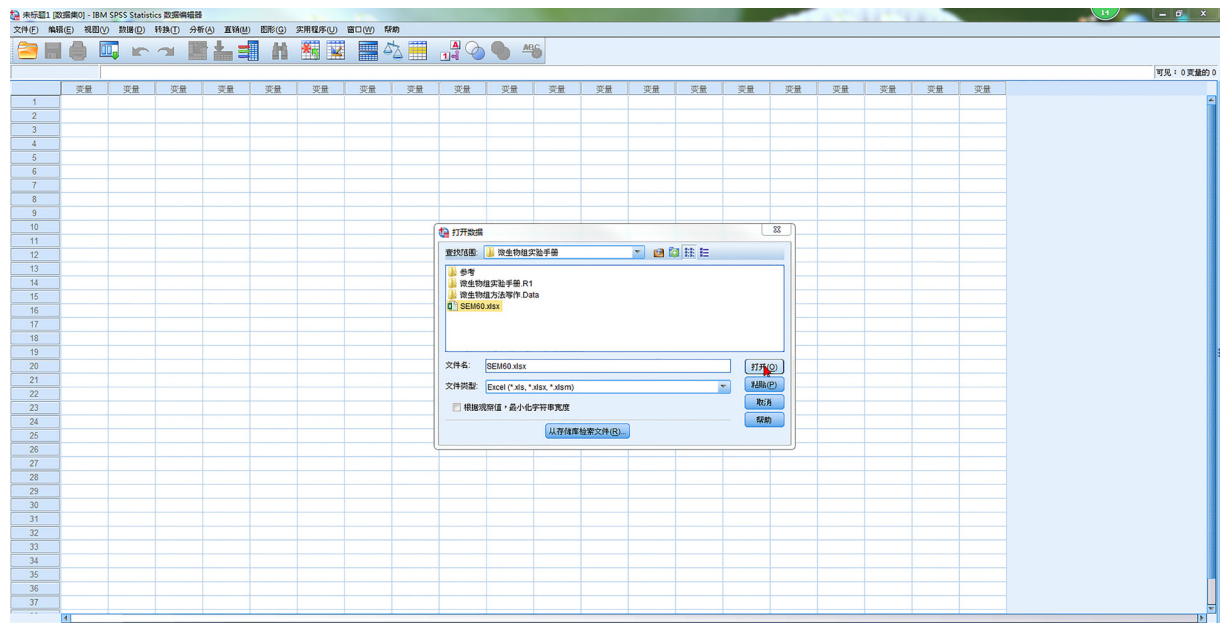


图 3. 向 SPSS 中导入 Excel 文件

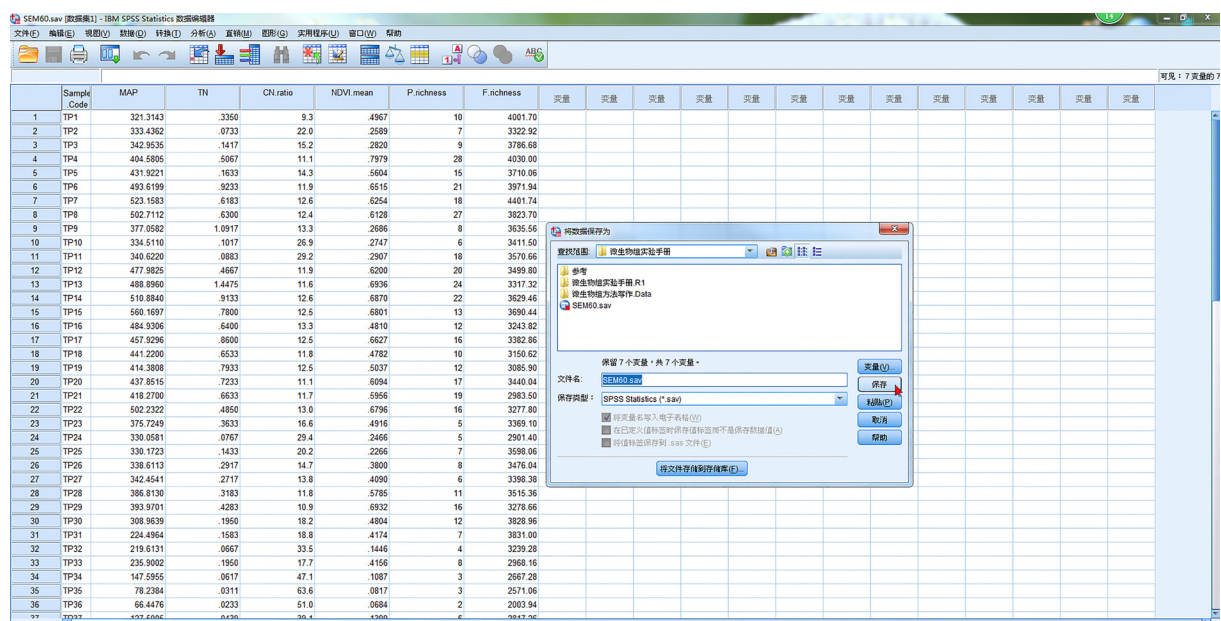


图 4. sav 格式文件的保存

3. Amos 操作—数据导入

打开 Amos 21 软件, 如图 5, 第一步点击左侧“Select data file(s)”图标弹出“Data Files”对话框, 第二步点击“File Name”找到 SEM60.sav 文件, 点击打开, 这时可见该文件名称出现在对话框中, 点击“OK”即可。此时, 点击左侧“List variables in data set”图标, 可见各变量名称均出现在“Variables in Dataset”的数据框内(图 6)。

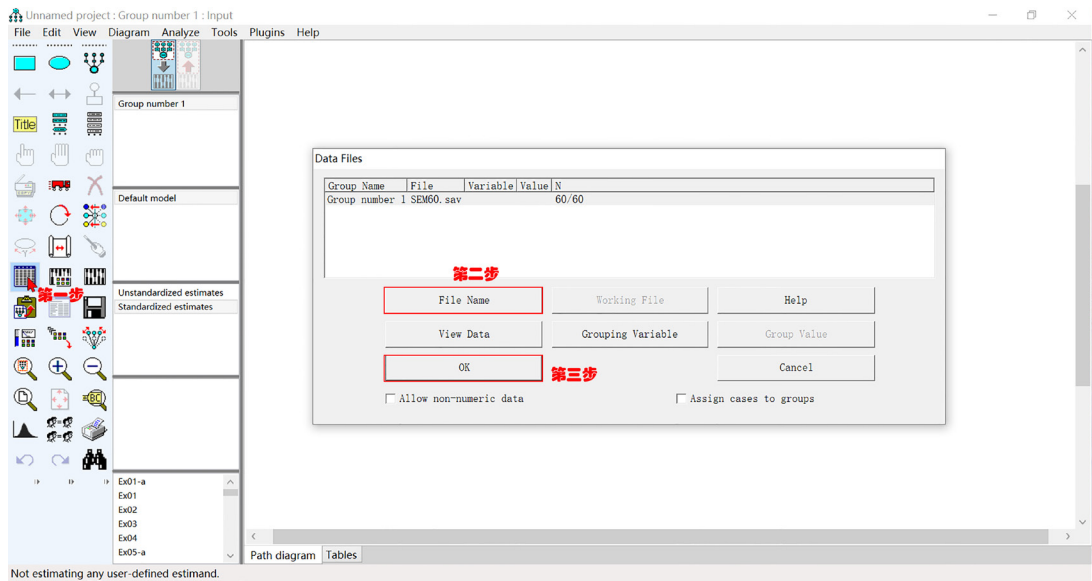


图 5. 向 Amos 中导入数据

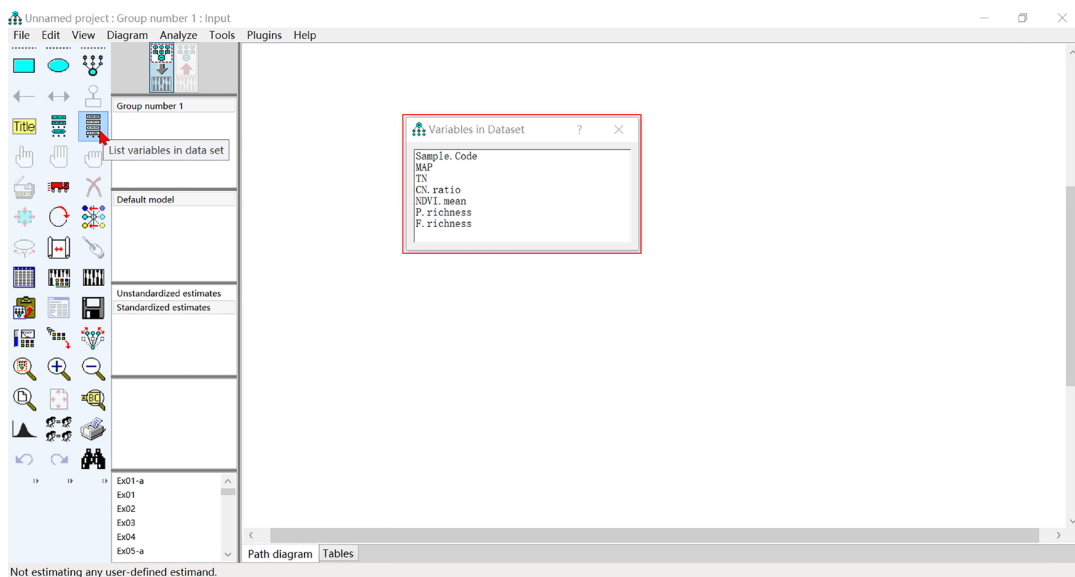


图 6. 导入数据的查看

4. Amos 操作—路径建构

根据图 1 SEM 理论模型构建路径。使用左侧“Draw observed variables”图标构建 1 个变量方格，使用“Duplicate objects”复制出另外 5 个，使用“Move objects”对以上 6 个变量方格在右侧画板上进行合理的布局调整。然后点击“List variables in data set”，将“Variables in Dataset”数据框内相应数据名称拖入对应的变量方格中，完成后可见图 7。

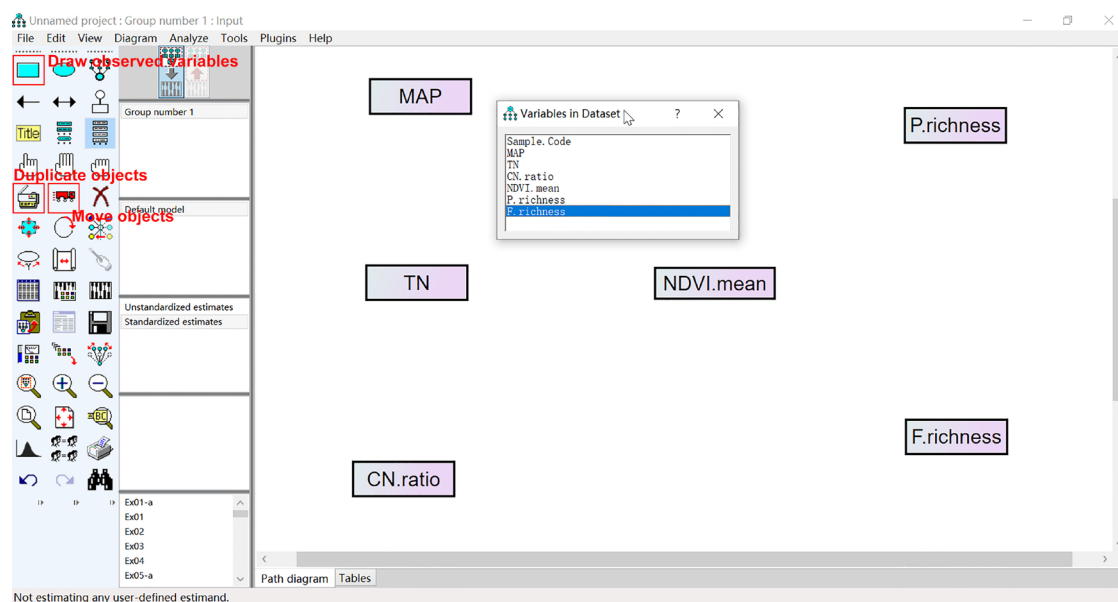


图 7. 变量的添加和布局

此时，使用左侧“Draw paths (single headed arrows)”图标，并根据图 1 构建变量之间的路径(图 8)。

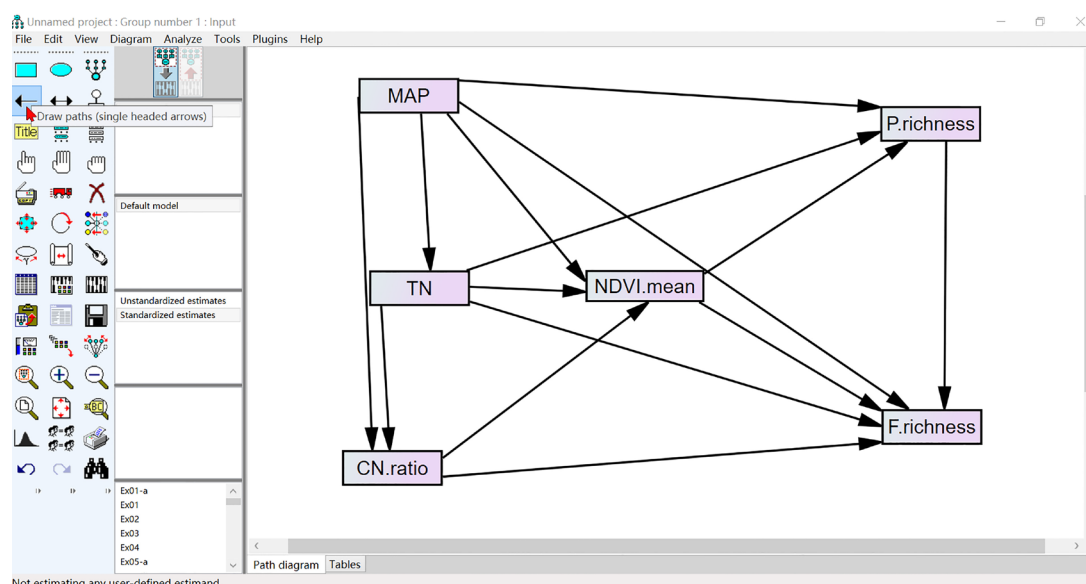


图 8. 路径的添加

5. Amos 操作—残差项的添加

使用左侧“Add a unique variable to an existing variable”图标对 TN, CN.ratio, NDVI.mean, P.richness 和 F.richness 设置残差项。点击上层菜单“Plugins-Name Unobserved Variables”完成对残差项的命名(图 9)。

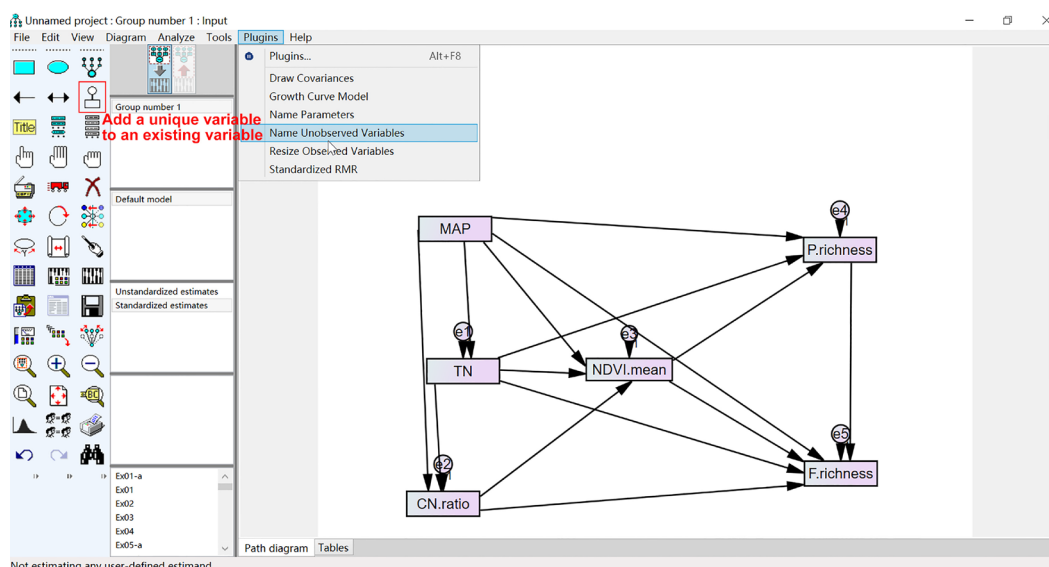


图 9. 残差项的添加

6. Amos 操作—数据运算

点击左侧“Analysis properties”图标，打开“Analysis properties”对话框，在“Estimation”选项卡中可见默认方法“Maximum likelihood”(图 10)。这里不改变任何默认选项，直接跳转到“Output”选项卡中，将所有可选标签勾上即可(图 11)。

注：“Maximum likelihood”法需要变量服从多变量正态分布，数据类型为区间或定量数据，也是最常用的模型评估方法之一；“Unweighted least squares”对变量分布无具体要求，但对测量数据的单位较为敏感。

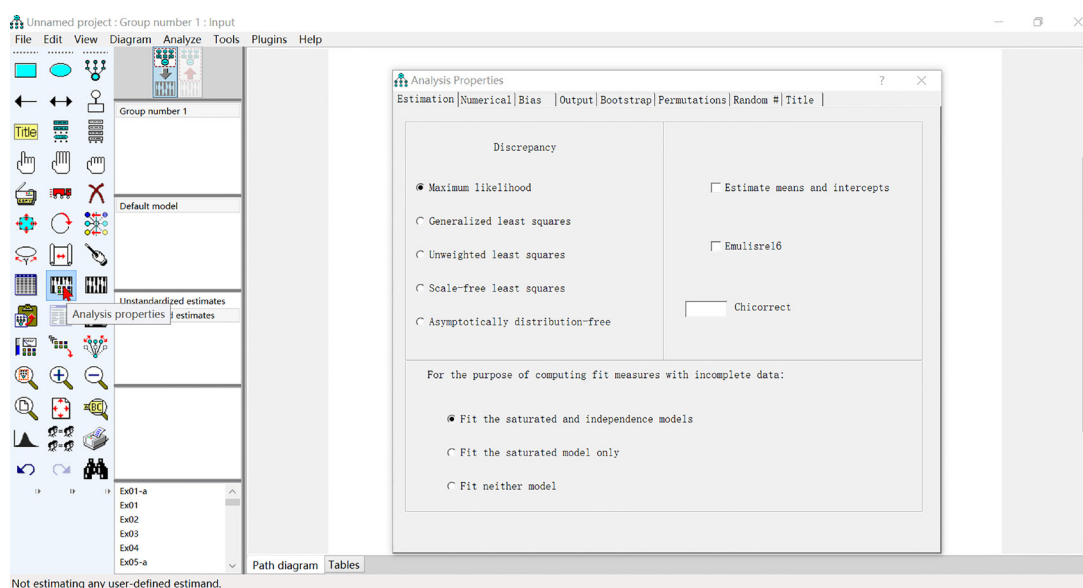


图 10. 打开“Analysis properties”对话框

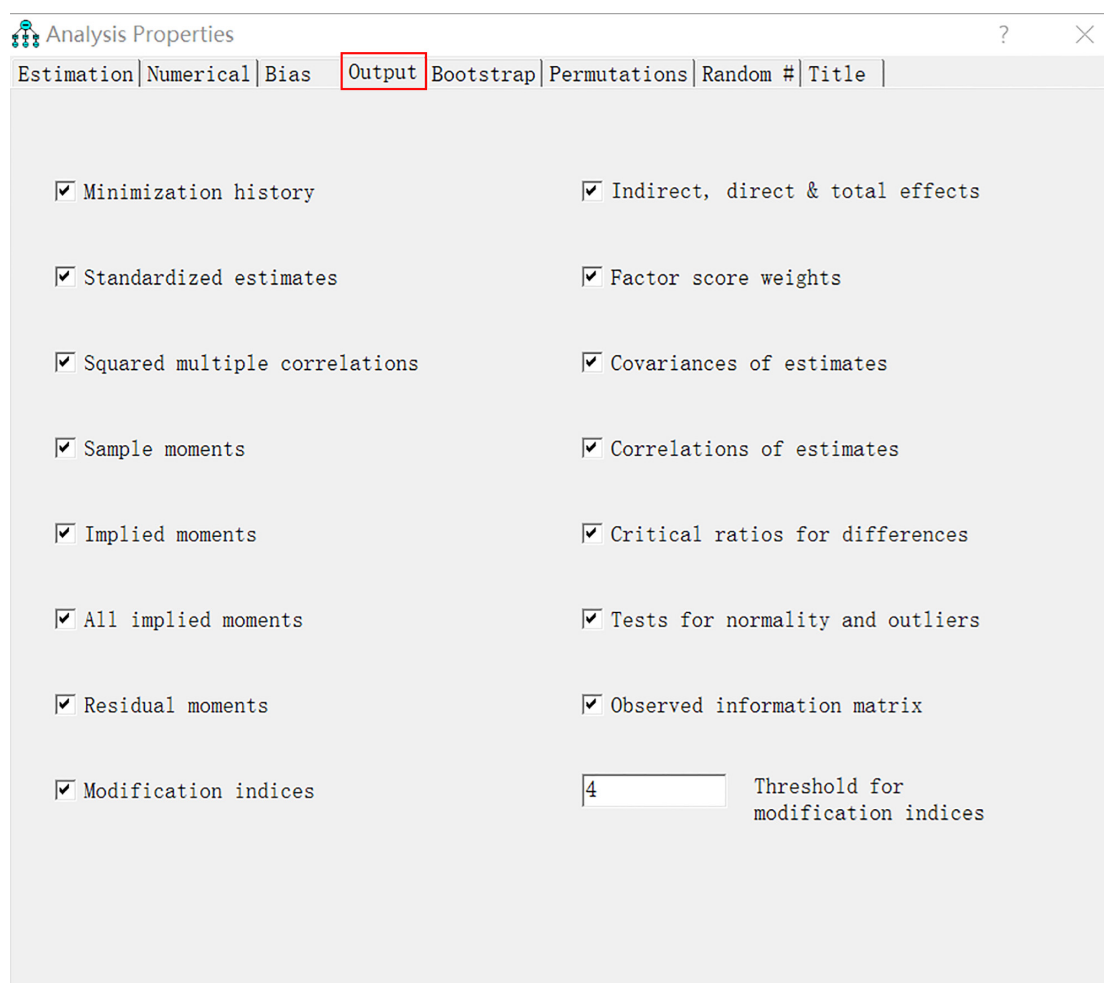


图 11. 勾选所有统计输出项

最后，关闭“Analysis properties”对话框，点击右侧“Calculate estimates”(位于“Analysis properties”图标的正右侧)，开始模型的统计运算。

7. Amos 操作—运算结果查看和重要参数解读

点击“View the output path diagram”图标，选择“Standardized estimates”选项后可见 SEM 的基本路径图结果(图 12)。该图中，路径上的系数为保留两位小数的标准化路径系数，反应了因果关系(路径)的强度，内因变量(TN, C:N ratio, NDVI.mean, P.richness 和 F.richness)框格边的数字则表示该变量在模型中的被解释率，即 R^2 。具体结果的查询，可点击左侧“View Text”图标，弹出“Amos Output”对话框。所有的可用参数都可在此对话框中找到。下面我们对一些重要参数进行解读，它们将是评判模型是否合理有效的重要依据。

7.1 卡方检验(Chi-square, Degrees of freedom 和 P 值)，这些参数可在“Notes for Model”标签下查阅。一般情况下， P 值应大于 0.05，Chi-square/Degrees of freedom 应小于 3，否则模型拟合欠佳。

7.2 近似误差均方根(root-mean-square error of approximation, RMSEA)可在“Model Fit”标签下查阅。通常其值接近或小于 0.05 被认为模型合理。

7.3 CFI, TLI 和 GFI 分别对应了比较拟合指数(Comparative fit index), Tucker-Lewis 指数(Tucker-Lewis coefficient)和拟合优度指数(Goodness-of-fit index)，它们均可在“Model Fit”标签下查阅。通常对于好的拟合模型，CFI 应接近或大于 0.95，TLI 和 GFI 应大于 0.90。

7.4 AIC 和 BIC 分别为赤池信息准则指数(Akaike information criterion)和贝叶斯信息准则指数(Bayesian information criterion)，它们也都在“Model Fit”标签下显示。该类型指数是一种相对测量值，在模型的比较和最佳模型的确定中有重要作用，其应用可参考文章(Lange *et al.*, 2015)。

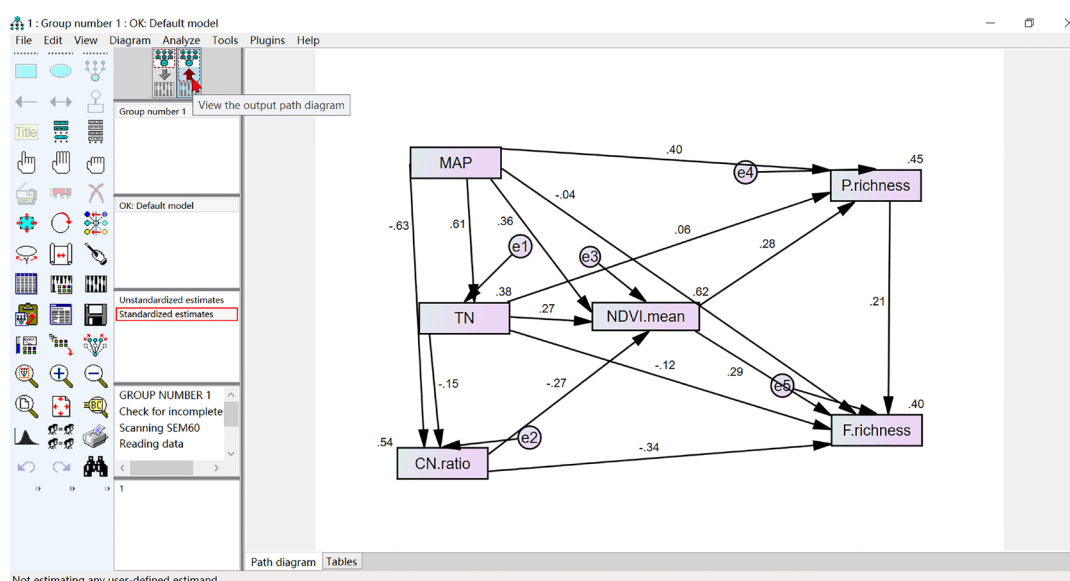


图 12. 运算结果的查看

8. 模型调整与最佳模型的呈现

在实际操作中，我们往往不可能一步就得到最佳的 SEM 模型，我们的理论模型也并非天生就与测量模型(即真实数据)完全契合。这时就需要操作者做出调整。Amos 对模型的调整，给出了简单且专业的指导，只需在“Amos Output”下的“Modification Indices”标签下查看即可，如果你的测量模型缺少关键的路径，这里就会显示。如果有一个以上的路径缺失，建议逐条添加路径，再查看“Modification Indices”。那么，除了缺失路径外，不合理的多余路径也是 SEM 分析中常常遇到的问题。不显著路径是保留(Jing *et al.*, 2015; Delgado-Baquerizo *et al.*, 2016)，还是去除(Delgado-Baquerizo *et al.*, 2013; Lange *et al.*, 2015)，不同文章给出不同的操作。对此，我们不给建议，只是告知读者，可在“Amos Output”下的“Estimates”标签中查看所有的路径系数及其对应的 *P* 值。

出于个人习惯，我们将 *P* 值不显著的路径删除(本测量模型中无缺失路径添加)，重新调整后再计算，根据 Amos 中的模型结果在 PowerPoint 中作图，最终效果见图 13a。此外，根据路径系数的加乘运算，我们以表格的形式展现出各因素对土壤真菌多样性分布的总的、直接和间接作用(图 13b)。

注：各因素对土壤真菌多样性的总的、直接和间接作用也可通过查看“Amos Output”下的“Estimates”标签中的“Standardized Total Effects”，“Standardized Direct Effects”和“Standardized Indirect Effects”三个表格获取。

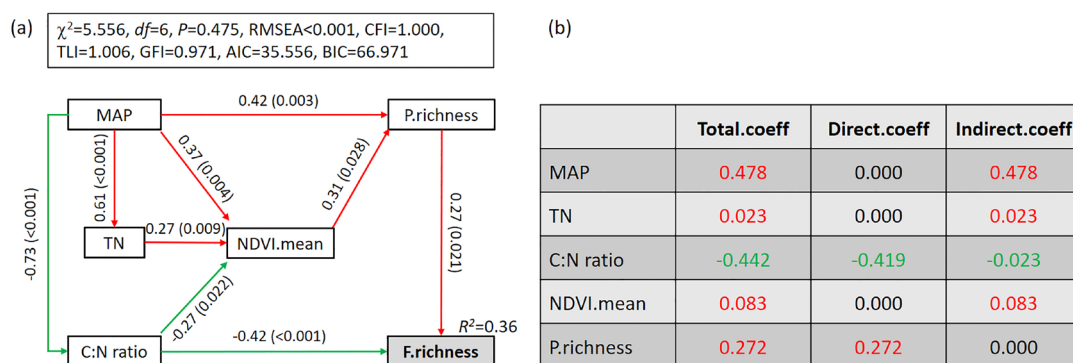


图 13. SEM 路径图(a)及各因子对土壤真菌多样性的直接和间接作用(b)。(a)路径旁边的数字为标准化的路径系数，括号内为对应 P 值。绿色为负效应，红色为正效应，只有显著的路径被保留。(b)所有系数为标准化的系数。

关于结构方程模型的后语和思考

由于笔者水平和文章篇幅的限制，本文只对结构方程模型及其应用做了初步的介绍，意在通过数据和案例的实操让读者对该方法有初步的理解，并通过 Amos 软件快速分析自己的数据。然而，SEM 的实际操作和理论构架是复杂的，包括潜变量、组成型变量的设置、样品数量的验证、多变量正态性检验等等。感兴趣的读者不妨参考下列文章(Grace, 2006; Eisenhauer *et al.*, 2015; Byrne, 2016; Fan *et al.*, 2016)。随着时代的发展，越来越多的研究人员已开始习惯使用 R 软件包处理日常的数据统计工作，lavaan 和 piecewiseSEM 是两个不错的选择 (<https://lavaan.ugent.be/>; <https://cran.r-project.org/web/packages/piecewiseSEM/vignettes/piecewiseSEM.html>)。值得注意的，SEM 方法应该基于严格的生态学假说，而非单纯的数字游戏，因此先验模型(理论模型)的构建是极为重要的，需要谨慎对待；机器学习技术与 SEM 的结合则有望促进假说驱动与数据驱动统计思想的融合，必将为今后结构方程模型的发展和应用注入新动力。

致谢

本实验得到国家自然科学基金项目(41907039)和国家重点基础研究发展规划(2014CB954002)的资助。同时也感谢华东师范大学张健教授在 SEM 生态学应用方面的一些建议和探讨。

参考文献

1. 王酋石, 储诚进. (2011). 结构方程模型及其在生态学中的应用. 植物生态学报 35 (3): 337–344.
2. Byrne, B. M. (2016). Structural equation modeling with AMOS: basic concepts, applications, and programming (Third Edition). Routledge, New York.
3. Delgado-Baquerizo, M., Maestre, F. T., Gallardo, A., Bowker, M. A., Wallenstein, M. D., Quero, J. L., Ochoa, V., Gozalo, B., Garcia-Gomez, M., Soliveres, S., Garcia-Palacios, P., Berdugo, M., Valencia, E., Escolar, C., Arredondo, T., Barraza-Zepeda, C., Bran, D., Carreira, J. A., Chaieb, M., Conceicao, A. A., Derak, M., Eldridge, D. J., Escudero, A., Espinosa, C. I., Gaitan, J., Gatica, M. G., Gomez-Gonzalez, S., Guzman, E., Gutierrez, J. R., Florentino, A., Hepper, E., Hernandez, R. M., Huber-Sannwald, E., Jankju, M., Liu, J., Mau, R. L., Miriti, M., Moneris, J., Naseri, K., Noumi, Z., Polo, V., Prina, A., Pucheta, E., Ramirez, E., Ramirez-Collantes, D. A., Romao, R., Tighe, M., Torres, D., Torres-Diaz, C., Ungar, E. D., Val, J., Wamiti, W., Wang, D. and Zaady, E. (2013). [Decoupling of soil nutrient cycles as a function of aridity in global drylands](#). *Nature* 502(7473): 672-676.
4. Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D., Berdugo, M., Campbell, C. D. and Singh, B. K. (2016). [Microbial diversity drives multifunctionality in terrestrial ecosystems](#). *Nat Commun* 7: 10541.
5. Eisenhauer, N., Bowker, M. A., Grace, J. B. and Powell, J. R. (2015). From patterns to causal understanding: Structural equation modeling (SEM) in soil ecology. *Pedobiologia* 58(2-3): 65-72.
6. Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H. and Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecological Processes* 5(1). <https://doi.org/10.1186/s13717-016-0063-3>
7. Grace, J. B. (2006). Structural equation modeling and natural systems. Cambridge

- University Press, New York.
8. Jing, X., Sanders, N. J., Shi, Y., Chu, H., Classen, A. T., Zhao, K., Chen, L., Shi, Y., Jiang, Y. and He, J. S. (2015). [The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate](#). *Nat Commun* 6: 8159.
 9. Lange, M., Eisenhauer, N., Sierra, C. A., Bessler, H., Engels, C., Griffiths, R. I., Mellado-Vazquez, P. G., Malik, A. A., Roy, J., Scheu, S., Steinbeiss, S., Thomson, B. C., Trumbore, S. E. and Gleixner, G. (2015). [Plant diversity increases soil microbial activity and soil carbon storage](#). *Nat Commun* 6: 6707.
 10. Yang, T., Adams, J. M., Shi, Y., He, J. S., Jing, X., Chen, L. T., Tedersoo, L. and Chu, H. Y. (2017). [Soil fungal diversity in natural grasslands of the Tibetan Plateau: associations with plant diversity and productivity](#). *New Phytologist* 215(2): 756-765.