

统计方法研究三叶虫演化初探

李柯臻、郑滕飞

关键词：三叶虫、统计分析、遗传进化

摘要：三叶虫的相关研究揭示了寒武纪大爆发时早期物种的演化规律，也能为早期物种演化模式的研究提供大量信息，因此一直是古生物学界的热门话题。然而，传统“将近论古”的研究方法在研究时遇到了巨大的障碍。由此，本文以统计学方法为基础，从三叶虫出发，理论建立了一种新的古动物演化的研究范式，并提出一定的演化数据模型，在不同的数据情况下估测模型的准确程度，并结合现有数据对三叶虫演化作出一定的分析，以此确定模型的实用性与在三叶虫情况下的可行性。研究发现，虽然模型建立还有不完备之处，但已经能针对数据做出一定的合理推测，也能对相关的结论进行验证。研究为统计学方法和古生物研究的结合提供了一个可能实现的新模式，在数学工具的引入上具有一定意义。

一、引言

动物起源是古生物学的重要研究课题之一，有着十分重要的意义。然而，传统古生物学“将今论古”的方法论在研究动物如何起源时，极易陷入思维困境。“将今论古”是将发现的化石置于现代动物分类体系中进行比对研究，但最早的动物及其祖先都早已灭绝，它们的形态结构、生理、生殖和发育过程均与现生动物有巨大的未知差异。再加上它们微小而柔软，保存为化石的概率极低，因此在地质记录中找到并正确地识别它们对传统古生物学的研究而言是空前挑战。故而该领域的古生物学探讨的问题实际上往往是“早期动物如何演化”而非“动物如何起源”（殷宗军，2021）。例如，寒武纪物种大爆发中，诞生了包括三叶虫在内的诸多与现代有巨大差异的早期后生生物（图1），其中的很多都难以被恰当放入今天的分类体系。因此，如果想要真正探讨动物起源的相关问题，就必须把化石放在当时的体系下进行研究。



图1：早期后生生物

一直以来，早期后生生物的研究始终是国际古生物学界解决为寒武纪大爆发重点关注的热点课题，寒武纪动物门类的爆发式出现及各门类之间的系统演化关系也是长期困扰古生物学家们的重要科学问题。而在研究这些问题时，需要一些聚焦好找、易识、分布广泛的动物化石。三叶虫化石就是这些化石的代表，也是

最主要的材料之一。三叶虫自寒武纪出现至二叠纪灭绝，曾在地球古生物中处于统治地位近亿年，是生命起源中的重要一环，揭开了地球生物走向物种多样化的序幕。三叶虫所处时代生态环境相对简单，竞争关系刚刚出现，因此三叶虫演化模型的建立可行性高，且对于其他生态环境复杂或难以建立演化模型的古生物研究有着指导作用。基于三叶虫的广泛分布，超长存在时长，以及较为简单的结构和遗传机制，三叶虫是进行遗传与演化研究的优质材料。此外，研究三叶虫化石（图2） also 具有重要的生态学价值。三叶虫化石在浅海地台区、斜坡相区及斜坡相再下的深水水域分布的属、种有很大差异。这种差异的原因是不同地区海水深度不同，水温不同，海水沉积物的不同及外部动力的不同等。这种差异还影响三叶虫化石保存状态的不同，由此，可以由三叶虫化石反推出当时的生态状态。



图2：三叶虫化石（拍摄）

因此，无论是在古生物地层学、演化生物学以及板块构造学、生态学等研究领域，三叶虫均具有重要意义。可以说，对三叶虫化石的研究对于推进对于寒武纪大爆发的研究乃至早期动物如何演化与动物如何起源的研究有着重要的作用。我们的这次研究，旨在初步寻找建立三叶虫演化模型的方法并对初步建立的模型做出分析，考察可靠性，如果成功，对早期生物的研究将有很大的参考意义。

二、化石数据的收集与初步处理

为建立演化模型，首先需要获得一定量的原始数据。在本方法中，理想的一组原始数据需要满足三个条件：数量不能过小、针对同一物种（最好为同一种群）、在地域与时间上具有一定的连续性。

对数量的要求保证了之后进行分析时的可靠性，也对同一物种则确保之后进行的时空上的模型建立是有意义的。最后的连续性要求中，空间连续性实际上和第二条需求一致，时间连续性则能更好地对模型进行验证，增加数据点的密度。三叶虫之所以出名，不仅是因为它们拥有奇特的结构，也是因为它们的化石大都保存得很好，并且遍布全球。所以，通过三叶虫化石分析进行三叶虫生态关系在通过化石研究分析物种情况中是相对可靠的。

然而，以化石为数据来源研究三叶虫仍存在许多问题：对统计学来说，数据量仍然偏小，特别是个别物种数量极少，同时各具体的种群具体特征仍常有更新，同时各种群分化遗传关系难以明确，研究这些关系对于我们来说过于困难，所以难以通过研究同一物种进行分析，据此，我们尝试以三叶虫各属为基本研究单位，研究各属三叶虫或者该属的主要大类之间的关系以研究三叶虫的遗传演化。至于

空间时间连续性，则一直是来自化石的数据的难题。

种间关系是演化的重要一部分，也是本文对研究三叶虫演化的切入点，当前对于种间关系的研究大多以观察为主，因为当代生物的种间关系往往有多种方法可以得到。但对于难以观察的古生物，常规方法都不再有效，这时我们期望去通过研究各种三叶虫数量的变化规律并与各种间关系特征做比较，试图刻划出各三叶虫种群间的竞争，共栖等关系及强度。

综上所述，在期望数据上，我们希望得到具体种群在具体时间具体地区的数量分布，而且，随时间推移，地质演变，化石群/簇数量对生物数量及分布的显示存在一个衰减。但鉴于本文考察的时间范围长，分布广，受影响较小，完全可以先进行模型建立与分析。而对数据的处理上，由于化石留存的不稳定性，我们只能通过同一化石群中各三叶虫的含量来显示各三叶虫的数量特征，难以做出不同化石群间的比较。

在同一地层中，往往会出现较集中的化石，因此首先需通过聚类分析确定种群。古生物地理中，为寻找古生物的分布规律与演化轨迹，一般通过对一定量样本的地理格局进行分析。具体而言，通常根据生物群之间的相似性状运用多元统计方法对古生物地理进行分区。其中聚类分析是最常使用的一种定量方法。聚类是将数据分类到不同的类或者簇的过程，因此同一个簇中的对象有很大的相似性，而不同簇间的对象有很大的相异性。从统计学的观点看，聚类分析通过数据建模简化了数据分布。

在聚类分析中，选取相似性测度（判断两个区域相似程度从而确定是否合并的标准，取值范围为 0 到 1，越大代表越相似）意味着选取一个合适的分类标准。原则上，聚类分析采用通用的 Jaccard 测度进行分析，公式为 $\frac{a}{a+b+c}$ 。其

中， a 代表选取的两个样本区域中共有的分类单元数， b 与 c 分别为两个区域所独有的分类单元数（图 3）。此测度在大部分情况下有较为良好的表现（Shi, 1993）。但是，当样本整体数据频率较小（低频项较多）时，Yule's Y 测度

会有更好的表现（公式为 $\frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$ ， d 表示总体中两个区域外的分类单元数）

（Chung and Lee, 2004），因此选取此测度（此段主要来自参考文献黄冰，2011，但根据的进一步的资料查找，参考文献中关于 $abcd$ 的解释有误，已更正）。

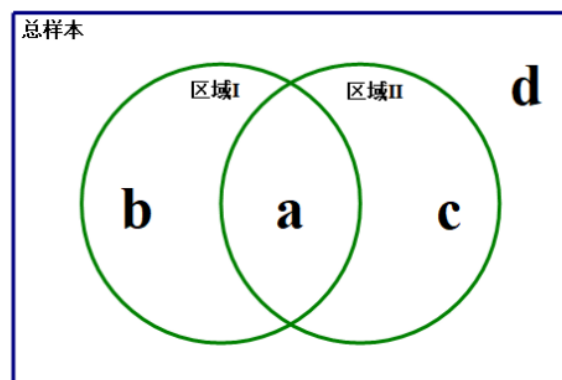


图 3：相似性测度计算公式中 a,b,c,d 对应范围示意图

但值得注意的是，在低频项较多时，进行时间上的演化模型建立容易更加不准确，因此实际应用中较多应用的仍然为 Jaccard 测度。

由于单纯的聚类分析无法进行统计评估，一般可使用非度量性标度变换（NMDS）进行辅助分析，通过计算应力值估算匹配程度，亦可以通过最小生成树（通过尽可能小的距离连接所有点，MST）（图 4）判断不同区域的距离关系。通过这两种方法综合判定后，即可知聚类分析结果的可靠性。有时，虽然聚类分析整体结果可靠性较强，但个别数据点分类存在问题，亦可通过此类方法调整（吴承义等，2021）。

```
void Graph::Kruskal(void) {
    QuickSort(edge[edgenum]);
    for (int i = 0; i < edgenum; i++) {
        if (!Connect(edge[i].vex1, edge[i].vex2)) {
            cout << edge[i].vex1 << edge[i].vex2 << ' ';
        }
    }
    cout << endl;
}
```

图 4：MST 的 Kruskal 算法示意

（edgenum 为边数，QuickSort 按边大小排序，vex1,2 为边连接的两顶点，输出选取的边）

在选取了可靠性较高的结果后，理想的情况是得到了某个种群的时间-数据关系散点图。下一步则为验证此散点图是否符合某个标准数据模型，若符合，则可一定程度推断出这段时间此种群所处的情况。

三、标准数据模型的建立

根据自然界中生物演化可能的情况，我们建立了七种考量同物种生物数量随时间变化的数据模型：基本保持在某一稳定值附近波动的情况、限制性自然增长的标准 S 型曲线、自然减少时的数量曲线、由于捕食关系而形成的动态平衡（分捕食与被捕食两种）、与其他物种竞争生态位的情况（分竞争强势与弱势两种）。接下来也将按照这样的分类进行数据模型的建立：

模型一为稳定平衡模型，代表种群数量基本稳定为某一数值的情况。

模型二、三分别为自然增长与自然减小的模型。由于两种模型的公式完全相同，差别只在初始值的取值，因此将两种模型放在一起处理。

模型四、五为捕食关系所形成的模型，用逻辑斯蒂映射处理。四代表捕食者的情况，五代表被捕食者的情况。

模型六、七为竞争关系所形成的模型，六代表竞争强势者的情况，七代表竞争弱势者的情况。

值得注意的是，一组数据可能在不同的段落处在不同的模型中，这暗示了新的生态条件的加入。并且，在数据改变速率过快时，外界的变化过于剧烈（如某些纪末尾的生物大灭绝），可能导致这些模型的不再适用。

以下的分析中，每个模型代表“随时间变化种群数量”的参数为 u_n, v_n ，默认等时间间隔采样，范围为 0 到 100，实际应用时，可根据实际数据放大/缩小比例或平移。所有数据均来自计算机根据公式模拟，由于公式中存在一定的随机性，这样的模拟可以给出数据的大致走向。

模型一：稳定平衡

涉及参数：稳定点 u ，回复率 p ，波动范围 r

公式： $u_{n+1} - u_n$ 在范围 $[-r + p(u - u_n), r + p(u - u_n)]$ 内等概率选择

初始值： u

以下为几种参数情况下作出的图像（图 5 - A, B, C）：

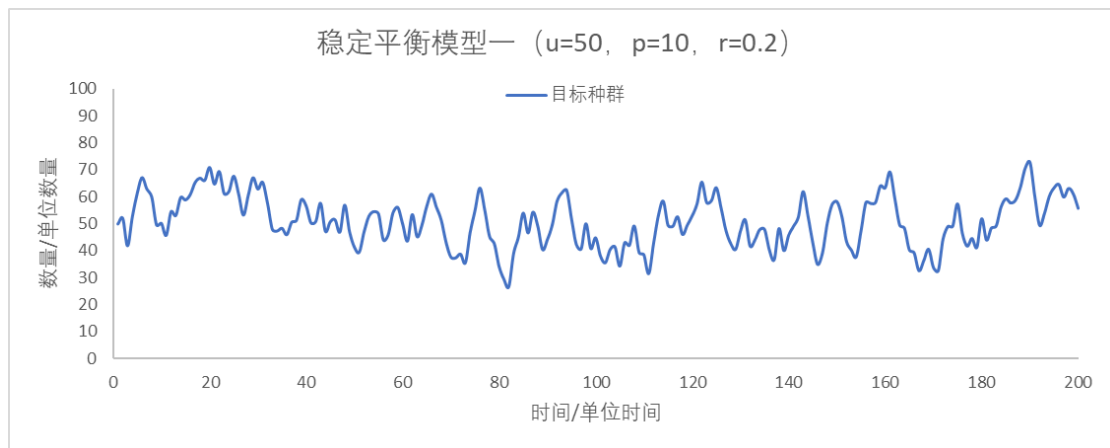


图 5 - A: 稳定平衡模型一

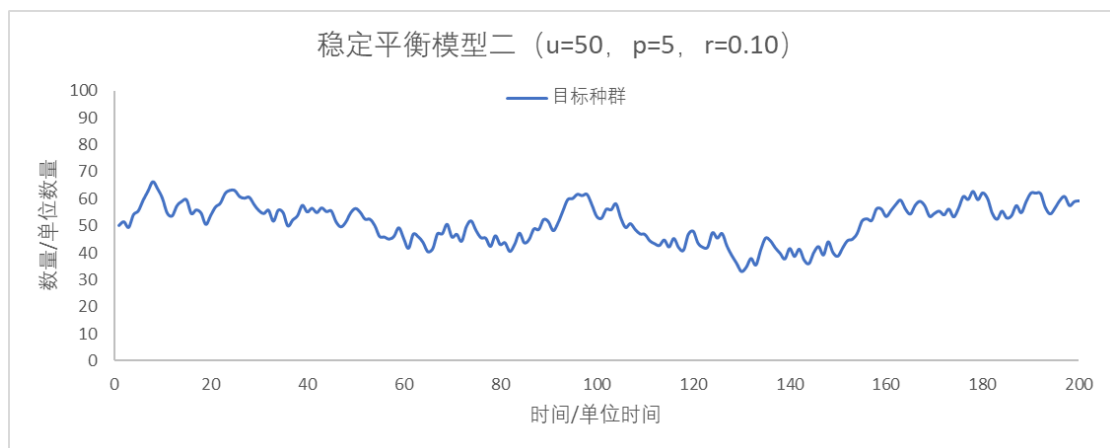


图 5 - B: 稳定平衡模型二

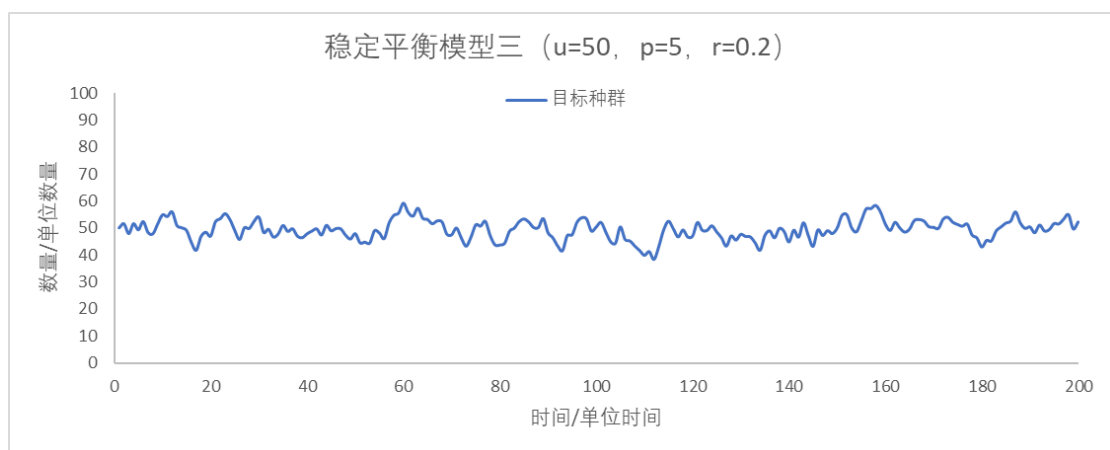


图 5 - B: 稳定平衡模型三

参数中， u 主要确定数据的中心位置（为方便设置为 50，实际可平移）， p 为回复力量的大小， r 为单次波动的大小。当 p 减小或 r 增大时，数据将变得更为紧凑。

稳定平衡模型的特点为：数据在稳定点附近波动，在其上与其下的数据量基本相同，增长与减少的长度基本相同。此三条即说明稳定平衡模式的特征，

也为确定种群数据是否基本处在稳定状态的标准。

若数据被判定在基本处在稳定状态，则意味着这段时间的生存环境相对较为稳定，且种群数量已达到适宜值。

模型二、三：自然增长&减少

涉及参数：稳定点 u ，增长率 p ，增长率波动范围 r

公式： $\frac{u_{n+1}-u_n}{u_n}$ 在范围 $[-r + p(u - u_n), r + p(u - u_n)]$ 内等概率选择

自然增长初始值：某较小的数值；自然减少初始值：某比 u 大较多的数值

以下为几种参数下自然增长的图像（图 6 - A, B）：

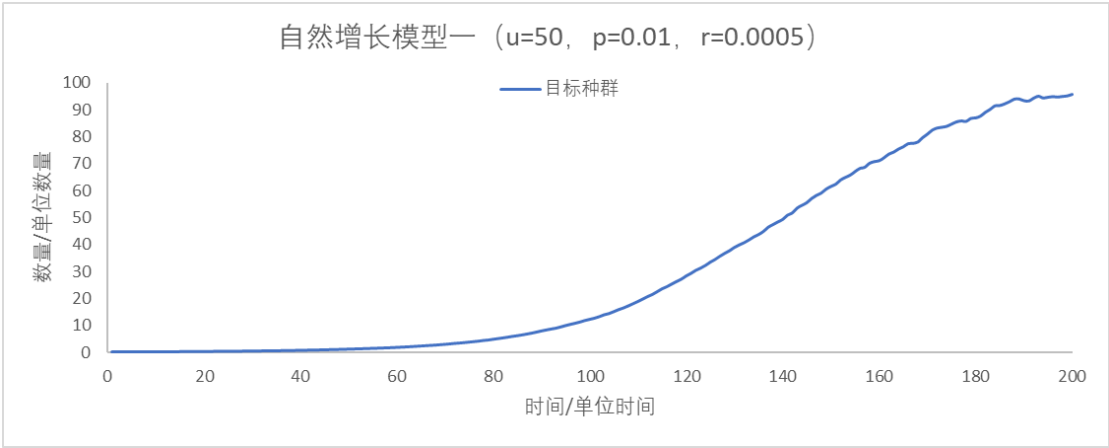


图 6 - A：自然增长模型一

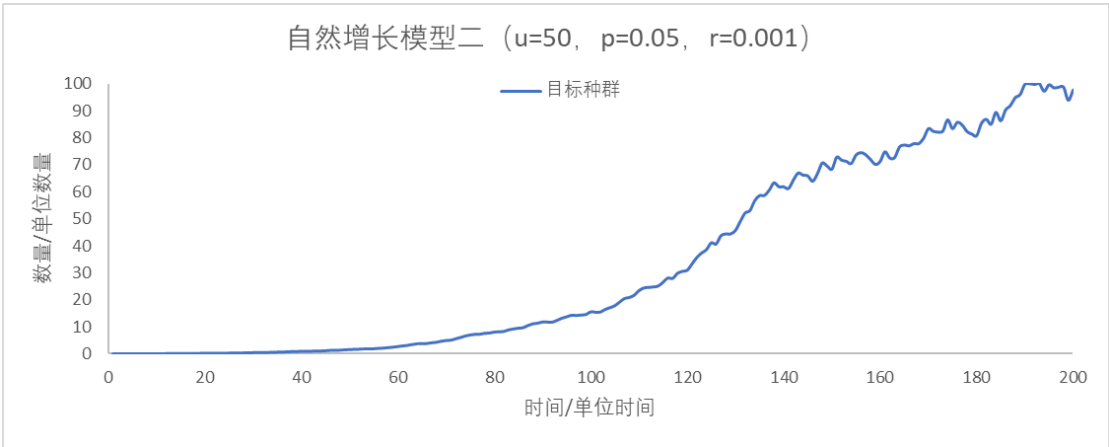


图 6 - B：自然增长模型二

参数中， u 主要设定为结果位置， p 为总体增长强度的大小， r 为单次增长率波动的大小。当 p 增大时，相对的增速会提升，而当 r 增大时，数据（尤其是增长后期的数据）将变得不稳定。

自然增长模型的特点为：总体而言，数据的增长速度先慢后快最后慢，最终形态为在稳定点附近类似稳定平衡模型。但注意，此为基本从 0 开始（两图的初始数据均取为 0.1）的自然增长模型，若初始时已有一定规模的数据，可能直接跳过前期的缓慢增长部分。

若数据被判定为处在自然增长模型中，存在两种可能。一种为初始值接近

0，此时数据解释了某一种群在某地区发源的过程。另一种可能为初始值并非 0，此时数据代表着此地区新的资源突然出现，或种群数量突然削减后的情况，种群将逐步朝着地区的最大容量增长。

以下几种参数下自然减少的图像（图 7 - A, B）：

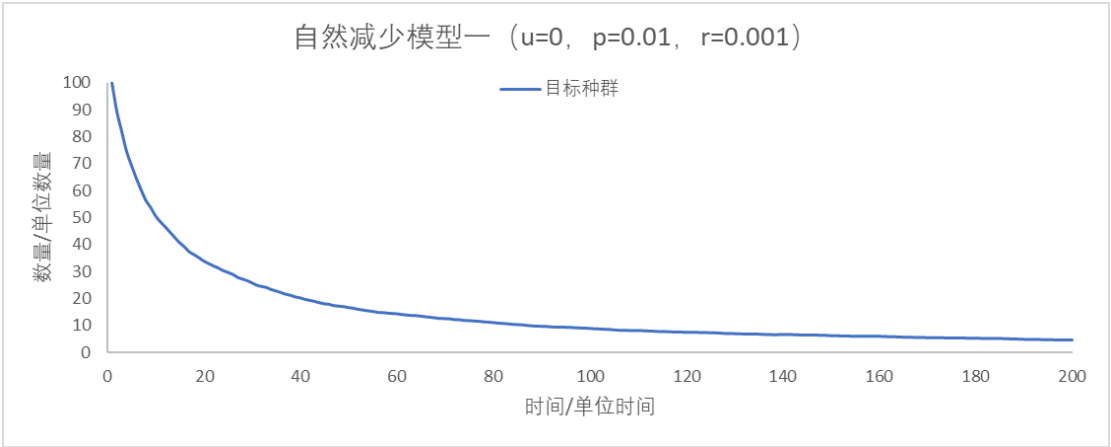


图 7 - A：自然减少模型一

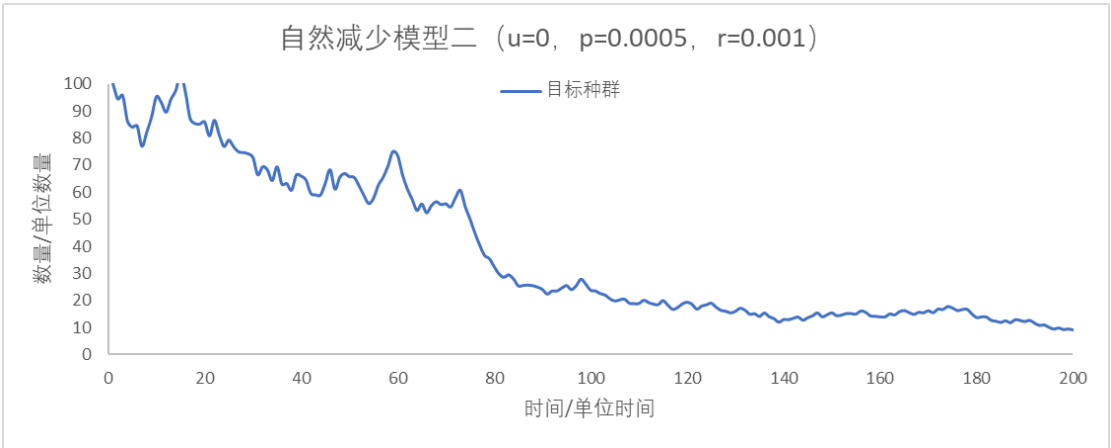


图 7 - B：自然减少模型二

参数中， u 主要设定为结果位置， p 为总体减少强度的大小， r 为单次减少率波动的大小。当 p 增大时，相对的降低速度速会提升，而当 r 增大时，前期的数据将变得不稳定，甚至间接影响前期减小速率与最终到达稳定值的时间。

自然减小模型的特点为：总体而言，数据的减小速度先快后慢，但越后期减小将越稳定，最终落在终值附近基本不再变化（注意到，由于波动的存在，最终值将比设定的结果位置 u 较大一些，若想要最终值为 0，实际 u 需为负值）。

若数据被判定为处在自然减小模型中，存在两种可能。一种为最终值接近 0，此时数据解释了某一种群在某地区最终绝灭的过程。另一种可能为最终值并非 0，此时数据代表着此地区旧的资源枯竭，或种群数量突然增加后的情况，种群将逐步朝着地区的最大容量减小。

模型四、五：捕食与被捕食动态平衡

涉及参数：被捕食者增长率 r_1 ，最大数量 a ，压力系数 e_1 ，捕食者死亡率 r_2 ，

捕食效率 e_2 ，增长率随机波动 r

公式：被捕食者的实际增长率 $\frac{u_{n+1}-u_n}{u_n}$ 为 $r_1 \left(1 - \frac{u_n}{a}\right) - e_1 v_n$ ，扰动（即在周围大小变化）为 r ，捕食者的实际增长率 $\frac{v_{n+1}-v_n}{v_n}$ 为 $-r_2 + e_2 u_n$ ，扰动（即在周围大小变化）为 r （王宁星等，2008）

初始值：两者均较小但不为 0

以下为几种参数下的图像（图 8 - A,B）：

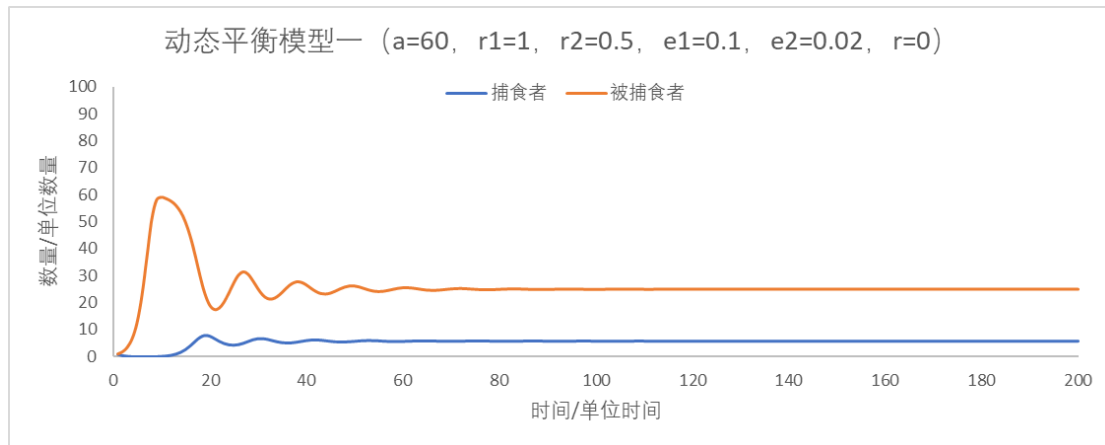


图 8 - A：动态平衡模型一

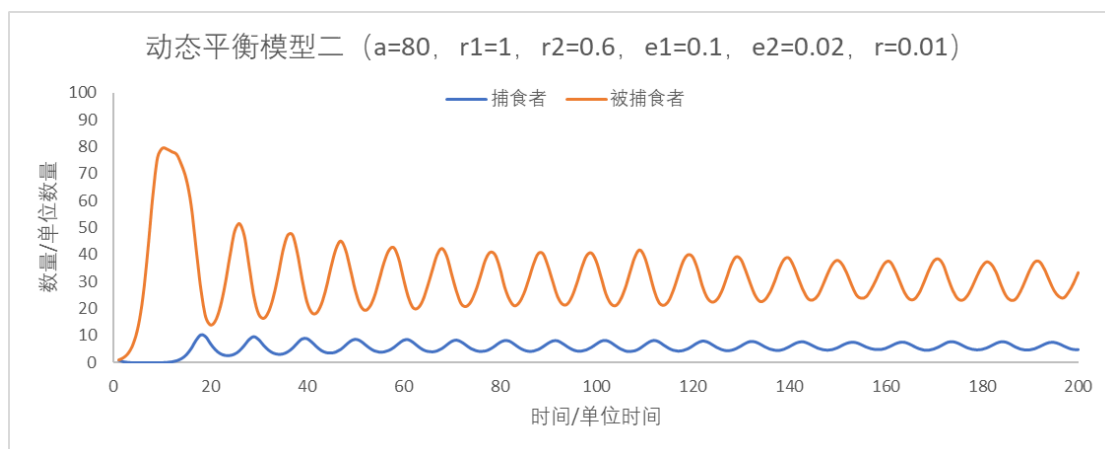


图 8 - B：动态平衡模型二

参数中， a 限制了被捕食者达到的最大数值， r 影响图像的随机波动，其余则均对图像形状有较大影响。

注意到，第一个图像实质为在波动中前往稳定点，而 r_1, r_2 相对于 e_1, e_2 的大小决定的稳定点的大小，稳定时捕食者与被捕食者的比值为 $e_2:e_1$ 。 r_1 与 r_2 的相对大小则影响波动的幅度，进而影响前往稳定点的速度。捕食与被捕食者的区别在于，被捕食者的每个峰值两侧均为凸起，捕食者则为凹陷。初始数据的不同会导致初始波动的差异，但后期会回到相似的状态，因此，判定是否处于此模型的关键在于是否有幅度较稳定降低的波动。

但是，在第二个图像中， r_2 超过了某一临界值，此时由单一稳定点变为了在两个极值之间波动，不过每个极值两侧的凸凹性仍然有效。继续增大 r_2 ，则

极值个数会增加至四个、八个直至无限，图像呈现混沌（这种峰值出现周期性质与采取的模型为 Logistic 映射有关，相对的临界值则涉及混沌映射的性质，由费根鲍姆在 1974 年提出的费根鲍姆常数确定（Bing 等，2012））。因此，此模型在 r_2 较小时具有更大的实际意义，本文则只考虑但稳定点与 2 个极值时的情况。

若数据被判定为捕食或被捕食者的变化模型，则暗示了另一方的存在。并且，动态平衡类的捕食与被捕食代表着能量来源较为稳定，不至于出现由于捕食者还有其他作为食物来源，被捕食者直接被捕食至灭绝等可能（否则模型将类似自然灭绝）。

模型六、七：竞争生态位

涉及参数：两物种扩散率 c_1, c_2 ，灭绝率 m_1, m_2

公式：物种 1 实际增长率 $\frac{u_{n+1}-u_n}{u_n}$ 为 $c_1(100 - u_n - v_n) - m_1 - c_2v_n$ ，物种 2

实际增长率 $\frac{v_{n+1}-v_n}{v_n}$ 为 $c_2(100 - u_n - v_n) - m_2 - c_1u_n$ ，（公式来源论文时培健等，

2009，论文中的公式经查找资料发现最后一项有误，已修正）

初始值：两者均较小但不为 0

以下为几种参数下的图像（图 9 - A, B）：

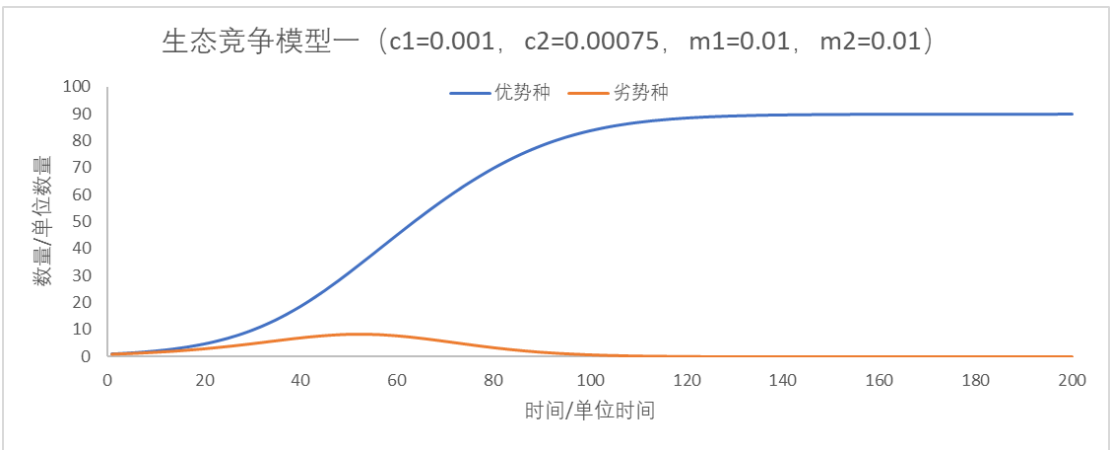


图 9 - A：生态竞争模型一

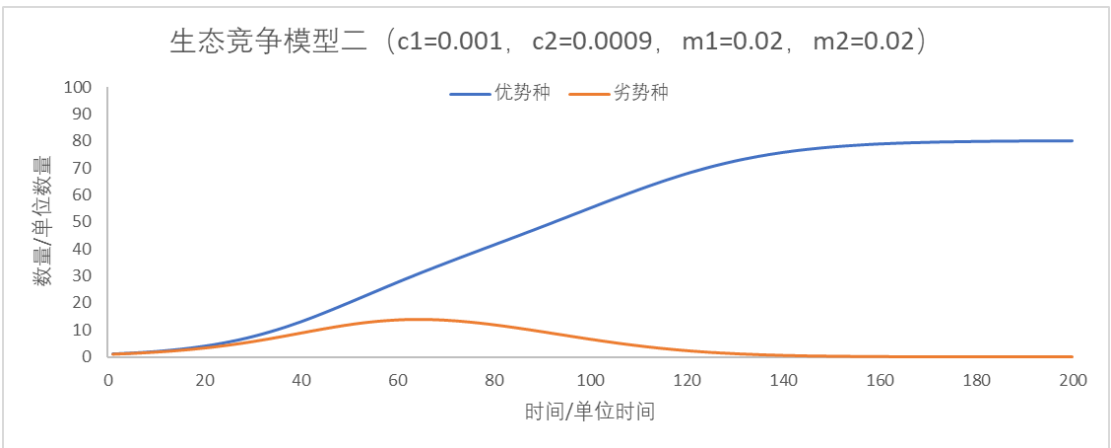


图 9 - B：生态竞争模型二

参数中，一般而言， c_1, c_2 的大小关系决定了竞争的成败，而且即使是十分微小的差距也会导致结果的巨大不同。失败者基本失去全部的生态位，成功者则占据全部。相对的差距则决定了达到最终稳定结果的速度。 m_1, m_2 的大小影响了最终稳定数量的值，也影响数据增长的快慢。

此模型中，竞争成功者的特点为先慢后快最后慢的上升，但与标准 S 型的自然增长不同，一方面是前期的速度相对较快，另一方面则是失败者初始占据较多生态位的情况（图 9 - C, D）：

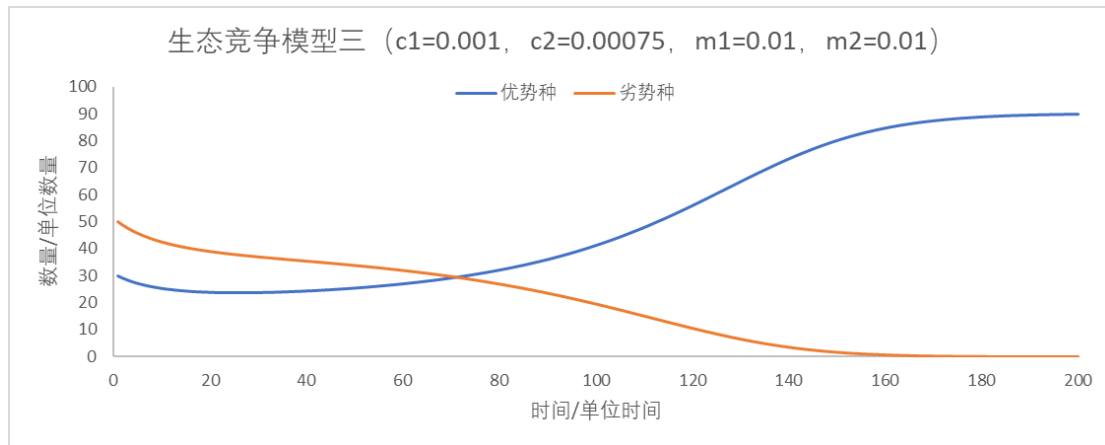


图 9 - C: 生态竞争模型三

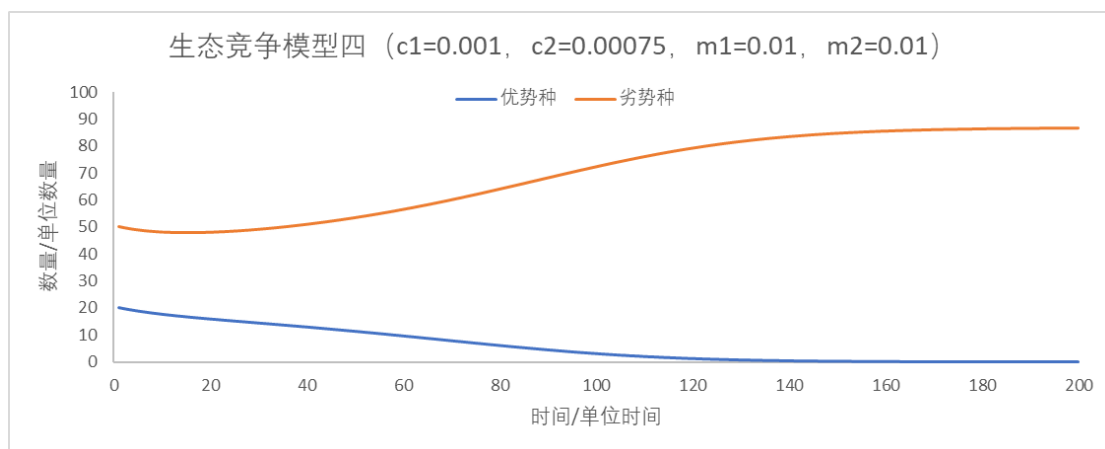


图 9 - D: 生态竞争模型四

这两张图中，事实上改变的只有初始数据。由此可见，竞争力强者如果没有达到一定量，仍然无法在竞争中取胜。

此时，若竞争成功，竞争力强者事实上经历了先下降后上升的过程，且图像的总体为前期凹后期凸，即能与其他物种区分。

多物种竞争时，依然为强者成功、强者失败、弱者成功、弱者失败四种情况，因此具有类似模型。若数据被判定为此模型，同样暗示了另一方或另几方的存在（当然，另一方可能就是其他种群）。注意到，不管具体哪一种，均是以两物种一种全灭一种全得为结局，因此，只要存在竞争力差异或初始值差异，便无法稳定共生。这就说明了，这个模型中并不存在动态平衡的结果。

模型总结

由上方所述的模型特点，即可将一些分立的散点归入具体的模型中，再由模型的性质推测或猜想此生物所具有的性质，从而诱导新的发现。模型之间的整体对照如下（表 1）：

	内容	特征	基本识别方式	图像
模型一	稳定平衡	环境适宜，种群稳定	基本稳定在某点附近	图 5
模型二	自然增长	环境与种群的变化规律单一	快-慢-快的上升过程	图 6
模型三	自然减少		先快后慢的下降过程	图 7
模型四	捕食者	与其他物种形成捕食关系，相互影响	波动中走向稳定或产生明确周期性	图 8
模型五	被捕食者			
模型六	竞争强势者	存在种内种外的竞争关系	一次小幅波动后在高点/低点趋于稳定	图 9
模型七	竞争弱势者			

表 1：模型间的对比

值得注意的是，以上理想模型都是对于单个个体面对单一变化因素的推测，这些推测的得出往往也运用了现代生物学的结论。虽然实际的三叶虫模型应是一个面向需求的考虑多种因素的极其复杂的模型，但本文要做的是尝试建立三叶虫数据模型并检验其可行性与可靠性，因此选择单一条件下对数量变化的预测的可靠性作为切入点是合理的。

从原始数据进行聚类分析，再根据数据随时间的变化建立演化模型，由此作出分析。这样的流程即为理想中模拟出的古动物研究方法。在实际操作中，由于样本量实际上并不多与实际中的种种随机变化，不会出现如此好看的结果，但仍然可以通过数据情况与模型链接，从而大致推测。至于更极端的无法归入任何模型的情况，则往往意味着生态环境的剧变。

四、讨论

1、方法的可行性分析

在上一部分中，我们已经讨论了前前后后共七种可能的生物演化进程中的数量变化模型，并从理论上检验了这些模型在数据存在各种可能噪音情况下的稳定程度。然而，为进一步验证其可靠性，还需将实际数据代入模型，并对预测结果与现有生物结论做比较。

在实际的比较过程中，我们发现理论数据模型在描述预测一些数据时表现欠佳，但在描述一些数据（例如下文所选取的来自贵州和湖北的数据时）表现异常良好。同时，我们也以观察数据为出发点，观察出了三叶虫数量变化的一些结论，建立了一些建议模型，惊喜地发现这些模型基本都可以在上文找到相似理论模型。

换言之，我们通过验证各模型对各数据的预测，证明了理论模型可以很好描述一些数据分布。虽然并不能普适于所有数据，且对数据的要求较高，但这已经是较大的成功了。

贵州把榔动物群

从寒武纪早期古板块分布来看，华南板块处在赤道附近。根据古地磁显示的证据，华南板块处于南半球的低纬度地区，当时华南板块还未与华北板块对撞拼接，华南板块独立出现在海洋中。贵州东部位于现今的“扬子准地台”边缘上，

而在寒武纪时则位于华南板块边缘被海水淹没的水域中。贵州东部杷榔组分布较广，地层出露较完整并呈北北东向条带状展布。杷榔组的沉积厚度与古地理位置具有一定联系，岩性主要为粉砂质泥岩、粉砂岩、含钙质粉砂岩组成，在陆棚—斜坡水体较深的沉积环境中形成。剑河一带，杷榔组厚度在 190 m—300 m 间，岩性主要为粉砂质泥岩、泥页岩组成，岩石中的粉砂质低，泥质含量较高，局部层理构造较为发育。凯里—都匀地区杷榔组未见底，仅出露杷榔组的上部，露头厚度在 60—190 m 间。都匀一带杷榔组的厚度在 150—360 m，而洛邦镇以北地区的杷榔组未见底、露头厚度大于 130 m，主要为泥岩、含粉砂质泥岩组成，薄层层理发育。（刘帅，2018）

综上所述，贵州杷榔动物群化石保存较好，量大分布广，空间时间连续性相对较好是良好，是研究古生物的材料。

根据前人的努力成果，我们选取了贵州杷榔组辣子寨剖面中的具有代表性和研究意义的 Lz-a 至 Lz-f 层。

其中 Lz-a 与 Lz-b 层在保藏其他生物化石方面在丰度，物种上都有较大区别，显示 Lz-a 与 Lz-b 层化石形成时间有较大差距，但令人震惊的是，两层在主要三叶虫种属上的丰度竟然高度相似，精度可达百分之一！两层在四种三叶虫的丰度上几乎完全相似，这契合了上文的模型 1，契合程度如此之高，这充分反映了三叶虫化石指示三叶虫种群状况的高度可靠性。同样的现象我们在搜集其他数据（伍孟银等，2019）时也有一定程度的出现。

Lz-a 与 Lz-b 层三叶虫保存的主要特征是：三叶虫主要由 *A. chauveui*、*Ar. xinzhaiheensis*、*C. elongata*、*D. duyunensis*、*R. (Pteroredlichia) chinensis* 组成，其中以 *A. chauveui* 丰度最高。这几个种的幼年个体一成年个体均有保存，保存质量较精美，保存状态以孤立、稀疏分布为主，压埋方式以背压保存为主，少量为侧背压的三叶虫，一般产生了埋藏变形。完整保存的三叶虫化石数量多，较合的三叶虫标本达 65%，野外统计发现三叶虫个体大小在 2—13 mm 范围内，且分属于不同种属，其中约 60% 的三叶虫个体长度在 3—6 mm ($N=160$)，长度 >6 mm 或长度 <3 mm 的个体频率分别为 17% 和 23%，化石分布无定向性，三叶虫多为浮游生活的类型。Lz-c 层三叶虫保藏极少，Lz-d 层 *D. duyunensis* 三叶虫保藏极好。Lz-e 三叶虫的数量很多，化石种类较丰富，三叶虫的群体同层保存的化石多在该层出现，如 *Ar. xinzhaiheensis* 群体同层保存、*R. (Pteroredlichia) chinensis* 群体同层保存等，还含有一些大型双瓣壳节肢动物等刺虫和吐卓虫、另外有少量三叶形虫、古蠕虫等。岩性为深灰绿色粉砂质泥岩，具韵律型层理，厚 65 cm。三叶虫保存具三个主要特点：多个三叶虫个体同层或近于同层共同保存较常见；三叶虫碎片较多三叶虫在层面壳体凸面的朝向无规律，个体和碎片的长轴方向具有弱的定向性；非较合的三叶虫个体很多，个体前活动颊前缘多已脱落，可能为三叶虫脱壳或死亡分解作用形成的标本。*I. acutangulus* 埋藏保存以单壳侧压并与藻类共同保存较为常见，其中 12% 的个体背边缘较合 ($N=50$)，壳体上及周缘可见藻类碎片。另外含保存了其它门类化石，见海绵动物、刺胞动物、腕足动物、蠕虫动物、软舌螺动物、棘皮动物等，化石组合基本与 Lz-b 层一致，保存质量较好。Lz-f 层保存特征：位于上化石层上部，围岩为具有韵律层理的泥岩，保存了种类及大小各异的三叶虫，主要有 *A. chauveui*、*Ar. xinzhaiheensis*、*C. elongata*、*R. (Pteroredlichia) chinensis*、少量 *D. duyunensis*，三叶虫个体及碎片立体形态保存，三叶虫分布较密，群体保存的方式较普遍；局部可见弱的定向排列，非较合的三叶虫壳体数量多于较合

个体，较合三叶虫的数量占三叶虫总数的比例达 33%。（刘帅，2018）
根据以上数据，我们作出了对应的数据模型图（图 10）：

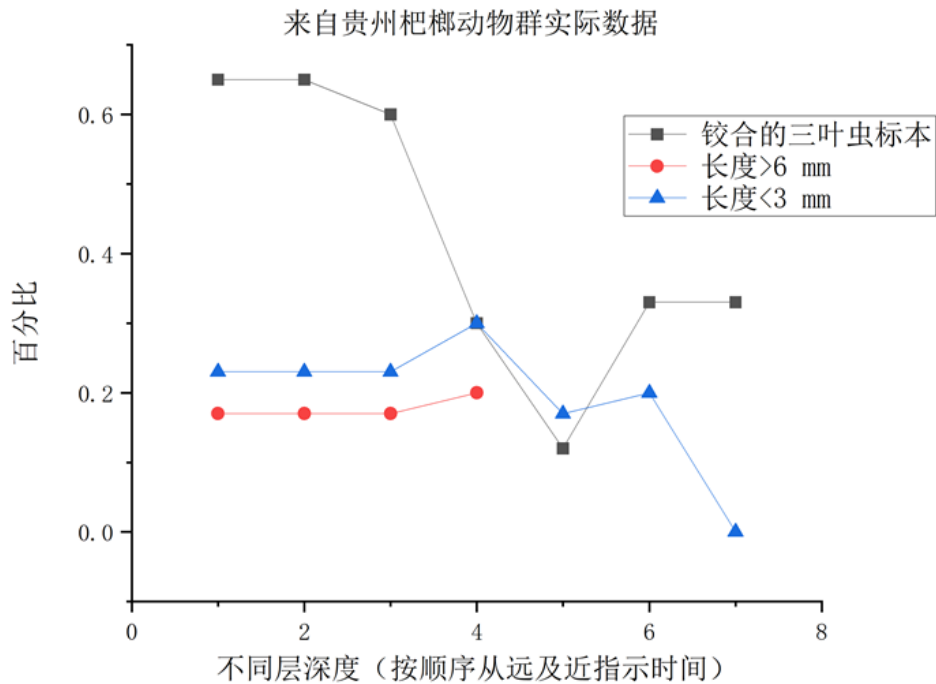


图 10：来自贵州杷榔动物群的实际数据

从模型图中可以看出，以层 3 与层 5 为断点，三叶虫各种标本都可以找到所符合的模型，而这正与由三叶虫及其他化石数量指示的各层关系相符。例如，长度小于 3mm 的种群情况基本符合自然减少对应的模型，而长度大于 6mm 的种群情况在可调查范围内与稳定平衡模型相符。此外，层 3 与层 5 中能看到较为偏离模型的点，可以推测环境发生了较大变化。事实上，这两层中其他化石组成及矿物组成都相较与其他层有大变化，指示环境的剧变或时间很长。背部较合作为三叶虫的种群特征，其分布基本符合捕食模型中捕食者的情况，由此可以推测捕食者的存在。不过，虽然在长度大于 6mm 和长度小于 3mm 确实存在一定的竞争关系，从中难以看到与竞争模型相似的特征，这可能是由于资源较为充足，因此并没有产生明显的竞争关系。

湖北宜昌市的桃源茅草铺剖面的紫台组顶部和牯牛潭组

此剖面大致情况为：在聚类分析后，已经分为了不同的组合，例如数量最多的 Microparia 组合，以及 Nlieus、Panderia、llaenus、Paratiresias 等。每一个组合内有着基本类似的三叶虫种群构成。由此，对此组进行划分时，先分为组合，再细分属为效率与准确性较高的方式。

具体来说，Microparia (Quadratapyde) 组合三叶虫多样性略低（9 个属/亚属），Microparia 占据绝对优势(81.9%)，Nlieus(9.29%)和 Panderia(5.13%)次之。Microparia 是典型的区域性分子，目前仅发现于华南、华北和塔里木板块，作为中层浮游的圆尾虫类，主要分布于较深水环境（深外陆棚和盆地），偶见于浅外陆棚环境，大致水深为 70-200m。Nlieus(9.29%)和 Panderia(5.13%)是浅外陆棚的代表分子它们在 Microparia 组合内出现，可能反映了深外陆棚内

侧沉积环境.。Nlieus 组合三叶虫多样性最高, 达 15 属 / 亚属, 以 Nlieus 为主 (39.2%), Microparia (25.1%)、Panderia (9.9%)、llaenus(8.0%) 和 Paratiresias (5.6%) 次之。值得一提的是, 该组合中 Rorringtonia 的虽然数量较少, 但却有重要的生态意义: 该属眼睛较小且常与笔石类及其它深水三叶虫共生, 能够指示相对深水的远岸环境。此外, 组合内的 Paratiresias s 是该属目前已知最早的种, 时代为达瑞威尔中期, 晚奥陶世该属分布于中低纬度地区, 如哈萨克斯坦地块 Alborz 地块、华北板块和滇缅马地块; 属于眼睛非常小的一类 Paratiresias 三叶虫, 该类群多见于灰泥丘相, 而 Paratiresias 却常与深水三叶虫共生 (除华南板块), 根据其分布规律和共生分子, 推断它可能指示水深大于 100m 的深外陆棚或斜坡环境, 本文所报道的最早分子也为该属相对较浅水起源提供了证据。与 Microparia 组合相比, 该组合内 Microparia 的数量大幅减少, 而 2Nlieus 所占比重明显增加, 反映了向上水体变浅的趋势。Panderia 组合仅包括 3Panderia (73.68%) 和 2Nlieus (26.32%) 两属 (魏鑫, 2020)。

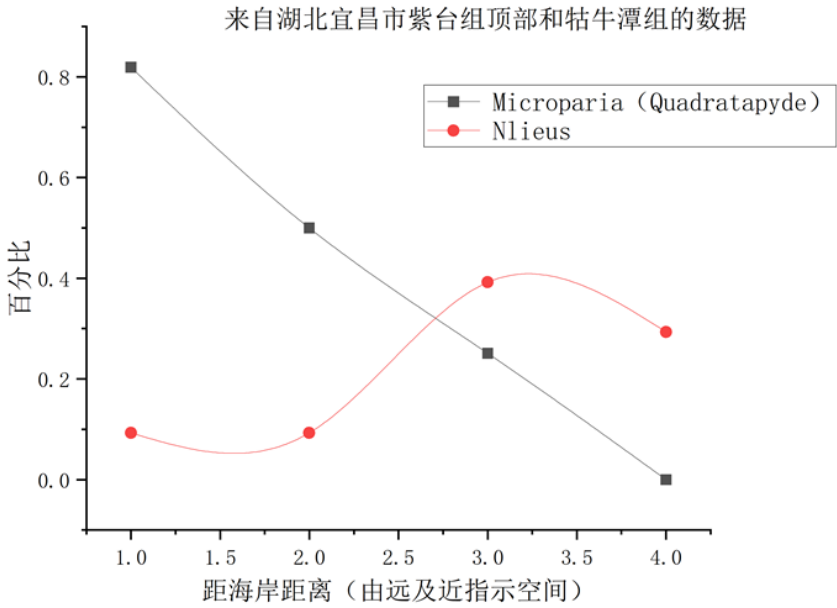


图 11: 来自湖北宜昌市的桃源茅草铺剖面的紫台组顶部和牯牛潭组的实际数据

由以上类别的三叶虫产生的数据模型中 (图 11), 由于整体的分布区域并不宽, 这里的空间实际上是时间的指示。但是, 由于无法精确确定地层的形成速度, 这个模型只能看出增长与减少的趋势, 而难以进行具体的数据拟合。我们可以发现, 在时间维度上, Microparia 满足自然减少模型, 而 Nlieus 则可能满足自然增长的模型。通过模型, 我们可以得到一些该论文作者得出的结论, 例如, Microparia 的减少与 Nlieus 的增加反映了对应的环境改变 (对应原文中作者解释为水体变浅)。对整体数据中, 当前的模型无法对 Panderia 的丰度突然增高作出解释, 原文作者则称可能与水体变浅有一定关系。

总结

从这两个符合较好的例子中可以看出, 在数据连续性较强时, 方案可以具有较好的可行性。但是, 当数据样本进一步缩小, 或是连续性减弱, 模型能起到的作用就更为有限了。因此, 模型的成立主要还是依赖于足够的连续性数据。

在对本方法的探究中, 我们也得到了其他附加的经验, 例如, 在古生物中,

三叶虫的演化过程相对不存在过多的噪音以及干扰信息。我们认为，这可能与三叶虫种群生态结构简单，所处环境变化缓慢有关。

2、模型的推广方向

从定性到定量

当前，数据模型只是作为定性的识别工具，因此最重要的推广就是将模型量化。当然，模型中有参数化的部分，在实际应用中可以通过最小二乘法调整参数进行拟合，但是，对拟合准确性的评估在当前的模型设计中是缺乏的。这一点上，可以借由卡方检验的思路设计对模型是否可以适用的检验方案与判断标准，并给出具体的可行概率。另一个量化思路则是增进模型自身的精细程度，并且根据实际应用对参数进行细化处理。这个思路依赖更多数据的检验，且可能需要新的看待现有数据的方式才能实现。

从时间到空间

在实际研究中，种群的数量往往并不只是和时间相关的一元函数，而是与空间分布亦有关系的（至少）二元函数。因此，能否将模型多元函数化也是推广所需要面临的挑战。这方面的想法可以借助元胞自动机的设计经验，通过只考虑局部影响的方式将种群切割出来，从而在更加复杂的条件下建立时空变化模型。在时空模型的基础上，可以对现有数据作出更精确的分析与讨论，也能分析出更多的影响因素。

从大量到少量

当前研究中，无法绕开的问题是对数据连续性和数据量的要求。某种程度上，无法进行定量研究也与过少的数据除了趋势无法判断其他情况有关。因此，为了进行模型的推广，小样本统计方法的引入是必要的。这依赖很多的统计专业性知识，因此在目前的研究中是难以进行的，但一旦能引入，模型的适用范围能大大加强，获得的结果也能比目前完备很多。

六、结论

综上所述，我们对统计方法研究三叶虫演化的可行性分析做出第一步探索，建立了理论模型。此后，我们假想了一些理论模型可能遇到的困难与扰动，并作出了可靠性可行性估计。最后，我们试着通过能否解释所观察到的现象来最终确定模型是否有效。

可喜的是，模型虽然不能解释所有现象，但它与一些拥有较好连续性的数据契合良好，的确可行。但研究至此，我们也发现了这个方法的一些短板，它对数据有着太高的要求，要求数据明确并且准确，然而，在通过化石进行研究的古生物考察时，结论与考察的进度往往时一同增长的，若没有通过其他方法得出一些前置的结论，对数据的认识就不够，本方法就无用武之地，但若对数据的认识充足，结论又往往早就可以由其他方式得出。但本方法仍有对前置结论做进一步确认的价值。因此，模型的推广方向中，最重要的是引入量化的识别工具，其次是将数量模型二维化，并增加小样本统计的手段。在加以合适的推广后，相信我们的研究方法可以更加完整、可靠。

有理由相信，未来的古生物学研究一定会结合更多的数学工具，从而进行更

为精细的分析。我们的方案虽然仍显粗糙,但为这样的引入创造了模式上的可能。可以期待,未来在更加完备的模型建立后,我们的分析可以得到切实可靠的结果,并为古动物发源等问题作出贡献。

参考文献

- 黄冰. 简论相似性测度的选择——以奥陶纪末大灭绝后全球腕足动物古地理为例[J]. 古生物学报, 2011, 50: 304-320.
- 林之, 徐阳. 生物的种间关系研究[J]. 实验教学与仪器, 2015, 32: 37-39.
- 刘帅. 贵州寒武纪杷榔动物群埋藏学研究[D]. 贵州大学, 2018.
- 时培建, 戈峰, 王建国, 郭世权. 外来物种入侵后的多物种竞争共存的集合种群模型[J]. 生态学报, 2009, 29: 1241-1250.
- 王宁星, 艾保全, 张学荣. 干扰条件下捕食者与被捕食者系统动态模型[J]. 生态学报, 2008, (03): 1059-1063.
- 魏鑫. 华南上扬子区中奥陶世达瑞威尔期三叶虫动物群[D]. 中国科学技术大学, 2020.
- 吴承羲, 陈哲, 庞科, 王霄鹏, 万斌, 周传明, 袁训来. 三峡地区埃迪卡拉纪石板滩生物群定量古生物学和生态空间分析[J]. 古生物学报, 2021, 60: 42-68.
- 伍孟银, 赵元龙, 杨兴莲, 陈争鹏, 杨光福. 贵州松桃黄莲寒武系都匀阶杷榔组三叶虫地层划分及意义[J]. 古生物学报, 2019, 58: 456-462.
- 殷宗军. 动物起源和寒武纪大爆发序言[J]. 古生物学报, 2021, 60: 1-9.
- Jia B, Gu H G, Li L, Zhao X Y. Dynamics of period-doubling bifurcation to chaos in the spontaneous neural firing patterns. [J]Cognitive Neurodynamics, 2012, (1): 89-106
- Chung Y M, Lee J Y. Optimization of some factors affecting the performance of query expansion. [J]Information Processing and Management, 2004, 40: 891-917.
- Shi G R. Multivariate Data-Analysis in paleoecology and paleobiogeography. [J]Paleogeography, Paleo-climatology, Paleoecology, 1993, 104:199-234.

致谢

首先, 感谢倪怀玮老师两学期的悉心指导与帮助。无论是研究过程中还是论文中都运用了大量课上所习得的知识。还要感谢所有参与交流的研讨课同学们, 这些交流让我们认识到了论文的不足, 也帮助我们进一步完善了论文。最后, 感谢所有在论文设计与撰写过程中提供过帮助或建议的人, 在你们的鼓励与支持下, 这篇文章才得以完成。