

# 应用 Excel 软件计算生物多样性指数

孔凡洲<sup>1,2</sup>, 于仁成<sup>1</sup>, 徐子钧<sup>3</sup>, 周名江<sup>1</sup>

(1. 中国科学院 海洋研究所 海洋生态与环境科学重点实验室, 山东 青岛 266071; 2. 中国科学院 研究生院, 北京 100049; 3. 国家海洋局 北海环境监测中心, 山东 青岛 266033)

**摘要:** 生物多样性指数是生物群落研究中的重要参数, 广泛应用于海洋生态系统生物多样性的评价和研究。作者利用微软的电子表格软件 Excel 中的数组函数、条件函数及其他函数, 提供了一种快速、简便、可靠的同时计算 Shannon、Simpson、Berger-Parker、Margalef 和 Pielou 均匀度指数等 8 种常用生物多样性指数的方法, 应用开发的 Excel 文件——Biodixcel.xlsx, 使用者只需在此文件列出的表格中一次性录入所要计算的原始数据即可在此表格中同时得到上述几个指数的计算结果, 本方法为群落生物多样性的计算提供了一种简单、实用的工具。

**关键词:** 生物多样性指数; 生物群落; 计算工具; Excel; Biodixcel

中图分类号: TP31; Q16      文献标识码: A      文章编号: 1000-3096(2012)04-0057-06

生物多样性(Biodiversity)对于维护生态系统稳定和生物资源持续利用具有重要作用, 是当前海洋生态学研究关注的热点问题之一。2000 年至 2010 年间实施的《国际海洋生物普查计划》, 首次尝试在全球尺度上评估和解释海洋生物的分布、丰度和多样性状况, 从种群、物种和基因三个层次建立起了海洋生物多样性研究体系<sup>[1]</sup>。在中国, 海洋生物多样性研究也受到密切关注, 多次大规模的海洋资源调查和专项调查为海洋生物多样性研究提供了大量资料<sup>[2]</sup>。同时, 中国海洋生态系统正承受着严峻的压力, 表现出物种减少、生物多样性下降等演变态势等<sup>[3]</sup>, 也需要大力加强海洋生物多样性的研究与保护工作。

在生物多样性研究中, 研究最多的是群落多样性, 并采用多样性指数来评价群落状况; 其中  $\alpha$  多样性指数用以测度群落内的物种多样性;  $\beta$  多样性指数用以测度群落的物种多样性沿着环境梯度变化的速率或群落间的多样性;  $\gamma$  多样性指数则是一定区域内总的物种多样性的度量。用于评价生物群落多样性的  $\alpha$  多样性指数使用最为广泛, 它包含两方面的含义, 即物种丰富度(表征群落中所含物种的多寡)和物种均匀度(表征群落中各个种的相对密度)。 $\alpha$  多样性指数包括丰富度指数(如  $d_{Gl}$ 、 $d_{Ma}$ 、 $d_{Me}$ 、 $d_{Mo}$ )、物种多样性指数(如 Simpson 指数、Shannon 指数、Brillouin 指数、McIntosh 指数)、均匀度指数(如 Pielou 指数、Heip 指数、Alatalo 指数、Hurlbert 指数)、用

于分析物种优势度的 Berger-Parker 指数及分析物种的多度分布规律的 Fisher  $\alpha$  指数等多种测度方法。利用 VB 等计算机高级语言或各种统计学软件(如 BIO-DAP, Biotools, r 软件等)可计算生物多样性指数, 但这些方法对于普通研究人员来说还相对繁琐。与上述方法相比, 微软公司生产的 Excel 软件操作简单、界面清晰、功能强大, 成为应用最广泛的数据处理和分析软件, 也已经应用于生物多样性指数的计算<sup>[4-5]</sup>。但是, 目前应用中往往根据多样性指数的定义, 将计算过程分解成不同的步骤, 分别利用简单的 Excel 函数如求和、平方、求平均数等先得到一些需要的中间结果, 再经过中间步骤的计算后得出最终结果。另外也有运用 VLOOKUP、TYPE 等较为复杂的函数, 或是采用 Excel “透视表”方法进行计算, 但这些计算方法基于无条件限制的简单数学函数, 计算繁杂, Excel 真正功能强大的条件函数和数组函数尚未得到应用。为简化多样性指数的计算、减少分步计算可能带来的错误, 本文利用 Excel 软件自带

收稿日期: 2011-10-12; 修回日期: 2012-02-21

基金项目: 国家基金委创新研究群体科学基金项目(40821004); 国家重点基础研究发展计划(973)项目(2010CB428705); 中国科学院知识创新工程重要方向项目群(KZCX2-YW-Q07-04); 中国科学院海洋研究所知识创新领域前沿项目

作者简介: 孔凡洲(1979-), 男, 博士研究生, 主要从事浮游植物生态学和有害赤潮研究, 电话: 0532-82898649, E-mail: fzkong@qdio.ac.cn; 于仁成, 通信作者, 电话: 0532-82898590, E-mail: rcyu@qdio.ac.cn

的条件函数和数组函数, 创建了一种含有条件函数的数组公式, 能同时计算一组样本的多种生物多样性指数。

## 1 计算准备

### 1.1 操作系统与软件版本

操作系统采用 Windows XP Service Pack 3(SP3) 操作系统, Excel 软件为 Microsoft Office Excel 2007 中文版。

### 1.2 生物多样性指数筛选与公式计算

可用于生物群落多样性分析的指数很多<sup>[6]</sup>, 根据生物群落中物种丰富度和相对优势度的高低, 各种生物多样性指数的适应性也有差别<sup>[7]</sup>。对此, 作者选择了最为常用的 Berger-Parker 指数、Margalef 指数、Simpson 指数( $\lambda$ 、 $D$  和  $D_r$  形式)、Shannon 指数(以 2 或 e 为底)和 Pielou 均匀度指数进行分析, 其中

Berger-Parker 指数用来表示优势度, Margalef 指数用来表示丰富度, Simpson 指数用来度量多样性的概率, Shannon 指数用来测度群落多样性的高低, Pielou 指数用来测度群落的均匀程度。各种生物多样性指数的计算公式如表 1 所示。

### 1.3 原始数据矩阵构建

在进行生物多样性指数计算前, 需要对数据进行处理, 构建一个样本与物种一一对应的原始数据矩阵, 记作  $A'_{m \times n}$  或  $(a'_{i,j})_{m \times n}$ ,  $a'_{i,j}$  ( $i=1, 2, \dots, m; j=1, 2, \dots, n$ ) 称为矩阵的第  $i$  行第  $j$  列的元素。在本文中用行( $m$ )表示来自不同总体的样本(sample), 列( $n$ )表示不同物种(species),  $A'_{m \times n}$  为  $m$  个样本的  $n$  个物种的集合,  $a'_{i,j}$  表示第  $i$  个样本中物种  $j$  的数量(这里的原始数据矩阵  $A'_{m \times n} = 2.2$  计算过程中图 1 所示的原始数据矩阵  $A_{m \times n}$ )。

表 1 用于各种生物多样性指数计算的公式列表

Tab. 1 Formulae for biodiversity indices

生物多样性指数	计算公式	计算说明	参考文献
Berger-Parker 指数 $d$	$d = 1 / \frac{n_{\max}}{N}$	$n_{\max}$ : 个体数量最多物种的个体数量	[8]
Margalef 指数 $d_{Ma}$	$d_{Ma} = \frac{S-1}{\ln N}$		[9]
Simpson 指数 $\lambda$ 形式	$\lambda = \sum_{i=1}^S P_i^2$	$P_i^2 = \frac{n_i(n_i-1)}{N(N-1)}$	[10]
Simpson 指数 $D$ 形式	$D = 1 - \sum_{i=1}^S P_i^2$	$P_i^2 = \frac{n_i(n_i-1)}{N(N-1)}$	[11]
Simpson 指数 $D_r$ 形式	$D_r = \frac{1}{\sum_{i=1}^S P_i^2}$	$P_i^2 = \frac{n_i(n_i-1)}{N(N-1)}$	[12]
Shannon 指数(以 e 为底) $H'_e$	$H'_e = -\sum_{i=1}^S P_i \ln P_i$	$P_i = \frac{n_i}{N}$	[13]
Shannon 指数(以 2 底) $H'_2$	$H'_2 = -\sum_{i=1}^S P_i \log_2 P_i$	$P_i = \frac{n_i}{N}$	[13]
Pielou 均匀度指数 $J_e$	$J_e = \frac{H'_e}{H'_{\max}}$	$H'_{\max} = \ln S$	[14]

注:  $S$  为物种数目;  $N$  为所有物种的个体数之和;  $n_i$  为第  $i$  个种个体数量

## 2 计算原理与计算过程

### 2.1 计算原理

依据表 1 所列各种生物多样性指数计算定义,

将 Excel 中的数组函数、条件函数及其他函数组合能满足上述计算要求的数组公式后分别对各种生物多样性指数进行计算。公式以等号“=”作为开始, 其内部可以包括上述各种函数、引用、运算符和常量

等。数组公式在录入后需要执行 Ctrl+Shift+Enter 操作,此时,Excel 会为输入的数组公式添加“{}”,自动执行该数组公式的计算。

数组是一组具有相同类型和名称的变量的集合,可以是一维的,也可以是多维的。如本文中每个样本的各个物种的个体数的集合即是一个一维数组,1.3 节所述的原始数据矩阵可以看作是相似的多个一维数组组成的多维数组。在数组公式中允许存在多个数组,但每个数组中的维数、行列数必须相同。数组公式中可以包括 Excel 数组函数,如本文中名称为 SUMPRODUCT 的数组函数,它应用于本文中的几个数组公式中,将数组间对应元素相乘,并返回乘

积之和。

数组公式默认数组中空单元格为“0”,当计算中出现“0”作分母或对“0”进行对数操作时会得到无意义的结果。另外,在计算某些指数(如 Simpson 指数)时,若原始数据矩阵中的元素值小于“1”,则与其定义不符,也会出现无意义的结果,这都需要使用条件函数(IF)将它们甄别出来分别进行处理,根据具体情况的不同分别赋值为 0 或用“(–)”来表示无意义的结果。条件函数(IF)会根据表达式是“TRUE”或“FALSE”返回不同的值或用不同的公式计算。表 2 列出了经条件函数甄别,在计算各种多样性指数时出现无意义情况时的处理方法。

表 2 在计算各种生物多样性指数时出现无意义情况的处理方法

Tab. 2 Processing methods of biodiversity indices while the results are meaningless

多样性指数	无意义情况	处理方法
$d$	当 $\sum_{j=1}^n a_{i,j} = 0$ 时, $N=0$ , 0 作分母无意义	结果为(–)
$d_{Ma}$	当 $\sum_{j=1}^n a_{i,j} < 1$ 时, $N < 1$ , $\ln N$ 为负数无意义	结果为(–)
	当 $\sum_{j=1}^n a_{i,j} = 1$ 时, $\ln N = 0$ , 0 作分母无意义	结果为(–)
$\lambda$ 、 $D$ 、 $D_r$	当 $\sum_{j=1}^n a_{i,j} = 0$ 或 1 时, 0 作分母无意义	结果为(–)
	当 $0 < a_{i,j} (j=1,2,\dots,n) < 1$ 时, 与定义不符	结果为(–)
$H'_e$ 、 $H'_2$	当 $\sum_{j=1}^n a_{i,j} = 0$ 时, $N=0$ , 0 作分母无意义	结果为(–)
	当 $a_{i,j} (j=1,2,\dots,n) = 0$ 时, $P_i = 0$ , 对 $P_i$ 求对数无意义	将 $\log_2 P_i$ 赋值为 0
	当 $S=1$ 时, $\ln S = 0$ , 0 作分母无意义	将 $\ln P_i$ 赋值为 0
$J_e$	当 $\sum_{j=1}^n a_{i,j} = 0$ 时, $N=0$ , 0 作分母无意义	结果为(–)
	当 $a_{i,j} (j=1,2,\dots,n) = 0$ 时, $P_i = 0$ , 对 $P_i$ 求对数无意义	将 $\log_2 P_i$ 赋值为 0
		将 $\ln P_i$ 赋值为 0

本文“其他函数”包括 ABS(求绝对值)、COUNTIF(返回满足给定条件的单元格的个数)、SUM(求和)、MAX(求最大值)、LN(返回一个数的自然对数)、LOG(按所指定的底数,返回一个数的对数)等,详细解释可以参考 Excel 2007 自带的帮助。

## 2.2 计算过程

计算过程的操作界面如图 1 所示, Excel 工作表自左向右由样本目录、计算结果区和原始数据区构

成。A3 正下方(包含 A3)为 1, 2, ...,  $m$  个样本的名称, 分别用 S001, S002.....来表示。C2–J2( $d-J_e$ )表示 8 种需要计算的多样性指数名称, 其下方为各个样本计算后的多样性指数值—记为计算结果矩阵  $B_{m \times 8}$ , 见图 1 计算结果区。L2(包含 L2)向右的单元格为 1, 2, ...,  $n$  个物种的名称, 分别用 s001, s002.....来表示, 其下方为各个样本对应的物种数—记为原始数据矩阵  $A_{m \times n}$ , 见图 1 原始数据区。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	样本 (1,2,...,m)		多样性指数									物种(1,2,...,n)				
2			$d$	$d_{Ma}$	$\lambda$	$D$	$D_r$	$H_s$	$H_2$	$J_e$		s001	s002	s003	s004	s005
3	S001		1.80	(-)	(-)	(-)	(-)	0.69	0.99	0.99		0	0.4	0.5	0	0
4	S002		1.57	10.49	(-)	(-)	(-)	0.66	0.95	0.95		0	0.4	0.7	0	
5	S003		2.64	11.50	(-)	(-)	(-)	1.08	1.56	0.99			0.45	0.44	0.3	0
6	S004		1.00	(-)	(-)	(-)	(-)	0.00	0.00	(-)			1	0	0	
7	S005		2.89	0.51	0.28	0.72	3.60	1.40	2.02	0.87		100	300	400	900	900
8	S006		2.89	0.72	0.28	0.72	3.63	1.40	2.02	0.87		10	30	40	90	90
9	S007		2.89	1.23	0.25	0.75	4.01	1.40	2.02	0.87		1	3	4	9	9
10	S008		2.89	4.19	(-)	(-)	(-)	1.40	2.02	0.87		0.1	0.3	0.4	0.9	0.9
11	S009		2.89	(-)	(-)	(-)	(-)	1.40	2.02	0.87		0.01	0.03	0.04	0.09	0.09

↑ 样本目录
↑ 计算结果区  
(B<sub>m×8</sub>)
↑ 原始数据区  
(A<sub>m×n</sub>)

图 1 操作界面

Fig.1 Operation Interface

现以样本(S001)为例,说明本文多样性指数的计算过程。C3-J3 单元格分别为样本(S001)的 8 个多样性指数的计算结果,系分别由 Excel 自动运行多样性指数的计算公式后得到,具体过程如下:

$$d_{S001} = C3 =$$

当群落所有物种的个体数之和不等于“0”时(TRUE),

输出  $1/\frac{n_{\max}}{N}$  的计算结果,否则(FALSE)输出“(-)”;  
即:

$$=IF(SUM(L3:SQ3)<>0,1/(MAX(L3:SQ3)/SUM(L3:SQ3)), "(-"))$$

以下各式类同,式中 SQ3 为 L3 向右(包括 L3)第 500 个单元格的名称。

$$d_{Ma\ S001} = D3 =$$

$$=IF(SUM(L3:SQ3)>1,(COUNTIF(L3:SQ3,">0")-1)/LN(SUM(L3:SQ3)), "(-"))$$

$$\lambda_{S001} = E3 =$$

$$=IF(ABS(SUM(L3:SQ3)-0.5)<0.5,IF(COUNTIF(L3:SQ3,">0")-COUNTIF(L3:SQ3,">=1")=0,SUM((L3:SQ3*(L3:SQ3-1))/(SUM(L3:SQ3)*(SUM(L3:SQ3)-1))), "(-")), "(-"))$$

$$D_{S001} = F3 =$$

$$=IF(ABS(SUM(L3:SQ3)-0.5)<0.5,IF(COUNTIF(L3:SQ3,">0")-COUNTIF(L3:SQ3,">=1")=0,1-SUM((L3:SQ3*(L3:SQ3-1))/(SUM(L3:SQ3)*(SUM(L3:SQ3)-1))), "(-")), "(-"))$$

$$D_r\ S001 = G3 =$$

$$=IF(ABS(SUM(L3:SQ3)-0.5)<0.5,IF(SUM(L3:SQ3)/COUNTIF(L3:SQ3,">0")=1, "(-)",IF(COUNTIF(L3:SQ3,">0")-COUNTIF(L3:SQ3,">=1")=0,1/SUM((L3:SQ3$$

$$*(L3:SQ3-1))/(SUM(L3:SQ3)*(SUM(L3:SQ3)-1))), "(-")), "(-"))$$

$$H'_e\ S001 = H3 =$$

$$=IF(SUM(L3:SQ3)<>0,-SUMPRODUCT((L3:SQ3/SUM(L3:SQ3)),(IF(L3:SQ3>0,(LN(L3:SQ3/SUM(L3:SQ3))))), "(-"))$$

$$H'_2\ S001 = I3 =$$

$$=IF(SUM(L3:SQ3)<>0,-SUMPRODUCT((L3:SQ3/SUM(L3:SQ3)),(IF(L3:SQ3>0,(LOG((L3:SQ3/SUM(L3:SQ3))),2))))), "(-"))$$

$$J_e\ S001 = J3 =$$

$$=IF(SUM(L3:SQ3)<>0,IF(COUNTIF(L3:SQ3,">0")=1, "(-)",-SUMPRODUCT((L3:SQ3/SUM(L3:SQ3)),(IF(L3:SQ3>0,(LN(L3:SQ3/SUM(L3:SQ3)))))/LN(COUNTIF(L3:SQ3,">0"))), "(-"))$$

其余样本(如 S002 等)的多样性指数计算公式与样本 S001 多样性指数的计算公式相同。

作者在提供的工作表(Biodixel.xls, 见下文 3)的公式中将矩阵  $A_{m \times n}$  预设为本数( $m$ )为 500, 物种数( $n$ )为 500 的大小,应可以满足一般海洋生态调查数据量的要求。当样本  $m$  大于 500 时,可以每次 500 个样本分多次计算,或按 2.2 介绍将 C3-J3 的多样性指数计算公式向下自动填充到大于 500 的项获得。当物种数  $n$  大于 500 时,需要参考本文 2.2 自行修订计算公式。

## 2.3 具体操作

拷贝原始数据矩阵  $A'_{m \times n}$ , 粘贴在工作表 L3 为首

行、首列的单元格上, Excel 会自动在 L3(包含 L3)的右下方将  $A'_{m \times n}$  矩阵填充在工作表的原始数据区中, 构成  $A_{m \times n}$  矩阵, 同时在计算结果区内自动生成包含  $m$  个样本的 8 组多样性指数的计算结果矩阵  $B_{m \times 8}$ 。

## 2.4 结果验证

经对比, 利用该方法计算的 2009 年 4 航次 24 个站位的长江口及邻近海域的浮游植物生物多样性结果与按照公式定义分步计算和利用 BIO-DAP 软件计算的结果完全一致, 表明应用本方法计算的生物多样性指数是正确的。

## 2.5 兼容性检验

本文的计算过程是基于 Windows XP 操作系统下的 Excel 2007 软件编写的, 可以在 Excel 2007 和更高的版本 Excel 2010 中使用, 但不适于在 Excel 2003 中运行(Excel 2003 的用户请见下文)。

另经检验, 本方法可以在 Windows Vista、Windows 7 操作系统中使用, 在目前很少使用的 Window 2000 等低版本的操作系统中未进行检验; 该文件也可以在安装有 OpenOffice.org 等软件的 Linux 操作系统下使用。

## 3 应用文件

为了方便研究人员应用, 作者制作了一个用于同时计算上述生物多样性指数的 Excel 文件 Biodixcel.xlsx(Biodiversity Excel), 研究人员可以在有害赤潮信息网 (<http://www.china-hab.cn/chinese/ccij/>) 下载或者通过 E-mail 向作者索取。

为方便 Excel 2003 用户, 作者另制作了一个可用于 Excel 2003 的 Excel 文件 - Biodixcel for Excel 2003.xls, 与上述 Biodixcel.xlsx 同时提供给研究者使用( $n \geq 245$ ), 该文件也可在 WPS 办公软件中应用。

## 4 讨论

关于 Shannon 指数和 Pielou 均匀度指数, 都分别有以 2 和  $e$  为底的两种算法, 国外多以  $e$  为底进行计算, 在中国《海洋监测规范》则将以 2 为底计算多样性指数作为标准方法<sup>[15]</sup>, 所以, 在 Biodixcel 中分别以 2 和  $e$  为底计算了两个 Shannon 多样性指数。以 2 和  $e$  为底计算的 Pielou 均匀度指数完全相同, 所以只取  $J_e$  进行计算。

条件函数的应用, 使得各种需要逻辑判断的计算得以进行, 理论上, 应该可以进一步应用条件函

数进行比本文更为复杂的运算, 读者可以自行尝试。

### 参考文献:

- [1] Tittensor D P, Mora C, Jetz W, et al. Global patterns and predictors of marine biodiversity across taxa [J]. Nature, 2010, 466: 1098-1107.
- [2] 周秋麟, 杨圣云, 陈宝红. 我国海洋生物物种多样性研究[J]. 科技导报, 2005, 23(2): 12-16.
- [3] 王斌. 我国海洋生物多样性保护的进展[J]. 海洋开发与管理. 2002, 3: 28-32.
- [4] 刘雨芳. EXCEL 在群落生物多样性参数计算中的应用[J]. 湘潭师范学院学报: 自然科学版, 2003, 25(2): 80-82.
- [5] 陈俊华, 文吉富, 王国良, 等. Excel 在计算群落生物多样性指数中的应用[J]. 四川林业科技, 2009, 30(3): 88-90.
- [6] 马克平, 刘玉明. 生物群落多样性的测度方法:  $1/\alpha$  多样性的测度方法(下)[J]. 生物多样性, 1994, 2(4): 231-239.
- [7] 孙军, 刘东艳. 多样性指数在海洋浮游植物研究中的应用[J]. 海洋学报(中文版), 2004, 26(1): 62-75.
- [8] Berger W H, Parker F L. Diversity of planktonic foraminifera in deep-sea sediments[J]. Science, 1970, 168: 1345-1347.
- [9] Margalef R. Information theory in ecology[J]. General Systematics, 1958, 3: 36-71.
- [10] Simpson E H. Measurement of diversity[J]. Nature, 1949, 163: 688.
- [11] Greenberg J H. The measurement of linguistic diversity[J]. Language, 1956, 32: 105-119.
- [12] Hill M O. Diversity and evenness: a unifying notation and its consequences[J]. Ecology, 1973, 54: 427-431.
- [13] Shannon C E. A mathematical theory of communication[J]. Bell System Technical Journal, 1948, 27: 379-423.
- [14] Pielou E C. The measurement of diversity in different types of biological collections[J]. Journal of Theoretical Biology, 1966, 13: 131-144.
- [15] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 中华人民共和国国家标准 GB17378.7-2007 海洋监测规范 第 7 部分: 近海污染生态调查和生物监测[S]. 2007.

# Application of excel in calculation of biodiversity indices

KONG Fan-zhou<sup>1, 2</sup>, YU Ren-cheng<sup>1</sup>, XU Zi-jun<sup>3</sup>, ZHOU Ming-jiang<sup>1</sup>

(1. Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; 2. Graduate University, Chinese Academy of Sciences, Beijing 100049, China; 3. North China Sea Environmental Monitoring Center, Qingdao 266033, China)

**Received:** Oct., 12, 2011

**Key words:** biodiversity index; biological community; calculation; excel; biodixel

**Abstract:** Biodiversity indices, which have been widely adopted in assessment and research of marine ecosystems, are important parameters in studies of biological communities. We developed a rapid, convenient and reliable method to calculate eight commonly used indices including Shannon diversity index, Simpson diversity index, Berger-Parker diversity index, Margalef diversity index and Pielou evenness index, simultaneously by using array function, condition function and other functions provided in the Microsoft Excel. An Excel file named “Biodixel.xlsx” was attached in this paper to facilitate the calculation process of the indices. The indices would be generated automatically once the original raw data being properly loaded into the provided programs of the Excel file, “Biodixel.xlsx”.

(本文编辑: 梁德海)