

МЕТОДИКА ИССЛЕДОВАНИЙ

УДК 56.07

А. П. РАСНИЦЫН и Г. М. ХОВАНОВ

УТОЧНЕННЫЙ МЕТОД ОЦЕНКИ ОБЪЕМА ЛОКАЛЬНЫХ ФАУН

На основании выведенного Макартом распределения численности видов определенной группы в биоценозе получена формула, позволяющая рассчитывать число видов, родов и т. д. в фауне по числу видов в выборке из этой фауны и размеру выборки: $M \approx mn / (n - m)$, где M — объем фауны, m — число видов в выборке, n — размер выборки. Кроме того, получена эмпирическая формула для расчета среднего квадратичного отклонения m , позволяющая указать вероятный интервал, заключающий действительное значение M , с заданным уравнением достоверности: $s_m \approx 1,5m^2 / n M^{0.55}$. Рассматриваются наиболее важные источники систематической ошибки M .

Предложенный в статье А. Г. Пономаренко и А. П. Расницына (1967) метод оценки общего числа видов (родов, семейств), образующих данную локальную фауну, по выборке из этой фауны основан на предположении, что распределение частот видов в выборке более или менее соответствует распределению частот в генеральной совокупности. Надежность такого метода, естественно, не слишком велика, тем более что он требует данных по двум выборкам существенно разного размера из одной фауны. При этом надежность данных по распределению частот видов в малой выборке оказывается особенно низкой. Все это заставило авторов искать способы уточнения методики расчета объема фауны.

Возможность такого уточнения и одновременно упрощения метода дает известное уравнение Р. Макатура (McArthur, 1957), описывающее распределение численности видов какой-либо группы животных в конкретном биоценозе:

$$N_r = \frac{N}{M} \sum_{i=1}^r \frac{1}{M - i + 1}, \quad (1)$$

где N — суммарная численность всех видов фауны, M — число этих видов, N_r — численность r -го редкого вида, т. е. вида, занимающего r -е место в ряду всех видов данной фауны, если эти виды расположить в порядке возрастания их численности; таким образом, N_1 — численность самого редкого вида, N_M — самого частого.

Уравнение Макатура выводится теоретически при условиях, что: 1) положение видов в ряду с течением времени не изменяется, т. е. номер каждого вида r остается постоянным, 2) доступные для данной группы местообитания используются более или менее полностью и 3) экологические ниши не перекрываются слишком сильно. Кроме того, самим Макартом было показано, что его уравнение хорошо описывает распределение современных птиц, а Л. Ван Вален (Van Valen, 1964) продемонстрировал то же самое на вымерших млекопитающих. Общие теоретические предпосылки распределения Макатура кажутся приложимыми к видам самых различных групп животных и, кроме того, не только к

видам, но и к любым другим таксонам при условии их сравнимости с экологической точки зрения. Поэтому нам кажется оправданной попытка использовать уравнение Макаурга для упрощения и уточнения метода оценки разнообразия локальных фаун.

Для решения поставленной задачи необходимо определить зависимость между общим числом видов фауны, обозначаемым через M , размером выборки, равным n экземплярам, и числом видов (родов и т. д.), представленных в этой выборке, равным m . Для этого выразим m через разность между общим числом видов M и числом видов, не попавших в выборку:

$$m = M - \sum_{r=1}^M (1 - p_r)^n,$$

где p_r — частота r -го вида, т. е.

$$p_r = \frac{N_r}{N} = \frac{1}{M} \sum_{i=1}^r \frac{1}{M - i + 1}.$$

Отсюда

$$m = M - \sum_{r=1}^M \left(1 - \frac{1}{M} \sum_{i=1}^r \frac{1}{M - i + 1} \right)^n. \quad (2)$$

Если M велико, то замена суммы, стоящей в скобках, интегралом дает небольшую относительную ошибку:

$$\sum_{i=1}^r \frac{1}{M - i + 1} \approx \ln \frac{M}{M - r + 1} \quad (1 \leq r \leq M).$$

Подставляя это выражение в (2), имеем:

$$m \approx M - \sum_{r=1}^M \left(1 - \frac{1}{M} \ln \frac{M}{M - r + 1} \right)^n.$$

Если M велико, мы можем использовать равенство $(1 + x)^n \approx e^{xn}$.

$$m \approx M - \sum_{r=1}^M e^{-\frac{n}{M} \ln \frac{M}{M - r + 1}} = M - \sum_{r=1}^M \left(\frac{M}{M - r + 1} \right)^{-\frac{n}{M}}.$$

Заменяя $M - r + 1$ на x и переходя от суммирования к интегрированию, получаем:

$$m \approx M - \sum_{r=1}^M \left(\frac{x}{M} \right)^{\frac{n}{M}} \approx M - \int_0^M \left(\frac{x}{M} \right)^{\frac{n}{M}} dx = M - M^{-\frac{n}{M}} \frac{M^{1+\frac{n}{M}} + 1}{1 + \frac{n}{M}}.$$

Поскольку $M^{1+\frac{n}{M}} \gg 1$, единицей в числителе пренебрегаем:

$$m \approx M - \frac{M}{1 + \frac{n}{M}} = \frac{Mn}{M + n} \quad (3)$$

или

$$1/m \approx 1/M + 1/n, \quad (4)$$

что означает гиперболическую зависимость m от n при постоянной M .

Предположение о гиперболической зависимости было высказано уже в работе Пономаренко и Расницына (1967), но там о форме гиперболы ничего не было известно, и для ее определения требовались две точки (две пары значений m и n); здесь же достаточно единственного значения. Теперь еще раз изменим форму записи равенства (3):

$$M \approx \frac{mn}{n - m} \quad (5)$$

Соотношение (5), позволяющее непосредственно рассчитывать M по m и n , дает меньшую относительную ошибку, как мы отмечали, при больших M . Ограничения эти не очень актуальны, поскольку большую точность при определении M получить невозможно из-за постоянной систематической ошибки, а при малой точности достаточно большим значением M оказывается уже 10.

Несмотря на благоприятные теоретические прогнозы, была предпринята попытка экспериментально оценить надежность формулы (5). Одновременно оказалось возможным обнаружить масштаб статистического разброса результатов и определить интервал вероятных значений M .

Эксперимент проводился по методу Монте-Карло. На листе миллиметровой бумаги выделялся участок, длина которого соответствовала N . Участок разбивался на M отрезков, длина каждого из которых определялась уравнением Макауртура. Затем с помощью таблицы случайных чисел получали наборы из n двузначных или трехзначных чисел (соответственно при $N = 100$ и $N = 1000$). Эти числа обозначали положение точек на отрезке и использовались как члены выборки, как «собранные экземпляры». По распределению этих точек на отрезке легко определить, сколько «видов» (отрезков) представлено в выборке. Результаты эксперимента представлены в табл. 1. Из таблицы следует прежде всего, что при всех значениях M , кроме $M = 6$, в том числе и при $n = 25$, рассчитанные значения объема «фауны» очень близки к истинным (для этих случаев в среднем $\bar{M}/M = 1,01$). Несколько большие отклонения получаются при $M = 6$, особенно при малых значениях n ($\bar{M}/M = 1,15 \div 1,40$). Однако даже максимальные ошибки метода кажутся не столь уж большими по сравнению с ожидаемой систематической ошибкой, к анализу которой мы вскоре перейдем, а во многих случаях и несущественными по сравнению со статистическими ошибками.

Определение статистической ошибки эксперимента обнаружило закономерную связь среднего квадратичного отклонения числа видов в выборке с другими переменными (табл. 1):

$$s_m \approx 1,5 \bar{m} / n(n - \bar{m}) / n \bar{M}^{0,45} = 1,5 m^2 / n \bar{M}^{0,55} = s_m' \quad (6)$$

Биологический смысл этой эмпирической формулы не ясен, но поскольку она соблюдается достаточно хорошо (в среднем $s_m'/s_m = 1,005$, максимальный разброс $0,88 \div 1,32$), ее можно использовать для приближенной оценки вероятного интервала, заключающего истинное значение M . Это тем более важно, что часто мы не имеем возможности получить достаточное число одинаковых по размеру выборок из одной и той же фауны и тем самым непосредственно определить s_m . В таких случаях, видимо, следует пользоваться формулой (6). Если же большой точности не требуется, можно ограничиться более простой, но более грубой формулой:

$$s_m \approx 1,5 m^2 / n \sqrt{M} \quad (6a)$$

Заканчивая изложение фактических данных, нужно отметить, что формула (5) использовалась для расчета объема фауны на том же материале, который послужил основой для разработки прежнего метода расчета (Пономаренко и Расницын, 1967). Результаты оказались до-

Таблица 1

Результаты статистического эксперимента по оценке объема локальной фауны

N	M	n	n'	$\bar{m}$	$\bar{M}$	$\frac{\bar{M}}{\bar{m}}$	s_m	s'_m	$\frac{s'_m}{s_m}$	$\text{marg } \frac{mn}{n-m}$	$\text{marg } M$ ($P \geq 0,99$)
1000	400	200	25	134,8	413	1,03	4,96	4,95	0,99	326—514	313—563
1000	400	100	50	80,1	402	1,00	3,38	3,55	1,06	285—810	250—785
1000	400	50	400	44,4	397	0,99	2,25	2,22	1,0	142—2450	174—∞
1000	100	100	50	49,8	99	0,99	3,21	2,99	0,93	72—144	70—140
1000	100	50	100	33,4	100	1,00	2,51	2,65	1,06	59—177	58—197
1000	25	50	100	16,7	25	1,00	1,42	1,29	0,91	18—31	18—34
100	25	24	25	12,5	26,1	1,05	1,85	1,62	0,88	14—48	11—62
100	6	24	25	5,4	6,95	1,16	0,65	0,63	0,97	5—8	4,4—10
100	6	12	50	4,6	7,4	1,24	0,95	0,88	0,93	4—12	2,6—17
100	6	6	100	3,5	8,35	1,39	0,71	0,94	1,32	3—∞	1,8—80

Условные обозначения: N — полная численность «фауны»; M — полное число «видов» в «фауне»; n — размер выборки; n' — число повторностей по каждому варианту; $\bar{m}$ — среднее число «видов» в выборке; $\bar{M}$ — объем «фауны», рассчитанный по $\bar{m}$; s_m — среднее квадратичное отклонение m ; s'_m — то же, но рассчитанное по формуле $s'_m = 1,5 \bar{m}^2/n \bar{M}^{0,5}$; $\text{marg } mn/(n-m)$ — крайние значения M , рассчитанные по крайним фактическим значениям m ; $\text{marg } M$ — крайние значения M , рассчитанные для 99%-ного уровня значимости по $t \pm 2,58 s_m$.

вольно близкими, во всяком случае старые цифры не выходят за границы вероятных значений M , полученные с помощью (6а). Это может служить косвенным подтверждением того, что распределение насекомых не отличается радикально от ожидаемого в соответствии с (1).

Использование выражений (5) и (6) — (6а) для расчета объема локальной фауны, как и прежняя методика расчета, может давать надежные результаты лишь при соблюдении ряда условий. Действительно, если распределение видов в фауне сильно отличается от (1) или распределение в выборке сильно отличается от распределения в исходной фауне, результат расчета будет лишь вводить нас в заблуждение, особенно при некритическом отношении к величинам, получаемым с помощью (5) и (6).

Поскольку доказательство соответствия распределения в реальных фаунах уравнению (1) требует постановки специальных весьма трудоемких исследований, мы на этом больше останавливаться не будем. Из других источников систематической ошибки выражения (5) особенно важными нам представляются следующие.

1. Непредвзятый отбор материала. Действительно, при обычных сборах исследователь с наибольшим рвением собирает или наиболее редкие, или наиболее интересные для него формы, а банальные очень быстро перестает брать вообще. Поскольку при этом объем выборки искусственно приуменьшается, объем фауны оказывается преувеличенным. Это, по нашему мнению, самый важный источник ошибки, поэтому расчет объема фауны следует вести по материалам, собранным беспристрастно, например, по сборам, сделанным специально для этой цели.

2. Особенности поведения разных организмов по отношению к данной методике сбора. Этот источник ошибки относится преимущественно к данным по современным организмам, так как особенности их поведения и экологии сильно влияют на распределение видов в пробах, сделанных разными методами, причем ни один из этих методов не дает распределения, точно соответствующего истинному. В отношении современных насекомых все это очень убедительно показано, например, в книге Р. Шовена (1970).

При расчете объема фаун прошлого существуют свои специфические источники ошибок. Частично они рассмотрены Ван Валеном (Van Valen, 1964). Наиболее важными мы считаем следующие факторы.

3. Влияние экологии и поведения на вероятность попадания в захоронение. На примере насекомых, остатки которых сохраняются обычно в водных отложениях, можно показать, что водные и околоводные насекомые во многих сборах имеют заведомо завышенную численность, а наземные нелетающие и особенно малоподвижные — сильно заниженную численность по сравнению с действительной. Учесть этот источник ошибки во многих случаях весьма трудно.

4. Вероятность сохранения остатков. Чем слабее скелет и (если речь идет об отпечатке насекомого) чем меньше развита окраска, тем меньшая доля остатков сохраняется и попадает в руки исследователя. Если анализируемая группа неоднородна в этом отношении, различная способность сохраняться в ископаемом состоянии также будет искажать исходное распределение численности видов.

5. Различная продолжительность генерации. Численность вымерших животных в сборах — это по существу численность их трупов, а при прочих равных условиях популяция, дающая два поколения в год, оставляет вдвое больше трупов, чем популяция с годичной генерацией. Поэтому, если члены анализируемой фауны сильно отличаются по продолжительности развития, этот фактор может оказать заметное влияние на результат.

6. При интерпретации результатов возникает также неопределенность, связанная с проблемой сборных видов и неравноценных таксонов. В палеонтологии гораздо в большей степени, чем в неонтологии, таксоны номинально одного ранга фактически могут не соответствовать друг другу (одни из них сборные, другие нет). Это, естественно, сильно изменяет распределение численности в такой «фауне» по сравнению с (1). Далее, даже если в пределах одной фауны таксоны одного ранга более или менее равноценны друг другу, в разных фаунах они могут не соответствовать. Поэтому выводы об изменении разнообразия данной группы в пространстве и во времени (для чего, собственно, и нужна предлагаемая методика) следует делать с осторожностью. В особенности это относится к сравнению объема вымерших и современных фаун. Расчеты дают нам число не истинных видов, а только того, что мы различаем как виды на данном материале.

Таким образом, источники систематической ошибки при оценке объема фауны многочисленны и разнообразны. Влияние их, вероятно, можно оценить с помощью специальных исследований, которые, по нашему мнению, нужно проводить на фаунах, состав которых уже достаточно полно изучен, и на специально собранном материале. Однако и без этого оценка объема фауны может принести определенную пользу. Объем древних фаун нередко не известен даже приблизительно, и определение хотя бы его порядка уже позволяет делать какие-то выводы. К тому же палеонтологический материал при всех его недостатках обладает и определенными преимуществами. Современные сборы иногда гораздо более чувствительны к деталям экологии и поведения животных, чем палеонтологические. Например, для наземных летающих насекомых, экологически не привязанных к воде, наиболее существенны в этом отношении только время, проводимое в полете, и сила, управляемость полета (возможность случайного попадания в воду, например при резких порывах ветра). Кроме того, на сборы современных организмов большое влияние оказывают суточные, сезонные и многолетние колебания численности и активности. Палеонтологические же сборы усреднены во времени за периоды во много сотен и тысяч лет, так что «мгновения» отклонений от распределения (1), которые очень существенны для современных сборов, на палеонтологическом материале могут усредниться.

Таким образом, представляется, что предлагаемый способ оценки разнообразия фауны, при своих недостатках и ограничениях, даже в его нынешней форме все же может быть полезным. Во всяком случае, еще более грубая методика, описанная Пономаренко и Расницыным (1967), принесла одному из авторов совершенно определенную пользу, позволив дать хотя бы гипотетическое объяснение прежде необъяснимому этапу геологической истории одной из групп насекомых (Расницын, 1969, стр. 19—20).

ЛИТЕРАТУРА

- Пономаренко А. Г. и Расницын А. П. 1967. К методике ориентировочной оценки разнообразия локальных фаун прошлого. Палеонтол. ж., № 3, стр. 98—105.
- Расницын А. П. 1969. Происхождение и эволюция низших перепончатокрылых. «Наука», стр. 1—196.
- Шовен Р. 1970. Мир насекомых. «Мир», стр. 1—240.
- McArthur R. H. 1957. On the relative abundance of bird species. Proc. Nat. Acad. Sci. U. S. A., vol. 45, p. 293—295.
- Van Valen L. 1964. Relative abundance of species in some fossil mammal faunas. Amer. Naturalist, vol. 98, № 899, p. 109—116.

Палеонтологический институт
Академии наук СССР

Статья поступила в редакцию
22 III 1971