

Структура разнообразия спорово-пыльцевых спектров: от математических моделей к палеоклиматическим выводам

А.В. Гоманьков

*Ботанический институт им. В.Л. Комарова РАН,
197376, Санкт-Петербург, ул. Профессора Попова, 2
Gomankov@mail.ru*

Поводом для написания настоящей статьи явилась подборка материалов о применении распределения Ципфа в естествознании, а также развернувшаяся по ее поводу дискуссия на страницах журнала «Природа» [Бялко, 1993, 1995; Карасёв, 1995; Трубников, 1993, 1995]. Мне также захотелось поделиться своим «опытом общения» с этим распределением, причем опытом скорее отрицательным, как будет видно из дальнейшего изложения.

Представляется довольно очевидным, что многие эмпирические распределения, в том числе с широким разбросом измеряемых величин, можно описывать с помощью разных математических моделей, то есть функций или «законов». И в смысле качества приближения, то есть отклонения наблюдаемых значений от значений, предсказываемых моделью, разные описания такого рода могут почти не отличаться друг от друга. Как выбрать в таком случае одну модель среди множества возможных? Каковы критерии, заставляющие нас предпочитать одну функцию распределения другой?

Цель настоящей статьи, которую можно рассматривать как подтверждение и развитие идей, высказанных Б.В. Карасёвым [1995] в ходе упомянутой дискуссии в «Природе», – проиллюстрировать решение указанных вопросов на конкретном примере из палеопалинологии. Традиционно палеопалинология – дисциплина довольно далекая от математики (если не считать элементарных арифметических подсчетов), и та важная роль, которую играют в ней широкие распределения, кажется совершенно неожиданной. Однако, коль скоро речь идет об универсальности некоторого явления, именно наиболее «экзотические» примеры и представляются самыми показательными и потому заслуживающими подробного рассмотрения.

Ввиду того, что палеопалинология не относится к числу широко популяризируемых областей знания, вероятно, имеет смысл начать изложение с основных понятий этой науки.

Что такое спорово-пыльцевой спектр?

Споры и пыльца (вместе они называются микроспорами) – это микроскопические образования, служащие для размножения высших растений. Они имеют характерные размеры порядка нескольких десятков микрон, заключены, как правило, в оболочку, состоящую из прочного и химически инертного биополимера – спорополленина, и могут легко переноситься по воздуху и воде на значительные расстояния. Благодаря физической и химической устойчивости спорополленина они могут сохраняться в осадочных горных породах в течение многих миллионов лет и служат важным источником информации о растительном мире Земли в прошлые геологические эпохи. Морфологическое разнообразие миоспор соизмеримо с разнообразием производящих их растений, так что в первом приближении можно считать, что всякий вид растений производит свой специфический вид спор или пыльцы и дисперсные (т. е. найденные в отрыве от материнских растений) миоспоры позволяют определять растения с точностью до вида.

Для изучения ископаемых дисперсных миоспор их надо прежде всего извлечь из породы. Для этого образец породы (обычно массой около 200 г) подвергается специальной физической и химической обработке, в результате которой все (или почти все) содержащиеся в нем миоспоры переводятся в какую-нибудь прозрачную жидкую среду (как правило, обычную дистиллированную воду). Пробирка с 1–2 мл жидкости, где во взвешенном состоянии находятся все миоспоры, извлеченные из одного исходного образца горной породы, называется палинологической пробой. Со многими оговорками (не все виды миоспор сохраняются в ископаемом состоянии, разные виды могут разноситься на разные расстояния и т. д.) можно считать, что палинологическая проба приблизительно отражает некоторый палеофитоценоз, то есть определенное со-

общество растений, обитавшее в прошлом на какой-то территории вокруг той точки, из которой взята проба.

Как правило, горные породы или не содержат спор и пыльцы вообще, или содержат их в массовом количестве. Поэтому число миоспор в каждой пробе настолько велико, что никто и никогда не изучает их все: из пробы делается выборка в виде одного или нескольких препаратов, которые затем просматриваются под микроскопом. Способ приготовления и последующего изучения палинологического препарата может рассматриваться как очень хорошая реализация той модельной ситуации, которая в математической статистике получила название случайной выборки без возвращения, причем объем этой выборки задается самим исследователем и может меняться им по ходу эксперимента.

Пусть, например, было изучено n миоспор и среди них выявлено t видов. Каждый из этих видов представлен некоторым числом экземпляров

– k_i ($i = 1, 2, \dots, t$; $\sum_{i=1}^t k_i = n$), и для каждого вида

вычисляется величина $q_i = \frac{k_i}{n}$, называемая его долей в спектре.

Таким образом, результатом всей описанной деятельности палинолога является список из t видов, каждому из которых поставлена в соответствие величина q_i (обычно выраженная в процентах). Такой список и называется спорово-пыльцевым спектром, а число n – объемом этого спектра.

Величины q_i в полном соответствии с канонами математической статистики рассматриваются как оценки величин Q_i – долей соответствующих видов в исходной пробе, т. е. величин, численно характеризующих наши представления об относительном изобилии того или другого вида миоспор. Спорово-пыльцевые спектры служат основным фактическим материалом, на котором основываются все приложения палеопалинологии: определения возраста горных пород, реконструкции растительности, реконструкции климата и т. д.

Структура разнообразия – распределение Ципфа?

Если считать, что разнообразие внутри какого-нибудь множества объектов определяется группировкой этих объектов в некоторые подмножества (таксоны), в пределах которых объекты рассматриваются как одинаковые, то под структурой разнообразия естественно понимать распределе-

ние таксонов по их численности. Многие распределения, анализировавшиеся в статье Б.А. Трубникова [1993], можно рассматривать как структуру разнообразия тех или иных множеств объектов. Ю.В. Чайковским [1990] была предложена новая наука – диатропика как общее учение о разнообразии. При этом распределение Ципфа рассматривается как универсальная структура разнообразия. Для биологических объектов эта универсальность была продемонстрирована английскими исследователями Д.Виллисом и Г.Юлом еще в 20-х годах прошлого века. «Виллис обнаружил гиперболы не только в распределении таксонов (т. е. в распределении биологических родов по числу видов, семейств по числу родов и т. д. – А. Г.), но и в распределении индивидов по видам, явившись тем самым одним из основоположников учения об *относительном обилии видов* <...>. Именно это обилие может быть сопоставлено с частотой встречаемости слова по Ципфу, так что понятие статуса слова <...>, означающее его значимость в речи, можно сопоставить со статусом вида (его значимости в экосистеме)» [Чайковский, 1990, с. 86–87].

Ясно, что структуру разнообразия можно проанализировать и для любого спорово-пыльцевого спектра. Для этого надо построить распределение входящих в него видов по значениям k_i (или, если угодно, q_i). На рис. 1 показаны такие распределения для нескольких спектров из пермских отложений Русской равнины. Видно, что все они, кроме одного, представленного на рис. 1, д (о нем отдельная речь впереди), имеют максимум в точке $k = 1$, общий характер монотонно убывающей вогнутой кривой и, вероятно, могут быть хорошо описаны с помощью гиперболы. Таким образом, структура разнообразия и для спорово-пыльцевых спектров носит, по видимому, характер ципфовского распределения. Кстати, одна из четырех основных разработанных в геоботанике моделей формирования разнообразия фитоценозов, так называемая геометрическая серия, также приводит к распределению Ципфа [Lamboy, Lesnikovska, 1988].

Казалось бы, все хорошо, все сходится, все совпадает. Однако...

«Антропный принцип» в статистике

Между числом видов t в спектре и функцией $f(k)$, описывающей структуру его разнообразия, существует, очевидно, простая зависимость

$$t = \int_{\frac{1}{2}}^{\infty} f(k) dk .$$

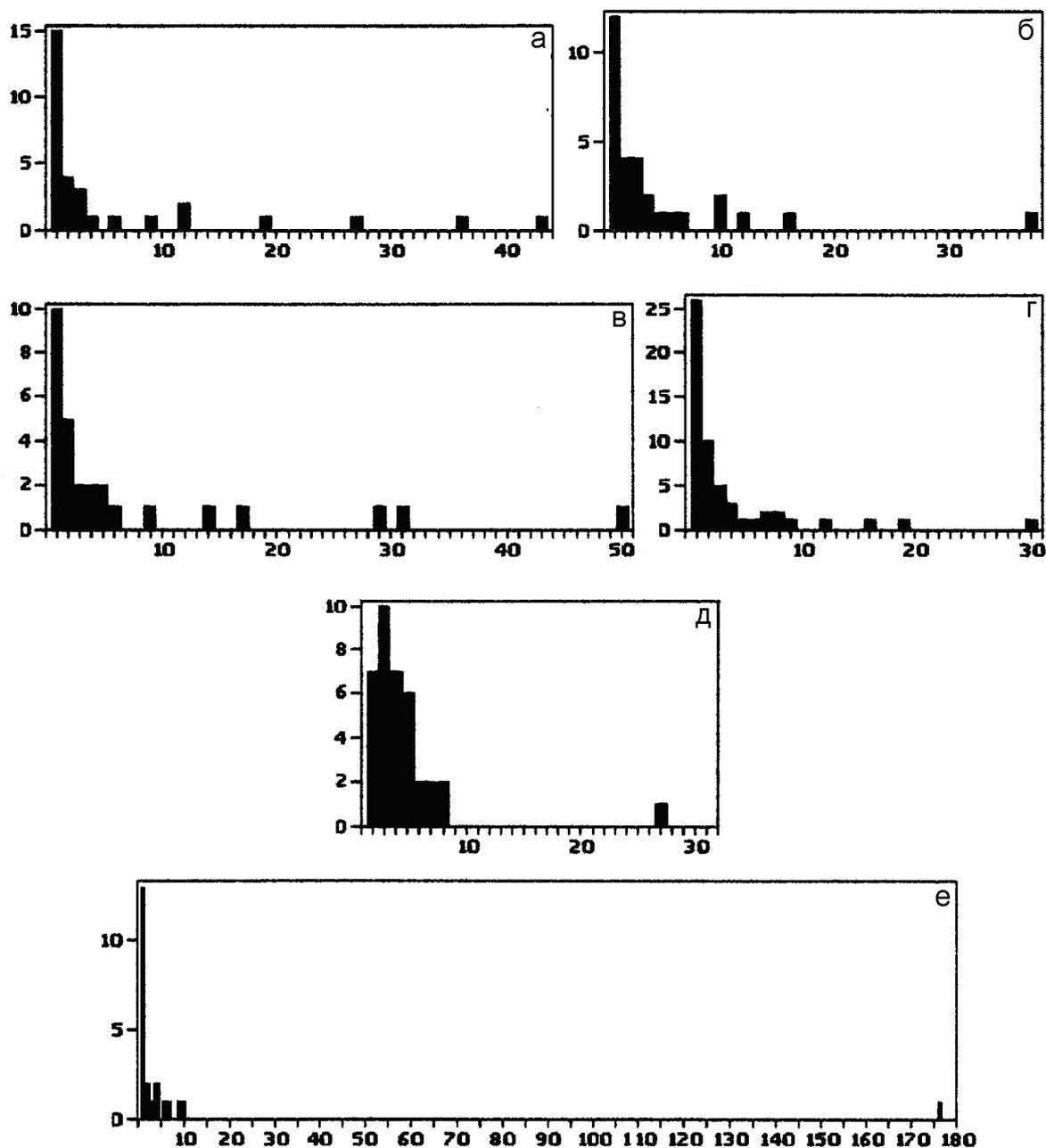


Рис. 1. Структура разнообразия некоторых спорово-пыльцевых спектров из отложений татарского яруса Русской равнины: а – Исады, б – Аристово II, в – Дор, г – Вострое, д – Луптюг, е – Титово

Однако, если моделировать функцию $f(k)$ обычной формулой распределения Ципфа $f(k) = ak^{-\gamma}$, то для того, чтобы данный интеграл сходился, необходимо, чтобы показатель степени γ был больше единицы. Это обстоятельство как-то сразустораживает: что это за ограничение на параметр распределения, вытекающее лишь из того факта, что некоторое число t реально существует? Заметим, кстати, что ограничение это довольно существенно, так как оно сразу выводит функцию $f(k)$ из числа «класси-

ческих»: ведь в большинстве случаев при рассмотрении распределения Ципфа принимается $\gamma = 1$. Но еще большая «несуразность» выявится, если мы зададимся вопросом о качественной репрезентативности спектра.

Если в исходной пробе общее число миоспор равно N , а общее число видов этих миоспор равно T , то с содержательной палеопалинологической точки зрения вполне справедливым представляется, что $N \gg T$ и число T должно считаться конечным даже в такой модели, где объем

пробы рассматривается как бесконечно большой. Это следует из того, что палеофитоценоз, послуживший для формирования данной пробы, занимал конечную территорию, на этой территории произрастало конечное число растений и уж тем более конечным было число видов этих растений. Практически же бесконечное число миоспор в пробе обусловлено тем, что по крайней мере некоторые из этих видов производили их в огромных количествах, что подтверждается наблюдениями над современными растениями.

Эти соображения делают возможной постановку вопроса о том, как соотносятся числа t и T ; все ли виды, присутствующие в пробе, попали в спектр, и если не все, то какая их часть? Ясно, что в спектр не попали виды с маленькими значениями Q_i . Поскольку при достаточно большом n $q_i \approx Q_i$ и, следовательно, $Q_i n \approx k_i$, то можно считать, что числом экземпляров k_i в спектре представлены все те и только те виды, для которых величина $Q_i n$ попадает в интервал от $k_i - \frac{1}{2}$

до $k_i + \frac{1}{2}$ и, следовательно, величина Q_i попадает

в интервал от $\frac{k_i - \frac{1}{2}}{n}$ до $\frac{k_i + \frac{1}{2}}{n}$. Таким образом,

в спектр не попадут (будут «представлены» в нем нулем экземпляров) те виды, для которых $0 < Q_i < \frac{1}{2n}$. Число таких видов можно рассматривать как значение функции f в точке $k = 0$.

Однако, гипербола $ak^{-\gamma}$ стремится к бесконечности при $k \rightarrow 0$. Фактически это означает, что существует бесконечное число видов, не принадлежащих данному спектру (представленных в нем нулем экземпляров), и по мере увеличения объема спектра n (вспомним, что эта величина может произвольно меняться, в частности, неограниченно возрастать по ходу эксперимента) в нем будут появляться все новые и новые виды. Другими словами, оказывается, что число T бесконечно велико, а это, как мы видели, противоречит содержательным палинологическим представлениям.

Исторически первые описания структуры разнообразия для спорово-пыльцевых спектров [Гоманьков, 1990] возникли позже, чем вопрос о их качественной репрезентативности, и именно в целях ответа на этот вопрос (оценку величины T можно получить, очевидно, проинтегрировав функцию $f(k)$ от нуля до бесконечности). Если

считать, что соответствие или несоответствие эмпирического распределения какой-либо теоретической модели есть объективная реальность (можно вспомнить слова из статьи Б.В. Карасёва [1995, с. 42]: «Покажем <...>, что часть приведенных авторами распределений *на самом деле* логнормальна <...>»; курсив мой – А.Г.), то в данном случае оказывается, что эта реальность объясняется не из другой объективной реальности, а из представлений о *корректности начальной постановки задачи*: мы принимаем за истинные такие значения параметров (в нашем случае – такой вид теоретического распределения), которые делают осмысленным поставленный перед нами вопрос.

Здесь можно увидеть аналогию с известным антропным принципом в физике, объясняющим наблюдаемые значения мировых констант на основе возможности существования во Вселенной разумной жизни и, следовательно, самой постановки вопроса о мировых константах (см., например, [Казютинский, Балашов, 1989]). «Антропный принцип» в статистике заставляет нас, таким образом, отказаться от моделирования структуры разнообразия в спорово-пыльцевых спектрах с помощью распределения Ципфа и поискать для этой цели каких-нибудь других распределений.

«Невидимые» виды

До сих пор мы рассматривали распределение $f(k)$ при постоянном объеме спектра. Как будет вести себя эта функция при увеличении n ? Если, согласно нашему предположению, величина T («потенциальное» число видов в спектре) конечна, то ясно, что функция $f(k)$ будет иметь максимум в точке $k = 1$ лишь при сравнительно маленьких n . При увеличении же n ее максимум начнет смещаться в область все больших значений k , а слева от него будет формироваться область малых значений $f(k)$, близких к нулю. Другими словами, функция $f(k)$ уже не будет монотонно убывающей, хотя сохранит свой асимметричный характер. Именно такой вид имеет эмпирическое распределение, изображённое на рис. 2, д. Общее число видов в соответствующей пробе было, очевидно, столь мало, что даже спектр сравнительно небольшого объема ($n = 200$) оказался близким к «насыщению» и максимум функции $f(k)$ сместился в точку $k = 2$.

На основании общего характера поведения функции $f(k)$ можно предложить логарифмически нормальное распределение для ее аналитического описания

$$f(k) = \frac{T}{k\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln k - M)^2}{2\sigma^2}\right],$$

где M и σ – обычные параметры логнормального распределения. Для того чтобы выделить в величине M компоненту, не зависящую от n , можно положить $M = m + \ln n$, и тогда соответствующая формула примет вид

$$f(k) = \frac{T}{k\sigma\sqrt{2\pi}} \exp\left[-\frac{\left(\ln \frac{k}{n} - m\right)^2}{2\sigma^2}\right],$$

где m и σ уже такие параметры, значения которых специфичны для каждой пробы, но не зависят от объема спектра n . Можно привести и еще один аргумент в пользу такого описания, аналогичный тому, с помощью которого Б.В. Карасёв [1995] обосновывает фундаментальность логнормального распределения.

Чем определяется численность каждого вида в спектре? Очевидно, она зависит от множества независимых природных факторов: от численности материнского вида в исходном палеофитоценозе, от количества миоспор, которое каждое растение этого вида производило за один сезон, от способности миоспор данного вида к разносу, от устойчивости их оболочек к разложению, от скорости осадконакопления в точке отбора образца, от размеров самого образца и т. д. Очевидно, что каждый из этих факторов будет влиять на итоговую численность в виде некоторого *коэффициента*, то есть сама эта численность будет равна *произведению* большого числа параметров, каждый из которых обуславливается каким-либо природным фактором. Таким образом, численность вида в спектре может рассматриваться как произведение, а логарифм этой численности – соответственно как сумма большого числа независимых случайных величин. Поэтому логарифм численности будет иметь, скорее всего, нормальное распределение, а сама численность, соответственно, – логнормальное.

В эмпирически наблюдаемых распределениях мы, однако, не знаем, чему равняется $f(0)$.

Виды, для которых $Q_i < \frac{1}{2n}$, остаются как бы «за кадром». Такого рода распределения в математической статистике называются односторонне усеченными. В нашем случае за «точку усечения» естественно принять величину $\frac{1}{2}$ (сердину отрезка $[0; 1]$).

Предлагаемая математическая модель позволяет легко оценить количество этих «невидимых» видов. В самом деле, если $f(k)$ – логнормальное распределение с параметрами M и σ , усеченное в точке $\frac{1}{2}$, то величина $\ln k$ будет иметь нормальное распределение с теми же параметрами, усеченное в точке $\ln \frac{1}{2} = -\ln 2$. Для нормальных односторонне усеченных распределений существуют специальные таблицы (см., например [Янко, 1961]), по которым легко можно оценить их параметры, в том числе и особенно важную для нас «степень усечения», то есть долю «невидимых» видов, а следовательно, и общее число видов в пробе. Если степень усечения равна P , то, очевидно, $T = \frac{t}{1-P}$.

Практика – критерий истины

Можно сказать, что с помощью аппроксимации структуры разнообразия спорово-пыльцевых спектров логнормальным распределением мы научились определять их новую характеристику – величину «истинного» разнообразия спектра T . Знание этой величины представляется чрезвычайно важным, поскольку она отражает (хоть и с искажениями) флористическое разнообразие на территории, расположенной вокруг той точки, из которой была взята палинологическая проба. Это разнообразие, в свою очередь, определяется в значительной степени климатическими факторами. Таким образом, анализируя изменения величины T во времени, можно сделать определенные выводы о динамике климата на рассматриваемой территории.

С помощью описанного метода я оценил «истинное» разнообразие в 17 палинологических пробах, происходящих из сравнительно небольшого стратиграфического интервала (верхи казанского и татарский яруса пермской системы) и региона (междуречье рр. Сухоны и Вятки; рис. 2). Все проанализированные пробы были взяты из литологически достаточно однородной толщ, которая накапливалась в условиях аллювиально-озерной равнины и аридного или семиаридного климата [История..., 1987; Твердохлебов, 1996; Gomanov, 1992]. Таким образом, они могут считаться происходящими из одной ландшафтной зоны, а колебания их разнообразия – обусловленными преимущественно климатическими причинами и в первую очередь, вероятно, изменениями влажности. Оценки общего числа видов

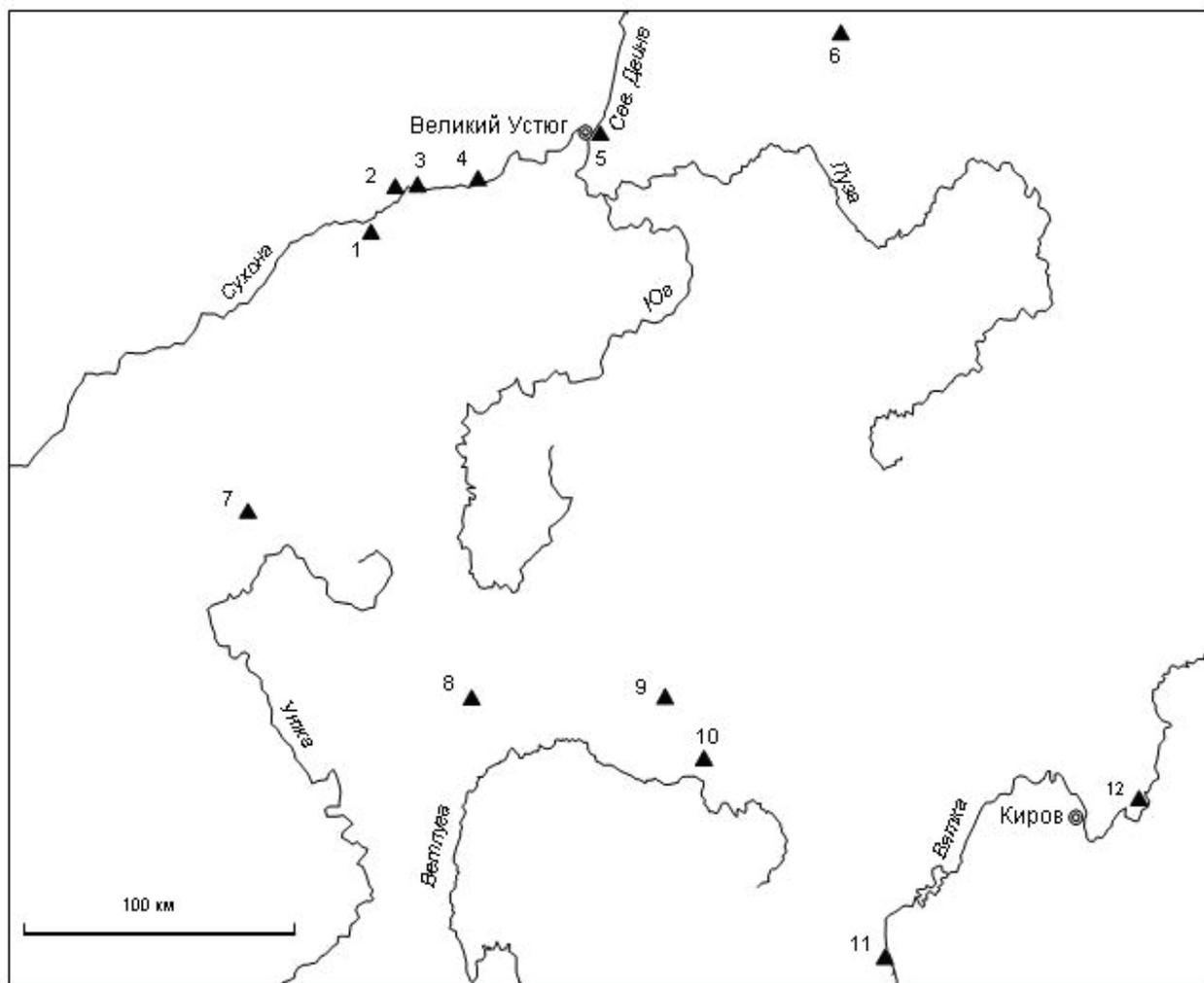


Рис. 2. Географическое положение проанализированных спорово-пыльцевых спектров: 1 – Бобровское; 2 – Вострое; 3 – Копылово; 4 – Исады; 5 – Аристово; 6 – Виледь; 7 – Титово; 8 – Калиновка; 9 – Дор; 10 – Луптюг; 11 – Чижы; 12 – Шихово-Чирки

в пробах приведены в таблице, а стратиграфическое положение соответствующих местонахождений – на рис. 3.

Можно заметить, что видовое разнообразие в палинологических пробах довольно закономерно изменяется по разрезу. Исключение составляет лишь одно местонахождение – Титово, где число видов существенно больше, чем в других пробах, имеющих примерно тот же возраст. Объяснение этого факта может заключаться в том, что миоспоры, определенные как однолучевые споры *Punctatisporites* sp. и доминирующие в данном спектре, на самом деле представляют собой цисты водорослей, которые, в отличие от миоспор, не приносились издалека в те водоемы, где формировались соответствующие осадочные породы, а «обитали» в самих этих водоемах. В результате механизм формирования численности доминирующего вида в данном спектре был существенно иным, чем во всех других, что прида-

ло структуре его разнообразия достаточно необычный облик и сказалось на оценке общего числа видов. Для большей наглядности общей картины динамики разнообразия в пермских спорово-пыльцевых спектрах местонахождение Титово не показано на рис. 3. Общая же закономерность в изменении разнообразия позволяет выделить внутри рассматриваемого временного интервала 5 этапов, которые интерпретируются как этапы относительного иссушения и увлажнения климата (рис. 3).

I этап, отвечающий концу казанского века, характеризуется весьма высоким палинологическим разнообразием ($T=272$) и, соответственно, очень влажным (по сравнению с последующими этапами) климатом. На II этапе, охватывающем раннетатарское время (уржумский горизонт), наблюдается иссушение климата и, соответственно, падение числа видов в пробах до значений, близких к 60. На III этапе, отвечающем са-

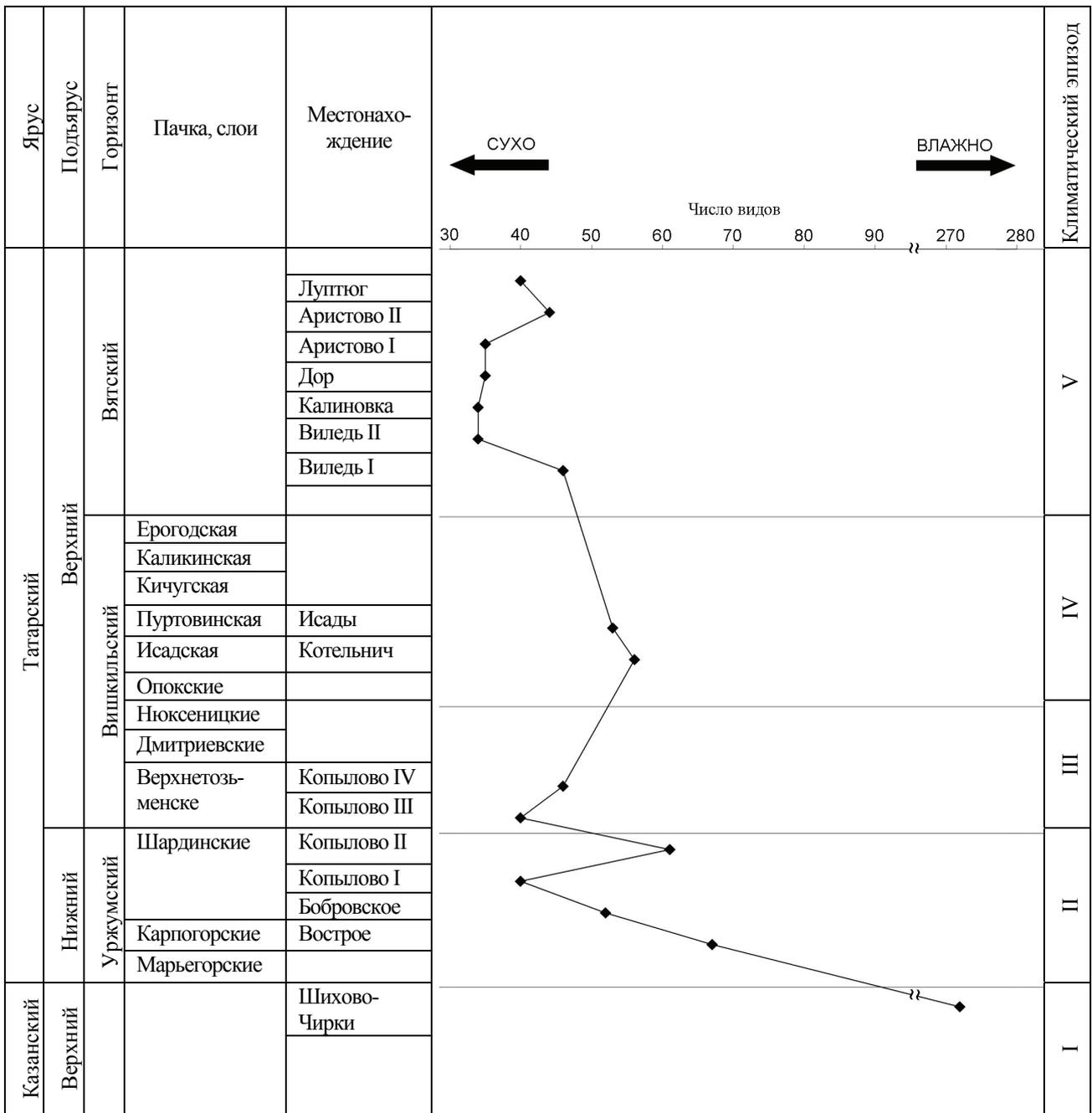


Рис. 3. Распределение «истинного» разнообразия миоспор в палинологических пробах из татарского и верхов казанского яруса

мым низам вишкильского горизонта (основание сухонской свиты), происходит дальнейшее иссушение климата и понижение разнообразия до 40–46 видов, хотя на последующем IV этапе, соответствующем верхней (и большей) части вишкильского горизонта, разнообразие восстанавливается примерно до уровня, характерного для II этапа ($T=53-56$). Наконец, V этап отвечает вятскому горизонту и характеризуется новым иссушением и уменьшением разнообразия в пробах до 40 видов и меньше.

Если научную деятельность представить себе как последовательность актов выбора, то «правильность» каждого такого выбора определяется, очевидно, богатством вытекающих из него методических и фактических следствий. В настоящей статье я стремился показать, что в ситуации выбора между логнормальным распределением и распределением Ципфа «правильный» выбор должен совершаться в пользу логнормального распределения.

Сведения о спорово-пыльцевых спектрах из верхов казанского и татарского яруса
Восточно-Европейской платформы

| № п/п | Местонахождения | Объем спектра (<i>n</i>) | Число обнаруженных видов (<i>t</i>) | Оценка «истинного» разнообразия (<i>T</i>) |
|-------|-----------------|----------------------------|---------------------------------------|--|
| 1 | Луптюг | 200 | 39 | 40 |
| 2 | Дор | 200 | 28 | 35 |
| 3 | Калиновка | 200 | 31 | 34 |
| 4 | Титово | 235 | 23 | 83 |
| 5 | Виледь II | 200 | 29 | 34 |
| 6 | Виледь I | 200 | 29 | 46 |
| 7 | Аристово II | 200 | 34 | 44 |
| 8 | Аристово I | 143 | 30 | 35 |
| 9 | Исады | 200 | 31 | 53 |
| 10 | Чижи | 186 | 40 | 56 |
| 11 | Копылово IV | 200 | 33 | 46 |
| 12 | Копылово III | 200 | 25 | 40 |
| 13 | Копылово II | 200 | 37 | 61 |
| 14 | Копылово I | 200 | 31 | 40 |
| 15 | Бобровское | 200 | 40 | 52 |
| 16 | Вострое | 200 | 55 | 67 |
| 17 | Шихово-Чирки | 200 | 30 | 272 |

Литература

- Бялко А.В.* Конструктивность закона конкуренции // Природа. – 1993. – №11. – С. 14–19.
- Бялко А.В.* Распределение коэффициентов // Природа. – 1995. – №11. – С. 51–57.
- Гоманьков А.В.* Об оценке разнообразия спорово-пыльцевых спектров // Палеофлористика фанерозоя. – 1990. – Вып. 2. – С. 16–25.
- История озер позднего палеозоя и раннего мезозоя. – Л.: Наука, 1987. – 280 с.
- Казютинский В.В., Балашов В.В.* Антропогенный принцип: история и современность // Природа. – 1989. – №1. – С. 23–32.
- Карасёв Б.В.* Логарифмически-нормальное распределение // Природа. – 1995. – №11. – С. 41–48.
- Твердохлебов В.П.* Континентальные аридные формации востока Европейской России на рубеже палеозоя и мезозоя. Диссертация в форме научного доклада на соискание ученой степени доктора геол.-мин. наук. – Саратов, 1996. – 57 с.
- Трубников Б.А.* Закон распределения конкурентов // Природа. – 1993. – №11. – С. 3–13.
- Трубников Б.А.* О законе распределения конкурентов // Природа. – 1995. – №11. – С. 48–50.
- Чайковский Ю.В.* Элементы эволюционной диатропики. – М.: Наука, 1990. – 271 с.
- Янко Я.* Математико-статистические таблицы. М.: Госстатиздат ЦСУ СССР, 1961. – 243 с.
- Gomankov A.V.* The relationship between micro- and megafossil assemblages: an environmental signal in the uppermost Permian of the Russian Platform // Organization Internationale de Paleobotanique, IV^{ème} Conference, Paris 1992. Resumes des communications OFP Informations N special 16-B. – Paris, 1992. – P. 68.
- Lambooy W., Lesnikovska A.* Some statistical methods useful in the analysis of plant paleoecological data // Palaios. – 1988. – Vol. 3. – No. 1. – P. 86–94.