АКАДЕМИЯ НАУК УКРАИНСКОЙ ССР
ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ
ИНСТИТУТ БИОЛОГИИ ЮЖНЫХ МОРЕЙ
им. А. О. КОВАЛЕВСКОГО

БИОЛОГИЯ МОРЯ

РЕСПУБЛИКАНСКИЙ
МЕЖВЕДОМСТВЕННЫЙ СБОРНИК

Основан в 1965 г.

Выпуск 40

ИССЛЕДОВАНИЕ МЕХАНИЗМОВ ФУНКЦИОНИРОВАНИЯ МОРСКИХ СООБЩЕСТВ И ЭКОСИСТЕМ С ПРИМЕНЕНИЕМ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

КИЕВ «НАУКОВА ДУМКА» 1977

УДК 62—506.22.001.57

Ю. В. Подвинцев, Ю. А. Горбенко

ФАКТОРНОЕ ПРЕДСТАВЛЕНИЕ МОРСКИХ МИКРОБИОЛОГИЧЕСКИХ СИСТЕМ

Проблема построения математических моделей морских микробиологических систем привлекает все возрастающее внимание исследователей. При общем подходе такие системы следует рассматривать со статистических позиций, предполагая случайный характер изменения внешних воздействий и свойств самих систем. В работе [1] была получена система регрессионных уравнений, описывающих статистическую взаимосвязь био- и абиотических параметров морской среды. Наиболее интересные результаты многомерного статистического анализа таких систем могут быть получены факторным методом [2, 3], являющимся одним из самых тонких и изящных приемов математической статистики.

Факторный метод предполагает, что поведение сложных систем определяется действием небольшого числа реально существующих или некоторых гипотетических факторов (например, рассмотрение морских экосистем связано с действием биотического, абиотического, физико-химического и других факторов). Влияние воздействия фактора на сложную систему может проявляться в изменении группы параметров. Поэтому фактор является более обобщенным понятием по отношению к наблюдаемым параметрам. Здесь и далее следует четко разделять понятия «параметр» и «фактор». Параметры наблюдаются и измеряются (например, O_2, pH, CO_3^{2-}, факторы непосредственно не наблюдаются и не измеряются — они появляются в процессе решения задач, а затем производится их интерпретация.

В данной работе используется приложение косоугольного факторного анализа в представлении параметров микробиологических систем. При этом множество измеренных параметров разбиваются на непересекающиеся подмножества, которые представляются коррелированными (косоугольными) факторами.

Постановка задачи факторного анализа. Пусть результаты N наблюдений за вектором состояния сложной системы $x^{т}=(x_1, x_2, \dots, x_n)$, где т — символ операции транспонирования, представлены матрицей наблюдений

$$X=\| x_{ij} \| \ (i=\overline{1,N};\ j=\overline{1,n})\ N>n. \qquad (1)$$

Здесь x_{ij} — численное значение наблюдаемого и измеряемого параметров системы x_j в i-м наблюдении.

Задача факторного анализа ставится следующим образом: для матрицы наблюдений X необходимо найти следующую линейную математическую модель

$$\overset{0}{x}=PF+DU, \qquad (2)$$

где $\overset{0}{x}$ — вектор-столбец нормированных и центрированных параметров сложной системы, компоненты которого определяются по формуле

$$\overset{0}{x}_i=\frac{x_i-\bar{x}_i}{\sigma_i} \qquad (3)$$

($\bar{x}_i$, σ_i — среднее значение и среднеквадратическое отклонение параметра x_i); $P=\|P_{ij}\|$ — матрица вкладов факторов размера $n\times m(n>m)$, m — число факторов; D — диагональная матрица ошибок факторного представления и измерений параметров,

$$D = \begin{Vmatrix} d_1 & & & & 0 \\ & d_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & d_n \end{Vmatrix}; \quad U = \begin{Vmatrix} U_1 \\ U_2 \\ \cdot \\ \cdot \\ \cdot \\ U_n \end{Vmatrix}; \quad F = \begin{Vmatrix} F_1 \\ F_2 \\ \cdot \\ \cdot \\ \cdot \\ F_m \end{Vmatrix}; \quad \overset{0}{x} = \begin{Vmatrix} \overset{0}{x}_1 \\ \overset{0}{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \overset{0}{x}_n \end{Vmatrix}; \tag{4}$$

F — вектор-столбец общих, а U — характерных факторов. В скалярной форме уравнение (2) можно записать в виде

$$x_i = p_{i1}F_1 + p_{i2}F_2 + \ldots + p_{im}F_m + d_iU_i \quad (i = \overline{1, n}). \tag{5}$$

Предполагаются следующие свойства факторов:

$$M\{F_i\} = 0; \quad M\{U_j\} = 0; \quad M\{U_kF_i\} = 0;$$
$$M\{U_kU_j\} = \begin{cases} 0 & k \neq j \\ 1 & k = j \end{cases}; \quad M\{F_iF_s\} = \begin{cases} 0 & i \neq s \\ 1 & i = s \end{cases} \tag{6}$$
$$(k, j = \overline{1, n}; \quad i, s = \overline{1, m}),$$

где M — символ операции математического ожидания. Следовательно, все факторы центрированы и имеют единичную дисперсию. Кроме того, отсутствует корреляция между общими и характерными факторами.

Основная задача факторного анализа сводится к нахождению матрицы факторных вкладов P и диагональной матрицы ошибок D, обеспечивающих статистически адекватное представление матрицы наблюдений (1).

Решение задачи факторного анализа. Используя математическое представление (2), с учетом (6) вычислим матрицу корреляций R параметров

$$R = M\{\overset{0}{x}\overset{0}{x}{}^{\text{т}}\} = M\{(PF + DU)(PF + DU)^{\text{т}}\} = PM\{FF^{\text{т}}\}P^{\text{т}} + DM\{UU^{\text{т}}\}D = P\Phi P^{\text{т}} + D^2. \tag{7}$$

Здесь Φ — матрица корреляций общих факторов.

Таким образом, задача факторного анализа сводится к поиску матриц P и D для заданных матриц R и Φ из уравнения

$$R - D^2 = P\Phi P^{\text{т}}. \tag{8}$$

Вычислим элемент d_i диагональной матрицы D. С этой целью уравнение (5) умножим слева и справа на характерный фактор U_i и, применив операцию математического ожидания, с учетом (6) получим

$$M\{\overset{0}{x}_iU_i\} = d_i. \tag{9}$$

Следовательно, величина d_i равна коэффициенту корреляции параметра x_i с характерным фактором U_i. Эта величина не связана с компонентами вектора F общих факторов. Матрица D^2, так же как и D, является диагональной с элементами d_i^2. Учитывая (9), имеем

$$0 \leqslant d_i^2 \leqslant 1 \quad (i = \overline{1, n}). \tag{10}$$

Поэтому d_i^2 — относительная дисперсия ошибки представления параметра x_i. На диагонали матрицы $R-D^2$ в выражении (8) стоят общности

$$h_i^2 = 1 - d_i^2, \tag{11}$$

представляющие собой относительную долю дисперсии параметра x_i, описываемого векторов общих факторов F. В качестве начальных оценок общностей $h^2{}_i$ (согласно [2]) выбирается квадрат коэффициента множественной корреляции величины x_i с остальными параметрами. Окончательно общности $h^2{}_i$ находятся в результате проведения итерационной процедуры, которая состоит из следующих шагов:

Шаг 1. Определяются начальные оценки общностей [2] по соотношению

$$h_i^2 = 1 - \frac{1}{r^{ii}}. \tag{12}$$

Здесь r^{ii} — диагональный элемент матрицы R^{-1}.

Шаг 2. Вычисление матрицы факторной структуры S размером $n \times m$.

Элемент матрицы S вычисляется по формуле

$$S_{jp} = \frac{\Sigma\,(r_{jk};\ k \in G_p)}{\sqrt{W_{pp}}},$$

где G_p — группа параметров, принадлежащих фактору F_p:

$$W_{pq} = \sum_j \sum_k (r_{jk};\ k \in G_p;\ j \in G_p),\ \ p, q = \overline{1, m}.$$

Шаг 3. Вычисление матрицы Φ корреляций факторов:

$$\Phi = \{r_{F_pF_q}\};\quad r_{F_pF_q} = \frac{W_{pq}}{\sqrt{W_{pp}W_{qq}}}.$$

Шаг 4. Вычисление матрицы факторного отображения P:

$$P = S\Phi^{-1}.$$

Шаг 5. Вычисление матрицы вычисленных корреляций по соотношению (8):

$$R^* = P\Phi P^{\mathrm{т}}.$$

На диагонали матрицы R^* находятся уточненные значения общностей. Это дает возможность организовать итерационную процедуру, используя уточненные общности и переходя ко второму шагу. Остановка производится на основании перечисленных ниже оценок.

Статистическое оценивание факторного решения. Предположим, что решена задача факторного анализа и в выражении (2) найдены матрицы P и D.

Достаточность m факторов проверяется по критерию Андерсона [4]

$$V_m = (N-1)\ln\frac{|P\Phi P^{\mathrm{т}} + D^2|}{|R|};\quad \nu = \frac{1}{2}\left[(n-m)^2 + (n-m)\right]. \tag{13}$$

Величина V_m распределена по закону χ^2 с числом степеней свободы ν.

Критерий Томпсона позволяет оценить близости матриц R и R^* по элементам матрицы остатков $\Delta r_{ij}=r_{ij}-r_{ij}^*$.
Для этой цели вычисляется среднеквадратическое отклонение внедиагональных элементов Δr_{ij}. Затем проверяется гипотеза о статистической независимости элементов Δr_{ij}. При

$$\sigma_r < \frac{\sqrt{2}}{\sqrt{n^2-n}} \tag{14}$$

остаточные корреляции считаются несущественными.

Факторное решение адекватно, если выполняются одновременно критерии Андерсона и Томпсона.

Пример факторного представления множества взаимосвязанных параметров в море. В качестве исходных данных (1) были взяты $N=20$ наблюдений в 1969 г. за $n=17$ параметрами в море. Наблюдения проводились на погруженных пластинах и в окружающей морской воде в течение года по плану эксперимента [5] с одним наблюдением в 5, 10 суток. Описание параметров: x_1 — палочковидные формы бактерий (ФБ); x_2 — кокковидные ФБ; x_3 — диатомовые водоросли (живые); x_4 — диатомовые (мертвые); x_5 — гетеротрофные ФБ; x_6 — растворенное органическое вещество (РОВ), трансформированное микроорганизмами; x_7 — pH метаболитов микроорганизмов обрастаний; x_8 — сухая масса слизистой пленки; x_9 — карбонаты, осаждаемые микроорганизмами; x_{10} — концентрация РОВ; морской воды: x_{11}—$t°$ С; x_{12}—pH; x_{13}—CO_3^{2-}; x_{14}—HCO_3^-; x_{15}—CO_2; x_{16}— соленость; x_{17} —содержание O_2.

В данном примере проверялась следующая нулевая гипотеза—механизм взаимодействия параметров является функцией четырех факторов: $F_1(x_1x_2x_3x_4x_5)$ — биотического фактора сообщества перифитонных микроорганизмов (СПМ), развивающихся на поверхности погруженных в море предметов; $F_2(x_6x_7x_8x_9)$— биокосного фактора метаболизма СПМ; $F_3(x_{10}x_{11}x_{12})$—фактор состояния морской воды; $F_4(x_{13}x_{14}x_{15}x_{16}x_{17})$ —фактора карбонатного равновесия, солености и кислорода. Последних два фактора абиотические.

В табл. 1 представлено решение (2) и значения статистических критериев Андерсона (13) и Томпсона (14). Нетрудно убедиться, что после умножения численных значений табл. 1 на общие и характерные факторы и сложения получим решение в виде (5).

Наибольший вклад в параметры x_7 и x_8 оказывает фактор $F_1(x_1x_2x_3x_4x_5)$, а в параметры x_{13} и x_{15} — фактор $F_3(x_{10}x_{11}x_{12}$ (табл. 1). Поэтому необходимо перенести указанные параметры в соответствующие факторы и свести задачу к проверке следующей гипотезы о факторном представлении: $F_1(x_1x_2x_3x_4x_5x_6x_7x_8$, $F_2(x_6x_9)$, $F_3(x_{10}x_{11}x_{12}x_{13}x_{15})$, $F_4(x_{14}x_{16}x_{17})$. Решение данной задачи приведено в табл. 2, из которой видно, что параметры x_6 и x_9 необходимо перевести в первую группу, x_7 — в третью и x_{10} — в четвертую.

После этого следует перейти к проверке гипотезы о существовании следующих факторов: $F_1(x_1x_2x_3x_4x_5x_8x_9x_6)$, $F_2(x_7x_{11}x_{12}x_{13}x_{15})$, $F_3(x_{10}x_{14}x_{16}x_{17})$.

Решение задачи представлено в табл. 3. Это окончательное решение, так как вклады факторов в параметры своих групп максимальны для всех параметров. Существенно повысилась адекватность решения в табл. 3 по сравнению с нулевой гипотезой в табл. 1. Окончательное решение находится на критическом уровне адекватности по критерию Томпсона и с вероятностью $P=0{,}98$ достаточно трех факторов для представления всех 17 параметров.

Анализ решения приводит к следующим выводам:

1. Не подтверждается нулевая гипотеза о существовании четырех факторов в представлении механизма взаимодействия параметров в море.

2. Установлено, что механизм взаимодействия этих параметров есть

функция трех факторов: $F_1(x_1x_2x_3x_4x_5x_8x_9x_6)$ — биотического фактора СПМ, $F_2(x_7x_{11}x_{12}x_{13}x_{15})$ — фактора карбонатного равновесия и температуры (условное название), а $F_3(x_{10}x_{14}x_{16}x_{17})$ —фактора состояния воды. Второй и третий факторы абиотические.

Таблица 1

Факторная модель

x_i	F_1	F_2	F_3	F_4	d_iU_i
$x_1=$	0,98	—0,25	0,06	—0,1	0,41 U_1
$x_2=$	0,96	—0,35	0,28	0,22	0,61 U_2
$x_3=$	0,52	0,21	—0,34	—0,23	0,73 U_3
$x_4=$	0,63	0,22	0,11	0,15	0,55 U_4
$x_5=$	0,64	0,05	—0,12	—0,04	0,74 U_5
$x_6=$	—0,45	0,99	—0,29	—0,24	0,67 U_6
$x_7=$	—0,55	0,26	0,32	0,38	0,81 U_7
$x_8=$	0,54	0,36	0,07	—0,12	0,48 U_8
$x_9=$	0,48	0,59	—0,1	—0,03	0,42 U_9
$x_{10}=$	0,19	—0,34	0,72	—0,28	0,52 U_{10}
$x_{11}=$	—0,03	0,01	0,82	—0,12	0,43 U_{11}
$x_{12}=$	—0,16	0,32	0,93	0,42	0,45 U_{12}
$x_{13}=$	—0,09	0,1	1,17	0,54	0,26 U_{13}
$x_{14}=$	0,06	—0,39	0,08	0,8	0,52 U_{14}
$x_{15}=$	0,09	—0,13	—0,92	—0,07	0,36 U_{15}
$x_{16}=$	0,2	0,03	0,43	—0,53	0,7 U_{16}
$x_{17}=$	—0,25	0,4	—0,03	0,46	0,79 U_{17}

Примечание. $\sigma_r=0{,}108$, $V_4=44$, $\gamma=91$.

Таблица 2

Факторная модель

x_i	F_1	F_2	F_3	F_4	d_iU_i
$x_1=$	0,91	—0,01	—0,27	—0,26	0,51 U_1
$x_2=$	0,74	—0,01	—0,08	—0,02	0,69 U_2
$x_3=$	0,84	—0,17	—0,23	—0,16	0,68 U_3
$x_4=$	0,67	0,19	0,09	0,12	0,57 U_4
$x_5=$	0,64	0,06	—0,23	—0,08	0,78 U_5
$x_7=$	—0,07	—0,4	0,72	0,56	0,84 U_7
$x_8=$	0,51	0,39	0,02	0,15	0,47 U_8
$x_6=$	—0,47	0,36	0,05	0,02	0,59 U_6
$x_9=$	0,54	0,45	—0,04	—0,02	0,44 U_9
$x_{10}=$	—0,03	0,04	0,37	—0,55	0,52 U_{10}
$x_{11}=$	0,07	—0,09	0,74	—0,26	0,39 U_{11}
$x_{12}=$	0,06	0,02	0,99	0,45	0,53 U_{12}
$x_{13}=$	0,02	0,05	1,12	0,43	0,25 U_{13}
$x_{15}=$	—0,12	0,07	0,91	—0,06	0,47 U_{15}
$x_{14}=$	—0,04	—0,27	0,07	0,74	0,61 U_{14}
$x_{16}=$	0,23	—0,08	—0,24	0,59	0,63 U_{16}
$x_{17}=$	—0,25	0,35	0,17	0,69	0,65 U_{17}

Примечание. $\sigma_r=0{,}109$, $V_4=40{,}8$, $\gamma—91$.

3. В состав биотического фактора F_1 (табл. 3) вошли параметры x_8 и x_9 (x_8 — масса слизистой пленки СПМ по сухой массе, а x_9 — количество карбонатов, осаждающихся в микроперифитоне микроорганизмами СПМ); x_8 и x_9 ранее были отнесены к фактору метаболизма СПМ $F_2(x_6x_7x_8x_9)$.

В действительности оба указанных параметра тесно связаны с жизнедеятельностью микроорганизмов СПМ, что было установлено ранее [5, 6]. Поэтому такой перенос параметров, по-видимому, закономерен.

Таблица 3

Факторная модель

x_i	F_1	F_2	F_3	d_iU_i
$x_1=$	0,82	—0,23	—0,04	0,54 U_1
$x_2=$	0,75	—0,14	0,08	0,66 U_2
$x_3=$	0,55	0,08	—0,15	0,81 U_3
$x_4=$	0,83	0,11	0,09	0,55 U_4
$x_5=$	0,67	—0,22	0	0,54 U_5
$x_8=$	0,89	0,03	—0,1	0,41 U_8
$x_9=$	0,91	0,05	0,04	0,42 U_9
$x_6=$	0,36	0,29	0,06	0,89 U_6
$x_7=$	—0,42	0,74	0,19	0,57 U_7
$x_{11}=$	0,18	0,59	—0,49	0,45 U_{11}
$x_{12}=$	0,26	0,86	0,07	0,42 U_{12}
$x_{13}=$	0,21	0,92	0	0,3 U_{13}
$x_{15}=$	—0,25	—0,76	0,23	0,44 U_{15}
$x_{10}=$	0,22	0,12	—0,42	0,85 U_{10}
$x_{14}=$	—0,32	—0,05	0,58	0,71 U_{14}
$x_{16}=$	0,08	—0,23	0,63	0,72 U_{16}
$x_{17}=$	0	0,15	0,57	0,82 U_{17}

Примечание. $\sigma_r=0{,}097$, $V_з=37$, $\gamma=105$.

4. Параметр x_6 (РОВ, трансформированное СПМ) тоже вошел в биотический фактор, однако его большая ошибка $d_6=0{,}89$ (табл. 3) объясняется тем, что фотометрический метод, которым его измеряли [5], по-видимому, не достаточно точен. Указанная ошибка относится и к параметру x_{10} в факторе $F_3(x_{10}x_{14}x_{16}x_{17})$, определяемому тем же фотометрическим методом.

5. Параметр x_7 (pH метаболитов СПМ), хотя и обусловливается метаболическими процессами в СПМ, но вместе с тем он вошел в абиотический фактор карбонатного равновесия $F_2(x_7x_{11}x_{12}x_{13}x_{15})$. Этот параметр, очевидно, теснее взаимодействует с карбонатным равновесием в море, чем с биотическим фактором СПМ.

6. Факторное представление параметров в микробиологических процессах в море дает возможность провести глубокий статистический анализ различных явлений в морской среде.

ЛИТЕРАТУРА

1. *Горбенко Ю. А., Подвинцев Ю. В.* Растворенное органическое вещество и pH как статистические функции био- и абиотических параметров Черного моря.— Тезисы докл. междунар. симпоз. «Взаимодействие между водой и живым веществом». Сессия 1. Одесса, 1975, с. 7.

2. *Харман Г.* Современный факторный анализ. М., «Статистика», 1972. 486 с.

3. *Подвинцев Ю. В.* Факторный метод представления биохимических и физических полей океанов.— Тезисы докл. 5-го всесоюз. симпоз. «Методы представления и аппаратурный анализ случайных процессов и полей». Секция 3. Л., 1975, с. 91—94.

4. *Anderson T. W., Rubin.* Statistical inference in factor analysis.— Proc. of the Third Berkeley Sump. on Math. Statistics and Probability, 1956, N 5, p. 111—150.

5. *Горбенко Ю. А.* Биологические основы борьбы с обрастанием. Киев, «Наукова думка», 1973, с. 27—45.

6. *Горбенко Ю. А.* Экология перифитонных микроорганизмов. Автореф. докт. дис. ИнБЮМ, Севастополь, 1973.

Севастопольский приборостроительный институт,
Институт биологии южных морей
АН УССР

Поступила в редколлегию
1.XII 1975 г.