УДК 556.16

Н.С. Лобода, *проф.,* **Нгуен Ву Ань**, *асп.*
Одесский государственный экологический университет

РАЙОНИРОВАНИЕ БАССЕЙНА Р.УССУРИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА

Выполнено районирование водосбора реки Уссури по синхронности колебаний годового стока с использованием методов многомерного статистического анализа – факторного и главных компонент.
Ключевые слова*: районирование, факторы, компоненты, синхронность*

Вступление. Бассейн р.Уссури характеризуется сложными условиями формирования стока. Климатические факторы обуславливают общее уменьшение осадков и усиление континентальности климата по мере удаления от побережья Японского моря внутрь континента, т.е. при движении в направлении с востока на запад. В то же время наибольший поверхностный сток рек Приморья формируется не на побережьи, а в высокогорных районах Сихотэ-Алиня [1]. Взаимодействие климатических факторов и рельефа создает сложный рисунок распределения годового стока в пространстве и времени. По характеру распределения норм годовых осадков с высотой в пределах только бассейна р.Уссури выделено 12 районов, а на основе гидрологического подхода (с использованием гидрологических и ландшафтно-морфологических характеристик) – 3 района [8]. При столь сложных условиях формирования годового стока особую значимость приобретает вопрос оценки синхронности колебаний годового стока при решении вопроса выбора опорных гидрологических постов для приведения коротких рядов стока к длинному периоду.

Анализ последних достижений и публикаций**.** Исследование циклических колебаний выполненное П.С. Кузиным и В.И. Бабкиным [5] по 30-ти летним рядам годового стока крупных рек бывшего СССР, позволило установить, что в пределах Приморья может быть выделен один район с синфазными колебаниями. Районирование территории Приморья по синхронности колебаний годового стока было выполнено М.Г. Васьковским [1] на основе анализа разностных интегральных кривых. Согласно этому районированию выделено три района. Первый из них включает в себя водосбор р.Уссури – пос. Кировский, второй - охватывает водосборы рек Иман и Бикин, третий включает в себя бассейн р.Хор. Это районирование выполнено по данным наблюдений за стоком рек до 70-х годов прошедшего столетия и не имеет строгого научного обоснования.

Целью статьи является научное обоснование районирования бассейна р. Уссури по синхронности колебаний годового стока с использованием методов многомерного статистического анализа.

Материалы и методы исследования. Синхронность колебаний годового стока исследована путем разложения полей стока по естественным ортогональным функциям и Q-модификации факторного анализа [4].

Основные положения факторного анализа базируются на гипотезе о том, что данные наблюдений являются лишь косвенными характеристиками изучаемого явления, которое можно описать при помощи небольшого числа неких параметров или свойств, называемых факторами [2,3,7]. Основная модель факторного анализа может быть представлена в виде:

$$\varphi_j = \sum_{p=1}^{k} l_{jp} f_p + \nu_j \,, \tag{1}$$

где φ_j - нормированная исходная переменная;

k - число факторов;

p - номер фактора;

l_{jp} - факторная нагрузка (коэффициент факторного отображения);

$f_1, f_2, ..., f_k$ - некоррелируемые между собой факторы;

ν_j - независимые остатки.

Факторы учитывают корреляцию между переменными, т.е. они представляют структуру корреляционной матрицы в терминах модели. Между собой факторы некоррелированны т.е. ортогональны. Остатки – случайные величины, не связанные ни между собой, ни с факторами. Элементы корреляционной матрицы, факторные нагрузки и дисперсии остатков связаны соотношениями

$$r_{jj} = \sum_{p=1}^{k} l_{jp}^2 + \nu_j \,, \tag{2}$$

$$r_{ji} = \sum_{p=1}^{k} l_{jp} l_{ip} \quad \text{при} \quad j \neq i \,, \tag{3}$$

где r_{ij} - элементы корреляционной матрицы.

Квадраты факторных нагрузок представляют доли дисперсии рядов наблюдений, приходящиеся на соответствующие факторы. Сумма квадратов факторных нагрузок по всем выделенным факторам может быть рассчитана следующим образом

$$h_j^2 = \sum_{p=1}^{k} l_{jp}^2 \tag{4}$$

Полученная величина определяет полноту отражения j-того ряда наблюдений в факторах f_p. Факторная дисперсия h_j^2 связана с остатками соотношением

$$h_j^2 = 1 - e_j^2 \,, \tag{5}$$

где e_j^2 - дисперсия остатков.

Полный вклад S_p (в процентах) фактора f_p в суммарную дисперсию нормированных рядов наблюдений (полей) определяется выражением

$$S_p = \frac{\sum_{j=1}^{m} l_{pj}^2}{m} 100\% \,, \tag{6}$$

где m - число рассматриваемых рядов (объектов).
Общий вклад всех выделенных факторов в суммарную дисперсию исследуемых рядов равен

$$S = \sum_{p=1}^{k} S_p \qquad (7)$$

Когда рассматриваются связи не между признаками, а между рядами, используется Q-техника факторного анализа, которую можно идентифицировать как вариант кластерного анализа [9,10]. При исследовании синхронности колебаний годового стока в практике гидрологических расчетов [2] применяют следующие графические построения. В случае, если первые два фактора описывают более 60 % общей дисперсии исходных данных, на графике, оси которого представляют два фактора, проводят векторы из начала координат в точку с координатами, соответствующими факторным нагрузкам. Длина вектора рассчитывается по выражению

$$d_j = \sqrt{l_{j1}^2 + l_{j2}^2}\,, \qquad (8)$$

где l_{j1} и l_{j2} - факторные нагрузки на первый и второй факторы.

Величина d определяет полноту отражения j-го ряда наблюдений в первых двух факторах, а косинус угла между j-тым и i-тым векторами есть коэффициент корреляции между ними. Таким образом, о степени связи между рядами можно судить по формирующимся группировкам точек на плоскости. В качестве меры сходства в данном случае используется мера расстояния: чем ближе расположены точки на графике и меньше косинус угла между ними, тем ближе значение коэффициента корреляции к 1. Применение Q- модификации факторного анализа позволяет "сжать" информацию, содержащуюся в корреляционной матрице, и интерпретировать её.

Метод главных компонент также, как и метод факторного анализа позволяет "сжимать" исходную информацию и анализировать ее. Однако, весовые коэффициенты разложения (факторные нагрузки) не несут смысловой информации. В методе главных компонент разложение полей по естественным ортогональным функциям приводит к системе ортонормированных функций, значение которых зависит от статистических свойств исследуемого поля, откуда и происходит их название – "естественные ортогональные функции".

Поиск собственных векторов и собственных значений достигается путем решения матричного уравнения вида

$$R_X - \lambda_i U_i = 0\,, \qquad (9)$$

где R_X - матрица коэффициентов корреляции размером $m \times m$ (m соответствует числу рассматриваемых объектов);

U_i - собственный вектор матрицы корреляций;

λ_i - соответствующее собственному вектору собственное значение.

Матрица R_X имеет m корней или m собственных чисел λ, которые являются действительными, положительными и простыми. Для нахождения m собственных векторов, соответствующих m собственным числам, необходимо решение m систем линейных уравнений. Процедура расчета осуществляется, как правило, при помощи итерационных методов, среди которых наиболее распространённым является метод Якоби [154,268].

Совокупность собственных векторов образует базис, в котором производится разложение полей исходных данных

$$U' \cdot \varphi_i = Z_i, \tag{10}$$

где U' - транспонированная матрица U размером $m \times m$;

φ_i - i-тый случайный вектор (поле) центрированных и нормированных исходных данных;

Z_i - вектор главных компонент, являющийся результатом линейного преобразования φ_i – того поля соответствующим собственным вектором.

Поскольку собственные векторы ортонормированны, главные компоненты поля являются статистически независимыми. Равенство (10) означает, что исходное поле разложено на m независимых компонент.

Составляющие вектора Z_i для p - той компоненты разложения определяются следующим образом

$$z_{ip} = \sum_{k=1}^{m} U_{pk} \varphi_{ik} \; ; \; p = \overline{1,m} \;, \tag{11}$$

где z_{ip} - составляющие p - той компоненты разложения;

U_{pk} - весовые коэффициенты, отражающие вклад k - того объекта в каждую p - тую компоненту (или вклад p-той компоненты в k-ый объект), которые являются составляющими собственных векторов матрицы корреляций;

φ_{ik} - i-тый случайный вектор (поле) центрированных и нормированных исходных данных.

Значения U_{pk} изменяются в пространстве при переходе от объекта к объекту, но не зависят от времени. Система функций U_{pk} часто представляется как функция координат (x_k, y_k) для k - того объекта и носит название базисной функции.

Если разложению в базисе собственных векторов подвергнуть не корреляционную, а ковариационную матрицу, то можно показать, что сумма дисперсий главных компонент будет равна сумме дисперсий исходных рядов, т.е.

$$\sum_{p=1}^{m} \sigma_{Zp}^2 = \sum_{j=1}^{m} \sigma_{Xj}^2 \tag{12}$$

Такое представление позволяет более наглядно понять суть метода главных компонент, так как эти некоррелированные линейные комбинации исходных переменных отражают в себе всю дисперсию, заключённую в m переменных исходного массива данных. Несколько первых собственных чисел ($\lambda_1 > \lambda_2 > \lambda_3 > ... > \lambda_m$) исчерпывают основную часть суммарной дисперсии поля, поэтому при анализе результатов разложения особое внимание уделяется первым собственным значениям и соответствующих им компонентам. А так как крупномасштабные процессы характеризуются большей дисперсией, то справедливо допущение, что именно они отражены в первых компонентах.

При использовании корреляционной матрицы сумма собственных чисел равна числу рассматриваемых переменных m, поэтому разделив каждое собственное число на m или $\sum_{s=1}^{m}\lambda_s$, можно получить долю от суммарной дисперсии, отвечающую каждой $k-той$ компоненте

$$S_k = \frac{\lambda_k}{\sum_{s=1}^{m}\lambda_s} = \frac{\lambda_k}{m}\ , \qquad (13)$$

Доля информации, заключенной в первых p компонентах по отношению ко всей суммарной информации о поле, оценивается с помощью соотношения

$$S = \frac{\sum_{k=1}^{p}\lambda_k}{\sum_{s=1}^{m}\lambda_s}\ , \qquad (14)$$

где числитель равен сумме дисперсий, приходящихся на p первых главных компонент, а знаменатель равен суммарной дисперсии поля.

Задавая необходимый уровень существенной информации, представленный значением S (например, $S=$0,70-0,80), можно установить число первых компонент, которые следует учитывать, чтобы сократить объем анализируемой информации и сохранить при этом ее основное содержание.

Для анализа синхронности колебаний стока были привлечены данные по 20 рядам годового стока за период совместных наблюдений с 1960 по 1986 гг, т.е. продолжительность наблюдений составляет 27 лет.

Результаты исследования и их анализ. В результате применения Q-модификации факторного анализа к массиву исходных данных установлено, что первый фактор описывает 66% исходной информации по годовому стоку, а второй – 20%. Таким образом, на первые два фактора приходится 86% исходной информации и это позволяет сделать вывод о возможности использования только первых двух факторов для описания закономерностей многолетних колебаний годового стока. Распределение точек на графике показывает, что в пределах бассейна р. Уссури можно выделить две группировки точек, которые можно рассматривать как два района с синхронными колебаниями стока. К первому относится северная часть водосбора р.Уссури: рр. Чирка, Хор, Подхоренок, верховья р.Тайдзибе, ко второму – южная часть бассейна. Водосборы р.Иман (под номерами 83,85,89,93,98), р.Танга (48), р.Шетуха (45), р.Тамга(48) занимают промежуточное положение между двумя районами. Значения характеристики d изменяются от 0,99 до 0,86. Полученные результаты соответствуют данным, полученным М.Г. Васьковским [1].

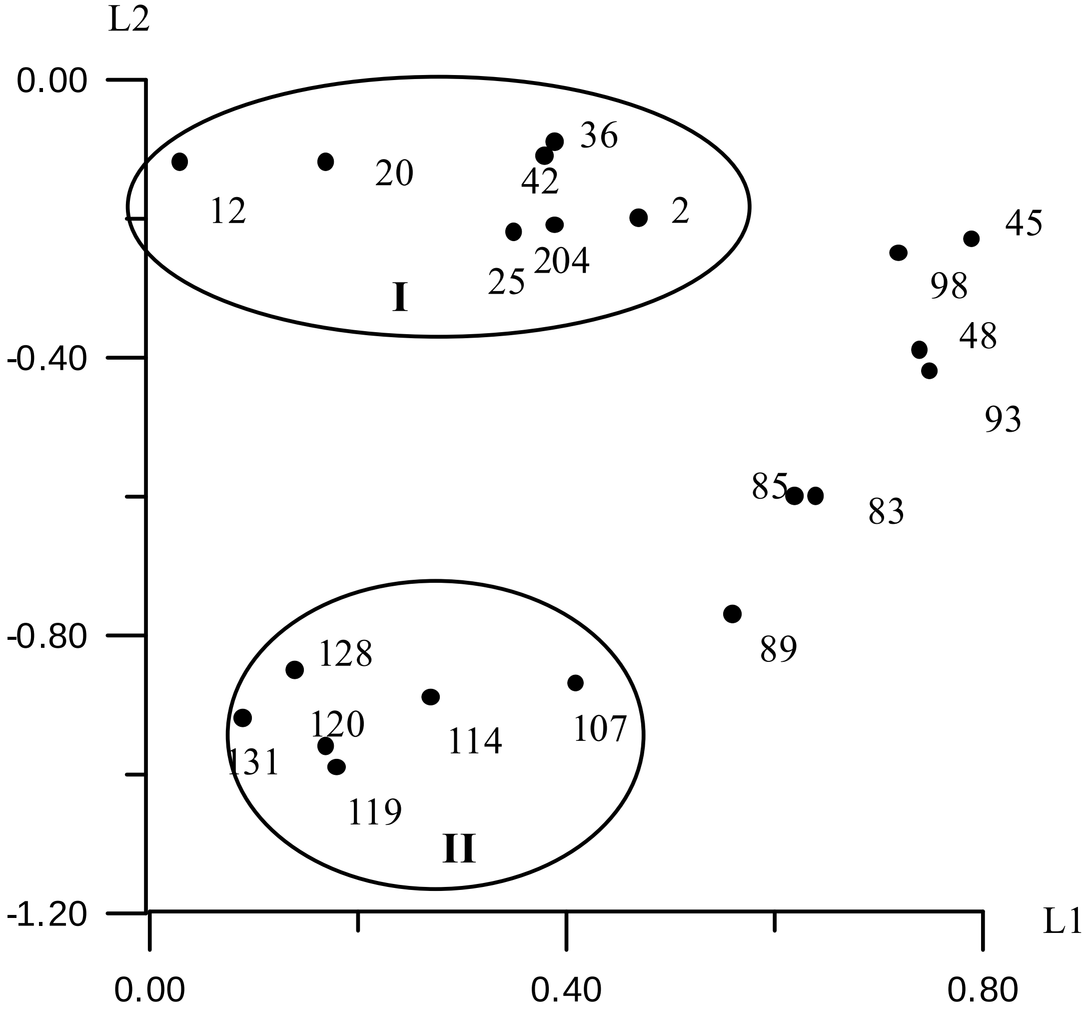

Рис. 1 – Выделение районов с синхронными колебаниями годового стока в бассейне р.Уссури на основе группировок точек

Определенную проблему представляет проведение границы районов с синхронными колебаниями стока. Для этой цели можно использовать метод главных компонент, в котором в качестве границы районов рекомендуется использовать нулевую изолинию для весовых нагрузок [6]. Вклад первой компоненты, соответствующей наиболее крупномасштабным процессам, в формирование полей годового стока составил 63% при разложении полей годового стока по естественным ортогональным функциям матрицы корреляций. Вклад второй компоненты равен 20%, а третьей – 5%. Таким образом, суммарный вклад первых двух компонент равен 83%, а первых трех – 88%, что является достаточным для анализа матрицы исходных данных.

Весовые нагрузки на первую компоненту разложения имеют положительные значения (табл.1), что интерпретируется как однонаправленное воздействие наиболее крупномасштабного физического процесса на формирование поля годового стока.

Таблица 1 – Значения первых базисных функций

Номер	Река –пост	U_1	U_2	U_3
2	р.Уссури-пос.Кировский	0,2582	-0,0745	0,0398
12	р.Улахэ-с.Березняки	0,2587	-0,0645	0,4140
20	р.Сыдагоу-с.Извилинка	0,2208	-0,22371	0,3047
25	р.Фудзин-с.Уборка	0,2566	0,0402	0,1200
36	р.Даубихэ-с.Яковлевка	0,2266	0,0836	-0,0678
42	р.Хонихеза-с.Варфоломеевка	0,2465	0,0027	0,0768
45	р.Шетуха-с.Крыловка	0,2359	0.1001	-0,3672
48	р.Тамга-с.Тамга	0,2446	0.2184	-0,3074
83	р.Иман-с.Картун	0,2317	0,2030	-0,1489
85	р. Иман-пос. Вагутон	0,2406	0,2193	-0,1916
89	р.Сибичи-с. Сибичи	0,2078	0,2933	-0,1818
93	р.Вака-с.Ракитное	0,1819	0.2592	-0,2741
98	р.Тудо-Вака-с.Ариадное	0,2067	0.2684	-0,3054
107	р.Бикин-с.Звеньевая	0,2290	-0,2318	0,0145
114	р.Горбун-с.Пушкино	0,1943	-0.3282	0,1602
119	р.Подхоренок-с.Дормидонтовка	0,1914	-0.3268	0,1762
120	р.Правый Подхоренок – лзу. Медвежий Ключ	0,2159	0,2414	0,1797
128	р.Сукпай-мет.ст.Сукпай	0,2337	-0,2414	0,2627
131	р.Кия-с.Марусино	0,2038	-0.2188	0,1932
204	р.Каменка-пос. Каменский	0,1548	-0,3472	0,1436

Знак второй компоненты меняется, что рассматривается как разнонаправленное воздействие второго по значимости физического процесса на формирование годового стока. По положению нулевой изолинии можно судить о принадлежности водосборов к первому или второму району. Нулевая изолиния делит водосбор р. Уссури на две части: северную и южную, медленно спускаясь к юго-востоку (рис.2).

При рассмотрении пространственного распределения весовых коэффициентов третьей компоненты разложения выделяется центральный район (под номерами 83,85,89,93, 98,48,45), в пределах которого размещены водосборы, занимающие промежуточное положение между группировками 1 и 2, показанными на рис.1.

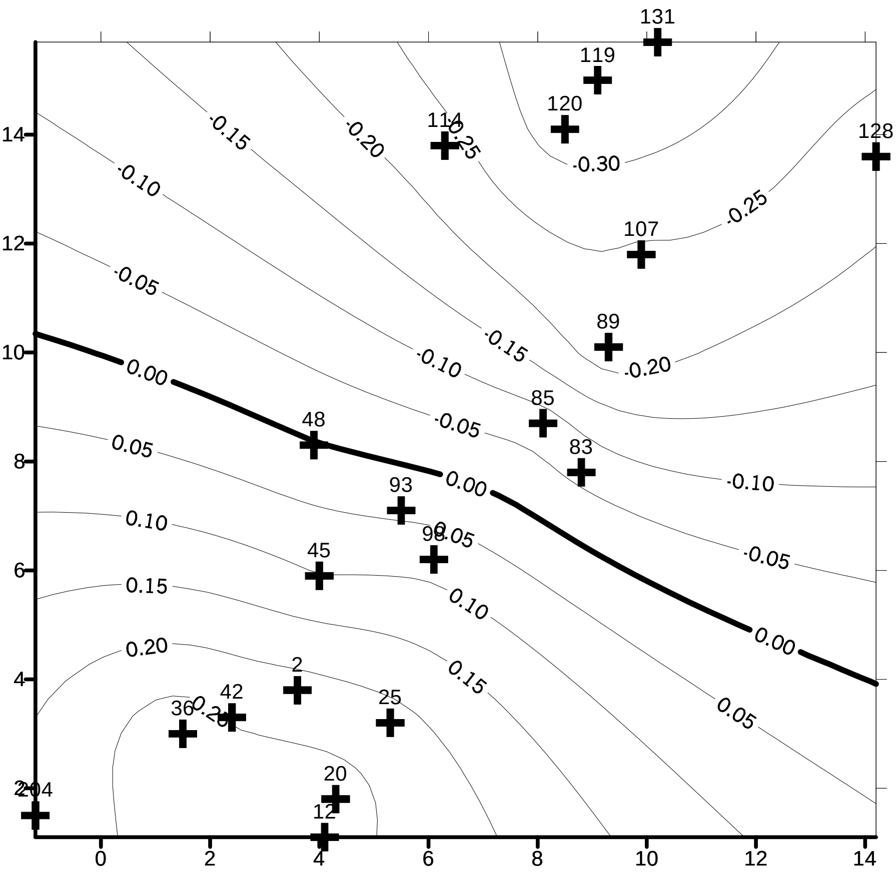

Рис. 2 – Карта-схема изолиний второй базисной функции U_2; у точек (+), соответствующих центрам тяжести водосборов, расположены номера водосборов

Выводы. На основе методов многомерного статистического анализа (факторного и главных компонент) установлено, что в результате воздействия крупномасшабных физических процессов на условия формирования годового стока, в пределах бассейна р.Уссури образовалось два района с синхронными колебаниями годового стока: северный и южный. Подсчет осредненных в пределах выделенных районов коэффициентов корреляции между рядами показал, что средний по бассейну р. Уссури коэффициент корреляции равен 0,52, в то время как осредненные по районам I и II коэффициенты корреляции равны соответственно 0,87 и 0,77. Водосборы рек Иман, Танга и Шетуха образуют отдельную группировку, которая выделяется при анализе третьей компоненты разложения, отражающей влияние процессов более мелкого масштаба.

Список литературы

1. *Васьковский М.Г.* Многолетние колебания годового стока рек Приморья. // Труды Дальневосточного научно-исследовательского гидрометеорологического института. – Л.: Гидрометеоиздат. – 1968. – Вып.27. – С. 64 – 75.
2. *Жук В.А., Евстигнеев В.М.* Исследование синхронности колебаний годового стока отдельных регионов приемами факторного анализа: Труды ВНИИГМИ-МЦД. - М.: Гидрометеоиздат. С. 78-91с.
3. *Иберла К.* Факторный анализ: Пер. с англ. - М.: Статистика,1980. - 397 с.
4. *Исследования и расчеты речного стока* / Под ред. В.Д.Быкова. - М.: Изд-во МГУ, 1981. - 228 с.
5. *Кузин П.С., Бабкин В.И.* Географические закономерности гидрологического режима рек. - Л.: Гидрометеоиздат, 1979. - 200 с.
6. *Лобода Н.С.* Расчеты и обобщения характеристик годового стока рек Украины в условиях антропогенного влияния. - Одесса: Экология, 2005. – 208 с.
7. *Лоули Д., Максвелл А.* Факторный анализ как статистический метод. Пер. с англ. М.: Мир.-1967. -144с.
8. *Ресурсы поверхностных вод СССР.* - Л.: Гидрометеоиздат, 1972.- т.18, вып.3: Приморье. – 628 с.
9. *Школьний Є.П., Лоєва І.Д., Гончарова Л.Д.* Обробка та аналіз гідрометеорологічної інформації: Підручник. - К.: Міносвіти України, 1999. - 600 с.
10. *Факторный, дискриминантный и кластерный анализ.* Пер. с англ. /Дж.-О.Ким,Ч.У.Мьюллер,У.Р. Клекка и др.- Финансы и статистика, 1989, - 215с.

Районування басейну річки Усурі з використанням методів багатовимірного статистичного аналізу. Лобода Н.С., Нгуєн Ву Ань

Виконано районування водозбору річки Усурі за синхронністю коливань річного стоку за допомогою методів багатовимірного статистичного аналізу – факторного та головних компонент.
***Ключові слова**: районування, фактори, компоненти, синхронність*

Zoning Ussuri river's basin by using of methods of multimeasured statistical analysis. Loboda N.S, Nguyen Vu Anh

To regionalize catchment of river Ussuri by the synchronism of annual runoff fluctuation, one may use the method of multimeasured statistical analysis – factorial and main components.
***Keyword**: zoning, factors, components, synchronism*