

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ЗАГРЯЗНЕНИЙ ПРИРОДНЫХ ПОВЕРХНОСТНЫХ ВОД

Колесников А.В., к.т.н, доцент,
Маковецкая Е.А., ассистент,

Одесская государственная академия строительства и архитектуры
moon.makelen@gmail.com

Аннотация. В статье рассматривается методика применения многомерного статистического анализа для исследования характера загрязнения природных поверхностных вод. Исходными переменными явились концентрации веществ-загрязнителей, микробиологические и санитарно-гигиенические показатели воды. Начальный этап исследования основывался на методе главных компонент, который позволил уменьшить размерность исследуемой системы показателей с 23 до 6 факторов комбинированного характера. Некоторым факторам дана интерпретация. Например, фактор 1 соответствует небιологическим загрязнениям. Рассматривается гипотеза, согласно которой состав суммарных загрязнений поверхностных вод можно представить как суперпозицию загрязнений из характерных источников. В результате проверки этой гипотезы методом кластерного анализа были выявлены 3 группы источников, соответствующим трем кластерам. По предположению, центроиды этих кластеров соответствуют трем характерным источникам загрязнений.

Ключевые слова: загрязнение природных вод, статистический анализ, метод главных компонент, кластерный анализ.

БАГАТОМІРНИЙ СТАТИСТИЧНИЙ АНАЛІЗ ЗАБРУДНЕНЬ ПРИРОДНИХ ПОВЕРХНЕВИХ ВОД

Колесников А.В., к.т.н, доцент,
Маковецька О.О., асистент,

Одеська державна академія будівництва та архітектури
moon.makelen@gmail.com

Анотація. У статті розглядається методика застосування багатомірного статистичного аналізу для дослідження характеру забруднення природних поверхневих вод. За вихідні змінні взято концентрації речовин-забруднювачів, микробиологічні та санітарно-гігієнічні показники води. Початковий етап дослідження ґрунтувався на методі головних компонент, який дозволив зменшити розмірність досліджуваної системи показників з 23 до 6 факторів комбінованого характеру. Деяким факторам дана інтерпретація. Наприклад, фактор 1 відповідає небіологічним забрудненням. Розглядається гіпотеза, згідно з якою склад сумарних забруднень поверхневих вод можна уявити як суперпозицію забруднень з характерних джерел. В результаті перевірки цієї гіпотези методом кластерного аналізу було виявлено 3 групи джерел відповідно трьом кластерам. За припущенням, центроїди цих кластерів відповідають трьом характерним джерелам забруднень.

Ключові слова: забруднення природних вод, статистичний аналіз, метод головних компонент, кластерний аналіз.

MULTIDIMENSIONAL STATISTICAL ANALYSIS OF POLLUTION OF NATURAL SURFACE WATER

Kolesnikov A.V., Ph.D., Associate Professor,
Makovetskaya E.A., Assistant,

Odessa State Academy of Civil Engineering and Architecture, Ukraine
moon.makelen@gmail.com

Abstract. The article deals with the methodology of applying multidimensional statistical analysis to study the nature of pollution of natural surface waters. The initial variables were concentrations of pollutants, microbiological and sanitary-hygienic indicators of water. The initial stage of the study was based on the method of the main components, which made it possible to reduce the dimensionality of the system of indicators under study from 23 to 6 factors of a combined nature. Some factors are interpreted. For example, factor 1 corresponds to non-biological contamination. The hypothesis is considered, according to which the composition of total surface water pollution can be represented as a superposition of impurities from characteristic sources. As a result of testing this hypothesis, three groups of sources corresponding to three clusters were identified by the cluster analysis method. By assumption, the centroids of these clusters correspond to three characteristic sources of pollution.

Keywords: pollution of natural waters, statistical analysis, main component method, cluster analysis.

Введение. Результаты анализа химического состава природных сред могут служить двум взаимосвязанным целям: оценке их качества, пригодности использования в хозяйственной деятельности и изучению характера воздействий, оказываемых на нее со стороны остальных элементов экосистемы, в том числе и антропогенных. Так, для загрязнений водных ресурсов характерным является не только его количественная мера, но и его структура – виды основных источников загрязнения, их подчиненность и взаимосвязь. Уточнение сведений об этой структуре является ключевым моментом для организации рационального использования водных ресурсов.

Анализ последних исследований. Средняя обеспеченность Украины качественной питьевой водой существенно ниже нормативов ВОЗ [1]. Из-за дефицита водных ресурсов многие населенные пункты и особенно города получают воду из удаленных поверхностных источников, подверженных интенсивному антропогенному загрязнению [2]. Для оптимизации водоподготовки и водопотребления важными являются факторы загрязнения природной воды, их типология и классификация, что может быть реализовано с помощью математической статистики. Один из методов статистического анализа на основе данных, полученных при исследовании состава поверхностных вод, основан на методах многомерной статистики – кластерного анализа и метода главных компонент [3, 4].

Цель и задачи исследования. Основой для статистического исследования могут служить данные анализа проб воды в изучаемой экосистеме. Целью исследования является разработка методики статистического анализа и классификации параметров загрязнения поверхностных природных вод. Реализация этой методики и результаты статистической обработки позволяют качественно оценить характер и структуру загрязнений водного бассейна.

Основной материал и результаты. В качестве исходных данных были использованы результаты анализов химического и микробиологического состава поверхностных вод [5], по перечисленным показателям (табл. 1).

Метод главных компонент, в основе которого лежит идея приведения эллипсоида рассеяния к главным осям, позволяет уменьшить число независимых переменных. Кроме этого он позволяет приближенно (эвристически) указать истолкование компонент, это полезное свойство более выражено в факторном анализе.

Таблица 1 – Анализируемые показатели состава поверхностных вод

1	Запах при 20°С, баллы	13	Сульфаты, мг/дм ³
2	Цветность, градусы	14	Бикарбонаты, мг/дм ³
3	Сухой/плотный/ остаток, мг/дм ³	15	Нитриты, мг/дм ³
4	Водородный показатель рН	16	Нитраты, мг/дм ³
5	Жесткость мг-экв/дм ³	17	Окисляемость перманганатная
6	Кальций, мг/дм ³	18	БПКполн, мгО ₂ /дм ³
7	Магний, мг/дм ³	19	ХПК, мгО ₂ /дм ³
8	Натрий, мг/дм ³	20	Нефтепродукты, мг/дм ³
9	Калий, мг/дм ³	21	Фенолы, мг/дм ³
10	Аммоний, мг/дм ³	22	СПАВ, мг/дм ³
11	Железо, мг/дм ³	23	Число ЛКП в 1 дм ³ воды
12	Хлориды, мг/дм ³		

Схема исследования отражена на рис. 1. В результате применения метода главных компонент формируются линейные комбинации исходных характеристик, приближенно соответствующих типам загрязнения поверхностных вод.

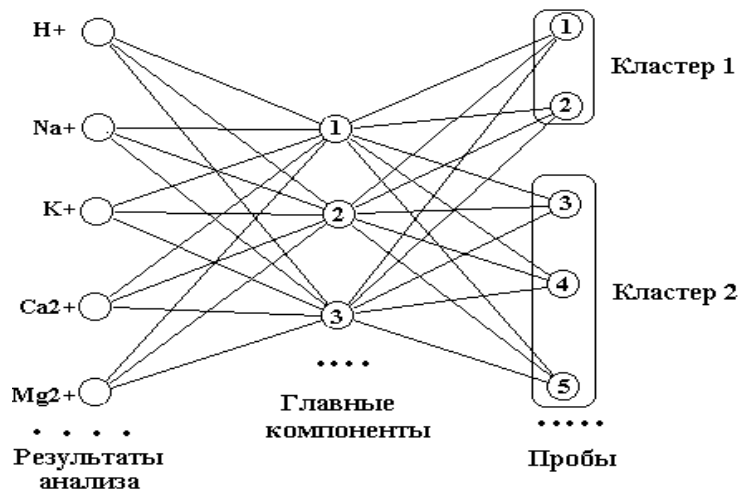


Рис. 1. Схема статистического исследования результатов анализа поверхностных вод

Анализируемые пробы воды при таком истолковании рассматриваются как результат смешения вод, состав которых соответствует главным компонентам. Первые три основных компонента отображены графически на рис. 2.

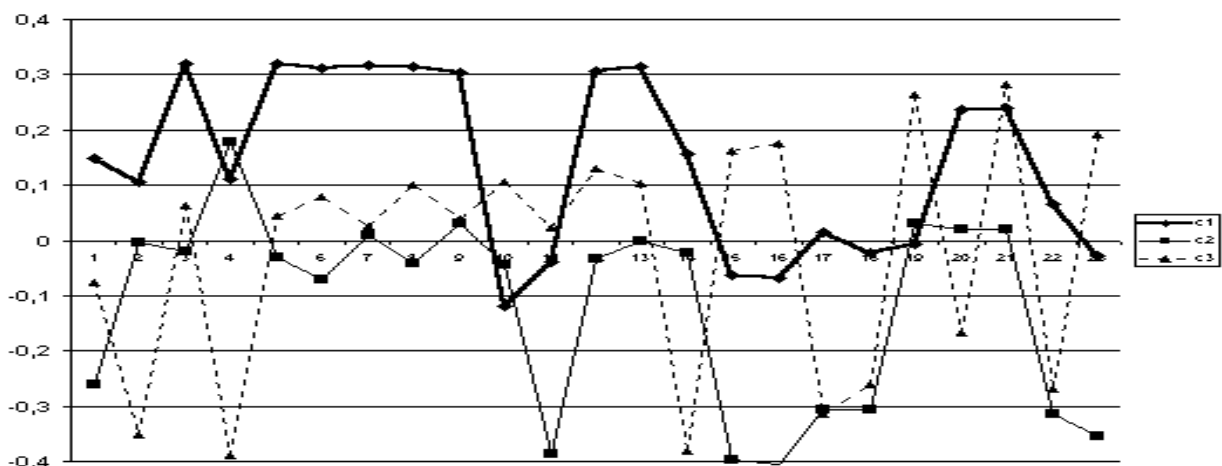


Рис.2. Коэффициенты первых трех главных компонент c1, c2, c3

Данные рис.2 позволяют дать возможные истолкования компонентам. Так, в первый компонент с положительными знаками вошли концентрации ионов (показатели 4–9, 12–13), концентрации фенолов и нефтепродуктов (20, 21).

Компонента 1 отражает, таким образом, небιологические виды загрязнений. Отрицательные значения коэффициентов (например, 9 и 10 – аммоний, железо) можно рассматривать как эффекты разбавления, добавления к источникам определенного объема воды с малыми концентрациями этих соединений относительно среднего уровня.

Исследование диапазона собственных значений приведенной корреляционной матрицы исходных данных, соответствующих главным осям эллипсоида рассеяния, показало, что из них 6 больше единицы, 6 главных компонент описывает 82% дисперсии (рис. 3).

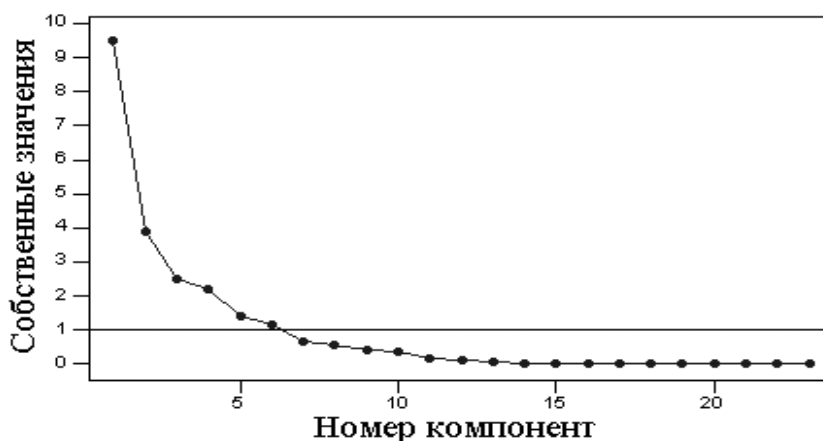


Рис. 3. Исследование собственных значений корреляционной матрицы

Поскольку применение метода главных компонент позволило уменьшить число эффективных переменных (с 23 до 6), относительно легко на основании полученных скоров проб (координат каждой пробы в новом базисе) реализовать метод кластерного анализа.

Для анализа группировки данных в кластеры использовался метод Варда и евклидово расстояние. Другие методы показали несколько худший результат. Вначале отдельные данные проб рассматриваются как отдельные группы, затем две группы с наименьшим взаимным расстоянием объединяются в новую группу, выполняется пересчет расстояний, этот процесс повторяется пока не осталось 3 группы. Ход процесса отражен на дендрограмме (рис. 4).

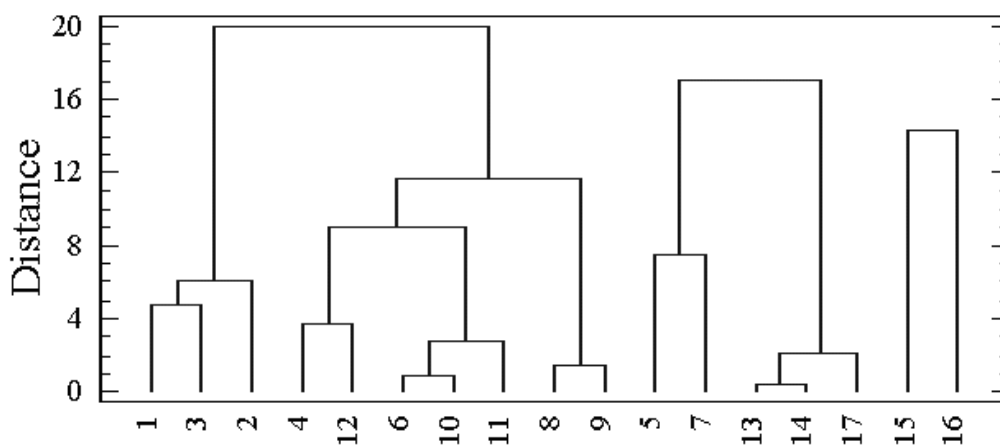


Рис. 4. Дендрограмма кластерного анализа

Анализ дендрограммы позволяет выделить пробу 16 и 15 в особую группу, остальные данные образуют еще две группы – всего три кластера. Структура кластеров хорошо просматривается уже для первых трех главных компонент (рис. 5).

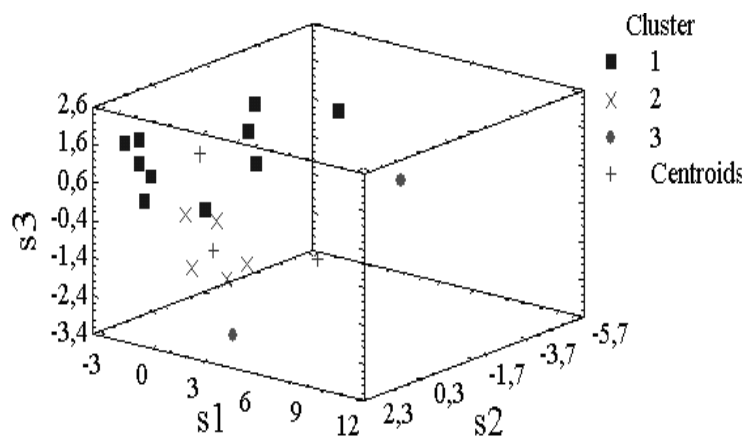


Рис. 5. Структура кластерів і групування результатів аналізу (S1, S2, S3 – три найбільш суттєвих комбінованих показателя)

Выводы и перспективы дальнейших исследований. Таким образом, применяемая процедура позволила выделить три основных вида поверхностной природной воды, со многими внутренними вариациями. Наиболее характерный состав загрязнений каждого из трех источников соответствует центроиду каждого из трех кластеров. Вода каждого из трех видов взаимодействует с другими компонентами экосистемы сходным образом. Характер антропогенных влияний здесь также подобен. Для воды каждого вида необходим индивидуальный способ водоподготовки при необходимости ее использования в бытовых и хозяйственных целях [6, 7]. Таким образом, у ингредиентных воздействий присутствует как количественный, так и качественный аспект – структура (рис.3, 4), что было показано на примере загрязнения гидросферы. Рассмотренная методика применима для исследования показателей загрязнения природной среды общего характера и представляется полезной для анализа ее экологического состояния. Природная вода из территориально различных источников будет характеризоваться различными количественными соотношениями выявленных факторов и характеризоваться различной активностью биологического и коррозионного действия. Перспективным представляется анализ этих пространственных вариаций и составление массива данных по различным видам загрязнения гидросферы для внесения в геоинформационные системы, связанные, в частности, с прогнозированием процессов коррозии строительных материалов.

Литература

1. Пашков А.П. Проблеми забруднення поверхневих, підземних і стічних вод та заходи щодо їх ліквідації і запобігання в Україні / А.П. Пашков // Безпека життєдіяльності. – 2011. – № 4. – С. 10-16.
2. Яцик А.В. Водні ресурси: використання, охорона, відтворення, управління / А.В. Яцик [та інші]. – К.: Генеза, 2007. – 360 с.
3. Справочник по прикладной статистике, ред. Л. Ллойд, У. Ледерман, т.2. – М.: Финансы и статистика, 1989. – 507 с.
4. Афифи А. Статистический анализ. Подход с использованием ЭВМ / А. Афифи, С. Эйзен. – М: Мир, 1982. – 486 с.
5. Медведева О.І. Практикум з охорони навколишнього середовища / О.І. Медведева, В.І. Жудіна, С.В. Красильникова, М.Н. Степанов, М.І. Стоянов, В.Г. Топуз, В.М. Шаригін. – К.: ІСДО, 1994. – 229 с.
6. ДСТУ 7525:2014 Вода питна. Вимоги та методи контролювання якості. – Мінекономрозвитку України. – Київ, 2014. – 30 с.
7. Гігієнічні вимоги до води питної, призначеної для споживання людиною. Державні санітарні правила і норми. ДСанПіН 2.2.4.171-10. – Офіційний вісник України. – 2010. – № 51. – С. 99-129.

Стаття надійшла 11.12.2017