

ОЦЕНКА КАЧЕСТВА ВОДЫ РЕК

С ИСПОЛЬЗОВАНИЕМ МЕТОДА АНАЛИЗА

ГЛАВНЫХ КОМПОНЕНТ

Целью настоящей работы является анализ физико-химических параметров воды для рек Вохча и Геха по данным 2005-2009 гг. методом анализа главных компонент (Principal Component Analysis). Для оценки ковариационной матрицы был применен метод анализа с недостающими данными (Missing Data Analysis). Проведен анализ отдельных значений главных компонент (ГК) и их весовых коэффициентов. Полученные результаты показывают, что расчетные оценки ГК имеют большой разброс и не существует четкой группируемости данных по времени. Анализ весовых коэффициентов показывает, что некоторые показатели образуют 4 группы. Выявлен характер и степень воздействия основных факторов, влияющих на качество воды рек.



Введение

В последнее время для обработки данных, полученных в результате мониторинга речных вод, широко применяется метод многомерного анализа [1-3].

В настоящей работе для обработки мониторинговых данных применен метод анализа главных компонент (ГК) [1]. Этот метод является современным стандартным инструментом для анализа данных, который имеет широкое применение в различных областях науки. Это связано с тем, что он относительно простой, непараметрический и позволяет выделить нужную информацию из существующей базы данных достаточно большого объема. Такой подход способствует обнаружению скрытых источников загрязнения [4], позволяет оценить характер и размер их воздействия на качество воды для выявления основных факторов загрязнения, обусловленных природным и антропогенным воздействием [5]. Этот подход может стать одним из основных методов для оценки качества воды и состояния речных экосистем и сыграть существенную роль для эффективного управления их качеством.

Г.П. Пирумян*,
доктор технических наук, профессор,
заведующий кафедрой
Экологическая химия,
химический факультет,
Ереванский государственный университет

Л.Э. Вардунян,
соискатель,
Ереванский государственный университет

Целью настоящей работы является анализ качества вод для рек южного речного бассейна Республики Армении рек Вохча и Геха, а именно:

- ♦ анализ ГК для загрязнителей, заранее выявленных с помощью Удельного Комбинаторного Индекса Загрязнения (УКИЗ) воды. При этом для оценки ковариационной матрицы применялся метод анализа с недостающими данными [6] за период 2005-2009 гг;
- ♦ анализ значений ГК (PC score) и их весовых коэффициентов (PC weight);

Материалы и методы исследования

Анализ главных компонент. До описания методологии анализа ГК необходимо четко определить структуру данных. Каждый раз данные интерпретируются как результат наблюдений, представленных в матричной форме, в которой каждый компонент является значением отдельных факторов.

* Адрес для корреспонденции: gevorg_pirumyan@mail.ru

Предположим, что матрица X $m \times n$ является начальной или исходной базой данных, где столбцы (m) соответствуют отдельным параметрам качества загрязнения, а строки (n) - времени наблюдения. Предположим, что Y является другой $m \times n$ матрицей, которая посредством P $m \times m$ матрицы является линейной модификацией X матрицы таким образом, чтобы выполнялось равенство:

$$PX = Y \quad (1)$$

При этом ГК вектора $\{p_1 \dots p_m\}$ выбираются ортонормальными.

Следовательно, необходимо найти такую P матрицу, чтобы $PX = Y$ и ковариационная матрица трансформированных данных

$$\Omega_Y = \frac{1}{n} YY^T \text{ были диагональными.}$$

В этом случае строки P матрицы станут основными компонентами X матрицы базы данных. Из определения Y и Ω_Y матриц получаем:

$$\begin{aligned} \Omega_Y &= \frac{1}{n} YY^T = \frac{1}{n} (PX)(PX)^T = \\ &= P \left(\frac{1}{n} XX^T \right) P^T = P \Omega_X P^T \quad (2) \end{aligned}$$

где Ω_X является ковариационной матрицей X . Из линейной алгебры известно, что любая симметричная матрица может быть представлена с помощью собственных векторов матрицы и собственных значений диагональной матрицы следующим образом: $\Omega_X = E \Lambda E^T$, где E - матрица собственных векторов, а Λ - диагональная матрица собственных значений.

Учитывая вышесказанное, получаем:

$$\begin{aligned} \Omega_Y &= P \Omega_X P^T = P (E \Lambda E^T) P^T = \\ &= (E^{-1} E) \Lambda (E^{-1} E) = \Lambda \quad (3) \end{aligned}$$

где $P = E^T = E^{-1}$.

Очевидно, что P матрица делает Ω_Y диагональной, что и является целью анализа основных компонентов. Подводя итог, можно отметить, что результатом анализа ГК является создание соответствующих P и Ω_Y матриц, где 1. основные компоненты X базы p_i данных являются собственными векторами Ω_X матрицы,

2. i -й компонент Ω_Y диагональной матрицы является вариацией X базы данных по p_i .

В соответствии с [1] для определения качества воды в ковариационной матрице P мы выбираем те ГК, собственные значения которых составляют порядка $>0,8$, а кумулятив больше 80 %.

С.Г. Минасян,

заместитель
директора
по научным
вопросам,
Центр мониторинга
воздействия
на окружающую
среду

Э.Э. Вардумян,

кандидат
экономических наук,
ассистент кафедры
Актуарная
математика,
факультет
математики
и механики,
Ереванский
государственный
университет

Результаты и их обсуждение

В бассейне р. Вохча за период 2005-2009 гг. пробоотбор проводился не регулярно. Отсутствие данных делает недопустимым анализ ГК для оценки ковариационной матрицы стандартным методом. Здесь наша цель двояка: с одной стороны необходимо заполнить недостающие данные, с другой стороны провести оценку таких параметров, как средняя величина и ковариационная матрица.

В работе [7] проведен анализ ГК физико-химических параметров воды рек Вохча и Геха за период 2005-2009 гг.. Для оценки ковариационной матрицы был применен метод анализа с недостающими данными [6], а столбцы матрицы X , т.е. основные параметры загрязнения, заранее выбирались с помощью УКИЗ. При расчете УКИЗ воды использовали рыбохозяйственные предельно допустимые концентрации (ПДК).

Анализ ГК и расчет отсутствующих в ряду данных был проведен с помощью пакета программного обеспечения «MATLAB R2010a».

Анализ отсутствующих в ряду данных

В естественных науках, в основном, используется предположение о нормальном распределении, как результат центральной предельной теоремы (ЦПТ). В настоящей работе также предполагается, что логарифмы вышеуказанных данных описываются совместным нормальным распределением.

Анализ отсутствующих данных был проведен для всех пунктов пробоотбора воды рек Вохча и Геха. В качестве примера на *рис. 1* представлены результаты по Cu для одного из пунктов пробоотбора воды р. Вохча.



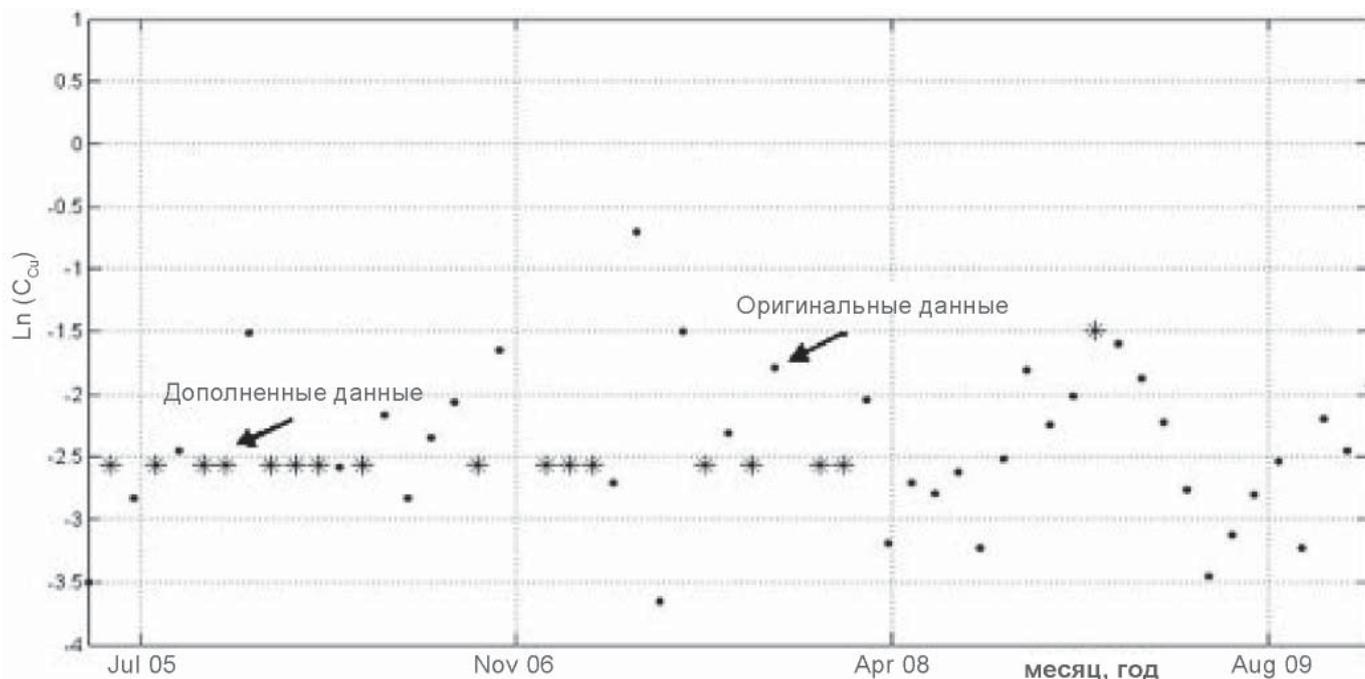


Рис. 1. Дополненные и оригинальные значения данных на примере Cu одного из пунктов пробоотбора воды р. Вохча.

Анализ основных компонент

Проведен анализ ГК в соответствии с указанными выше методами. В качестве примера в *табл. 1* представлены собственные значения и векторы ковариационной матрицы для р. Геха.

Как видно из таблицы первый главный компонент (ГК 1) имеет самое большое собственное значение 6,17, что объясняет вариацию 38,6 % данных. Второй главный компонент (ГК 2) имеет наибольшее собственное значение 2,28 и объясняет вариацию 14,3 % данных. Следующие ГК в соответствии с убыванием собственных значений распределены следующим образом: третий, имея собственное значение 1,64, объясняет 10,3 % вариации данных; четвертый, имея собственное значение 1,38, объясняет 8,6 % вариации данных; пятый, имея собственное значение 1,09, объясняет 6,8 % вариации данных; шестой, имея собственное значение 0,97, объясняет 6,1 % вариации данных; седьмой, имея собственное значение 0,83, объясняет 5,2 % вариации данных. Приведенные в таблице главные компоненты ГК 8-ГК 16 в соответствии с [1] не могут быть учтены.

Кумулятивные первые семь главные компоненты: ГК 1-ГК 7 объясняют 89,9 % общей вариации и показывают, что семь главных компонентов достаточны для объяснения закономерности и структуры данных. Таким образом, в результате анализа ГК удалось сократить размерность данных с шестнадцати до семи.

Уменьшение количества ГК пропорционально степени корреляции данных между собой. Аналогичные расчеты проведены для остальных пунктов пробоотбора. Результаты представлены ниже.

В верхней части р. Вохча 33,98 % загрязнения объясняется количеством сульфат иона, магния, меди, селена и бора, которые имеют почти одинаковое весовое значение. 15,53 % загрязнения связано с алюминием, ванадием, кремнием и бором; 13,97 % связано с количеством БПК₅, цинка и марганца. Ниже по течению реки от г. Каджаран 45-50 % загрязнения воды связано с присутствием ионов сульфата, нитрита, аммония, магния, селена, бора, кремния, фосфора, кислорода, БПК₅ и взвешенных частиц. Это свидетельствует о значительном антропогенном воздействии на качество воды реки. Марганец, медь, железо, алюминий, ванадий, хром и цинк объясняют 25-30 % загрязнения. До г. Капан в результате расчетов из 14 показателей были выбраны 6 ГК, которые объясняют 80,78 % от общей загрязненности. Около 60 % загрязнения обусловлено геохимическими особенностями региона и высокими естественными значениями фоновых концентраций, а 20 % загрязнения обусловлено антропогенным фактором. В речной воде вблизи аэропорта г. Капан из 17 показателей были учтены всего 6 ГК, которые объясняют 81,82 % загрязненности. Анализ показателей, включенных в ГК, весовые значения которых существенно отличны от нуля, показывает, что в этом пункте пробоотбора качество воды в значительной степени обусловлено антропогенным фактором и наличием промышленных сточных вод. Наиболее существ-

Таблица 1

Анализ собственных значений и векторов ковариационной матрицы (источник - р. Гехи)

Параметры качества	ГК 1	ГК 2	ГК 3	ГК 4	ГК 5	ГК 6	ГК 7	ГК 8	ГК 9	ГК 10	ГК 11	ГК 12	ГК 13	ГК 14	ГК 15	ГК 16
БПК ₅	-0,03	-0,12	-0,61	0,16	0,13	-0,49	0,05	0,22	-0,23	-0,20	-0,04	-0,08	0,09	-0,20	0,19	0,31
Нитриты	0,18	-0,31	-0,30	0,17	0,26	0,32	0,32	-0,48	-0,12	0,14	-0,18	-0,21	-0,04	-0,19	0,06	-0,33
Аммония	0,04	0,48	-0,18	0,24	-0,38	-0,22	0,01	-0,28	0,19	0,40	-0,31	-0,28	0,04	0,12	-0,09	0,12
Алюминий	0,34	0,13	0,10	0,16	0,06	-0,16	-0,19	0,40	-0,02	-0,21	-0,34	-0,16	0,21	-0,01	-0,03	-0,62
Железо	0,34	0,03	-0,01	0,11	0,34	-0,02	-0,19	0,07	-0,44	0,26	0,21	-0,15	-0,20	0,38	-0,43	0,16
Медь	0,34	0,24	-0,06	-0,03	0,03	0,08	-0,11	0,04	0,29	0,06	0,65	-0,34	0,03	-0,38	0,18	-0,02
Цинк	0,22	0,28	-0,13	0,29	0,05	0,55	0,07	0,16	-0,06	-0,03	-0,08	0,37	0,43	0,02	0,08	0,29
Марганец	0,36	-0,05	0,22	0,18	0,16	-0,04	-0,02	-0,01	0,22	-0,09	-0,22	0,00	-0,47	0,22	0,57	0,25
Ванадий	0,36	0,06	-0,02	-0,16	-0,03	-0,19	-0,07	-0,07	0,01	0,17	-0,18	0,54	-0,28	-0,54	-0,26	0,03
Хром	0,21	0,09	-0,38	-0,36	-0,06	0,01	-0,42	-0,44	0,04	-0,38	0,04	0,15	0,08	0,33	0,05	-0,10
Взвешенные вещества	0,03	-0,40	0,33	0,33	-0,06	-0,07	-0,54	-0,30	-0,02	-0,01	-0,06	-0,10	0,34	-0,23	-0,03	0,24
Селен	0,28	-0,03	0,21	-0,33	-0,33	-0,14	0,20	-0,06	-0,55	0,23	0,05	0,00	0,27	0,03	0,41	-0,02
pH	0,15	-0,17	-0,06	0,47	-0,62	0,03	0,14	0,00	-0,13	-0,30	0,30	0,13	-0,27	0,08	-0,14	-0,14
Суммарный фосфор	0,32	-0,10	0,13	-0,30	-0,11	0,07	0,35	0,01	0,13	-0,41	-0,21	-0,37	0,09	-0,06	-0,37	0,37
Кремний	0,03	-0,35	-0,31	-0,23	-0,34	0,37	-0,34	0,40	0,07	0,30	-0,19	-0,17	-0,17	-0,01	0,05	0,05
Бор	0,24	-0,41	-0,09	0,01	0,04	-0,27	0,22	0,07	0,49	0,30	0,16	0,24	0,33	0,33	-0,08	-0,05
Собственное значение	6,17	2,28	1,64	1,38	1,09	0,98	0,83	0,51	0,31	0,24	0,18	0,16	0,11	0,09	0,02	0,01
Доля, %	38,6	14,3	10,3	8,6	6,8	6,1	5,2	3,2	1,9	1,5	1,1	1,0	0,7	0,6	0,1	0,0
Кумулятивный, %	38,6	52,9	63,1	71,7	78,6	84,7	89,9	93,0	94,9	96,5	97,6	98,6	99,3	99,9	99,9	100,0

венный вклад в степень загрязнения воды имеют показатели БПК₅, ионы нитрита и сульфата, магний, медь, цинк, марганец и бор (собственное значение ГК равно 5,22 а доля 30,70 %). Следует отметить также вклад алюминия, железа, ванадия и селена (собственное значение ГК равно 3,37, а доля 19,81 %).

В истоке р. Геха главные шесть компонент объясняют 89,85 % загрязнения. При этом 38,59 % загрязнения связано с присутствием алюминия, железа, меди, марганца, ванадия, селена и фосфора. На качество воды также влияет антропогенный фактор. Об этом свидетельствует значительное весовое значение ионов нитрита, аммония, БПК₅, фосфора, взвешенных веществ. В устье р. Геха загрязнение, обусловленное антропогенной нагрузкой, составляет около 25 %, из которых 30 % обусловлено присутствием алюминия, железа, ванадия, марганца, меди и цинка.

Результаты показывают, что применение метода анализа ГК позволяет сократить размерность данных в 2,0-2,8 раза. Проведен также анализ отдельных значений ГК и их весовых коэффициентов. Полученные результаты показывают, что расчетные оценки ГК для рек Вохча и Геха имеют большой разброс и не существует четкой группировки данных по времени. Из анализа весовых коэффициентов следует, что некоторые показатели образуют 4 группы: 1) магний, бор и ион сульфата; 2) ванадий, хром, взвешенные частицы и значение pH; 3) растворенный кислород и БПК₅; 4) марганец, медь и цинк. Третья и четвертая группы обуславливают коммунально-бытовые и промышленные воздействия, соответственно, в то время как концентрации показателей первой

группы обусловлены открытой эксплуатацией ближайших шахт. Концентрации показателей второй группы обусловлены использованием неоправданно строгих ПДК, которые не имеют ничего общего с гидрогеохимическими особенностями региона и подлежат пересмотру.

Заключение

Исследования показали, что для бассейна р. Вохчи оценку качества воды методом анализа главных компонент можно провести даже при наличии отсутствующих (используя специальный метод анализа [6]).

Полученные результаты показывают, что в результате анализа удалось сократить размерность данных в 2,0-2,8 раза. Это обусловлено тем, что для анализа ГК учитывались те загрязнители, которые заранее отбирались с помощью УКИЗ. Из анализа параметров, имеющих значительное весовое значение в ГК, следует, что загрязнение в истоках рек обусловлено геохимическими особенностями региона, а ниже населенных пунктов добавляются коммунально-бытовые загрязнения. На качество воды р. Вохча влияют также промышленные сточные воды.

Анализ весовых коэффициентов показывает, что некоторые показатели можно объединить в 4 подгруппы.

Из анализа отдельных значений ГК следует, что расчетные оценки ГК для рек Вохча и Геха имеют большой разброс и не существует четкой группировки данных по времени. Полученные результаты указывают,



что в бассейне р. Вохча должен быть проведен двенадцатимесячный пробоотбор для обеспечения необходимых данных для полного статистического анализа.

Литература

1. Gardiner W.P. Statistical Analysis Methods for Chemists, A Software-based Approach // The Royal Society of Chemistry. 1997, № 8. P. 293-329.
2. Suresh S., Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River basin, Japan / Suresh S., Kazama F. // Environ. Model. Software, 2007. V. 22. P. 464-475.
3. Juahir H. Using chemometrics in assessing Langat River water quality and designing a cost-effective water sampling strategy / Juahir H., Zain S.M., Khan R.A., Yusoff M.K., Mokhtar M.B., Toriman M.E. // Maejo International, J. Science and Technology. 2009. V. 3. № 1. P. 26-42.

Ключевые слова:

оценка качества
воды,
весовые
коэффициенты,
анализ,
главные компоненты

4. Singh K.P. Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques: A case study / Singh K.P., Malik A., Sinha S. // Anal. Chim. Acta, 2005. V. 35. P. 3581-3592.
5. Mendiguchia C. Using chemometric tools to assess anthropogenic effects in river water. A case study: Guadalquivir River (Spain) / Mendiguchia C., Moreno C., Galindo-Riano D.M., Arcia-Vargas M. // Anal. Chim. Acta. 2004. V. 515. P. 143-149.
6. Little R.J.A. Statistical Analysis with Missing Data / Little R.J.A., Rubin D.B // Wiley, 1987. Parts 8 and 11. P. 292-326.
7. Бюллетени Армэкомониторинга, 2005-2009.

Авторы выражают благодарность всем сотрудникам Армэкомониторинга за техническую помощь.



G.P. Pirumyan, L.E. Vardumyan, S.G. Minasyan, E.E.Vardumyan

RIVER WATER QUALITY ASSESSMENT BY METHOD OF PRINCIPAL COMPONENTS

This work was aimed at physico-chemical parameter analysis of water taken from Vokhchi and Gekhi rivers for the period of 2005-2009 y. by method based on principal components (Principal Component Analysis). Missing Data Analysis was applied for covariance matrix

estimation. Individual values of principal components (PC) have been analysed with estimation of their weighting coefficients. The results show that calculated values of PC have big time and data spread. Weighting coefficient analysis shows that some of the parameters form 4 groups.

Essential factors that affect water quality of rivers have been revealed.

Key words: water quality assessment, weighting coefficients, analysis, principal components