



ИССЛЕДОВАНИЕ АВТОКОРРЕЛЯЦИОННОЙ И КОРРЕЛЯЦИОННОЙ ЗАВИСИМОСТЕЙ НА ПРИМЕРЕ ГИДРОХИМИЧЕСКИХ ХАРАКТЕРИСТИК ЗАГРЯЗНЕНИЯ МАЛЫХ ВОДОТОКОВ

Приведена методика статистического анализа на примере данных природных гидрохимических наблюдений. Выполнены автокорреляционный и корреляционный анализы биогенных загрязняющих веществ по малой реке Верхняя Ерга – притоку реки Сухоны Двинско-Печерского водного бассейна. Дан анализ полученным результатам с учетом определения существенности автокорреляционных и корреляционных связей.

Гидрохимическое загрязнение малых водотоков, биогенное загрязнение, нечерноземная зона РФ, корреляционный и автокорреляционный анализы, анализ существенности корреляционных и автокорреляционных связей.

Целью представленных в статье исследований является подбор методостатистических подходов для проведения анализа результатов обработки массивов данных на примере природных гидрохимических наблюдений.

Статистический анализ включает в себя оценку характеристик стохастических данных наблюдений, в том числе проверку на однородность, наличие корреляционных и автокорреляционных связей с учетом оценок параметров регрессионных зависимостей [1–6].

Результаты анализа позволяют с помощью критериев согласия подобрать сглаживающие кривые, характеризующие законы распределения концентрации загрязняющих биогенных веществ в речных водах, а также расходов воды малых рек-водоприемников осушительных мелиоративных систем, и определить их характеристики. Математическое моделирование с целью расширения рядов наблюдений реализуется методом Монте-Карло. Полученные с помощью математического моделирования ряды наблюдений анализируются с точки зрения закона распределения, оценки его характеристик проверяются на смещенность, состоятельность, робастность и эффективность [3]. Эти ряды и соответствующие им законы распределения дают возможность построить кривые обеспеченности, по которым определяется вероятность превышения предельно-допустимой концентрации биогенных веществ в речных водах в створе полного смешения их со сточными водами. Предлагается назначать предельно-допустимое содержание биогенных загрязняющих веществ в этом створе в зависимости от категории реки. Предельно-допустимые сбросы в реку сточных вод из дренажных коллекторов и магистральных каналов определяются с учетом предельно-допустимого содержания загрязняющих веществ в створе полного смешения. В периоды превышения предельно-допустимых сбросов предполагается ряд мероприятий, позволяющих аккумулировать часть сильно загрязненных биогенами сточных вод с целью их повторного использования на сельскохозяйственных угодьях.

Корреляционные связи отражают междурядовую зависимость случайных величин. В данном случае речь идет о поисках зависимости в хронологических рядах наблюдений концентраций биогенных загрязняющих веществ (соединений азотной и фосфорной групп) от расхода воды, а также друг от друга. В азотную группу входят азот аммонийный, нитритный и нитратный. В фосфорную группу – минеральный и общий фосфор. Для исследования корреляционных и автокорреляционных связей между ними и расходом воды выбран типичный для Северо-Западной зоны РФ бассейн рек Двинско-Печерского водосбора. Основу электронного банка составляют многолетние данные природных наблюдений, полученные на стационарных постах Росгидромета [7].

В среднем отбор проб на малых реках производится не более 7–10 раз в год. Для формирования однородных рядов природных наблюдений выделены два расчетных периода: весенний (март–май) и осенний (август–октябрь). Корреляционный анализ проводился как для всего ряда наблюдений, так и для выбранных расчетных периодов. Результаты расчетов сведены в таблицу 1.

Коэффициент корреляции определяется по формуле [1]:

$$r_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}, \quad (1)$$

где, $\text{cov}(x, y)$ – ковариация x, y ; σ_x и σ_y – средние квадратические отклонения сравниваемых выборок.

Значимость коэффициента корреляции оценивалась по следующим показателям:

1. Визуальная оценка. Если r_{xy} стремится к 1 или к (-1), то корреляция значима. Причем, положительное значение коэффициента корреляции определяет прямую зависимость между случайными величинами, отрицательное – обратную.

2. Средняя квадратическая ошибка коэффициента корреляции определяется по формуле [4]:

$$\Delta R = 0.674 \cdot \frac{1 - r_{xy}^2}{\sqrt{N - 2}}. \quad (2)$$

Если учетверенная ошибка коэффициента корреляции будет меньше значения коэффициента корреляции, то он значим.

3. Приближенная оценка значимости коэффициента корреляции на уровне $\alpha=0,05$ определяется по следующей зависимости [3]:

$$|r_{xy}| \cdot \sqrt{N-1} \geq 2.5. \quad (3)$$

Если условие выполняется, то корреляция значима.

4. Расчетное значение критерия Стьюдента определяется по формуле [1, 3]:

$$t_{\text{рас}} = \frac{r_{xy}}{\sqrt{1-r_{xy}^2}} \cdot \sqrt{N-2}. \quad (4)$$

Критическое значение определяется по таблицам [1] в зависимости от уровня значимости ($\alpha=0,05$) и числа степеней свободы ($m=N-2$). Если расчетное значение превышает критическое, то корреляция значима.

Таблица 1

Результаты корреляционного анализа хронологических натуральных данных для общего ряда и расчетных периодов

Корреляция	Q-Nh ₄ ⁺	Q-NO ₂ ⁻	Q-NO ₃ ⁻	Q-P _{мин}	Q-P _{общ}	Nh ₄ ⁺ -NO ₂ ⁻	Nh ₄ ⁺ -NO ₃ ⁻	NO ₂ ⁻ -NO ₃ ⁻	P _{мин} -P _{общ}
1	2	3	4	5	6	7	8	9	10
Река Верхняя Ерга, натурные данные за весь период наблюдения									
Количество значений, N	143	146	144	136	122	152	148	152	131
Коэффициент корреляции, r _{xy}	0,03	-0,02	-0,08	0,01	0,13	0,17	-0,03	0,20	0,33
Ошибка r _{xy} ΔR	0,06	0,06	0,06	0,06	0,06	0,05	0,06	0,05	0,05
4ΔR	0,23	0,22	0,22	0,23	0,24	0,21	0,22	0,21	0,21
Значимость	-	-	-	-	-	-	-	-	+
Примерная оценка	0,34	-0,20	-1,01	0,11	1,39	2,08	-0,40	2,44	3,82
Значимость >2,5 при α=0,05	-	-	-	-	-	-	-	-	+
Критерий Стьюдента, t _{рас}	0,34	-0,20	-1,01	0,11	1,39	2,11	-0,40	2,48	4,04
Значимость	-	-	-	-	-	+	-	+	+
Река Верхняя Ерга, весенний период									
Количество значений, N	83	84	80	79	69	87	83	85	73
Коэффициент корреляции, r _{xy}	0,02	-0,01	-0,21	-0,03	0,03	-0,01	-0,14	0,20	0,48
Ошибка r _{xy} ΔR	0,07	0,07	0,07	0,08	0,08	0,07	0,07	0,07	0,06
4ΔR	0,30	0,29	0,29	0,30	0,32	0,29	0,29	0,28	0,24
Значимость	-	-	-	-	-	-	-	-	+
Примерная оценка	0,16	-0,11	-1,84	-0,25	0,23	-0,05	-1,27	1,84	4,05
Значимость >2,5 при α=0,05	-	-	-	-	-	-	-	-	+
Критерий Стьюдента, t _{рас}	0,16	-0,11	-1,87	-0,25	0,22	-0,05	-1,27	1,87	4,58
Значимость	-	-	-	-	-	-	-	-	+
Река Верхняя Ерга, осенний период									
Количество значений, N	59	61	61	58	52	64	64	66	57
Коэффициент корреляции, r _{xy}	-0,02	-0,03	0,11	-0,04	0,30	0,36	0,19	0,28	0,11
Ошибка r _{xy} ΔR	0,09	0,09	0,09	0,09	0,08	0,07	0,08	0,08	0,09
4ΔR	0,35	0,34	0,34	0,35	0,34	0,29	0,32	0,31	0,35
Значимость	-	-	-	-	-	+	-	-	-
Примерная оценка	-0,18	-0,24	0,84	-0,32	2,15	2,87	1,55	2,22	0,80
Значимость >2,5 при α=0,05	-	-	-	-	-	+	-	-	-
Критерий Стьюдента, t _{рас}	-0,18	-0,24	0,84	-0,32	2,23	3,06	1,56	2,30	0,80
Значимость	-	-	-	-	+	+	-	+	-

- значимость коэффициента корреляции отсутствует;
 + коэффициент корреляции значим;
 α – уровень значимости.

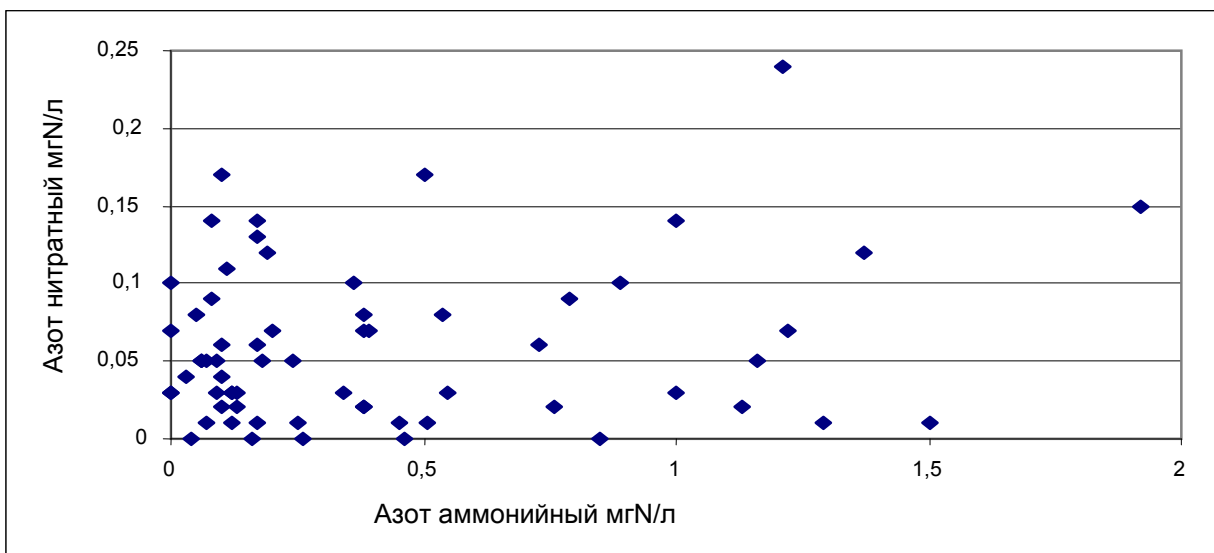


Рис. 1. График, свидетельствующий об отсутствии корреляции между азотом аммонийным и нитратным в осенний период наблюдений

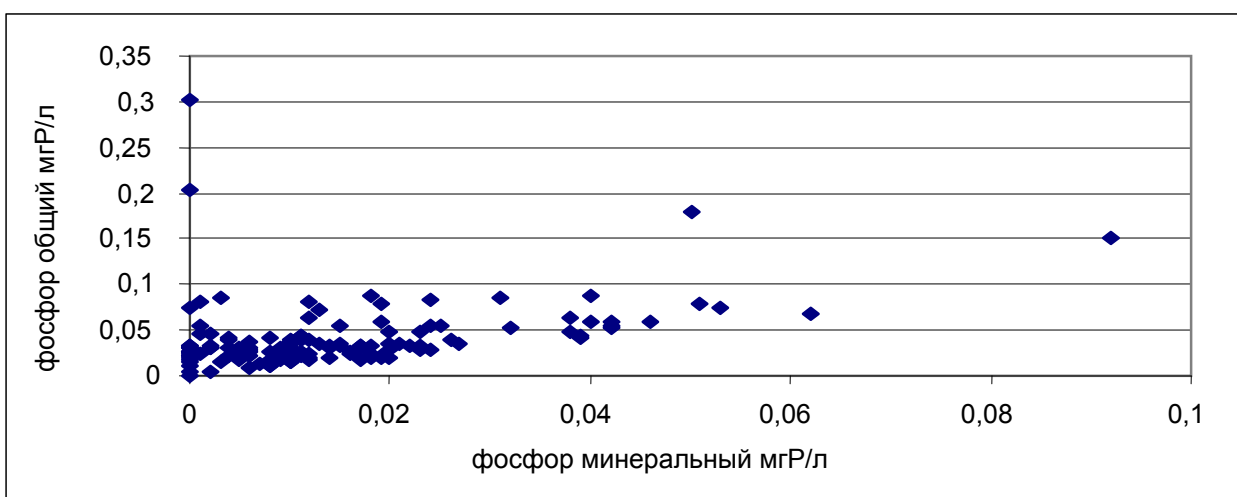


Рис. 2. Зависимость соединений фосфора общего и минерального в общем ряду наблюдений

По результатам анализа натурной информации можно сделать следующие выводы:

1. Зависимость концентраций биогенных загрязняющих веществ от расхода воды не обнаружена.
2. Корреляционная связь установлена между соединениями фосфора. Однако при этом значимость коэффициента корреляции для фосфора минерального и фосфора общего подтверждена лишь для общего ряда наблюдений и весеннего периода.
3. Коэффициент корреляции оказался значимым для азота аммонийного и нитритного в весенний период наблюдений.
4. Значимость коэффициента корреляции только по одному из трех рассмотренных оценок установлена в общем ряду наблюдений между азотом аммонийным и нитритным, азотом нитритным и нитратным; в осеннем периоде наблюдений – между расходом и фосфором общим, азотом нитритным и нитратным.

На рис. 1 представлен график, характерный для большинства сравниваемых массивов. Разброс точек в виде «облака» графически подтверждает отсутствие корреляционных связей между рассматриваемыми

случайными величинами. Рис. 2, напротив, свидетельствует о наличии корреляционной связи между минеральным и общим фосфором.

Следующим этапом анализа является выявление автокорреляционных связей между соседними случайными величинами в хронологических рядах наблюдений.

Результаты автокорреляционного анализа приведены в таблице 2. Автокорреляционный анализ выполнен для хронологических рядов натуральных наблюдений по реке Верхняя Ерга как для общего ряда, так и для расчетных периодов.

Коэффициент автокорреляции определяется по формуле [1, 5]:

$$r_{aem} = \frac{cov(x_1, x_2)}{\sqrt{\sum_{i=1}^{N-1} (x_i - x_{cp1})^2 \cdot \sum_{i=1}^N (x_{i+1} - x_{cp2})^2}}, \quad (5)$$

где $cov(x_1, x_2)$ – ковариация смежных значений ряда; x_{cp1} , x_{cp2} – средние арифметические значения из элементов ряда $(i - N - 1)$ и $(i + 1 - N)$.

**Результаты автокорреляционного анализа хронологических натуральных данных
для общего ряда и расчетных периодов**

Автокорреляция	Q	NH ₄ ⁺	NO ₂ ⁻	NO ₃ ⁻	P _{мин}	P _{общ}
1	2	3	4	5	6	7
Река Верхняя Ерга, натурные данные за весь период наблюдения						
Количество значений, N	147	153	156	152	148	131
Коэффициент автокорреляции, r _{авт}	-0,03	0,36	0,11	0,09	0,24	-0,02
По Нейману	2,05	1,26	1,78	1,77	1,50	1,92
Значимость	-	+	-	-	+	-
По Андерсону	-	+	-	-	+	-
По Стъденту	-0,33	4,67	1,34	1,13	3,05	-0,22
Значимость	-	+	-	-	+	-
Река Верхняя Ерга, весенний период						
Количество значений, N	85	88	89	85	84	73
Коэффициент автокорреляции, r _{авт}	-0,16	0,29	-0,03	-0,05	0,12	-0,11
По Нейману	2,30	1,40	2,06	2,02	1,69	1,97
Значимость	-	+	-	-	-	-
По Андерсону	-	+	-	-	-	-
По Стъденту	-1,44	2,76	-0,32	-0,41	1,11	-0,95
Значимость	-	+	-	-	-	-
Река Верхняя Ерга, осенний период						
Количество значений, N	61	64	66	66	63	57
Коэффициент автокорреляции, r _{авт}	-0,03	0,29	0,0005	-0,05	0,29	-0,05
По Нейману	2,06	1,37	2,00	2,10	1,18	2,08
Значимость	-	+	-	-	+	-
По Андерсону	-	+	-	-	+	-
По Стъденту	-0,23	2,38	0,00	-0,42	2,36	-0,36
Значимость	-	+	-	-	+	-

Значимость коэффициента автокорреляции определяется тремя способами:

1. По зависимости Неймана, которая представляет собой отношение среднего квадрата последовательных разностей к дисперсии [2, 5]:

$$\frac{\delta^2}{\sigma^2} = \frac{1}{N-1} \frac{\sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}{\sum_{i=1}^N (x_i - x_{cp})^2}. \quad (6)$$

Для определения значимости коэффициента автокорреляции в зависимости от отношения Неймана существуют специальные таблицы [5].

2. Значимость коэффициента автокорреляции устанавливается также по кривой связи Андерсона [5]. Кривая представляет собой зависимость коэффициента автокорреляции от количества значений в выборке, при этом на графике указана область, в которой коэффициент автокорреляции значим при заданном уровне значимости.

3. Проверка значимости автокорреляции по критерию Стъдента аналогична проверке значимости коэффициента корреляции.

По полученным результатам автокорреляционного анализа можно сделать следующие выводы:

1. Коэффициент автокорреляции значим в хронологических общих натуральных рядах наблюдений и в осенний период для азота аммонийного и фосфора минерального; в весенний период – для азота аммонийного.

2. Коэффициент автокорреляции не значим для расхода воды, азота нитритного, нитратного и фосфора общего.

Полученные статистические характеристики о наличии корреляционных связей между случайными величинами, какими являются расход воды в малых реках и концентрации биогенных загрязняющих веществ, а также автокорреляционные связи внутри хронологических рядов наблюдений могут быть использованы при организации математического моделирования упомянутых случайных величин с целью получения репрезентативных рядов этих величин.

Литература

1. Бронштейн, И. Н. Справочник по математике / И. Н. Бронштейн, К. А. Семенов. – Москва: Наука, 1980. – 977 с.
2. Гавич, И. К. Методы обработки гидрогеологической информации с вариантами задач / И. К. Гавич, С. М. Семенова, В. М. Швец. – Москва: Высшая школа, 1981. – 159 с.
3. Максимум, Ю. Д. Математика. Вып. 8: Математическая статистика: опорный конспект / Ю. Д. Максимум. – Санкт-Петербург: Изд-во СПбГПУ, 2004. – 96 с.
4. Михалев, М. А. Инженерная гидрология / М. А. Михалев. – Санкт-Петербург: СПбГПУ, 2003. – 360 с.
5. Гидрологические основы гидротехники / А. Ш. Резниковский, А. Ю. Александров, В. В. Атурин и др. – Москва: Энергия, 1979. – 232 с.

6. Общая теория статистики / Т. В. Рябушкин, М. Р. Ефимова, И. М. Ипатова, Н. И. Яковлева. – Москва: финансы и статистика, 1981. – 279 с.

Федеральная служба России по гидрометеорологии и мониторингу окружающей среды. Северное территориальное управление по гидрометеорологии и

мониторингу окружающей среды. Государственный водный кадастр. Раздел 1: Поверхностные воды. Серия 2: Ежегодные данные Ежегодные данные о качестве поверхностных вод. Часть 1: Реки и каналы. Том 1 (28) РФ (Бассейны рек на территории Архангельской, Вологодской и республики Коми).

S.N. Kovalenko
Vologda State University

STUDY OF AUTO-CORRELATION AND CORRELATION DEPENDENCE USING THE EXAMPLE OF HYDROCHEMICAL CHARACTERISTICS OF SMALL WATERCOURSES POLLUTION

The technique of statistical analysis of data of field hydrochemical observations is given. The autocorrelation and correlation analysis of biogenic pollutants were carried out along the small river Verkhnyaya Erga of the Sukhona Dvinsko-Pechersk water basin. The analysis of the obtained results is given, taking into account the determination of the significance of autocorrelation and correlation links.

Hydrochemical pollution of small watercourses, biogenic pollution, nonchernozem zone of the Russian Federation, correlation and autocorrelation analysis, analysis of the significance of correlation and autocorrelation links.