

ВИЗУАЛИЗАЦИЯ ДАННЫХ КАК МЕРА ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ АНАЛИЗА КАЧЕСТВА ПОВЕРХНОСТНЫХ ВОД

Е.Г. Ларина

ФГБУН «Институт водных проблем Российской академии наук», Москва, Россия

E-mail: ekl84@yandex.ru

АННОТАЦИЯ: Представлены возможности визуализации данных по различным выборкам на примерах гидрологической и водохозяйственной информации. Продемонстрирована эффективность использования некоторых приемов (графическое представление данных, контрольные карты Шухарта, диаграмма «ящик с усами», диаграмма рассеяния) для исследования результатов контроля качества воды р. Москвы и ее притоков. Показано, что методы визуального анализа позволяют наглядно отобразить изменения качества поверхностных вод и могут стать действенным инструментом при организации водопользования.

КЛЮЧЕВЫЕ СЛОВА: визуализация результатов наблюдений, качество воды, контрольный створ, загрязняющее вещество, р. Москва, контрольные карты Шухарта, карта рассеяния.

Финансирование: Работа выполнена в рамках темы № 0147-2019-0004 (№ государственной регистрации АААА-А19-119040990079-3) государственного задания Института водных проблем РАН.

Для природных вод характерна изменчивость состава, свойств и других признаков в пространстве и во времени в результате растворения, выщелачивания, осадкообразования и других процессов, происходящих в гидроэкосистеме и при ее взаимодействии с окружающей средой. Особенности трудности для нормирования, контроля и управления водопользованием создает изменчивость поверхностных вод селитебных территорий, поскольку поверхностный сток с городских территорий и площадок предприятий является одним из интенсивных источников загрязнения окружающей среды продуктами техногенного происхождения. Однако результаты измерительного контроля обычно используются исключительно для получения усредненных результатов [1, 2]. Такова распространенная практика, обусловленная невозможностью охватить массивы данных из огромного количества цифр.

В этой ситуации для анализа данных и более эффективного их изучения необходима визуализация – представление информации в форме, ориен-

© Ларина Е.Г., 2021

тированной на зрительное восприятие (диаграммы, графики и другие наглядные образы). Это позволяет обнаружить связь величин друг с другом, особенности их трансформации, неочевидные при табличном представлении, например, тренды, повторяющиеся структуры и т. п. [3, 4]. Именно поэтому визуализация данных находит широкое применение в научных и статистических исследованиях.

Примером недооценки визуализации результатов наблюдений за качеством воды могут служить таблицы концентрации загрязняющих веществ, элемент которых приведен в табл. 1.

Таблица 1. Концентрация нефтепродуктов в створах русла р. Москвы (в пределах МКАД) в феврале 2014 г., мг/л (ПДК=0,3 мг/л [5, 6])
Table 1/ Petrol products concentration in the Moscow River bed ranges (within the limits of MRR) in February, 2014, mg/l (MPC = 0.3 mg/l [5,6].

Номер контрольного створа	1	2	3	4	5	6	7	8	9	10	11	12	13
Концентрация нефтепродуктов, мг/л	0,07	0,04	0,11	0,07	0,22	0,31	0,13	0,19	1,45	0,06	0,83	0,05	0,2
Номер контрольного створа	14	15	16	17	18	19	20	21	22	23	24	25	26
Концентрация нефтепродуктов, мг/л	0,74	0,94	0,54	0,22	0,16	0,09	0,07	0,13	0,07	0,1	0,14	0,12	0,13

Примечание: Названия контрольных створов в соответствии с номерами, указанными в табл. 1: 1 – ниже Рублево; 2 – Спасский мост; 3 – Керосиновый р.; 4 – р. Сходня, устье; 5 – Ниже р. Сходни; 6 – руч. Соболевский; 7 – р. Таракановка; 8 – р. Филька; 9 – руч. Ваганьковский Студенец; 10 – р. Пресня, устье; 11 – выше р. Сетуни; 12 – р. Сетунь, устье; 13 – Бабьгородская плотина.

Очевидно, что достаточно сложно сделать какие-либо значимые выводы из приведенной записи, кроме заключения о том, что концентрация нефтепродуктов во многих случаях превышает ПДК. Однако графическое представление (рис. 1) позволяет легко убедиться в том, что концентрация нефтепродуктов значительно увеличивается к центру города, достигая значений в 3–5 раз превышающих норматив (вода в устьях рек Сетуни, Пресни, Яузы и Чуры содержит наибольшее количество загрязняющих веществ). По мере удаления от центра Москвы концентрация нефтепродуктов постепенно снижается, приближаясь к уровню ПДК. Эти выводы могут оказаться весьма важными при организации водопользования.

В отличие от табл. 1, табл. 2 содержит информацию, по которой можно констатировать, что концентрации нефтепродуктов не превышают ПДК. Однако анализ графической формы данных этой таблицы (рис. 2) наглядно показывает, что концентрация загрязняющего вещества изменяется скачкообразно в точке 5 (в мае) почти в два раза и остается высокой до

точки 8 (в августе). Это может свидетельствовать о наличии некоторого тренда или изменения каких-то существенных условий. Указанная информация может показаться важной для последующих действий и ее трудно извлечь из таблицы.

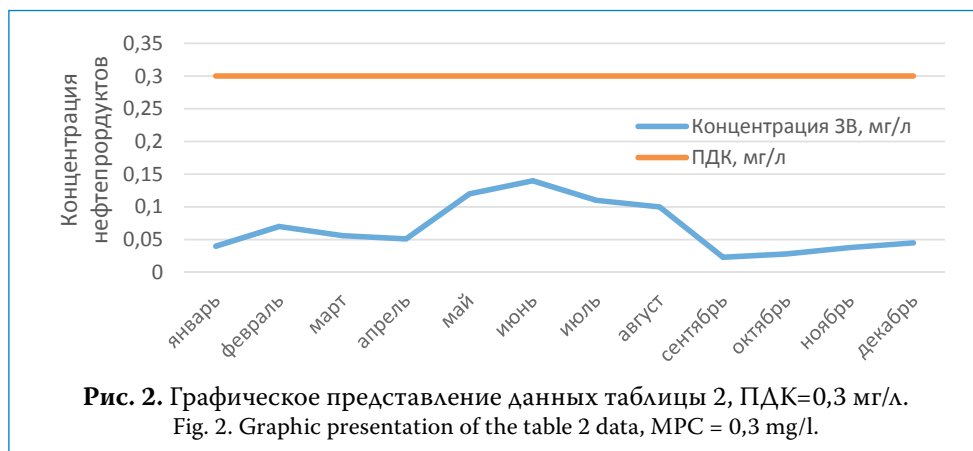
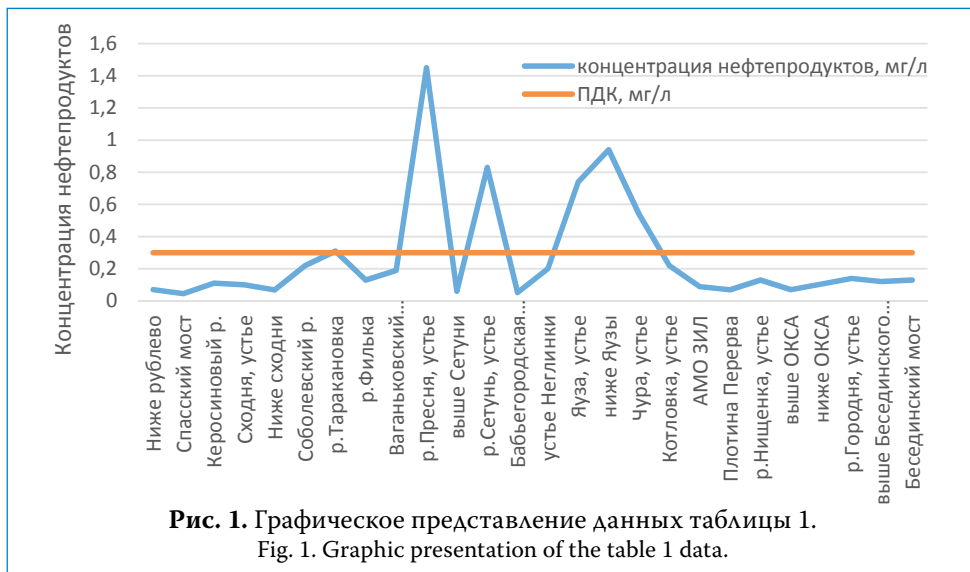


Таблица 2. Концентрации нефтепродуктов в створе ниже Рублево в 2014 г., мг/л

Table 2. Petrol products concentration in the range downstream Rublevo in 2014, mg/l.

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
Концентрация	0,04	0,07	0,056	0,051	0,12	0,14	0,11	0,1	0,023	0,028	0,038	0,045

Важные выводы можно сделать, используя контрольные карты Шухарта (ККШ). Согласно [7], контрольная карта (control chart) – график, для построения которого наносят в установленном порядке значения статистического показателя в последовательности выборок. ККШ используются для управления процессом, в т. ч. для снижения его изменчивости. Карта Шухарта (Shewhart control chart) – контрольная карта с контрольными границами Шухарта, предназначенная для разделения причин изменчивости контролируемой характеристики на случайные и специальные. ККШ помогут ответить на вопрос о стабильности того или иного процесса. Рассмотрим пример их использования.



По данным службы ГПБУ «Мосэкомониторинг» концентрация нефтепродуктов в контрольных створах, расположенных на основном русле р. Москвы и наиболее крупных притоках первого порядка – реках Сходне, Сетуни, Яузе (в пределах МКАД), обозначена на рис. 3 голубой линией, оранжевая линия – среднее по 39 точкам (от контрольного створа ниже Рублево до контрольного створа Бесединский мост), серая – ПДК=0,3 мг/л. В нашем случае оранжевая и серая линии являются контрольными границами Шухарта. Таким образом, как отражает рис. 3, концентрация нефтепродуктов превышала ПДК в некоторых точках. При этом выясняется, что максимальное количество загрязняющих веществ в русло реки добавляли притоки: Керосиновый ручей, р. Сходня, Соболевский ручей, реки Сетунь и Яуза. Данные по створу Электроводский мост на р. Яузе вызывают сомнения в достоверности (измеренная концентрация равна 2,2 мг/л, что превышает ПДК более чем в 7 раз, значение не умещается на рисунке при выбранном масштабе).

Использование ККШ перспективно, прежде всего, в целях обнаружения случаев потери стабильности водной системы, возникающей, к примеру, из-за несанкционированных или сверхнормативных сбросов загрязняющих веществ в природный водный объект. В этой ситуации контролирующие структуры могут совершать следующие ошибки [7]:

– принимать обычную флуктуацию за особую причину нарушений водоохраных требований и на этом основании приступить к широкомасштабному контролю водопользования и необоснованным запретам, что неизбежно приведет к дополнительным расходам и дестабилизации производственного процесса водопользователей;

– принимать реальное отклонение системы за обычную флуктуацию и в этом случае не проводить своевременных работ по выявлению источника нарушений водоохраных требований и пресечению сверхнормативного водопользования.

Применение контрольных карт Шухарта позволяет сократить количество подобных ошибок, поскольку дает возможность однозначно и практически просто определить, когда флуктуация значима и требует вмешательства, а когда это вариабельность внутри стабильной системы.

В промышленности рекомендуется обращать внимание на несколько таких конфигураций, перечисленных ниже [8].

1. Если хотя бы одна точка на контрольной карте выходит за пределы 3σ (как на рис. 3), то вероятность такого события составляет 0,00135. Следовательно, выход точки за пределы данной контрольной границы при воздействии исключительно случайных причин изменчивости возможен примерно в одном-двух случаях из двух тысяч. Поэтому каждая точка вне трехсигмового интервала, вероятнее всего, указывает на наличие неслучайных причин изменчивости и на нарушение статистически устойчивого состояния системы.

2. Две точки из трех последовательных лежат в области между 2σ и 3σ . Вероятность этого равна 0,0015, что может говорить о нарушении статистической стабильности системы.

3. Четыре из пяти последовательных точек находятся за пределами σ . Вероятность попадания одной точки в эту зону равна 0,079325, а вероятность размещения четырех из пяти последовательных точек за этими пределами – 0,001075, т. е. крайне мала. Эта ситуация также свидетельствует о выходе системы из стабильного состояния.

4. Шесть точек подряд характеризуются возрастающим значением ординаты, что указывает на неслучайные причины, приводящие к усилению изменчивости.

5. Восемь последовательных точек располагаются по обе стороны центральной линии так, что ни одна из них не попадает внутрь односигмового диапазона. Вероятность такой ситуации составляет 0,0001, что может быть обусловлено как воздействием неслучайных причин изменчивости, так и ошибками оператора или отсутствием репрезентативности выборки.

6. Десять точек подряд находятся по одну сторону от центральной линии. Для одной точки такая вероятность равна 0,5, а для десяти – 0,00195. Следовательно, в статистически стабильном состоянии такая ситуация возможна только в одном случае из двух тысяч и ее появление, скорее всего, указывает на возникновение неустойчивости.

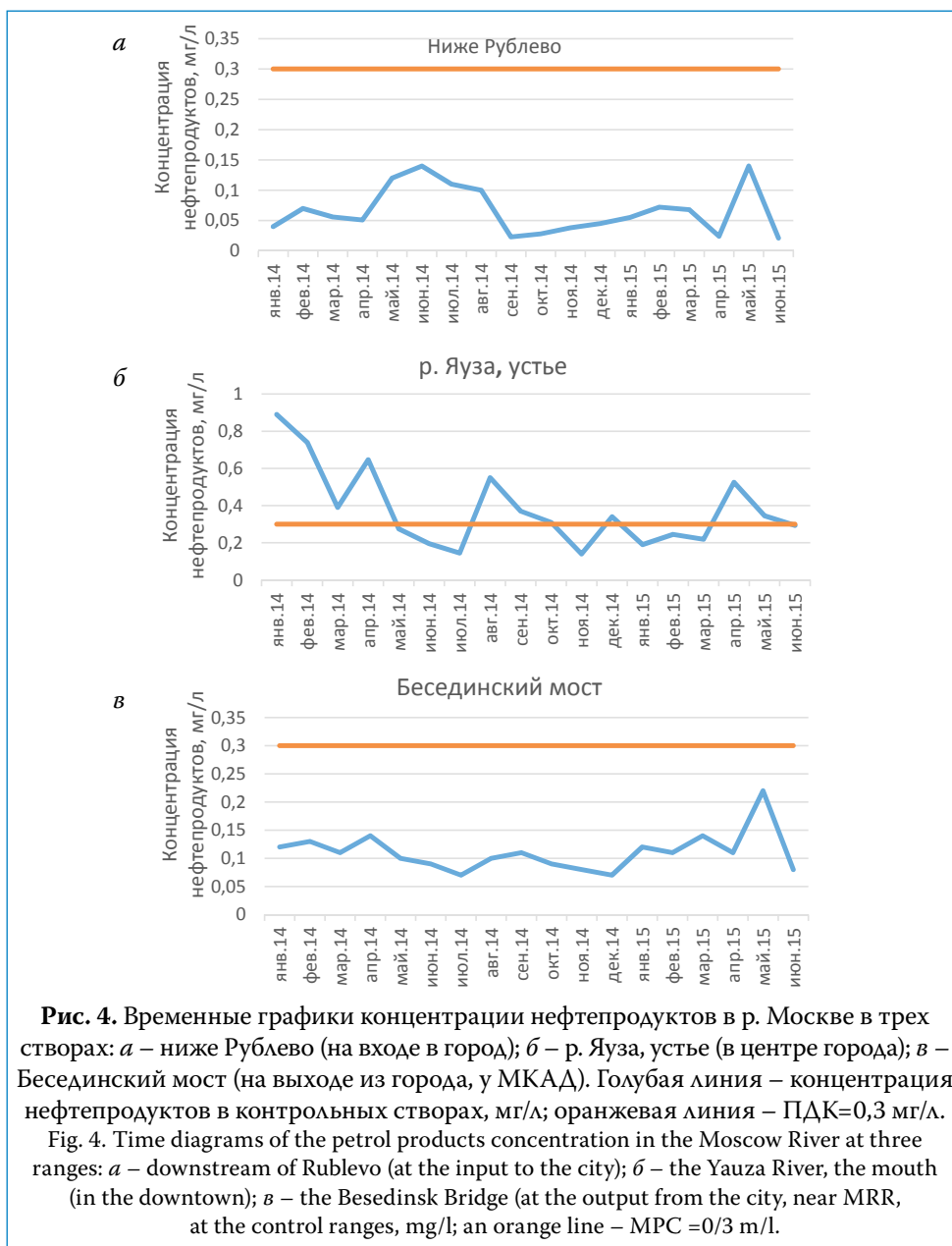
7. Четырнадцать точек характеризуются попеременно возрастающей и убывающей ординатой. Такие периодические циклы могут быть сигналом наличия неслучайных причин изменчивости, создающих периодические изменения значений выборочных характеристик.

8. Пятнадцать последовательных точек располагаются в зоне между центральной линией и σ . Вероятность попадания одной точки в односигмовые границы – 0,170675, а вероятность расположения пятнадцати последовательных точек в этих границах – 0,000825. Это свидетельствует, в первую очередь, о неустойчивости, хотя может быть обусловлено ошибкой оператора, а также тем, что используемые для построения контрольных карт выборки не являются представительными.

Приведем еще один пример эффективной визуализации массивов данных качества воды для анализа результатов контроля одновременно на нескольких контрольных створах. Такой анализ необходим для принятия стратегических решений о развитии системы водопользования территории.

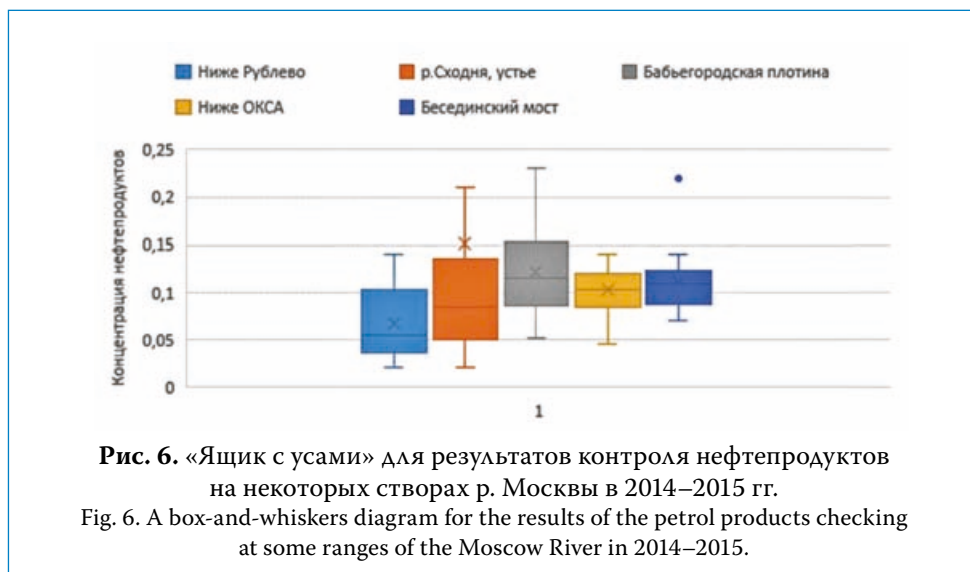
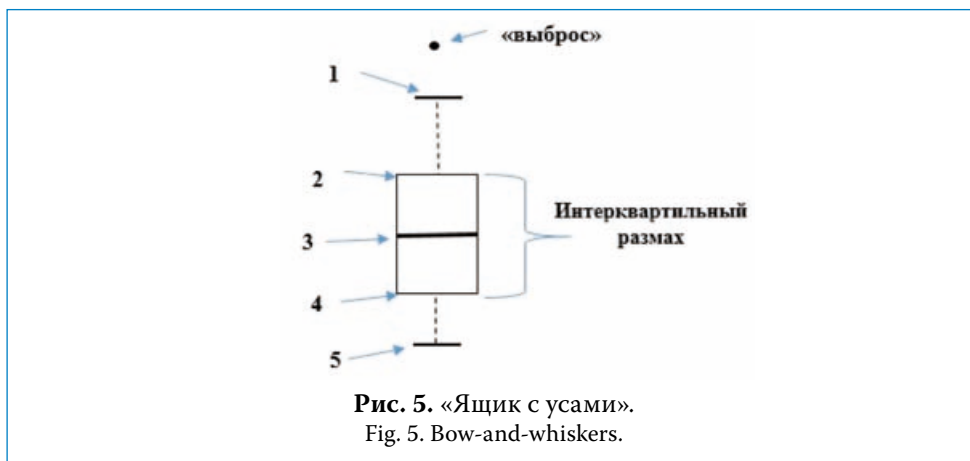
На участке р. Москвы в пределах МКАД расположены 26 контрольных створов наблюдения за качеством воды, принадлежащих ГПБУ «Мосэкомониторинг». Сравнительный анализ результатов ежемесячных наблюдений, полученных на этих створах за 1,5 года (с 2014 г. по 2 квартал 2015 г.) малоэффективен. И даже если массивы данных представить графически в виде временных рядов, все равно трудно из-за высокой изменчивости контролируемых показателей сделать определенные выводы, необходимые для перспективного планирования восстановления и охраны речных вод. В этом легко убедиться, взглянув на три таких графика (рис. 4 а, б, в).

Для придания наглядности статистическим данным рассмотрим специальные приемы, позволяющие визуально оценить ситуацию. Во-первых, используется метод визуализации с помощью блочной диаграммы, известной также как «ящик с усами», предложенный известным статистиком Дж. Тьюки в 1970-х годах. Он позволяет эффективно и устойчиво отражать основные характеристики любой выборки по пяти точкам (рис. 5): 1 – максимальное значение выборки; 2 – верхний квартиль (0,75-квантиль: значение концентрации, выше которого находится 25 % данных); 3 – медиана (0,5-квантиль: значение, выше и ниже которого находится одинаковое количество данных); 4 – нижний квартиль (0,25-квантиль: значение концентрации, ниже которого находится 25 % данных); 5 – минимальное значение выборки. Интерквартильный размах – разность между третьим и первым квартилями.



На рис. 6 показаны «ящики с усами» для концентрации нефтепродуктов в различных сечениях р. Москвы. Этот вид диаграмм позволяет в наглядной форме представить результаты измерений на створах для их совместного анализа. На диаграмме хорошо заметно, что на входе в город концен-

трации загрязняющих веществ находятся на уровне норматива. Очевидно, что центральный участок испытывает максимальную техногенную нагрузку, а крупные притоки первого порядка несут в своих водах большое количество загрязнений. На выходе из города концентрация загрязняющих веществ немного снижается. Возможно, это связано с работой Курьяновских станций аэрации, а также с процессами самоочищения (разбавление, седиментация, минерализация и т. д.). Точка синего цвета (у «ящика с усами» Бесединский мост) соответствует концентрации, равной 0,22 мг/л и является точкой выброса. Согласно [9], выброс представляет собой наблюдение, несовместимое с остальными наблюдениями ряда данных.



Еще один пример – диаграмма рассеяния. Она используется для демонстрации наличия или отсутствия корреляции между двумя переменными. На рис. 7 представлена диаграмма рассеяния для концентраций нефтепродуктов и фенола в створе р. Сходня, устье в 2014 г. По конфигурации точек можно предположить, что между содержанием нефтепродуктов и фенола в воде существует отрицательная регрессионная связь.

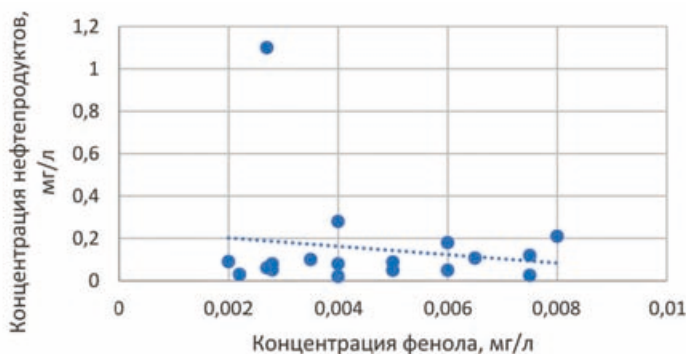


Рис. 7. Диаграмма рассеяния для концентраций нефтепродуктов и фенола, створ р. Сходня, устье, 2014–2015 гг.

Fig. 7. A scatter plot for petrol products and phenol concentrations, a reach of the Skhodnya River, 2014–2015.

ВЫВОДЫ

В работе проанализированы преимущества применения графических методов при исследовании результатов наблюдения за качеством воды р. Москвы в пределах МКАД.

Диаграммы изменения концентрации нефтепродуктов позволяют сделать выводы о повышении уровня загрязнения реки к центру мегаполиса и снижении концентрации загрязняющих веществ по мере удаления от него. Графический анализ ежемесячных наблюдений за концентрацией нефтепродуктов в 2014 г. наглядно продемонстрировал скачкообразный рост содержания в воде загрязняющих веществ в летние месяцы.

Применение контрольных карт Шухарта показало, что в апреле 2014 г. максимальные концентрации нефтепродуктов в русло р. Москвы добавляли притоки: руч. Керосиновый, р. Сходня, Соболевский ручей, реки Сетунь и Яуза. При этом одна из точек на контрольной карте (створ Электрозаводский мост) указывает на ситуацию потери стабильности водной системы. В перспективе подобные случаи требуют дополнительных исследований на наличие несанкционированных или сверхнормативных сбросов. В ходе анализа диаграммы «ящик с усами», отображающей одновременно концентрацию нефтепродуктов в нескольких контрольных створах, было установ-

лено, что в 2014–2015 гг. центральный участок русла р. Москвы испытывал значительную техногенную нагрузку. Диаграмма рассеяния выявила возможное наличие отрицательной регрессионной связи между содержанием нефтепродуктов и фенола в водном объекте.

Таким образом, в данной работе продемонстрировано, что визуализация – это действенный способ интерпретации данных наблюдений в виде графиков, диаграмм и др., позволяющий повысить эффективность анализа и обнаружить скрытую информацию, увидеть связь различных величин и параметров друг с другом.

СПИСОК ЛИТЕРАТУРЫ

1. Данилов-Данильян В.И., Розенталь О.М. Парадоксы экологического нормирования // Стандарты и качество, № 5, 2007. С. 42–44.
2. Доклад «О состоянии окружающей среды в городе Москве в 2018 году» / под ред. А.О. Кульбачевского. М.: ДПиООС; НИиПИ ИГСП: ООО «Студио Арроу», 2019. 247 с.
3. Миттаг Х.-Й., Ренни Х. Статистические методы обеспечения качества. М.: Машиностроение, 1995. 616 с.
4. Лунина В.В., Шпер В.А. Визуализация данных – инструмент роста знаний. Методы менеджмента качества, 2011, № 12. С. 28–32.
5. Приказ Минсельхоза России от 13.12.2016 № 552 «Об утверждении нормативов качества воды водных объектов рыбохозяйственного значения, в том числе нормативов предельно допустимых концентраций вредных веществ в водах водных объектов рыбохозяйственного значения».
6. Готовцев А.В., Данилов-Данильян В.И., Ларина Е.Г. Оценка самоочищающей способности русловой и коллекторной сети реки Москвы в пределах МКАД // Вода: химия и экология. 2016. № 10. С. 16–24.
7. ГОСТ Р ИСО 7870-1-2011 Статистические методы. Контрольные карты. Ч. 1. Общие принципы.
8. Булдакова Т.В., Розенталь О.М., Шпер В.А. Как повысить эффективность анализа качества воды // Методы оценки соответствия, № 3, 2012. С. 16–20.
9. ГОСТ Р ИСО 16269-4-2017 Статистические методы. Статистическое представление данных. Ч. 4. Выявление и обработка выбросов.

Для цитирования: Ларина Е.Г., Визуализация данных как мера повышения эффективности анализа качества поверхностных вод // Водное хозяйство России: проблемы, технологии, управление. 2021. № 2. С. 72–82.

Сведения об авторе:

Ларина Екатерина Геннадьевна, младший научный сотрудник, лаборатория управления водными ресурсами, ФГБУН «Институт водных проблем Российской академии наук», Россия, 119333, Москва, ул. Губкина, 3; e-mail: ekl84@yandex.ru

DATA VISUALIZATION AS A MEASURE TO IMPROVE THE EFFECTIVENESS OF SURFACE WATER QUALITY ANALYSIS

Ekaterina G. Larina

Water Problems Institute of the Russian Academy of Sciences, Moscow, Russia

Abstract: The article shows the possibilities of data visualization for various samples using examples of hydrological and water management information. The efficiency of using some techniques (graphical representation of data, Shewhart control maps, box-and-whiskers diagram, and scattering diagram) for analyzing the results of water quality control of the Moscow river and some of its tributaries is demonstrated. It is shown that the methods of visual analysis allow you to visually display any changes in the quality of surface water and can become an effective tool in the organization of water use in the metropolis.

Key words: visualization of observation results, water quality, control target, pollutant, the Moscow River, Shewhart control maps, box-and-whiskers diagram, scatter plot.

About the author:

Ekaterina G. Larina, Junior Researcher, Laboratory of Water Resources Management, Water Problems Institute of the Russian Academy of Sciences, ul. Gubkina, 3; Moscow, Russia; e-mail: ekl84@yandex.ru

For citation: Larina E.G. *Data Visualization as a Measure to Improve the Effectiveness of Surface Water Quality Analysis* // *Water Sector of Russia*. 2021. No. 2. P. 72–82.

REFERENCES

1. Danilov-Danil'yan V.I., Rozental' O.M. Paradoxy ekologicheskogo normirovaniya [Paradoxes of environmental normalization] // *Standarty i kachestvo*, № 5, 2007. P. 42–44.
2. Doklad «O sostoyaniyi okruzhayushchey sredy v gorode Moskve v 2018 godu» [Report “On the environmental situation in the city of Moscow”] / Pod red. A.O. Kul'bachevskogo. – M.: DPiOOS; NIiPI IGSP: OOO «Studio Arrou», 2019. 247 p.
3. Mittag H.-J., Renni H. Statisticheskie metody obespecheniya kachestva. [Statistical methods of the quality securing] M.: Mashinostroenie, 1995. 616 p.
4. Lunina V.V., Shper V.L. Vizualizatsiya dannyh – instrument rosta znaniy [Data visualization as a tool of knowledge growth]. – *Metody menedzhmenta kachestva*, 2011, № 12. P. 28–32.
5. Prikaz Minsel'hoza RF ot 13.12.2016 №552 “Ob utverzhdeniyi normativov kachestva vody vodnykh ob’ektov rybokhozyajstvennogo znacheniya, v tom chisle normativov predel’no dopustimyykh koncentratsiy vrednykh veshchestv v vodakh vodnykh obyektov rybokhozyajstvennogo znacheniya” [Resolution of the Ministry of Agriculture of the Russian Federation dated 13.12.2016 №552 “About approval of the water quality norms of the fishery water bodies including norms of the maximal permissible concentration of adverse pollutants in the fishery water bodies’ waters”].
6. Gotovcev A.V., Danilov-Danil'yan V.I., Larina E.G. Otsenka samoochishchayushchey sposobnosti ruslovy i kollektornoy seti reki Moskvy v predelakh MKAD [Assessment of the self-purification ability of the bed and collector network of the Moscow River within the limits of the Moscow Ring Road] // *Voda: himiya i ekologiya*, 2016, № 10. P. 16–24.
7. GOST R ISO 7870-1-2011 Statisticheskie metody. Kontrol'nye karty. Chast' 1. Obshchie printsipy [Statistic methods. Control maps. Part 1. General principles].
8. Buldakova T.V., Rozental' O.M., Shper V.L. Kak povysit' effektivnost' analiza kachestva vody [How to improve the effectiveness of the water quality analysis] // *Metody otsenki sootvetstviya*, № 3, 2012. S.16–20.
9. GOST R ISO 16269-4-2017 Statisticheskie metody. Statisticheskoe predstavlenie dannykh. Chast' 4. Vyyavlenie i obrabotka vybrosov [Statistic methods. Statistical presentation of data. Part 4. Ejections revealing and processing].