

УДК 551.578.46:550.42
DOI: 10.18799/24131830/2025/9/4754
Шифр специальности ВАК: 03.00.27
Научная статья

Эффективность и валидность кластерного анализа содержания микроэлементов пыли снегового покрова

А.В. Захарченко✉, А.А. Тигеев

Тюменский научный центр СО РАН, структурное подразделение Институт проблем освоения Севера,
Россия, г. Тюмень

✉avzakh@gmail.com

Аннотация. Актуальность. Кластеризация как метод анализа данных нашла широкое применение в различных областях знания, где требуется классификация объектов исследования. Очевиден поиск алгоритмов, способствующих максимально эффективному использованию метода. Успешность формирования классификационного дерева иерархического кластерного анализа зависит от применяемых методов стандартизации данных. **Цель:** провести сравнительный анализ методов стандартизации состава химических элементов пыли снегового покрова для оценки экологической опасности загрязнения атмосферы и валидности результатов иерархического кластерного анализа. **Объекты и методы.** Для примера использован микроэлементный состав твердой фазы снега г. Тюмени и фоновых точек, удаленных от города более чем на 10 км. Содержание в снеговом покрове загрязняющих веществ отражает загрязнение атмосферного воздуха. На примере анализа содержания химических элементов в твердой фазе снегового покрова г. Тюмени обоснованы наиболее простые методы предварительной обработки данных с целью их стандартизации для последующего статистического анализа. Рассмотрено четыре метода стандартизации данных в сравнении с исходными данными. Результативность кластеризации оценена с помощью интегрального показателя экологического загрязнения, а ее валидность – с помощью индекса Калински–Харабаша. Для подтверждения основных выводов результаты сопоставлены с аналогичными данными по Томской области. **Результаты.** Показано графическое отображение геохимических спектров при использовании разных методов стандартизации данных, а также проведен анализ различий результатов кластеризации. Для их сравнения использованы данные микроэлементного состава снегового покрова Томской области. **Выводы.** Метод весов (%) «Weight», позволяющий выявить различия относительного содержания микроэлементов в городе и в условиях фона, оказался наиболее эффективным при графическом отображении геохимического спектра. Полагалось, что чем выше их значения, тем эффективней кластеризация; контролем служили те же показатели для снегового покрова Томской области, которые оказались согласованными с показателями для Тюмени. Установлено, что наибольшей эффективностью обладает метод стандартизации медианой и квантилями 0,25 и 0,75 «Median».

Ключевые слова: микроэлементы, пыль снегового покрова, стандартизация данных, кластерный анализ, валидность, загрязнение атмосферы

Благодарности: Работа выполнена в рамках государственного задания Министерства науки и высшего образования РФ (№ FWRZ-2021-0006).

Для цитирования: Захарченко А.В., Тигеев А.А. Эффективность и валидность кластерного анализа содержания микроэлементов пыли снегового покрова // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2025. – Т. 336. – № 9. – С. 97–108. DOI: 10.18799/24131830/2025/9/4754

UDC 551.578.46:550.42
DOI: 10.18799/24131830/2025/9/4754
Scientific paper

Efficiency and validity of cluster analysis of trace elements content in snow cover dust

A.V. Zakharchenko✉, A.A. Tigeev

*Tyumen Scientific Center SB RAS, structural unit Institute of the Problems of Northern Development,
Tyumen, Russian Federation*

✉avzakh@gmail.com

Abstract. Relevance. Clustering, as a method of data analysis, has found wide application in various fields of knowledge where classification of research objects is required. The search for algorithms that facilitate the most efficient use of the method is obvious. The success of forming a classification tree of hierarchical cluster analysis depends on the data standardization methods used. **Aim.** To conduct a comparative analysis of methods for standardizing the composition of chemical elements of snow dust for assessing the environmental hazard and validity of the results of hierarchical cluster analysis. **Objects and methods.** As an example, we used the microelement composition of the solid phase of snow in the city of Tyumen and background points more than 10 km away from the city. The content of pollutants in the snow cover reflects atmospheric air pollution. Using the example of analyzing the content of chemical elements in the solid phase of snow cover in the city of Tyumen, the simplest methods of preliminary data processing are substantiated in order to standardize them for subsequent statistical analysis. The paper considers four methods of data standardization in comparison with the original data. The effectiveness of clustering was assessed using the integral indicator of environmental pollution, and its validity – using the Kalinski-Harabash index. To confirm the main conclusions, the results are compared with data for the Tomsk region. **Results.** The paper shows a graphical display of geochemical spectra using different methods of data standardization, as well as an analysis of the differences in clustering results. To compare them, the data on the microelement composition of the snow cover in the Tomsk region were used. **Conclusions.** The “Weight” method of weights (%) turned out to be the most effective in graphically displaying the geochemical spectrum, allowing us to identify differences in the relative content of trace elements in the city and in background conditions. It was believed that the higher their values, the more effective the clustering; the control was the same indicators for snow cover in the Tomsk region, which turned out to be consistent with the indicators for Tyumen. It was established that standardization with a median and quantiles of 0.25 and 0.75 “Median” is most effective.

Keywords: trace elements, snow dust, data standardization, cluster analysis, validity, air pollution

Acknowledgements: The work was performed within the framework of the state assignment of the Ministry of Science and Higher Education of the Russian Federation (no. FWRZ-2021-0006).

For citation: Zakharchenko A.V., Tigeev A.A. Efficiency and validity of cluster analysis of trace elements content in snow cover dust. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2025, vol. 336, no. 9, pp. 97–108. DOI: 10.18799/24131830/2025/9/4754

Введение

Формирование пула пыли снегового покрова в каждой локации происходит уникальным способом. Химический состав и концентрация микроэлементов твердой фазы снегового покрова зависят от глобального фонового поступления в атмосферу природных и техногенных загрязняющих атмосферу веществ как из удаленных, так и из региональных источников местных промышленных предприятий [1–5]. Снежный покров как природный накопитель в течение длительного периода, от залегания до снеготаяния, аккумулирует химические вещества, и его состав может служить индикатором загрязнений атмосферных осадков [6–9], что зави-

сит от дальности распространения, метеоусловий, уровней загрязнения атмосферы, мощности источников, условий выброса и т. д. [10, 11]. Исходя из анализа снегового покрова, загрязнение атмосферного воздуха в Тюменском регионе имеет средний и умеренно опасный уровень, изредка достигая опасных величин [12–14].

Стандартизация – это процесс предварительной обработки данных для их преобразования в единый формат. Она позволяет устранить исходную неоднородность данных и выровнять разномасштабные различия между переменными для их сравнительного анализа [15, 16]. Данный метод широко используется в медицине и во множестве других об-

ластей знаний и имеет ключевое значение для повышения совместимости данных в геохимии, экологии, экологической безопасности.

Одним из вариантов опосредованной оценки загрязнения территории является суммарный показатель загрязнения снегового покрова (Z_c) при сравнении данных из зон антропогенного влияния с условным фоном, удаленным на достаточное расстояние от источника загрязнения [17]. Для этого отобраны пробы на удалении от городской черты. На практике выбор условного фона может являться проблематичной задачей ввиду неоднородности снегового покрова в силу случайных особенностей циклонического переноса и выпадения загрязняющих веществ [1]. Для решения проблемы авторами предложен выбор фонового участка с использованием иерархического кластерного анализа. Ранее показана перспективность его использования для повышения точности анализа и обоснованности выводов об уровне загрязненности воздуха, выявления местной специфики и локализации источников загрязнений [1] с учётом выбранного алгоритма кластеризации [18, 19].

Содержание микроэлементов в геохимическом спектре пыли снегового покрова из-за их природных особенностей меняется в десятки, иногда в сотни раз от главных микроэлементов (Ni, Zn, Cr, Cu, Co, Pb, V) до редких и рассеянных. Результаты кластерного анализа могут существенно различаться из-за структуры геохимических данных, в том числе отсутствия нормального, логнормального и симметричного распределения данных [20–22]. Эффективное использование статистических методов анализа зависит от предварительной стандартизации данных [23]. В зависимости от выбранной предварительной стандартизации данных результаты классификации кластеров (дендрограммы) существенно различаются. Надежность кластеризации оценивается показателями валидности с использованием различных методов [24, 25].

Цель работы – сравнительный анализ методов стандартизации состава химических элементов снеговой пыли для выявления наиболее эффективного варианта оценки степени экологической опасности и валидности результатов кластерного анализа.

Применение экологически эффективных и валидных статистических методов анализа данных позволяет выявить загрязненные опасными веществами территории и повысить экологическую безопасность городских агломераций.

Объекты и методы

Изучена пыль снегового покрова Тюмени (57°15' с. ш., 65°54' в. д.). Региональными источниками пыли являются предприятия, производящие нефтегазовое оборудование, аккумуляторный завод, а также неспециализированные источники –

две ТЭЦ [26]. В анализе использовано 60 проб, которые были взяты в последней декаде февраля. 50 точек опробования расположены в черте города. Схема опробования представлена на рис. 1, а. Остальные точки расположены на некотором удалении от городской черты и относятся к фоновым (рис. 1, б). Разделение твердой и растворенной фракций произведено путем фильтрации. В аналитическом сертификационном испытательном центре РАН (г. Черноголовка, Московская область) определен элементный химический состав твердофазных выпадений (51 элемент). В статистическом иерархическом кластерном анализе использованы 24 микроэлемента: Li, Sc, V, Cr, Co, Ni, Cu, Zn, Ga, As, Rb, Sr, Y, Zr, Nb, Mo, Ag, Cd, Sn, Sb, Cs, W, Ba, Pb. Также использованы пробы лаборатории почвоведения Биологического института Томского государственного университета, отобранные в Томской области (1997–1999 гг.) [27], где в одной базе представлены содержания макроэлементов (NH_4 , P_2O_5 , K, Ca, Mg) талой снеговой воды и микроэлементов (Ni, Co, Pb, Cu, Zn, Mn, Cr, Cd) твердой фазы снегового покрова. Территория удалена от Тюмени, но имеет сходные условия формирования снегового покрова [28]. Схема точек отбора проб в Томской области представлена на рис. 1, в.

Кластерный анализ используется для анализа геохимических данных [22, 29]. Содержание химических элементов в геохимическом спектре различаются на порядки, что делает некорректным их сравнение в едином статистическом комплексе данных без предварительной стандартизации и перехода от абсолютных значений к относительным. Расчеты проведены четырьмя способами подготовки исходных данных к статистической обработке.

Первый метод наиболее простой (MinMax). Он осуществляется при нормировании отклонений исходных данных x_{ij} от минимальных значений $x_{ij} - \min(x_{ij})$, деленных на размах шкалы содержания микроэлемента $\max(x_{ij}) - \min(x_{ij})$. Для каждого случая наблюдения он рассчитывался по формуле [23]:

$$a_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}, \quad (1)$$

где x_{ij} – содержание химического элемента, $\min(x_{ij})$ – минимальное значение, $\max(x_{ij})$ – максимальное значение, a_{ij} – безразмерная величина, отражающая нормированное значение содержания элемента.

Второй метод Weight реализуется через расчет относительных вероятностей (P_{ij}) содержания химических элементов, где i – порядковый номер точки отбора пробы, j – порядковый номер химического элемента. Он является положительной безразмерной величиной с минимальным значением, равным нулю.

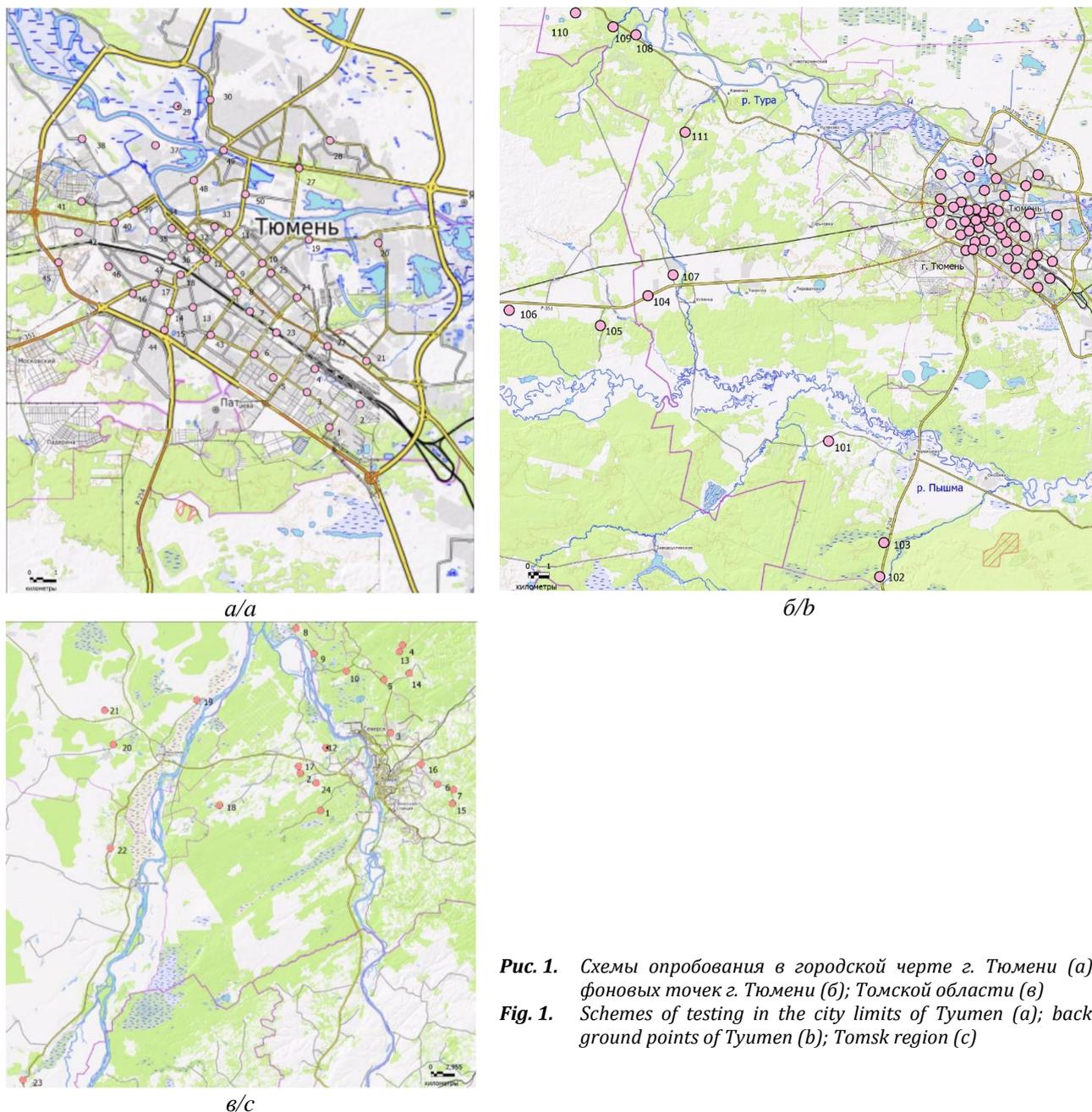


Рис. 1. Схемы опробования в городской черте г. Тюмени (а); фоновых точек г. Тюмени (б); Томской области (в)
Fig. 1. Schemes of testing in the city limits of Tyumen (a); background points of Tyumen (b); Tomsk region (c)

Его значение рассчитывается по отношению a_{ij} (1) к сумме значений содержания химических элементов ($j=1 \dots n$):

$$P_{ij} = \frac{a_{ij}}{\sum_{j=1}^n (a_{ij})}, \quad (2)$$

P_{ij} представляет доли или вероятность присутствия микроэлемента, их сумма равна единице. Известны и другие способы стандартизации. В случае, если данные имеют нормальное распределение, обычно используют отклонение от среднего значения с делением на среднеквадратичное отклонение (классическая стандартизация) (Mean):

$$Smean_{ij} = \frac{x_{ij} - mean_j}{s_j}, \quad (3)$$

где $Smean_{ij}$ центрированное относительно среднего значения, отнесенное к среднеквадратичному отклонению, x_{ij} – значение содержания элемента в точке наблюдения, $mean_j$ – среднее по j -му элементу, s_j – среднеквадратичное отклонение. Величина отношения изменяется от отрицательных значений к положительным, среднее значение стремится к нулю.

В случае, если данные существенно отклоняются от нормального распределения, что характерно для микроэлементов, значения можно центрировать

относительно медианы, соотнося с расстоянием между квантилями (Median):

$$S_{\text{median}_{ij}} = \frac{x_{ij} - \text{median}_j}{\text{Quantile } 75 - \text{Quantile } 25}, \quad (4)$$

где $S_{\text{median}_{ij}}$ – стандартизованные значения относительно медианы, нормированные на разницу между квантилями, x_{ij} – значение содержания элемента в точке наблюдения, median_j – медиана по j -му элементу, $\text{Quantile } 75$ – квантиль 75 % выборки, $\text{Quantile } 25$ – квантиль 25 %, которые изменяются от отрицательных до положительных значений.

Для каждого кластера рассчитан суммарный показатель загрязнения снеговой пыли Z_c согласно общепринятым методикам [28, 30]. Коэффициент концентрации загрязняющих веществ определен по отношению содержания микроэлементов к фоновому значению:

$$Kc_{ij} = \frac{c_{ij}}{C_{\Phi j}}, \quad (5)$$

где C_{ij} – концентрация j -микроэлемента в i -точке наблюдения, среднее значение $C_{\Phi j}$ – фоновая концентрация, которые взяты из литературных источников [3]. В случае томских проб использованы данные фоновых наблюдений лаборатории почвоведения НИИ биологии и биофизики.

Интегральный показатель загрязнения микроэлементами (Z_c) определяется по формуле:

$$Z_c = \sum_{j=1}^n Kc_{ij} - (n - 1), \quad (6)$$

где Kc_{ij} – коэффициент загрязнения, представляющий безразмерную величину, n – число учитываемых элементов с $Kc > 1,5$.

При оценке загрязнения снегового покрова показатель Z_c считается низким при значении < 32 ; средним и умеренно опасным – 32–64; высоким и опасным – 64–128; очень высоким и опасным – 128–256 [30]. Чрезвычайно опасный уровень имеет значение более 256.

Дальнейший анализ проведен нами в пяти вариантах по соответствующим формулам:

- 1) без подготовки («Data»),
- 2) «MinMax» (1),
- 3) «Weight» (2),
- 4) «Mean» (3),
- 5) «Median» (4).

Статистическая подготовка данных для иерархического анализа проведена с использованием пакетов программ Статистика, Excel как для тюменских, так и для томских данных. Для кластеризации использованы Евклидово расстояние и метод Ward в программе Orange 3.29.3. Для упрощения оценки результатов кластерного анализа установлено одинаковое число кластеров – 4.

Значения интегрального показателя загрязнения различаются для точек (5), объединенных в кластеры. Кластеры имеют отрицательные значения показателя загрязнения (6), когда отсутствуют загрязнения микроэлементами, а значение более 32 указывает на его наличие. Чем больше разница между кластерами, тем лучше работает алгоритм, отделяя загрязненные точки от сравнительно чистых. Тогда оценка экологической эффективности кластеризации определяется по разнице между максимальным и минимальным значением показателя Z_c . Оценка валидации дендрограмм иерархического кластерного анализа проведена с помощью индекса Калински–Харабаша (Calinski–Harabasz index – CH), который изменяется от 0 (худшая) до десятков и сотен (лучшая) [31].

Результаты

Идея статьи заключается в том, что для каждой точки наблюдения характерен уникальный геохимический спектр. Он является идентификатором химического состава и позволяет выделять совокупность точек разной степени загрязнения техногенными веществами. Можно найти близкие по форме геохимические спектры, но одинаковые отсутствуют. Это можно использовать для их географической привязки к местам загрязнения, а также для выявления относительно чистых фоновых территорий. Например, в центральной части Тюмени в точке наблюдения 8 (далее т. 8) обнаружен геохимический спектр, который не коррелирует по коэффициенту Спирмена ни с одним другим спектром в других точках. В т. 101, расположенной на удалении от городской черты Тюмени (фон), обнаружен аномально высокий уровень содержания загрязняющих веществ. Точка удалена из данных как артефакт при их подготовке для кластерного анализа. В Томске аналогичная ситуация сложилась с т. 16 в пяти километрах к юго-востоку от ТНХК, где в 1998 г. обнаружены высокие уровни загрязнения Ni, Co, Cu, Zn, Cr с аномально высоким содержанием Mn. Для пыли снегового покрова Тюмени в положение фона попадают точки с высокими значениями содержания микроэлементов и интегрального показателя загрязнения Z_c .

Исходные значения, используемые в геохимическом спектре без стандартизации, в масштабе существенно отличаются по величине, так, что на фоне главных элементов содержание многих микроэлементов исчезающе мало (рис. 2, а) (изменены на маленькие буквы согласно названию рисунка). Для сравнительного анализа геохимических спектров между собой необходим их перевод в относительные и стандартизованные значения (рис. 2, б, в).

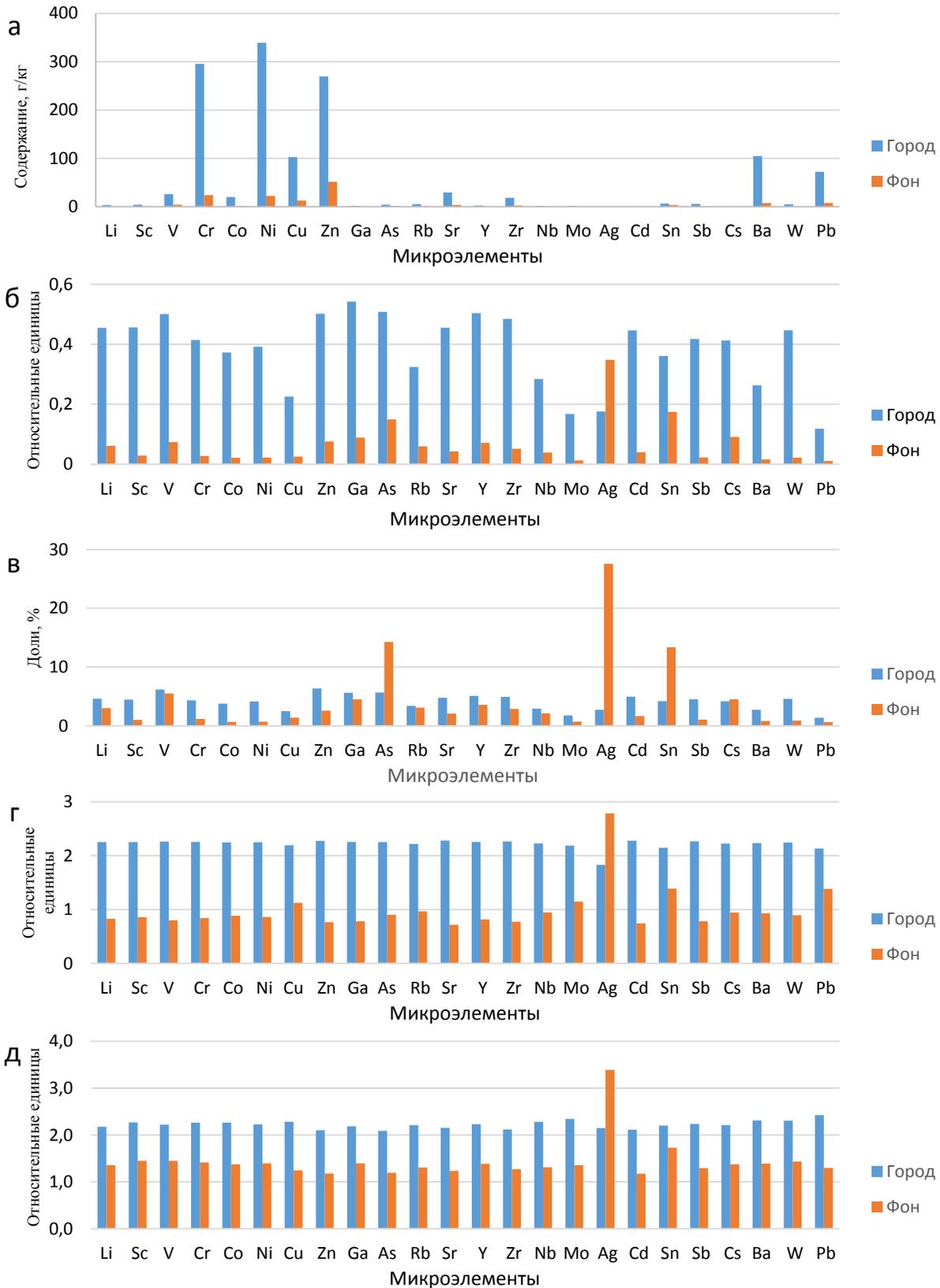


Рис. 2. Среднее содержание микроэлементов твердой фазы снега: без стандартизации, (мг/кг) (а), MinMax (1) (б), в % Weight (2) (в), Mean (3) (г), Median (4) (д) в черте г. Тюмени (Город) и на удалении от него (Фон). Для перевода отрицательных величин в положительную область значения Mean и Median увеличены на 2

Fig. 2. Average content of trace elements of the solid phase of snow: without standardization, (mg/kg) (a), MinMax (1) (б), in % Weight (2) (в), Mean (3) (г), Median (4) (д) within the city of Tyumen (City) and at a distance from it (Background). To convert negative values to a positive area, the Mean and Median values are increased by 2

Анализ исходных данных без стандартизации показывает, что преобладают основные микроэлементы, а рассеянные на их фоне не выражены на рис. 2, а. Их значения в черте города и на удалении от него неразличимы на графике. В случае стандартизации значений содержания микроэлементов пыли в снегу a_{ij} (1) весь геохимический спектр микроэлементов присутствует, и хорошо выражены те элементы, которые не видны без пересчета. Различаются геохимические спектры в городской черте и на фоновом участке, особенно для Ag и Sn (рис. 2, б). В случае стандартизованных взвешенных значений P_{ij} (2) фоновые значения превышают городские для As, Ag, Sn, Cs (рис. 1, в). Геохимический спектр таких относительных значений изменяется в более узком диапазоне, но проявляет особенности содержания одних элементов относительно других. После стандартизации величины содержания элементов выравниваются между основными и рассеянными микроэлементами как в городской черте, так и в условиях фона. Взвешивание выравнивает содержание элементов между городскими и фоновыми значениями. В случае стандартизации при центрировании относительно среднего значения (3) или медианы (4) появляется область отрицательных значений, которая делает графики не сравнимыми с предыдущими вариантами на рис. 2. Для перевода в положительную область все значения увеличены на две единицы. Стандартизация значений (3) ведет к схожести значений микроэлементов в городских условиях (рис. 2, г) за исключением Ag. Различия по величине относительно фона существенны из-за малой величины выборки фона (11 точек). В случае центрирования относительно медианы (4) геохимический спектр имеет большее разнообразие фоновых значений по сравнению с таковыми, рассчитанными по формуле (3) (рис. 1, д). При использовании для стандартизации формулы (2) наиболее наглядны различия как между содержанием микроэлементов, так и между городом и фоном.

Несмотря на то, что априори взяты четыре кластера, при разных способах стандартизации проведение сравнения дендрограмм иерархического кластерного анализа оказывается достаточно сложным. Используемый алгоритм кластеризации ранжирует кластеры по величинам, присваивая им номер, и оценивает их близость на основе Евклидова расстояния между ними, располагая кластеры с высокими и низкими значениями по своему алгоритму. При выделении дендрограмм по форме геохимических спектров получаем кластеры, которые объединяют высокие, средние и низкие содержания микроэлементов, случайно присваивая им номера кластеров (С1–С4) в разном порядке для методов стандартизации. В базе тюменских данных имеют-

ся девять точек, которые можно использовать как опорные для оценки фоновых содержаний микроэлементов. В них ожидается существенно более низкое содержание микроэлементов в твердой фазе снега. Использование четырех способов стандартизации данных на основе одинаковых значений исходного содержания микроэлементов дает разные результаты иерархического кластерного анализа (рис. 3).

В случае исходных данных без подготовки (рис. 3, а) кластер С1 включает в себя все фоновые точки наблюдения и многие городские, в которых тоже очень низкое содержание микроэлементов. Этот кластер максимально удален от остальных. В кластере С2 объединяются точки вблизи автомагистралей и дорог. Наиболее высоким содержанием Cr и Ni отличается т. 31. Кластер С3 включает в себя две точки (т. 23 и т. 33) в зоне воздействия аккумуляторного завода, которые выделяются максимальным содержанием Ni и Cu. Кластер С4 содержит загрязненные точки наблюдения (т. 40) с высоким содержанием Pb. Он занимает городские территории вдоль дорог и попадает в промежуточное положение между точками со сравнительно низким содержанием микроэлементов в С1 и их очень высокими значениями в С2 и С3.

Данные, посчитанные по формуле (1) (MinMax), содержат фоновые точки в кластере С4, тогда как точки с высоким содержанием микроэлементов т. 23 и т. 33 объединены с другими подобными в С1. Кластер С2 нагружен тяжелыми металлами (Ni, Zn и Cr), тогда как С1 имеет повышенное содержание рассеянных элементов (рис. 3, б). Кластер С3 попадает в промежуточное положение между кластерами с промышленным загрязнением и фоновыми точкам. Его точки отличаются некоторым удалением от аккумуляторного завода. Кластер С3 схож с кластером С1 по отношению к исходным данным без стандартизации. В кластер С4 входят фоновые точки и точки внутри городской черты с низким содержанием микроэлементов.

При использовании для подготовки данных метода Weight (2) в результатах кластерного анализа точки делятся на две большие группы: в одну попадают фоновые точки и низким содержанием микроэлементов С1 и С2, в другую – как фоновые, так и городские точки с высоким содержанием микроэлементов (рис. 3, в). Отдельным кластером С3 выделилась точка 8, геохимический спектр которой не коррелирует ни с одной другой изученной точкой. Он имеет высокие содержания Mo, Ag, Cd и Sn. В итоге наиболее детально на кластеры разложены фоновые участки, тогда как все городские точки, подверженные воздействию, сосредоточены в одном кластере С4.

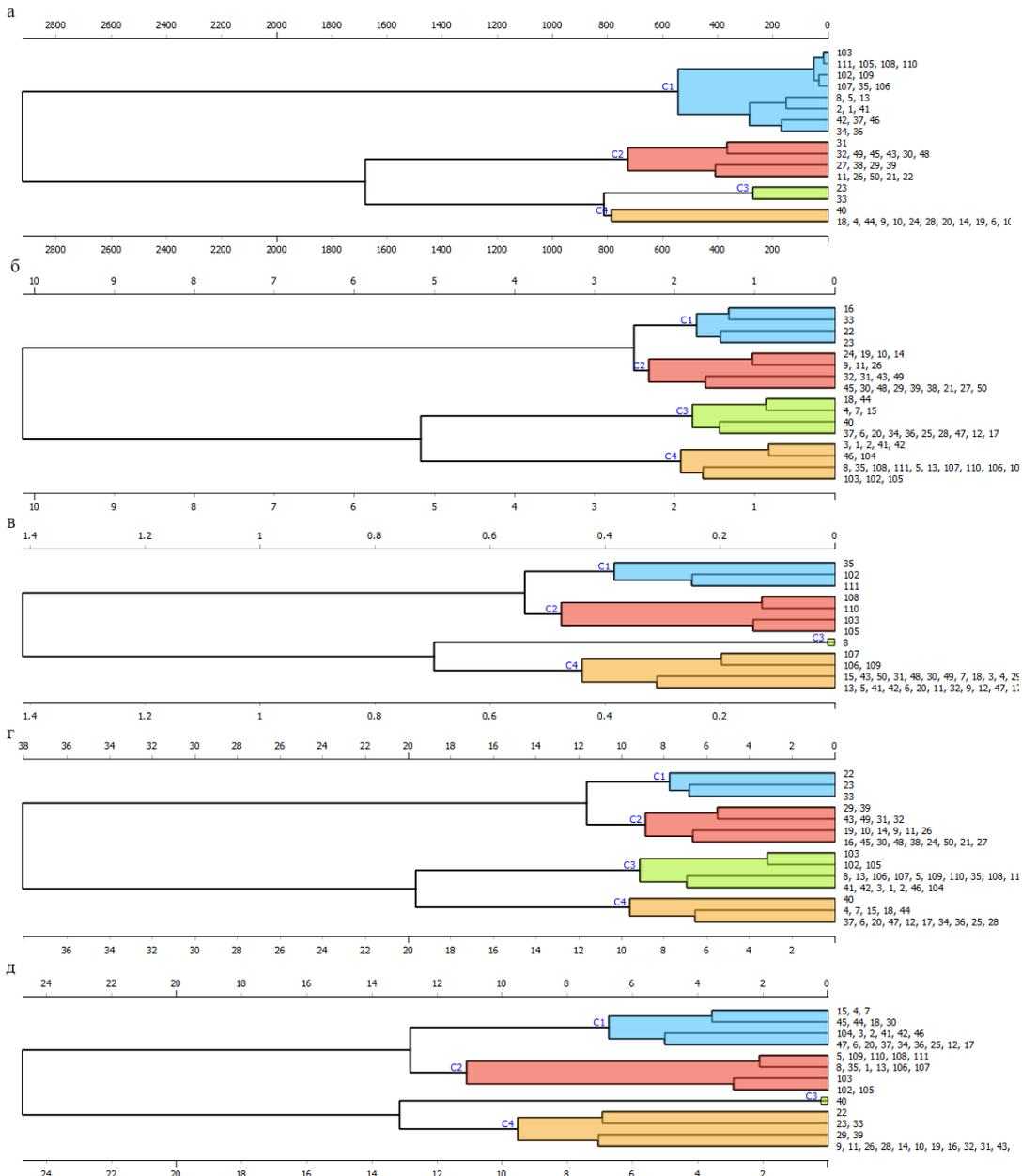


Рис. 3. Кластеризация исходных и стандартизованных данных: без стандартизации (а), MinMax (1) (б), взвешенные значения Weight (2) (в), стандартная нормализация Mean (3) (г) и Median (4) (д) точек наблюдения в черте г. Тюмени – 1–50, на удалении (фон) – 102–110

Fig. 3. Clustering of initial and standardized data: without standardization (a), MinMax (1) (б), weighted values of Weight (2) (в), standard normalization of Mean (3) (г) and Median (4) (д) observation points within the city of Tyumen – 1–50, at a distance (background) – 102–110

Использование способа «Mean» (3) приводит также к разделению данных на две большие группы: C1 и C2 с высоким содержанием тяжелых металлов, а также C3 и C4 с низким их содержанием (рис. 3, г). Фоновые точки и городские с низким содержанием микроэлементов попадают в кластер C3, а C4 занимает промежуточное положение по уровню загрязнения пыли снегового покрова техногенными выбросами, включает т. 40 с высоким содержанием Pb.

Содержание микроэлементов в большинстве случаев не соответствует закону нормального распределения случайных величин, поэтому наиболее предпочтительным способом их предварительного преобразования является метод «Median» (4) (рис. 3, д). В результате кластерного анализа выделены группы C1 и C2, включающие фоновые точки, удаленные от города, и т. 35, расположенную в пойме р. Туры. В C3 попала одна единственная уникальная т. 40. Кластер C4 объединяет точки с

повышенным уровнем загрязнения тяжелыми металлами и рассеянными микроэлементами.

Аналогичные расчеты проведены для территории Томской области. Большинство дендрограмм для рассмотренных методов подготовки данных к иерархическому кластерному анализу схожи по своему виду. Так как томские данные имеют вспомогательный характер относительно тюменских, описание их дендрограмм опускаем, чтобы не перегружать текст, а ограничиваемся оценкой интегрального показателя загрязнения и валидности кластеризации.

Обсуждение

Стандартизация геохимических данных – это приведение их к некому виду, позволяющему сравнивать микроэлементы, качественно различающиеся по содержанию химических свойств в геохимическом спектре, и сами спектры между собой. Для этого используют стандартные методы. Геохимические данные имеют специфику, выраженную в степенном распределении случайной величины содержания микроэлементов. Возникает необходимость использовать непараметрические методы, основанные на ранжировании исходных значений.

Данные микроэлементного состава пыли снегового покрова Тюмени, стандартизованные по формулам (1)–(4), были оценены по двум направлениям: 1) экологическая эффективность по разнице между максимальными и минимальными значениями показателя загрязнения, 2) качество кластеризации (валидность). В первом случае использовали интегральный показатель загрязнения Z_c (6), который применяют в отечественной научной литературе при экологической оценке территорий, подверженных антропогенному загрязнению микроэлементами (табл. 1). Объективность показателя загрязнения снегового покрова обусловлена его широким применением [30]. Расположение точек экологически опасных кластеров на изученной территории можно использовать при ее районировании.

Валидация дендрограмм кластеризации оценивается с помощью индекса СН. Он оценивает качество разделения кластеров на основе отношения внутренней неоднородности данных (дисперсии) к дисперсии кластеров относительно среднего значения выборки. Чем меньше дисперсия внутри кластера и больше между общим центром и центрами кластеров, тем выше значения индекса, тем более качественно проведена кластеризация.

В Тюмени наибольшее различие между максимальным и минимальным значениями Z_c отмечается без использования методов стандартизации. Высокую экологическую эффективность кластеризации показали методы (4) и (2). Метод (4) дает хорошие результаты оценки экологической эффективности с выделением двух экологически опасных кластеров

С3 ($Z_c=71,3$) и С2 ($Z_c=66,4$). Метод (2) выявляет сравнительно безопасную экологическую ситуацию для всех кластеров ($Z_c<32$), но различия между кластерами не столь велики в сравнении с другими методами. Наихудшие результаты получены при использовании методов (2) и (3). Максимально высокая оценка валидности кластеризации получена при использовании метода (4) Median ($C-H=82$). Следующими по величине индекса являются исходные данные без их стандартизации ($C-H=44$).

Таблица 1. Значение интегрального показателя загрязнения Z_c в кластерах с использованием разных методов стандартизации и нормирования, различия между максимальным и минимальными значениями в кластерах С1–С4 (Max-Min), валидность оценена с помощью индекса СН

Table 1. Value of the integral indicator of Z_c pollution in clusters using different methods of standardization and rationing, differences between the maximum and minimum values in clusters С1–С4 (Max-Min), validity was estimated using the CH index

Метод стандартизации (формула) Standardization method (formula)	Кластеры Clusters				Мин-Макс Max-Min	Индекс СН CH Index
	C1	C2	C3	C4		
Тюмень/Тyumen (N=60)						
Исходные данные (мг/кг) Data (mg/kg)	-12,5	65,2	75,4	26,6	87,9	44
$\frac{x_{ij}-\min(x_{ij})}{\max(x_{ij})-\min(x_{ij})}$ (1)	-15,5	-6,8	62,3	29,3	77,8	36
$\frac{a_{ij}}{\sum_{j=1}^n(a_{ij})}$ (2)	-13,7	-13,3	15,7	28,0	41,7	12
$\frac{x_{ij}-\text{mean}_j}{s}$ (3)	-13,7	-7,5	53,3	32,7	67	39
$\frac{x_{ij}-\text{median}_j}{\text{Quantile } 75-\text{Quantile } 25}$ (4)	-11,3	66,4	71,3	32,7	82,6	82
Томск/Tomsk (N=58)						
Исходные данные (мг/кг) Data (mg/kg)	82,4	1,9	24,8	39,6	80,5	45
$\frac{x_{ij}-\min(x_{ij})}{\max(x_{ij})-\min(x_{ij})}$ (1)	25,9	40,2	-1,7	21,8	41,9	14
$\frac{a_{ij}}{\sum_{j=1}^n(a_{ij})}$ (2)	-2,0	3,9	39,0	38,7	41	16
$\frac{x_{ij}-\text{mean}_j}{s}$ (3)	83,4	25,9	31,9	2,4	81	44
$\frac{x_{ij}-\text{median}_j}{\text{Quantile } 75-\text{Quantile } 25}$ (4)	83,4	44,8	-0,7	30,0	84,1	46

Использованные показатели (Z_c и С-Н) зависят от структуры данных, поэтому в дополнение проведена аналогичная подготовка данных по формулам (1)–(4) для Томской области. На основании оценки экологической эффективности (Z_c) алгоритмов кластеризации содержания микроэлементов пыли снегового покрова в Томской области наиболее успешными оказались методы: Median, данные без стандартизации и Mean. Значения Z_c для некоторых кластеров находятся в пределах 82–83, что указывает на высо-

кий и опасный уровень загрязнения. Наибольшая разница между кластерами отмечается для метода Median (4), наихудшая – для Weight. Кластеризация данных, подготовленных методом Median, имеет наибольшую валидность (46), как и в случае тюменских данных. Наименьшая валидность отмечается для метода MinMax (1) – 14. Данные анализа без использования методов стандартизации показали высокую валидность – 45.

Для стандартизации данных содержания микроэлементов снеговой пыли рекомендуем метод по формуле (4) (Median).

Выводы

1. Показаны наиболее простые и эффективные методы предварительной обработки и стандартизации данных содержания микроэлементов пыли снегового покрова для кластерного анализа. Их использование положительно сказывается на сопоставлении данных зоны воздействия антропогенных факторов и фона в случае графического отображения геохимического спектра.
2. Кластеризация позволила выделить группы точек наблюдения с разной интенсивностью воздействия антропогенных факторов. Максимальную эффективность оценки экологической опасности в Тюмени и Томской области на основе интегрального показателя Z_c выявили методы стандартизации: 1) различие между исходными значениями и медианой, отнесенное к разности квартилей 0,25 и 0,75 (Median), 2) отклонения от средней величины, отнесенной к стандартному отклонению (Mean), 3) исходные данные без стандартизации (Data). Валидность, оцененная по индексу Калински–Харабаша, позволила установить, что наилучшая кластеризация происходит при стандартизации с использованием медианы (Median).
3. Сравнительный анализ показателей для Томской и Тюменской областей дал в общих чертах сходные результаты, что позволяет рекомендовать использование медианы (Median) для стандартизации данных химического состава пыли снегового покрова.

СПИСОК ЛИТЕРАТУРЫ

1. Кластерный анализ стандартизованных значений содержания микроэлементов твердой фазы снега городов Тобольск и Тюмень / А.В. Захарченко, А.А. Тигеев, О.А. Пасько, Д.В. Московченко // Экология и промышленность России. – 2023. – Т. 27. – № 4. – С. 61–65. DOI: 10.18412/1816-0395-2023-4-61-65.
2. Иванов В.Б., Мухаметдинова Э.А., Королик В.С. Распределение загрязнения тяжелыми металлами в снежном покрове г. Нижневартовск // Вестник ТюмГУ. – 2010. – № 3. – С. 148–153.
3. Геохимическая характеристика снежного покрова Тюмени / Д.В. Московченко, Р.Ю. Пожитков, А.Н. Курчатова, Р.И. Тимшанов // Вестник московского университета. Серия 5. География. – 2021. – № 3. – С. 13–26.
4. Dissolution and solubility of trace metals from natural and anthropogenic aerosol particulate matter / K. Desboeufs, A. Sofikitis, R. Losno, J. Colin, P. Ausset // Chemosphere. – 2005. – Vol. 58. – P. 195–203. DOI: 10.1016/j.chemosphere.2004.02.025.
5. Pizarro J., Vergara P.M., Cerdá S. Contaminant emissions as indicators of chemical elements in the snow along a latitudinal gradient in southern Andes // Scientific reports. – 2021. – Vol. 11. – № 1. – Article no. 14530. DOI: 10.1038/s41598-021-93895-1.
6. Быков Н.И., Попов Е.С. Наблюдения за динамикой снежного покрова в ООПТ Алтае-Саянского экорегиона (методическое руководство). – Красноярск: Наука, 2011. – 64 с.
7. Rudnick R.L., Gao S. Composition of the continental crust // Treatise on Geochemistry. – 2003. – Vol. 3. – P. 1–64. DOI: 10.1016/B978-0-08-095975-7.00301-6.
8. The atmospheric transport and deposition of smelter emissions: evidence from multi element geochemistry of snow, Quebec, Canada / K. Telmer, G.F. Bonham-Carter, D.A. Kliza, G.E.M. Hall // Geochimica et Cosmochimica Acta. – 2004. – Vol. 68. – № 14. – P. 2961–2980. DOI: 10.1016/j.gca.2003.12.022.
9. Estimating pollution loads in snow removed from a port facility: snow pile sampling strategies / A. Vijayan, H. Österlund, J. Marsalek, M. Viklander // Water, Air, & Soil Pollution. – 2021. – Vol. 232. – № 2. – Article no. 75. DOI: 10.1007/s11270-021-05002-9.
10. Таловская А.В. Экогеохимия атмосферных аэрозолей на урбанизированных территориях юга Сибири: автореф. дис. ... д-ра. геол.-минерал. наук. – Томск, 2022. – 46 с.
11. Адильбаева Т.Е., Таловская А.В., Языков Е.Г. Применение статистических методов анализа эколого-геохимических измерений содержания элементов в снеговом покрове в районе расположения теплоэлектростанции (г. Караганда, Республика Казахстан) // Снежный покров, атмосферные осадки, аэрозоли: Материалы V Байкальской международной научной конференции – стратегической сессии. – Иркутск, 19–23 июня, 2023. – Иркутск: Репроцентр А1, 2023. – С. 140–144.
12. Московченко Д.В., Бабушкин А.Г. Особенности формирования химического состава снежного покрова на территории Ханты-Мансийского автономного округа // Криосфера Земли. – 2012. – Т. XVI. – № 1. – С. 71–81.
13. Пожитков Р.Ю., Московченко Д.В., Кудрявцев А.А. Геохимия снежного покрова г. Нижневартовска // Вестник Тюменского государственного университета. Экология и природопользование. – 2018. – Т. 4. – № 1. – С. 6–24. DOI: 10.21684/2411-7927-2018-4-1-6-24
14. Moskovchenko D.V., Pozhitkov R.Yu., Tigeev A.A. Assessment of the concentration of metals and metalloids in the snow cover at oil production sites in the Middle Ob Region // Water Resources. – 2023. – Vol. 50. – № S1. – P. S57–S67. DOI: 10.1134/s009780782370032x.
15. Sidak J.G. The value of a standard versus the value of standardization // Baylor Law Review. – 2016. – Vol. 68. – № 1. – P. 26.
16. Gal M., Rubinfeld D.L. Data standardization // NYU Law Review. – 2019. – Vol. 94. – № 4. – 30 p. DOI: 10.2139/ssrn.3326377
17. Assessment of snow cover pollution of the mining area. / Y.T. Suyundukov, I.N. Semenova, R.F. Khasanova, Y.S. Rafikova, M.B. Suyundukova, G.R. Ilbulova, E.E. Papyan // IOP Conference Series: Materials Science and Engineering. 2019. – Vol. 663. DOI: 10.1088/1757-899X/663/1/01205.

18. Templ M., Filzmoser P., Reimann C. Cluster analysis applied to regional geochemical data: problems and possibilities // *Applied Geochemistry*. – 2008. – Vol. 23. – № 8. – P. 2198–2213. DOI: <https://doi.org/10.1016/j.apgeochem.2008.03.004>.
19. Factor and cluster analysis of water quality data of the groundwater wells of Kushtia, Bangladesh: implication for arsenic enrichment and mobilization / Md. Golzar Hossain, A.H.M. Selim Reza, Mst. Lutfun-Nessa, Syed Samsuddin Ahmed // *Journal of Geological Society of India*. – 2013. – Vol. 81. – № 3. – P. 377–384. DOI: 10.1007/s12594-013-0048-0.
20. Прохоренков П.А., Репер Т.В., Гудкова Н.В. Методы кластерного анализа в региональных исследованиях // *Фундаментальные исследования*. – 2022. – № 3. – С. 100–106. DOI: 10.17513/fr.43221.
21. Гололобова А.Г. Вариационная статистика при эколого-геохимических исследованиях почв, подверженных техногенному воздействию // *Проблемы региональной экологии*. – 2022. – № 5. – С. 54–59. DOI: 10.24412/1728-323X-2022-5-54-59.
22. Zhang W., Xu H., Zheng F. Classifying aerosols based on fuzzy clustering and their optical and microphysical properties study in Beijing, China // *Advances in Meteorology*. – 2017. – P. 1–18. DOI: 10.1155/2017/4197652.
23. Гилязов А.Ф. Кластерный анализ как инструмент районирования территории по крупности речных наносов (на примере бассейна Волги) // *Вестник Удмуртского университета. Биология. Науки о Земле*. – 2015. – Т. 25. – Вып. 2. – С. 149–158.
24. Корячко В.П., Орешков В.И. Выбор числа кластеров в задачах кластеризации с использованием метода силуэтов // *BIG DATA и анализ высокого уровня: сборник научных статей IX Международной научно-практической конференции*. – Минск, 17–18 мая 2023. – Минск: Белорусский государственный университет информатики и радиоэлектроники, 2023. – С. 333–340.
25. Журавлева В.В., Маничева А.С. Упрощенный показатель силуэта для определения качества кластерных структур // *Известия АлтГУ. Математика и механика*. – 2022. – № 4 (129) – С. 110–114. DOI: [https://doi.org/10.14258/izvasu\(2022\)4-17](https://doi.org/10.14258/izvasu(2022)4-17).
26. Concentrations of major and trace elements within the snowpack of Tyumen, Russia / D.V. Moskovchenko, R.Yu. Pozhitkov, A.V. Zakharchenko, A.A. Tigeev // *Minerals*. – 2021. – Vol. 11. – № 7. – P. 1–21. DOI: 10.3390/min11070709.
27. Пространственное распределение геохимических компонентов снегового покрова на удалении от Томск-Северской промышленной агломерации / А.В. Захарченко, А.А. Тигеев, О.А. Пасько, Л.Г. Колесниченко, Д.В. Московченко // *Криосфера Земли*. – 2021. – Т. XXV. – № 6. – С. 16–27. DOI: 10.15372/KZ20210602.
28. МР 5174-90. Методические рекомендации по оценке степени загрязнения атмосферного воздуха населенных пунктов металлами по их содержанию в снежном покрове и почве. – М.: ИМГРЭ, 1990. – 16 с.
29. Impact of snow deposition on major and trace element concentrations and elementary fluxes in surface waters of the Western Siberian Lowland across a 1700 km latitudinal gradient / V.P. Shevchenko, I.V. Krickov, R.M. Manasypov, N.V. Politova, S.G. Kopysov, O.M. Dara, Y.V. Auda, L.S. Shirokova, L.G. Kolesnichenko, V.A. Zemtsov, S.N. Kirpotin // *Hydrol. Earth Syst. Sci.* – 2017. – Vol. 21 (11). P. – 5725–5746. DOI: 10.5194/hess-21-5725-2017.
30. Геохимия снежного покрова в восточном округе Москвы / Н.С. Касимов, Н.Е. Кошелева, Д.В. Власов, Е.В. Терская // *Вестник Московского университета. Серия 5. География*. – 2012. – № 4. – С. 14–24.
31. Calinski-Harabasz Index for K-means clustering evaluation using Python // *Pyshark.com*. – URL: <https://pyshark.com/calinski-harabasz-index-for-k-means-clustering-evaluation-using-python/> (дата обращения 08.09.2023).

Информация об авторах

Александр Викторович Захарченко, доктор биологических наук, главный научный сотрудник, Тюменский научный центр СО РАН, структурное подразделение Институт проблем освоения Севера, Россия, 625026, г. Тюмень, ул. Малыгина, 86. avzakh@gmail.com; <https://orcid.org/0000-0001-9201-5733>

Александр Анатольевич Тигеев, кандидат географических наук, старший научный сотрудник, Тюменский научный центр СО РАН, структурное подразделение Институт проблем освоения, Россия, 625026, г. Тюмень, ул. Малыгина, 86. trruubbaa@mail.ru; <https://orcid.org/0000-0001-5449-5086>.

Поступила в редакцию: 26.06.2024

Поступила после рецензирования: 18.12.2024

Принята к публикации: 17.07.2025

REFERENCES

1. Zakharchenko A.V., Tigeev A.A., Pasko O.A., Moskovchenko D.V. On the background values of the geochemical spectra of microelements of the solid phase of snow during their comparative analysis in the cities of Tobolsk and Tyumen. *Ecology and Industry of Russia*, 2023, vol. 24, no. 4, pp. 61–65. (In Russ.). DOI: 10.18412/1816-0395-2023-4-61-65.
2. Ivanov V.B., Mukhametdinova E.A., Korolik V.S. Allocations of heavy metal pollution in the snow cover of Nizhnevartovsk City. *Tyumen State University Herald. Natural Resource Use and Ecology*, 2010, no. 3, pp. 148–153. (In Russ.)
3. Moskovchenko D.V., Pozhitkov R.Yu., Kurchatova A.N., Timshanov R.I. Geochemistry of snowcover within the Tyumen City. *Moscow University Bulletin. Series 5, Geography*, 2021, vol. 3, pp. 13–26. (In Russ.)
4. Desboeufs K., Sofikitis A., Losno R., Colin J., Ausset P. Dissolution and solubility of trace metals from natural and anthropogenic aerosol particulate matter. *Chemosphere*, 2005, vol. 58, pp. 195–203. DOI: 10.1016/j.chemosphere.2004.02.025.
5. Pizarro J., Vergara P.M., Cerda S. Contaminant emissions as indicators of chemical elements in the snow along a latitudinal gradient in southern Andes. *Scientific reports*, 2021, vol. 11, no. 1, Article no. 14530. DOI: 10.1038/s41598-021-93895-1.
6. Bykov N.I., Popov E.S. *Observations on the dynamics of snow cover in protected areas of the Altai-Sayan ecoregion (methodological guide)*. Krasnoyarsk, Nauka Publ., 2011. 64 p. (In Russ.)
7. Rudnick R.L., Gao S. Composition of the continental crust. *Treatise on Geochemistry*, 2003, vol. 3, pp. 1–64. DOI: 10.1016/B978-0-08-095975-7.00301-6.
8. Telmer K., Bonham-Carter G.F., Kliza D.A., Hall G.E.M. The atmospheric transport and deposition of smelter emissions: evidence from multi element geochemistry of snow, Quebec, Canada. *Geochimica et Cosmochimica Acta*, 2004, vol. 68, no. 14, pp. 2961–2980. DOI: 10.1016/j.gca.2003.12.022.

9. Vijayan A., Österlund H., Marsalek J., Viklander M. Estimating pollution loads in snow removed from a port facility: snow pile sampling strategies. *Water, Air, & Soil Pollution*, 2021, vol. 232, no. 2, Article no. 75. DOI: 10.1007/s11270-021-05002-9.
10. Talovskaya A.V. *Ecogeochemistry of atmospheric aerosols in urbanized territories of southern Siberia*. Dr. Diss. Abstract. Tomsk, 2022. 46 p. (In Russ.)
11. Adilbayeva T.E., Talovskaya A.V., Yazikov E.G. Snow cover, atmospheric precipitation, aerosols: chemistry and climate. *V Baikal international scientific conference – strategic session endorsed*. Irkutsk, June 19–23, 2023. Irkutsk, Ltd «Reprocenter A1», 2023. pp. 224. (In Russ.)
12. Moskovchenko D.V., Babushkin A.G. Peculiarities of formation of chemical composition of snow waters (on example of khanty-mansi autonomous district. *Earth's Cryosphere*, 2012, vol. 16, no. 1, pp. 71–81. (In Russ.)
13. Pozhitkov R.Yu., Moskovchenko D.V., Kudryavtsev A.A. The geochemistry of snow cover in Nizhnevartovsk. *Tyumen State University Herald. Natural Resource Use and Ecology*, 2018, vol. 4, no. 1, pp. 6–24. (In Russ.) DOI: 10.21684/2411-7927-2018-4-1-6-24.
14. Moskovchenko D.V., Pozhitkov R.Yu., Tigeev A.A. Assessment of the concentration of metals and metalloids in the snow cover at oil production sites in the Middle Ob Region. *Water Resources*, 2023, vol. 50, no. S1, pp. S57–S67. DOI: 10.1134/s009780782370032x.
15. Sidak J.G. The value of a standard versus the value of standardization. *Baylor Law Review*, 2016, vol. 68, no. 1, pp. 26.
16. Gal M., Rubinfeld D.L. Data standardization. *NYU Law Review*, 2019, vol. 94, no. 4, 30 p. DOI: 10.2139/ssrn.3326377.
17. Suyundukov Y.T., Semenova I.N., Khasanova R.F., Rafikova Y.S., Suyundukova M.B., Ilbulova G.R., Papyan E.E. Assessment of snow cover pollution of the mining area. *IOP Conference Series: Materials Science and Engineering*, 2019. Vol. 663. DOI: 10.1088/1757-899X/663/1/01205.
18. Templ M., Filzmoser P., Reimann C. Cluster analysis applied to regional geochemical data: problems and possibilities. *Applied Geochemistry*, 2008, vol. 23, no. 8, pp. 2198–2213. DOI: <https://doi.org/10.1016/j.apgeochem.2008.03.004>.
19. Golzar Hossain Md., Selim Reza A.H.M., Lutfun-Nessa Mst., Syed Samsuddin Ahmed. Factor and cluster analysis of water quality data of the groundwater wells of Kushtia, Bangladesh: implication for arsenic enrichment and mobilization. *Journal of Geological Society of India*, 2013, vol. 81, no. 3, pp. 377–384. DOI: 10.1007/s12594-013-0048-0.
20. Prokhorenkov P.A., Reger T.V., Gudkova N.V. Methods of cluster analysis in regional studies. *Fundamental Research*, 2022, vol. 3, pp. 100–106. (In Russ.) DOI: 10.17513/fr.43221.
21. Gololobova A.G. Variation statistics in ecological and geochemical studies of soils causing technogenic pollution. *Regional Environmental Issues*, 2022, vol. 5, pp. 54–59. (In Russ.) DOI: 10.24412/1728-323X-2022-5-54-59.
22. Zhang W., Xu H., Zheng F. Classifying aerosols based on fuzzy clustering and their optical and microphysical properties study in Beijing, China. *Advances in Meteorology*, 2017, pp. 1–18. DOI: 10.1155/2017/4197652.
23. Gilyazov A.F. Cluster analysis as a tool of zoning according to the size of river sediments (by the example of Volga basin). *Bulletin of Udmurt University. Series Biology Earth Sciences*, 2015, vol. 25, no. 2, pp. 149–158. (In Russ.)
24. Koryachko V.P., Oreshkov V.I. Choice of the number of clusters in clustering problems using the silhouette method. *BIG DATA and Advanced Analytics*. Minsk, May 17–18, 2023. Minsk, Belarusian State University of Informatics and Radioelectronics Publ., 2023. pp. 333–340. (In Russ.)
25. Zhuravleva V.V., Manicheva A.S. Simplified silhouette parameter for assessing the quality of cluster structures. *Altai State University*, 2022, no. 4 (129), pp. 110–114. (In Russ.) DOI: [https://doi.org/10.14258/izvasu\(2022\)4-17](https://doi.org/10.14258/izvasu(2022)4-17).
26. Moskovchenko D.V., Pozhitkov R.Yu., Zakharchenko A.V., Tigeev A.A. Concentrations of major and trace elements within the snowpack of Tyumen, Russia. *Minerals*, 2021, vol. 11, no. 7, pp. 1–21. DOI: 10.3390/min11070709.
27. Zakharchenko A.V., Tigeev A.A., Pasko O.A., Kolesnichenko L.G., Moskovchenko D.V. Spatial distribution of geochemical characteristics of snow cover within and outside Tomsk-Seversk industrial agglomeration. *Earth's Cryosphere*, 2021, vol. 25, no. 6, pp. 16–27. (In Russ.) DOI: 10.18412/1816-0395-2023-4-61-65.
28. *MR 5174-90. Methodological recommendations for assessing the degree of atmospheric air pollution of settlements with metals based on their content in snow cover and soil*. Moscow, IMGRE, 1990. 16 p. (In Russ.)
29. Shevchenko V.P., Krickov I.V., Manasyrov R.M., Politova N.V., Kopysov S.G., Dara O.M., Auda Y.V., Shirokova L.S., Kolesnichenko L.G., Zemtsov V.A., Kirpotin S.N. Impact of snow deposition on major and trace element concentrations and elementary fluxes in surface waters of the Western Siberian Lowland across a 1700 km latitudinal gradient. *Hydrol. Earth Syst. Sci.*, 2017, vol. 21, no. 11, pp. 5725–5746. DOI: 10.5194/hess-21-5725-2017.
30. Kasimov N.S., Vlasov D.V., Kosheleva N.E., Nikiforova E.M. Geochemistry of snow cover within the eastern district of Moscow. *Moscow University Bulletin. Series 5, Geography*, 2012, vol. 4, pp. 14–24.
31. Calinski-Harabasz Index for K-Means Clustering Evaluation using Python. *Pyshark.com*. Available at: <https://pyshark.com/calinski-harabasz-index-for-k-means-clustering-evaluation-using-python/> (accessed 8 September 2023).

Information about the authors

Alexander V. Zakharchenko, Dr. Sc., Chief Researcher, Tyumen Scientific Center SB RAS, structural unit Institute of the Problems of Northern Development, 86, Malygin street, Tyumen 625026, Russian Federation. avzakh@gmail.com; <https://orcid.org/0000-0001-9201-5733>.

Alexander A. Tigeev, Cand. Sc., Senior Researcher, Tyumen Scientific Center SB RAS, structural unit Institute of the Problems of Northern Development, 86, Malygin street, Tyumen 625026, Russian Federation. ttrruubaa@mail.ru; <https://orcid.org/0000-0001-5449-5086>.

Received: 26.06.2024

Revised: 18.12.2024

Accepted: 17.07.2025