

Подходы для оценки и управления качеством компонентов урбоэкосистемы с помощью методов нейросетевого моделирования

© 2019. Ю. А. Тунакова¹, д. х. н., профессор, зав. кафедрой,
С. В. Новикова¹, д. т. н., профессор,
А. Р. Шагидуллин², к. ф.-м. н., с. н. с., В. С. Валиев², с. н. с.,
А. Х. Мораис³, профессор, зав. кафедрой,

¹Казанский национальный исследовательский технический университет
им. А. Н. Туполева (КАИ),

420126, Россия, г. Казань, ул. Четаева, д. 18,

²Институт проблем экологии и недропользования АН РТ,
420087, Россия, г. Казань, ул. Даурская, д. 28,

³Лиссабонский университет,

1349-063, Португалия, г. Лиссабон, ул. Са Нугейра, д. 1,

e-mail: juliaprof@mail.ru, artur.shagidullin@tatar.ru, ajmorais@iol.pt

Разработана методология оценки и управления качеством компонентов урбоэкосистемы, основанная на использовании новых информационных методов, обоснования маркерных биологических объектов и способов оценки их отклика на антропогенное воздействие, путей свёртки массивов данных, используемых для оценки. В представленной методологии учтены процессы поступления, накопления и отклика биологических объектов на воздействие маркерных веществ. Представлен подход для одновременного учёта и анализа разнородных и разноразмерных данных, имеющих различную степень влияния на состояние исследуемых компонентов урбоэкосистемы с использованием метода кластерного анализа. Зонирование территории урбоэкосистемы осуществлялось с помощью нейронных самоорганизующихся сетей Кохонена, реализованных в авторской модели нейросетевой фильтрации. С использованием вероятностных подходов предложен способ определения вероятностей превышения пороговых значений концентраций маркерных веществ-металлов в исследуемых компонентах урбоэкосистемы. Оценивалась вероятность того, что на территории, к которой относится то или иное наблюдение, будет отмечаться превышение заданного порога хотя бы одним из исследуемых показателей. В качестве порогов этих событий приняты медианы непрерывных рядов наблюдений концентраций металлов. В результате нами были рассчитаны вероятности превышения в каждой зоне половины исследуемых параметров заданных для них порогов. Факты превышения установленных таким образом порогов (события) регистрировались во всей совокупности наблюдений, выделенных в виде кластеров, сформированных с помощью нейросетевой классификации. При этом пространственное распределение выбранных наблюдений, имеющих конкретные адресные привязки, формирует соответствующие территориальные зоны. Даны рекомендации по использованию полученных результатов для оперативного и планового управления качеством компонентов урбоэкосистемы на локальном уровне.

Ключевые слова: урбоэкосистема, оценка риска, отклик организма человека, биосубстраты, нейросетевые технологии, металлы.

Approaches for evaluation and quality management of urban ecosystem components using neural network modeling methods

© 2019. Y. A. Tunakova¹ ORCID: 0000-0002-8826-8639[?]

S. V. Novikova¹ ORCID: 0000-0001-8207-1010[?], A. R. Shagidullin² ORCID: 0000-0003-3837-6110[?]

V. S. Valiev² ORCID: 0000-0002-8848-5326[?], A. J. Morais³ ORCID: 0000-0002-5673-0315[?]

¹Kazan National Research Technical University named after A. N. Tupolev – KAI,
18, Chetaeva St., Kazan, Russia, 420126,

²Research Institute for Problems of Ecology and Mineral Wealth Use
of Tatarstan Academy of Sciences,
28, Daur'skaya St., Kazan, Russia, 420087,

³University of Lisbon,

1, Sa Nugeira St., Lisbon, Portugal, 1348-063,

e-mail: juliaprof@mail.ru, artur.shagidullin@tatar.ru, ajmorais@iol.pt

The methodology of evaluation and quality management of components of the urban ecosystem, based on the use of new information methods, substantiation of marker biological objects and methods for assessing their response to anthropogenic impact, and ways of data arrays folding used for evaluation has been developed. The presented methodology takes into account the processes of receipt, accumulation and response of biological objects to the effects of marker substances. The approach for the simultaneous recording and analysis of heterogeneous and different-sized data having different degrees of influence on the state of the studied urban ecosystem components using the cluster analysis method is presented. The territory of the urban ecosystem was zoned using the neural self-organizing networks of Kohonen, implemented in the author's model of neural network filtering. A method for determining the probabilities of exceeding the threshold values of concentrations of marker substances-metals in the studied components of the urban ecosystem is proposed. We estimated the probability that in the territory to which a particular observation belongs, an excess of a given threshold will be marked by at least one of the studied indicators. As thresholds for these events, we have taken the medians of continuous series of observations of metal concentrations. As a result, we calculated the probabilities of exceeding half of the studied parameters of the specified thresholds in each zone. The facts of exceeding the thresholds (events) established in this way were recorded in the totality of observations selected in the form of clusters formed by neural network classification. At the same time the spatial distribution of selected observations with specific address bindings forms the corresponding territorial zones. Recommendations on the use of the obtained results for operational and planned quality management of the urban ecosystem components at the local level are given.

Keywords: urban ecosystem, risk assessment, response of the human body, biosubstrates, neural network technologies, metals.

В настоящее время работа Комитета по экологическим индикаторам и Комиссии по устойчивому развитию при ООН направлена на способы систематизации разнородной экологической информации, а также путей оценки и снижения экологического риска, в том числе и для экосистем, находящихся под интенсивным антропогенным воздействием – урбоэкосистем [1–2]. Согласно современным подходам, урбоэкосистему принято рассматривать как системный набор компонентов, взаимосвязанных между собой потоками вещества в процессе функционирования: 1) приземный слой атмосферы; 2) биологические объекты; 3) депонирующие среды. Наибольшие перспективы для исследования урбоэкосистем имеет объединение различных научных областей знания для решения данной задачи, особенно на локальном уровне. Антропогенные источники воздействия в урбоэкосистеме имеют свои локальные особенности, факторы экологического риска специфичны на каждом антропогенно нагруженном участке территории, системы принятия природоохранных решений их реализуют на локальном уровне, поэтому система оценки уровня экологического риска на урбанизированной территории должна создаваться именно на локальном уровне [3–6].

Любая урбоэкосистема является не простой совокупностью различных компонентов, а представляет собой сложную систему, в которой доминирует аэрогенный привнос загрязняющих веществ, происходит накопление в депонирующих компонентах и формируется отклик биологических компонентов на воздействие [3–5]. Поэтому в качестве маркерных веществ были выбраны металлы, которые без трансформации мигрируют в ди-

намичных и депонирующих компонентах урбоэкосистемы, обеспечивая взаимосвязь между ними и формируя отклик биологических компонентов на воздействие [6]. Поступление металлов на территорию урбоэкосистем происходит обычно с малой интенсивностью в течение длительного времени с накоплением в депонирующих компонентах. Оценку полиметаллической нагрузки в урбоэкосистеме необходимо проводить с учётом показателей, характеризующих как антропогенное полиметаллическое воздействие, так и отклик компонентов на воздействие. Количественные характеристики статичных и динамичных биосубстратов организма человека являются чувствительными маркерами состояния компонентов урбоэкосистемы, особенно на локальном уровне. Локализованное местонахождение детей-подростков и их большая чувствительность к антропогенному воздействию с откликом составов биосубстратов позволяет учитывать воздействие антропогенной полиметаллической нагрузки с территориальной дифференциацией [7, 8].

Элементы-металлы постоянно поступают в организм человека по водно-пищевому пути и с вдыхаемым воздухом и имеют значительный по времени период полувыведения из организма, что приводит к их накоплению в биосубстратах. В среднестатистическом выражении, если водно-пищевой путь поступления металлов в организмы детей-школьников в пределах единой территории можно принять в качестве фактора постоянного и равномерного действия, то качество воздуха варьирует в значительных пределах, обеспечивая постоянно действующему фоновому фактору значительный градиент.

Анализ снежного и почвенного покровов характеризует длительно формирующийся уровень загрязнения атмосферного воздуха, охватывающий значительный временной промежуток (сезон). Более того, почва является источником пыли, которая содержит металлы, формируя дополнительное загрязнение воздуха жилых зон [6]. Поэтому нами выбраны в качестве объектов исследования почвенный и снежный покровы, кровь и волосы детей-школьников для интегральной оценки и управления качеством компонентов урбоэкосистемы, а также для решения задач зонирования территории урбоэкосистемы.

Зонирование пространственной неравномерности показателей, характеризующих полиметаллическую нагрузку в урбоэкосистеме, является нетривиальной задачей, а обобщение многомерных массивов и их классификация по совокупности признаков – важнейшей проблемой системного анализа [8–10].

Поскольку стоит задача одновременного учёта и анализа разнородных и разноразмерных данных, имеющих различную степень влияния на состояние исследуемого компонента урбоэкосистемы, то идея построения аналитической модели расчёта была отвергнута. Был выбран другой подход, заключающийся в учёте всех факторов, характеризующих состояние объектов исследования, в процессе кластерного анализа комплексных данных. При фиксации изменений проводится агрегация информации с кластерным анализом разнородных данных.

Для решения данной проблемы предлагается использовать разработанную авторами методику поэтапной (каскадной) кластеризации, названную методикой каскадной фильтрации. Разработанная методика отличается от стандартной кластеризации повышением точности, которое достигается многократной декомпозицией. Суть методики изложена в работе [9], заключается в многократном повторении процедуры декомпозиции данных неким алгоритмом кластеризации для тех групп данных, детализация в которых недостаточна с точки зрения эксперта. В процессе исследований было рассмотрено несколько кластеризующих алгоритмов, пригодных для решения поставленной задачи. Точность кластеризации оценивалась по критерию SWC (Silhouette Width Criterion – коэффициент силуэта), коэффициенту FM (Folkes and Mallows), а также на основе экспертного оценивания. Тесты показали, что и для данных, изначально топологически образующих

плотные группы, и для данных с изначально смешанной, существенно неоднородной структурой, наилучшие результаты по точности демонстрирует алгоритм нейронных самоорганизующихся сетей Кохонена (SOM-карт). Именно этот алгоритм был нами отобран для построения модели каскадной нейросетевой фильтрации. Данный алгоритм анализирует ряды распределений исследуемых параметров и выделяет зоны (локации), различающиеся интенсивностью исследуемых параметров.

Методы исследования

Нами проведены химико-аналитические определения содержания металлов (Cd, Pb, Co, Cu, Ni, Zn, Cr, Mn, Fe) в снежном (суммарно, мкг/л) и почвенном покрывах (подвижные формы, мг/кг) (косвенные характеристики уровня загрязнения приземного слоя атмосферного воздуха), динамичных (кровь (мкг/мл)) и аккумулирующих (волосы (мкг/г)) биосубстратах детей-подростков (общая выборка составляла 342 человека) 9–16 лет (в качестве отклика биологических компонентов урбоэкосистемы на полиметаллическое загрязнение) с математическим анализом полученных результатов, которые подробно описаны в работах [11–12]. Для учёта поступления металлов от стационарных источников выбросов производился расчёт концентраций указанных металлов (мг/м³) в приземном слое атмосферного воздуха, поступающих от стационарных источников выбросов на территории исследования в пределах г. Казани. Для пробоотбора были выбраны 170 точек урбоэкосистемы на территории г. Казани, находящихся под воздействием стационарных и передвижных источников поступления металлов.

На первом этапе исследования была проведена экспертная оценка исходных данных для подготовки адекватных для исследования массивов и определения направления исследований. Поскольку при оперативном управлении качеством атмосферного воздуха в первую очередь используются данные автоматизированных постов наблюдений, определяющих приоритетность и программу управления выбросами, необходимо учесть территориальную привязку участков пробоотбора к зонам охвата действующими автоматическими станциями контроля загрязнения атмосферы (АСКЗА) на территории исследуемой урбоэкосистемы. При помощи нейросетевых SOM-карт производилась группировка имеющихся векторов

данных в 3 кластера. Кластеризующие нейросетевые SOM-карты имеют размерность 16 x 12 нейронов, обучение производится согласно модифицированному алгоритму WTA. Для моделирования использован пакет Deductor Studio 4.3 (оформляется свидетельство о государственной регистрации программы для ЭВМ).

Нами предложена авторская методика, изложенная в работе [13], позволяющая рассчитать вероятности превышения параметрами их пороговых значений в пределах разных зон, а в качестве обобщающего показателя мы предлагаем использовать вероятность превышения порога хотя бы по одному из оцениваемых параметров. Массив экспериментальных данных представлен как D , а событие достижения порога используемых параметров как S . В этом случае условную вероятность при заданном наборе данных обозначали как $p(S|D)$. Полная вероятность события S равна сумме всех вероятностей, учитывая как вероятность $p(D)$, так и условную вероятность $p(S|D)$:

$$P(S) = \sum_{i=1}^n p(D_i)p(S|D_i).$$

Если события D_i ($i = 1, 2, 3... n$) образуют полную группу событий

$$\sum_{i=1}^n p(D_i) = 1,$$

при этом событие S происходит обязательно и происходит в одном из D_i , то в этом случае вероятность определяется формулой:

$$P(D_i|S) = \frac{p(D_i)p(S|D_i)}{p(S)}.$$

Вероятность $p(S|D)$ называют апостериорной вероятностью, а $p(S)$ – априорной вероятностью. Следует отметить, что знаменатель дроби является интегралом (или, в случае дискретных случайных величин, суммой) числителя по всем возможным значениям параметра, нормализующим апостериорную вероятность.

Для практического применения вероятностной модели, сформулируем соответствующие события S , которые фиксируют факт превышения содержания некоторого показателя его порогового значения (например, значения, соответствующие 0,5-квантили или предварительно установленные региональные нормативы), а также события D_i – которые отмечают, что превышение произошло в i -той

зоне, предварительно сформированной с помощью нейросетевой классификации факторов воздействия, которую мы рассмотрим ниже. Тогда $p(D_i) = n_i/N$, где N – сумма всех n ($N = n_1 + n_2 + ... n_i$), а $p(S|D_i) = 1$.

При этом вероятность отдельного события в конкретной зоне будет равна отношению произведения доли всех событий в этой зоне с долей относительного числа наблюдений зоны к сумме всех подобных произведений по всем участкам. Таким образом, мы можем рассчитать вероятности превышения параметрами их пороговых значений в пределах разных зон, а в качестве обобщающего показателя мы предлагаем использовать вероятность превышения порога хотя бы по одному из оцениваемых параметров.

Так как мы рассматриваем независимые в совокупности события S_1, S_2, S_3 , то вероятности событий, противоположных событиям S_1, S_2, S_3 , соответственно равны $qS_1 = 1 - p(S_1)$, $qS_2 = 1 - p(S_2)$, $qS_3 = 1 - p(S_3)$, так как общая вероятность того, что событие S_i произойдет или не произойдет, равна 1, то есть $p_i + q_i = 1$.

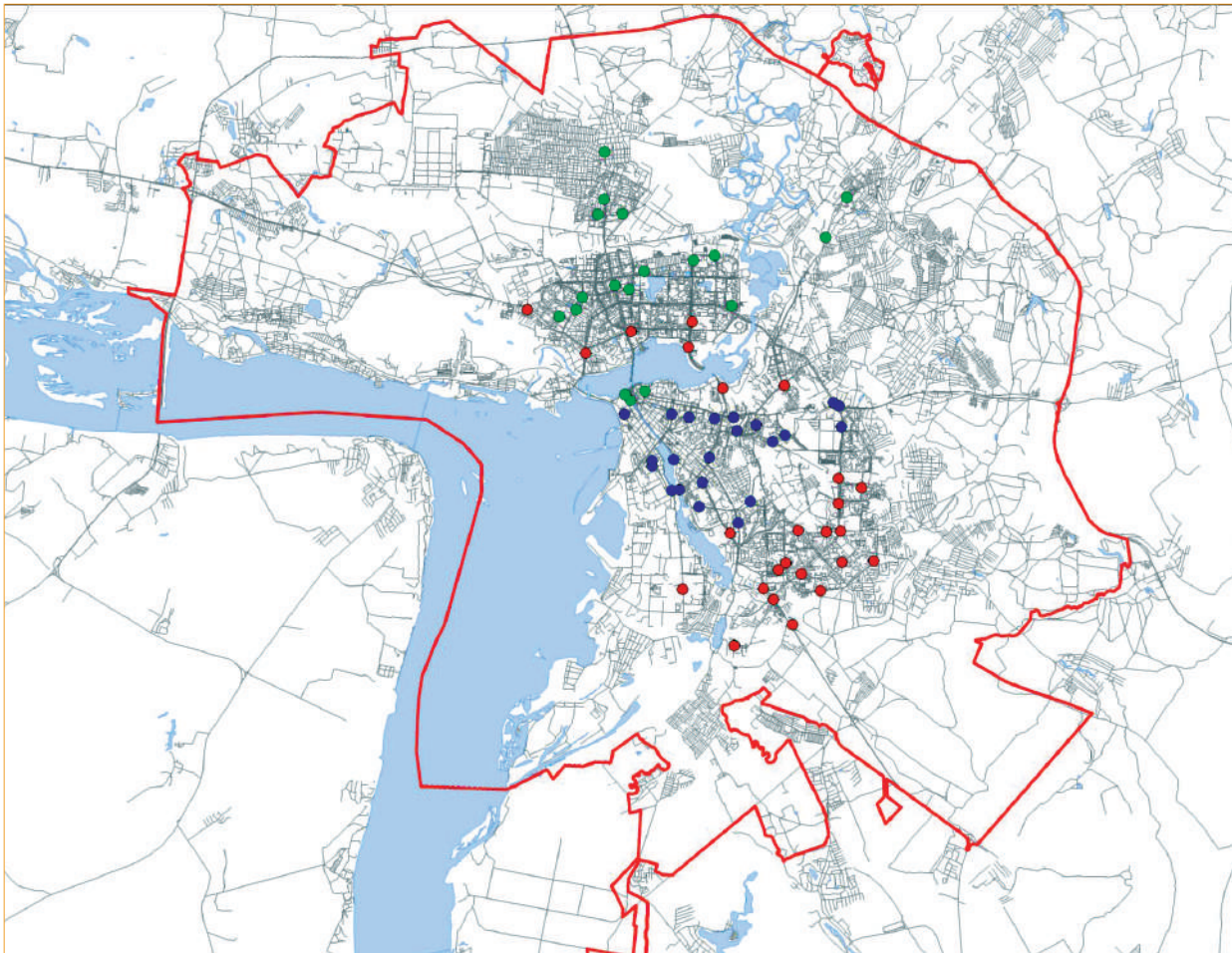
Согласно теореме умножения вероятностей для независимых событий и теореме сложения вероятностей для совместных событий, искомая вероятность $P(S) = 1 - q_1q_2q_3$.

В качестве пороговых значений мы использовали величины концентраций металлов, соответствующих 0,5-квантили распределения (медиане). Факты превышения установленных таким образом порогов (события) регистрировались во всей совокупности наблюдений, выделенных в виде кластеров, сформированных с помощью нейросетевой классификации. При этом пространственное распределение выбранных наблюдений, имеющих конкретные адресные привязки, формирует соответствующие территориальные зоны – кластеры.

Результаты и обсуждение

Результаты кластеризации демонстрируют компактное расположение точек, отнесённых системой к одному и тому же кластеру (см. рис. на цв. вкладки). Кластер 0 включает большие открытые территории в сочетании с высотной застройкой и высокой плотностью автомобильных потоков. Кластер 1 характеризуется невысокой неплотной застройкой с широкими улицами с расположением крупных промышленных предприятий, вынесенных за городскую черту. Кластер 2 характеризуется плотной невысокой застройкой, нагруженным авто-

**Ю. А. ТУАКОВА, С. В. НОВИКОВА, А. Р. ШАГИДУЛЛИН,
В. С. ВАЛИЕВ, А. Х. МОРАИС
«ПОДХОДЫ ДЛЯ ОЦЕНКИ И УПРАВЛЕНИЯ
КАЧЕСТВОМ КОМПОНЕНТОВ УРБОЭКОСИСТЕМЫ С ПОМОЩЬЮ
МЕТОДОВ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ», С. 31**



● кластер 0 / cluster 0

● кластер 1 / cluster 1

● кластер 2 / cluster 2

Рис. Топологическое расположение точек пробоотбора на территории г. Казани с распределением по трём кластерам
Fig. Topological location of sampling points on the territory of Kazan with the distribution of three clusters

Вероятности (*P*) неблагоприятных событий, рассчитанные для содержания металлов в почве, снежном покрове и биосубстратах
 Probabilities (*P*) of adverse events calculated for the content of metals in the soil, snow cover and biosubstrates

Кластер Cluster	Снег Snow	Биосубстраты / Biosubstrates		Почва Soil
		кровь / blood	волосы / hair	
0	0,27	0,23	0,35	0,15
1	0,34	0,24	0,38	0,33
2	0,39	0,30	0,42	0,56

транспортным потоком с невысокой скоростью передвижения.

В результате проведённых расчётов получены вероятности проявления неблагоприятного события, по крайней мере, у одного из оцениваемых параметров и вероятности неблагоприятного события у половины оцениваемых параметров, рассчитанные в пределах выделенных кластеров. Если сопоставить вероятности превышения пороговых содержаний металлов, рассчитанных для крови и волос с вероятностями загрязнения снежного покрова и почвы, то можно заметить, что эти вероятности чётко коррелируют друг с другом (табл.).

Из представленной таблицы следует, что наибольший риск накопления металлов в организме детей-подростков отмечен среди точек, отнесённых к кластеру 2. Причём для этого кластера характерны наибольшие вероятности высоких концентраций металлов, как в волосах, так и в крови. Вместе с этим, следует отметить, что кластер 1 также имеет неблагоприятную вероятностную характеристику, хотя и несколько лучшую, чем кластер 2. Наиболее благоприятная ситуация складывается на точках, объединённых кластером 0, в котором отмечены относительно невысокие вероятности накопления металлов в волосах и крови детей.

Выводы

Таким образом, можно с высокой степенью надёжности утверждать, что выделенный нами кластер 0 объединяет наблюдения, в которых доля высоких концентраций металлов в биосубстратах детей значительно ниже, чем в кластере 1 и, особенно, в кластере 2. Кроме того, указанные распределения вероятностей превышения металлами их пороговых значений свидетельствуют о значительном отклике урбоэкосистемы на воздействие и позволяют использовать выделенные кластеры для её зонирования.

Полученные результаты можно использовать для планового управления качеством компонентов урбоэкосистемы, которое поз-

воляет в зависимости от результатов зонирования территории осуществить регулирование планировки и застройки и варьировать стратегические планы дальнейшего хозяйственного освоения территории.

Работа выполнена при финансовой поддержке РФФИ и Правительства Республики Татарстан в рамках научного проекта № 18-41-160020.

References

1. Sornette D., Maillart T., Kroger W. Exploring the limits of safety analysis in complex technological systems. RiskCenter, Zurich. 2013. [Internet resource] <http://arxiv.org/pdf/1207.5674.pdf> (Accessed: 18.04.2019).
2. Sprenger J. Environmental risk analysis: robustness essential for precaution // Philosophy of Science. 2012. No. 79 (5). P. 881–892. doi: 10.1086/667873
3. Suter G.W. Ecological risk assessment in the United States Environmental Protection Agency: a historical overview // Integrated Environmental Assessment and Management. 2008. V. 4. No. 3. P. 285–289. doi: 10.1897/IEAM_2007-062
4. Janssen C. Putting ecological realism in environmental risk assessment // Report on 14th Annual CEFIC-LRI Workshop “Evolution or Revolution – Research priorities for future risk assessment”. 2012.
5. Yayli E.A., Muzalevskiy A.A. The methodology and way for assessing the quality of environmental components of urbanized territories based on indicators, indices and risk // Ekologicheskie sistemy i pribory. 2006. No. 12. P. 23–29 (in Russian).
6. Elpatyevsky P.V. Geochemistry of migration flows in natural and natural-technogenic geosystems. Moskva, 1993. 253 p. (in Russian).
7. Blaurock-Busch E., Amin O.R., Dessoki H.H., Rabah T. Toxic metals and essential elements in hair and severity of symptoms among children with Autism // Maedica (Buchar). 2012. No. 7 (1). P. 38–48.
8. Zainullin V.G., Bodnar I.S., Kondratenok B.M. Features of accumulation of chemical elements in the hair of the children's population of the Komi Republic // Izvestiya Komi nauchnogo tsentra UrO RAN. 2014. No. 2 (18). P. 24–31 (in Russian).

9. Novikova S.V., Tunakova Yu.A., Shagidullin A.R., Valiev V.S., Gabdrakhmanova G.N. Synergistic neuro-expert model based on SOM MAPS for integrated analysis of environmental objects // Chemistry and environmental engineering XVIII: Sbornik trudov mezhdunarodnoy nauchnoy konferentsii. Kazan: Izd-vo KNITU-KAI, 2018. P. 318–322 (in Russian).

10. Guojun G., Ma Ch., Wu J. DataClustering: theory, algorithms, and applications // ASA-SIAM series on statistics and applied probability. SIAM, Philadelphia, ASA, Alexandria, VA, 2007. 466 p.

11. Tunakova Y.A., Ivanov D.V., Shagidullin A.R., Valiev V.S. Soil cover as an indicator of polymetallic pollution

of a ground layer of atmospheric air // Rossiyskiy zhurnal prikladnoy ekologii. 2019. No. 1 (17). P. 53–58 (in Russian).

12. Tunakova Yu.A., Novikova S.V., Faizullin R.I., Valiev V.S. Approximation of the dependency of trace elements concentrations in internal media upon their contents in environment objects // BioNanoScience. 2018. No. 8. P. 288–295. doi: 10.1007/s12668-017-0487-y

13. Tunakova Yu.A., Novikova S.V., Shagidullin A.R., Valiev V.S. Methodology of evaluation of regional ecological risk in the urban ecosystem (on the example of Kazan) // Transformation of ecosystem under the influence of natural and anthropogenic factors: Materialy mezhdunarodnoy nauchnoy konferentsii. 2019. P. 263–267 (in Russian).