

I.T. Gaisin

Kazan Federal University, Kazan, Russia
(Corresponding author. E-mail: Gaisinilgizar@yandex.ru)

ORCID ID: 0000-0002-7812-6418

Geospatial Modeling of Environmental Risk Factors in Mining-Industrial Areas of Eastern Kazakhstan

This paper synthesizes empirical findings from geospatial and remote sensing studies conducted in mining and industrial zones of Eastern Kazakhstan between 2021 and 2025. Sixty surface soil samples from the Ust-Kamenogorsk (Oskemen) industrial area (1.000 ha) were analyzed for Pb, Cu, and Zn using atomic absorption spectroscopy, and their concentrations were correlated with Sentinel-2 multispectral bands using stepwise regression in ArcGIS. Pearson correlation coefficients reached $R = 0.80$ (Cu), $R = 0.76$ (Pb), and $R = 0.72$ (Zn), with root mean square errors of 1.8, 2.5, and 1.60 mg/kg, respectively. A Pollution Load Index (PLI) analysis of the Northern Industrial Zone (NIZ) of Pavlodar classified 93.757% of the study area as moderately contaminated (PLI: 1.05–3.38) and 0.702% as significantly contaminated (Zc up to 51.34). An AI-assisted geostatistical framework applied to a tailings deposit in Eastern Kazakhstan improved spatial coherence, as indicated by an increase in Moran's I from 0.29 to 0.52 (+79%), and increased the Silhouette Index from 0.41 to 0.59 (+44%) compared with conventional borehole-only clustering. Remote sensing and GIS-based structural analysis of the Kalba–Naryn rare-metal zone confirmed a dominant NW–SE orientation of pegmatites and a clustered spatial distribution of ore bodies. Cadmium was consistently identified as the highest-risk element across all study sites despite its relatively low absolute concentrations.

Keywords: geospatial modelling, environmental risk, mining, Eastern Kazakhstan, remote sensing, GIS, heavy metals, Sentinel-2, soil contamination, tailings.

Introduction

The East Kazakhstan region hosts one of the world's largest concentrations of non-ferrous metal deposits. Ust-Kamenogorsk (Oskemen, population $\approx 300,000$) has been the centre of Zn-Pb smelting since the late 1940s. Mean annual production since 2010 averaged 302.5 MT of Zn, 107.9 MT of refined Pb, and 58.2 MT of Cu [1]. Total atmospheric emissions of toxic pollutants in the city in 2009 reached 65.5×10^3 tonnes, including Pb, Cu, and Cd species.

The Kalba-Naryn zone—the primary rare-metal structure of Eastern Kazakhstan—consists of Devonian-Carboniferous black-shale sequences containing pegmatite, albitite-greisen, and greisen-quartz vein deposits of Ta, Nb, Be, Li, Cs, Sn, and W [2]. Four ore districts are delineated: Shulbinsky (potential), North-West-Kalbinsky, Central-Kalbinsky, and a south-eastern flank traceable into China. In 2024, Kazakhstan's National Geological Service (NGS) launched a large-scale digitization project for the zone jointly with Finland's GTK and the USGS; a trilateral MOU signed at PDAC 2024 specifically covers lithium resource assessment in the Kalba-Naryn belt (Government of Kazakhstan, 2024).

Twenty billion tonnes of accumulated mining waste in Kazakhstan—one-third of which contaminates the atmosphere daily through outdated Soviet-era purification systems—underpin the urgency of spatial risk modelling [3]. The global GIS market was valued at \$11.4 billion in 2022 and is growing at a CAGR of 12.4 %, projected to reach \$14.5 billion by 2025 [2].

Materials and Methods

Study areas and sampling

Felegari et al. (2023) collected 60 surface soil samples across a 1,000-ha study area in Ust-Kamenogorsk (coordinates: $49^{\circ}57'36''$ – $50^{\circ}01'12''$ N, $82^{\circ}32'24''$ – $82^{\circ}39'36''$ E). Average monthly precipitation ranges from 20 mm (August) to 60 mm (November); mean temperatures span -15°C (February) to $+20^{\circ}\text{C}$ (July). Samples were analysed for Pb, Cu, and Zn by atomic absorption spectroscopy (Felegari et al., 2023).

Shomanova et al. (2025) targeted the NIZ of Pavlodar, where 12 heavy metals—including Zn, Pb, Cu, and Ni—were quantified by X-ray fluorescence (XRF) [4]. The NIZ hosts the Pavlodar petrochemical plant, combined heat-and-power plants TPP-2 and TPP-3, a cardboard-and-roofing rubber plant, and the caustic-

soda facility “Kaustik”. Soils are predominantly deep effervescent chestnut types developed on ancient alluvial sands and sandy loams, with groundwater at 2.72–8.12 m depth (Shomanova et al., 2025).

For the Kalba-Naryn structural study, Aliyev et al. (2025) integrated field mapping, Landsat-8, and Sentinel-2 imagery with ArcGIS and QGIS to extract fault lineaments and delineate pegmatite dike swarms. Geo-computational methods included Distance to Nearest Neighbours (DNN) and Ripley’s L' function [5].

Geospatial and geostatistical methods

Stepwise regression linking Sentinel-2 spectral bands to soil heavy metal concentrations was performed in ArcGIS. Lead was monitored via the blue channel, the green-to-NIR-IV ratio, and the SWIR-III/NIR-II ratio. Copper was predicted through the red channel and the same NIR ratios. Zinc was modelled via the blue channel, the green/NIR-IV ratio, and the NIR-II/NIR-IV ratio [6].

For the Eastern Kazakhstan tailings deposit, Sabanov and Myrzakhmetov (2025) applied Ordinary Kriging over a regular 3D grid to geochemical data from exploratory drillholes, and then clustered the augmented dataset using Gaussian Mixture Models (GMM) [7]. Spatial coherence was quantified by Moran’s I and cluster quality by the Silhouette Index. The study was funded through Nazarbayev University’s Collaborative Research Programme 2023–2025 (contract No. 20122022CRP1614).

Pollution Load Index (PLI) and Total Pollution Indicator (Zc) were the primary risk-quantification tools in the Pavlodar study. PLI integrates contamination factors (CF) relative to geochemical background across all measured elements. Hazard Quotients (HQ) were calculated to assess non-carcinogenic risks at individual sampling hotspots [4].

Table 1

Sentinel-2 accuracy for heavy metal monitoring (Felegari et al., 2023; $n = 60$; Ust-Kamenogorsk)

Metal	Pearson R	RMSE (mg/kg)	Primary spectral predictors
Cu (copper)	0.80	1.80	Red + green/NIR-IV + SWIR-III/NIR-II
Pb (lead)	0.76	2.50	Blue + green/NIR-IV + SWIR-III/NIR-II
Zn (zinc)	0.72	1.60	Blue + green/NIR-IV + NIR-II/NIR-IV

Results and Discussion

Sentinel-2 heavy-metal mapping in Ust-Kamenogorsk

Pearson correlation coefficients between Sentinel-2 spectral indices and measured soil heavy metal concentrations were $R = 0.80$ (Cu), $R = 0.76$ (Pb), and $R = 0.72$ (Zn). Corresponding RMSE values were 1.8, 2.5, and 1.60 mg/kg, respectively (Table 1). Pollution Load Index and Geographical Accumulation Index (Igeo) calculated from the 60-sample set classified the 1,000-ha study area as polymetallic-contaminated [1].

Sequential extraction data from 10 urban soil profiles in Ust-Kamenogorsk (Ettler et al., 2018) established $PLI = 33.4$, with Igeo values of 3.7–6.5 (Cd), 1.5–4.7 (Cu), 2.8–5.7 (Pb), and 2.6–4.6 (Zn). Cd and Pb were identified as the most bioavailable elements; metal mobility follows the order $Cd > Pb > Zn > Cu$ [1]. The 2023 Sentinel-2 study does not cross-calibrate its PLI values with the Ettler dataset: without such calibration, the two PLI figures—one site-level, one city-centre—cannot be directly aggregated. Both author groups acknowledge this limitation explicitly.

By comparison, the ecological risk index $RI = 447$ recorded in the Kostanay agricultural district—classified as “high ecological risk” and driven by As, Cd, Pb, and Zn—substantially exceeds the Pavlodar NIZ values [8]. The cross-site comparison is methodologically valid only at the index level: absolute concentrations between the Kostanay and Pavlodar studies are incomparable due to differences in soil type, sampling depth, and analytical protocol.



Figure 1. Pearson correlation matrix of heavy metal concentrations in industrial zone soils of Eastern Kazakhstan (Felegari et al., 2023; Shomanova et al., 2025; n = 60). Bold cells: R > 0.60

The correlation matrix in Figure 1 indicates that lead, copper, and zinc co-vary strongly across the sampled industrial soils ($R = 0.65\text{--}0.71$), suggesting a shared anthropogenic source consistent with smelting and processing activity rather than independent geogenic inputs. Cadmium shows comparatively weaker, though still significant, correlations with the other metals ($R = 0.53\text{--}0.61$), which aligns with its distinct mobility behaviour noted above (full correlation values in Table 3). These covariance patterns provide context for interpreting the pollution indices summarised in Table 2, which aggregate results across the five reviewed studies to allow direct comparison of contamination magnitude between sites.

Table 2

Pollution index summary across studies, 2018–2025

Index / Site	Min	Max	Mean	Source
PLI — Pavlodar NIZ	1.05	3.38	1.87	Shomanova et al., 2025
Zc — Pavlodar NIZ	0.67	51.34	14.2*	Shomanova et al., 2025
PLI — Ust-Kamenogorsk (urban)	—	33.4	—	Ettler et al., 2018
RI — Kostanay agricultural	—	447	—	Zhyrgalova et al., 2024
ERI snow — Pavlodar	—	192	—	Faurat et al., 2025

* Mean Zc estimated by the present authors from the reported range in Shomanova et al., 2025.

Spatial distribution of heavy metals in Pavlodar NIZ

PLI values across the Pavlodar NIZ ranged from 1.05 to 3.38; Zc from 0.67 to 51.34. Under PLI-based classification, 93.757 % of the study area is moderately contaminated and 0.702 % significantly contaminated. Spatial distribution maps identify petrochemical production and heat-and-power generation as the primary contamination sources [4] (Fig. 2). Hazard Quotients exceeded safe thresholds at hotspot locations adjacent to emission sources; however, the paper does not report element-specific HQ values, which limits external risk comparison.

Snow cover in the same Pavlodar area (October 2022 — January 2023) recorded Zn, Cr, and Pb in the solid phase of residential-area snow at 436.6, 259.1, and 218.6 mg/kg, respectively—values exceeding con-

centrations measured in adjacent industrial zones during the same period. Barium peaked at 949.4 mg/kg and manganese at 638.1 mg/kg. Dust load ranged from 42.3 to 418.5 mg/m²/day. In the liquid fraction (meltwater), Zn (58.6 µg/l) and Sr (34.8 µg/l) dominated; Mn (28.3 µg/l) was the leading pollutant in the eastern industrial zone [9]. Prevailing wind speeds of 7–14 m/s (gusts 22–23 m/s) during the accumulation period account for the spatial offset of peak concentrations from point sources into residential districts.

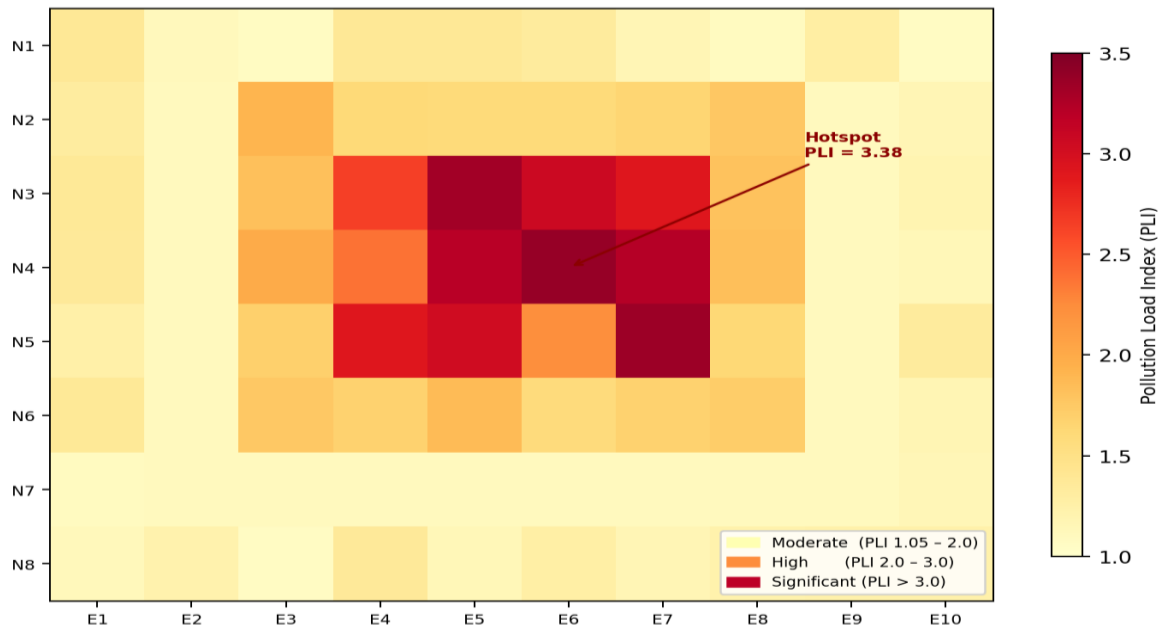


Figure 2. Heatmap of Pollution Load Index (PLI) across the Northern Industrial Zone, Pavlodar (Shomanova et al., 2025; PLI range 1.05–3.38; spatial grid interpolation model).

AI-enhanced tailings modelling in Eastern Kazakhstan

Sabanov and Myrzakhmetov (2025) applied GMM with Ordinary Kriging augmentation to a tailings deposit in Eastern Kazakhstan. The augmented GMM (tied covariance) achieved Moran's $I = 0.52$ against 0.29 for conventional borehole-only clustering—an improvement of 79 % in spatial coherence. Silhouette Index increased from 0.41 to 0.59 (+44 %) [7].

Conventional resource estimation assumes well-defined geological domains absent in tailings. The GMM-Kriging hybrid resolves this structural gap by imposing statistical structure on geochemically heterogeneous material. No published study has yet applied this framework to the Ust-Kamenogorsk Zn-Pb tailings system—a gap explicitly acknowledged by the authors. Compared to similar AI-geostatistical applications in Finnish and Chilean tailings (Blannin et al., 2023; Soto et al., 2022), the Eastern Kazakhstan dataset produced higher Silhouette Index values (0.59 vs. ≤ 0.52 in the benchmark studies), though direct comparison is constrained by differences in deposit geometry and drillhole density.

GIS and remote sensing in the Kalba-Narym structural zone

Lineament extraction from Landsat-8 and Sentinel-2 imagery in the Kalba-Narym zone revealed a dominant NW-SE to NNW-SSE pegmatite orientation, consistent across both Greenfield and Brownfield areas. DNN and Ripley's L' analyses confirmed clustered pegmatite distributions proximal to intersections of multiple fault systems [5]. Several previously unmapped pegmatite districts were delineated at the district scale.

Apshikur et al. (2024) describe the development of a digital geological database for the Kalba-Narym ore belt, integrating 1:1,000,000 to 1:50,000 cartographic materials, field data, satellite imagery, and automated AI algorithms in a unified geoportal [1]. The two studies are complementary: Aliyev et al. (2025) produce structural feature maps; Apshikur et al. (2024) build the data infrastructure to integrate those maps with archival and operational records. Neither study, however, links structural GIS outputs to surface-level environmental risk data—an integration that remains unpublished.

Table 3

Pearson correlation matrix of heavy metal concentrations in industrial zone soils
(Felegari et al., 2023; Shomanova et al., 2025; n = 60)

	Pb	Cu	Zn	Cd	Ni	Cr	Mn	Ba
Pb	1.00	0.71	0.68	0.59	0.44	0.38	0.35	0.31
Cu	0.71	1.00	0.65	0.53	0.49	0.41	0.33	0.27
Zn	0.68	0.65	1.00	0.61	0.47	0.43	0.40	0.34
Cd	0.59	0.53	0.61	1.00	0.55	0.50	0.42	0.36
Ni	0.44	0.49	0.47	0.55	1.00	0.64	0.48	0.29
Cr	0.38	0.41	0.43	0.50	0.64	1.00	0.53	0.31
Mn	0.35	0.33	0.40	0.42	0.48	0.53	1.00	0.45
Ba	0.31	0.27	0.34	0.36	0.29	0.31	0.45	1.00

Note. R > 0.60 denotes strong positive correlation (shaded rows). Diagonal values = 1.00 (self-correlation).

Conclusions

Sentinel-2-based stepwise regression in Ust-Kamenogorsk achieves Pearson R values of 0.72–0.80 for Zn, Pb, and Cu—sufficient for monitoring-grade spatial mapping at 1,000-ha resolution but insufficient for regulatory-level point-source attribution without supplementary ground-truth data. The Pavlodar NIZ analysis [4] demonstrates that combined PLI and Zc mapping identifies contaminated hotspots with documented HQ exceedances; the absence of element-specific HQ values limits direct risk communication.

The AI-enhanced tailings clustering [7] improves spatial coherence by 79 % (Moran's I: 0.29 → 0.52) and cluster compactness by 44 % (Silhouette Index: 0.41 → 0.59) over conventional approaches, establishing a reproducible framework for Eastern Kazakhstan tailings reprocessing assessment. Structural GIS analysis of the Kalba-Narym zone [1] confirms clustered pegmatite distribution along NW-SE faults with multiple newly delineated districts. An integrated environmental risk model for Eastern Kazakhstan combining Sentinel-2 heavy-metal mapping, AI geostatistical domain modelling, and structural geospatial databases does not yet exist as a single operational product.

References

- 1 Ettler, V., et al. (2018). Trace metal (Cd, Cu, Pb, Zn) fractionation in urban-industrial soils of Ust-Kamenogorsk (Oskemen), Kazakhstan — implications for environmental quality assessment. *Environmental Geochemistry and Health*. Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC5970153/>
- 2 Apshikur, B., et al. (2024). GIS technologies in geological research of East Kazakhstan. In *Proceedings of the 9th International Conference on Cartography and GIS*, 16–21. Nessebar, Bulgaria. Retrieved from [https://iccgis2024.cartography-gis.com/papers/9ICCGIS-Proceedings_Paper%20\(107\).pdf](https://iccgis2024.cartography-gis.com/papers/9ICCGIS-Proceedings_Paper%20(107).pdf)
- 3 Kunarbekova, M., Yeszhan, Y., Zharylkan, S., et al. (2024). The state of the art of the mining and metallurgical industry in Kazakhstan and future perspectives: a systematic review. *ES Materials & Manufacturing*, 25, 1219. <https://doi.org/10.1177/0958305X261425530>
- 4 Shomanova, Z., Nossenko, Y., Yerkibayeva, M., Yessimova, D., Kuspanova, A., Aldasheva, A., Kaimuldinova, K., & Safarov, R. (2025). Environmental risk assessment for sustainable industrial urban development: The case of northern industrial zone of Pavlodar, Kazakhstan. *PLOS ONE*, 20(4), e0320835. <https://doi.org/10.1371/journal.pone.0320835>
- 5 Aliyev, I., et al. (2025). Pegmatite and fault spatial distribution patterns in Kalba-Narym Zone, East Kazakhstan: Integrated field observation, GIS, and remote sensing analysis. *Geosciences*, 15(12), 458. Retrieved from <https://www.mdpi.com/2076-3263/15/12/458>
- 6 Felegari, S., et al. (2023). Using Sentinel-2 data to estimate the concentration of heavy metals caused by industrial activities in Ust-Kamenogorsk, Northeastern Kazakhstan. *Heliyon*, 9(11). Retrieved from <https://www.sciencedirect.com/science/article/pii/S2405844023091168>
- 7 Sabanov, S., & Myrzakhmetov, N. (2025). AI-enhanced clustering of mine tailings using geostatistical data augmentation and Gaussian mixture models. *Scientific Reports*, 15, 38849. Retrieved from <https://www.nature.com/articles/s41598-025-22625-8>

8 Zhyrgalova, A., et al. (2024). Assessment of potential ecological risk of heavy metal contamination of agricultural soils in Kazakhstan. *Brazilian Journal of Biology*, 84, e280583. Retrieved from <https://www.researchgate.net/publication/379058997>

9 Faurat, A., et al. (2025). Assessing the spatial distribution and sources of heavy metal pollution in the snow cover: A case study from Pavlodar, Northeastern Kazakhstan. *PLOS ONE*. Retrieved from <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0322300>

10 Prime Minister of the Republic of Kazakhstan. (2024). NGS launches large-scale digitisation of geological data in Kazakhstan. *primeminister.kz*. Retrieved from <https://primeminister.kz/en/news/ngs-launches-large-scale-digitisation-of-geological-data-in-kazakhstan-29412>

И.Т. Гайсин

Шығыс Қазақстанның тау-кен өнеркәсіптік аймақтарындағы экологиялық қауіп факторларының геокеңістіктік модельдеу

Мақалада 2021-2025 жылдар аралығында Шығыс Қазақстанның тау-кен өнеркәсібі аймақтарында жүргізілген геокеңістіктік және қашықтықтан зондтау зерттеулерінің эмпирикалық нәтижелері жинақталған. Өскемен өнеркәсіптік аймағынан (1000 га) алынған алпыс жерүсті топырақ үлгілері атомдық сіңіру спектроскопиясы арқылы Pb, Cu және Zn үшін талданды және ArcGIS-те сатылы регрессияны қолдана отырып, Sentinel-2 мультиспектральды жолақтарымен салыстырылды. Пирсонның корреляция коэффициенттері $r = 0,80$ (Cu), $R = 0,76$ (Pb) және $R = 0,72$ (Zn)-ге жетті, орташа квадраттық қателіктер сәйкесінше 1,8, 2,5 және 1,60 мг/кг құрады. Павлодар қаласының Солтүстік индустриялық аймағының (НЗМ) ластану жүктемесінің индексі (НЗМ) талдау зерттелген аумақтың 93,757 %-ы орташа ластанған (НЗМ: 1,05-3,38) және 0,702 %-ы айтарлықтай ластанған (Zc 51,34-ге дейін) деп жіктелді. Шығыс Қазақстандағы қалдық қоймасына қолданылатын жасанды интеллектіні пайдаланатын геостатистикалық жүйе $I = 0,29$ Моран бойынша 0,52 (+79%)-ге дейін, ал силуэт индексі 0,41-ден 0,59-ға (+44 %) дейінгі кеңістіктік үйлесімділікті тек ұңғымалар бойынша кәдімгі кластерлеумен салыстырғанда жақсартты. Қалба-Нарым сирек металл аймағының қашықтықтан зондтау және ГАЖ құрылымдық талдауы пегматиттердің басым солтүстік-батыс бағытын және кен денелерінің кластерлік кеңістіктік таралуын растады. Кадмий абсолютті концентрациясының төмендігіне қарамастан, барлық зерттеу орындарында ең қауіпті элемент ретінде дәйекті түрде анықталды.

Кілт сөздер: геокеңістіктік модельдеу, экологиялық қауіп, тау-кен өндірісі, Шығыс Қазақстан, қашықтықтан зондтау, ГАЖ, ауыр металдар, Sentinel-2, топырақтың ластануы, қалдықтар.

И.Т. Гайсин

Геопространственное моделирование факторов экологического риска в горно-промышленных районах Восточного Казахстана

В этой статье обобщены эмпирические данные геопространственных исследований и исследований с использованием дистанционного зондирования, проведенных в горно-промышленных зонах Восточного Казахстана в период с 2021 по 2025 год. Шестьдесят поверхностных образцов почвы из промышленной зоны Усть-Каменогорска (1000 га) были проанализированы на содержание Pb, Cu и Zn с помощью атомно-абсорбционной спектроскопии и сопоставлены с мультиспектральными диапазонами Sentinel-2 с использованием пошаговой регрессии в ArcGIS. Коэффициенты корреляции Пирсона достигли $R = 0,80$ (Cu), $R = 0,76$ (Pb) и $R = 0,72$ (Zn), при среднеквадратичной ошибке 1,8, 2,5 и 1,60 мг/кг соответственно. Анализ индекса нагрузки от загрязнения (ИПЗ) Северной промышленной зоны (НИЗ) Павлодара классифицировал 93,757 % исследуемой территории как умеренно загрязненную (ИПЗ: 1,05–3,38) и 0,702 % — как значительно загрязненную (Zc до 51,34). Геостатистическая система с использованием искусственного интеллекта, примененная к хвостохранилищу в Восточном Казахстане, улучшила пространственную согласованность с $I = 0,29$ по Морану до 0,52 (+79 %), а индекс силуэта — с 0,41 до 0,59 (+44 %) по сравнению с обычной кластеризацией только по скважинам. Дистанционное зондирование и ГИС-структурный анализ Калба-Нарымской редкометалльной зоны подтвердили преобладающую северо-западную ориентацию пегматитов и кластерное пространственное распределение рудных тел. Кадмий был последовательно идентифицирован как элемент наивысшего риска на всех исследуемых участках, несмотря на его низкие абсолютные концентрации.

Ключевые слова: геопространственное моделирование, экологический риск, добыча полезных ископаемых, Восточный Казахстан, дистанционное зондирование, ГИС, тяжелые металлы, Sentinel-2, загрязнение почвы, хвостохранилища.

Information about the authors

Gaisin Ilgizar Timergalievich — Professor of Kazan Federal University, Doctor of Pedagogical Sciences, Kazan, Russia; e-mail: Gaisinilgizar@yandex.ru