

# Cluster and correlation analyses of the chemical composition of groundwater nearby MSW landfills

*Alexander Nevzorov*<sup>1</sup>, *Mikhail Slesarev*<sup>2</sup> and *Alexey Korshunov*<sup>1\*</sup>

<sup>1</sup> Northern (Arctic) Federal University named after M.V. Lomonosov, 163002, Arkhangelsk, Russia

<sup>2</sup> Moscow State University of Civil Engineering (MGSU), 129337, Moscow, Russia

**Abstract.** In areas of the European North with a low population density and poorly developed transport infrastructure, it is difficult to organize waste removal and storage at large landfills, therefore many settlements have small landfills of municipal solid waste (MSW), and the practice of placing them in swamps is very common, which leads to the migration of pollutants into adjacent rivers. Environmental monitoring nearby such landfills is hampered by high costs of periodic water sampling and chemical composition analyses. It is possible to decrease the number of substances under analysis and thus reduce costs by determining the marker substances, the concentration of which in groundwater near the landfill has stable correlations with the content of other pollutants. The article provides an analysis of the pollutant concentrations in 45 MSW landfills around the world. The results of cluster and correlation analyses made it possible to identify marker substances and establish the target correlation dependences.

## 1 Introduction

In areas of the European North with a low population density and poorly developed transport infrastructure, it is difficult to organize waste removal and storage at large landfills, therefore a small waste landfill is generated near almost every settlement, and the practice of placing them in swamps is very common. This leads to the migration of pollutants into adjacent rivers.

The processes of sorption and desorption of pollutants by peat have not been studied enough yet. According to various authors, some substances are securely settled in peat, others are transferred by groundwater. Moreover, the effect of peat composition, temperature and other factors on the experimental results was not usually taken into consideration during the research [1, 2, 3].

So far, researchers do not have achieved any common opinion on what processes prevail in peat after the termination of waste dumping, and monitoring was carried out in the vicinity of existing landfills, i.e. under conditions of the constant supply of pollutants.

---

\* Corresponding author: [a.a.korshunov@yandex.ru](mailto:a.a.korshunov@yandex.ru)

According to the research data, previously one of the authors of this article determined that in the Arkhangelsk region in the swamps subjected to man-made effects from hydrolysis and forest industry enterprises such pollutants as oil products, phenols, ammonium salt spread into the entire depth of the peat deposit, and their content can exceed the maximum permissible concentration (MPC) by ten and hundred times. Long-term monitoring showed that after the termination of waste storage and wastewater discharge, the release of pollutants from the swamp was not over for 16 years, and the prevailing factor in reducing their concentration in the peat deposit was removal by groundwater [2].

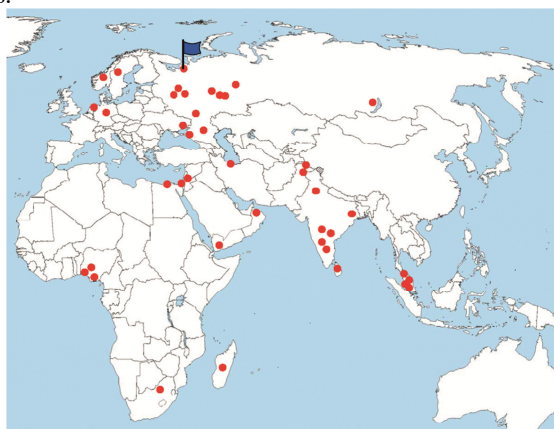
Thus, the study of the chemical composition of groundwater and the identification of specific pollutant substances will make it possible to establish the dependences necessary for predicting the distribution of pollutants at the base of the MSW landfills in small settlements.

## 2 Analysis of scientific publications on groundwater pollution at the base of MSW landfills

In order to determine the specific pollutants in groundwater at the base of MSW landfills, a review of scientific publications by Russian and foreign researchers was carried out. Data on 45 locations in different parts of the world were summarized, including Africa – 6, Europe – 5, the Middle East – 4, Asia – 16, and Russia – 14 (Figure 1).

Analysis of the data made it possible to identify 22 substances, the presence of which was found in groundwater in concentrations significantly exceeding the standard values. The widest range of substances at the base of one storage facility was observed in Sri Lanka – 19 substances, the least number of substances (4) was detected in landfills in Sweden, Iran and Pakistan. This can apparently be explained by the specific features of domestic waste management in a given country, as well as differences in monitoring systems and requirements for chemical analysis of water samples.

The concentration of pollutants in groundwater also varies over a wide range. So, for example, the concentration of zinc varies from  $0.4 \cdot 10^{-3}$  mg/l to 632 mg/l, i.e. it differs by 1.6 million times, the concentration of cobalt changes from  $6 \cdot 10^{-3}$  mg/l to 1.12 mg/l – differs by 187 times.



**Fig. 1.** Locations of the landfills considered in the analysis. The City of Arkhangelsk is marked with a flag.

### 3 Cluster analysis of the climatic factors effect on groundwater pollution

The study of the presence or absence of statistically significant dependence of the pollutant concentration in groundwater on climatic factors was started with a cluster analysis in the SPSS Statistics 23.0 software, the first stage of which was the selection of groups of objects with similar values of the factors.

The classification features for hierarchical clustering are as follows: average annual temperature, °C (attribute X) and average annual precipitation, mm, (attribute Y).

Clustering was performed on the basis of quantitative assessment of similarity of objects as a measure of distance between them. The Euclidean distance was taken as a measure of distance between the objects:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{1}$$

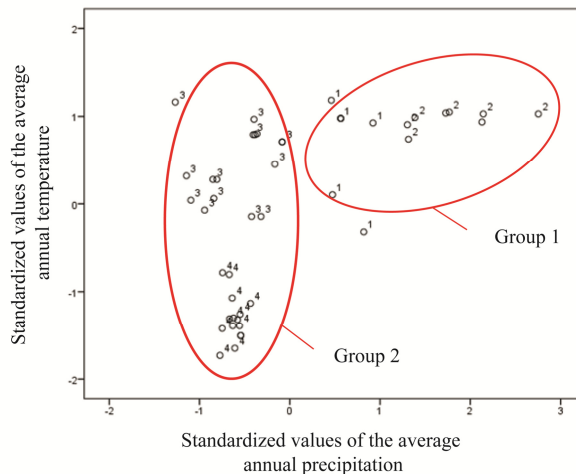
where  $x_{i,j}$  and  $y_{i,j}$  are the coordinates (values of attributes X and Y respectively) of  $i$ -th and  $j$ -th objects.

At the first step of the cluster analysis, the elements having the smallest measure of distance were combined, and they formed primary clusters. Then, at each step, the nearest element (or a cluster) was attached to each primary cluster. The process ended when all the elements merged into one cluster.

The results of clustering, which allowed distinguishing two groups of drives according to the climatic conditions of the territories of their location, are presented in Figure 2 and Table 1.

Statistical processing of data on the pollutant concentration and clustering thereof as per two parameters –the area of the MSW landfill and the average annual amount of precipitation – did not give statistically reliable dependences and did not allow identifying the appropriate groups. The methodology for statistical analysis was the same as for clustering by the temperature and the amount of precipitation.

Probably, in addition to the influence of the above-mentioned specific features of municipal waste management and climatic conditions, the essential effect came from the geological and hydrogeological conditions of the storage location. It is not possible to estimate the contribution of the latter factor judging by the available publications.



**Fig. 2.** Clustering on climatic parameters.

**Table 1.** Average values of the classification attributes.

Group	Average annual temperature, °C	Average annual atmospheric precipitation, mm	Number of landfills
1	25.4	2041	14
2	13.4	606	31

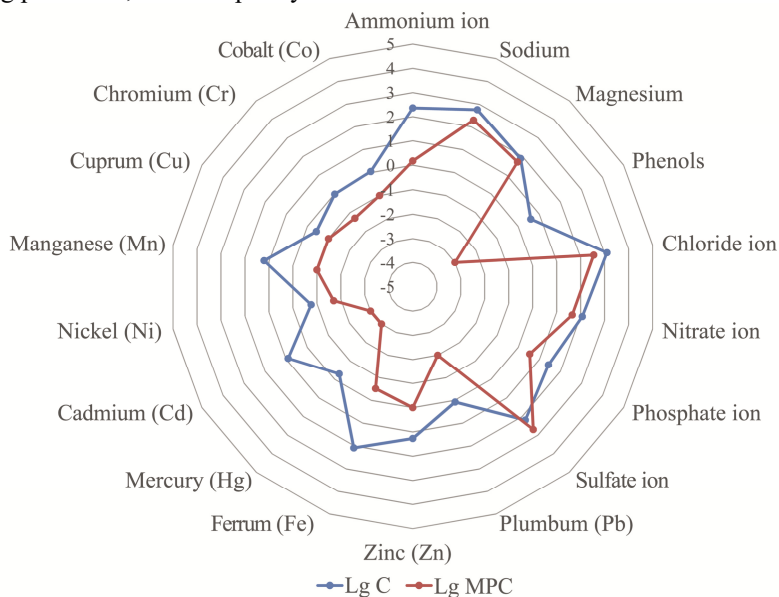
## 4 Statistical analysis of pollutant concentration by climate groups

In both groups, statistical data processing was performed for each of the pollutants, and the average values of their concentration, as well as the multiplicity of MPC exceedance, were calculated. The results are shown in Figures 3 and 4. Hereinafter, by the MPC, we will mean the maximum permissible concentrations of substances in the water of water bodies of household, drinking, and cultural and domestic water use in accordance with GN 2.1.5.1315-03. It is clear that these values are very conditional in this case, but they are used to assess the extent of contamination of the swamps considered to be water bodies.

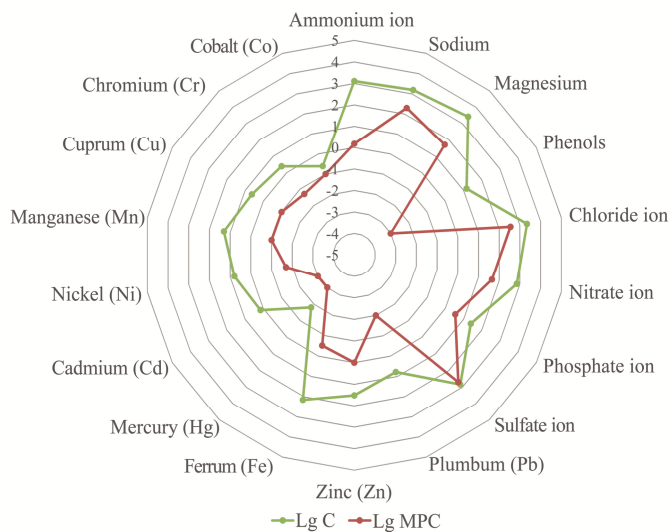
The average concentrations of phenols, Fe, Hg, Cd in Group 1 are characterized by the MPC exceedance by 404 ... 8125 times. For Na, Mg, chloride ion, nitrate ion, Cu, the MPC is exceeded by 1.5 ... 3.78 times. At that, the concentration of sulfate ion is 3.2 times lower than the MPC. For the remaining pollutants, the multiplicity of MPC exceedance varies from 7.8 to 158.8 times.

The average concentrations of ammonium ion, phenols, Pb, Fe, Cd in Group 2 are characterized by the MPC exceedance by 474 ... 14 640 times.

For chloride ion, sulfate ion, Co, the MPC exceedance is 1.38 ... 6.29 times. For the remaining pollutants, the multiplicity of the MPC exceedance varies from 7.2 to 313 times.



**Fig. 3.** Average logarithms of pollutant concentrations (lg C) for Group 1.



**Fig. 4.** Average logarithms of pollutant concentrations (lg C) for Group 2.

In order to compare the extent of groundwater contamination at the base of the landfills in two climatic groups, 18 of the 22 pollutants were selected that are present in groundwater samples in both groups. In each group, the data on the average pollutant concentration was statistically processed, and the logarithms of the multiplicity of MPC exceedance were calculated (Table 2).

The groundwater contamination index was determined for each group, which was calculated as the arithmetic mean value of the pollutants' MPC exceedance multiplicity logarithms. It turned out that a slightly lower extent of groundwater pollution is observed in Group 1. The decrease in the pollutant concentration and the multiplicity of the MPC exceedance respectively is probably caused by a stronger dilution of groundwater with abundant atmospheric precipitation. So, the average annual precipitation in Group 1 is 2041 mm, and in Group 2 it is equal to 606 mm.

**Table 2.** Logarithms of the pollutants' MPC exceedance multiplicity at the base of landfill.

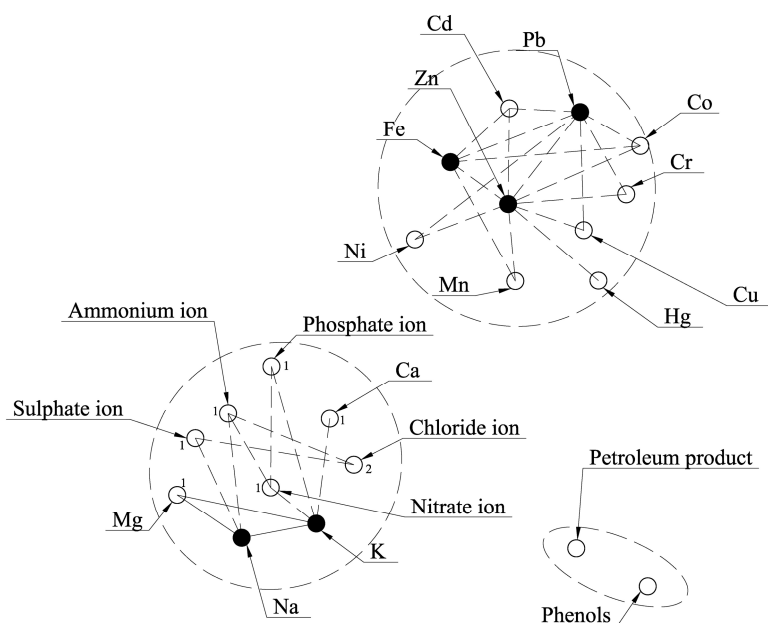
Substance	Climate group	
	1	2
Ammonium ion	2.19	2.93
Sodium	0.46	0.89
Magnesium	0.19	1.73
Total phenols	3.59	4.17
Chloride ion	0.56	0.80
Nitrate ion	0.42	1.20
Phosphate ion	0.89	0.86
Sulfate ion	-0.51	0.14
Plumbum (Pb)	2.08	2.78
Zinc (Zn)	1.27	1.51
Ferrum (Fe)	2.61	2.68
Mercury (Hg)	2.71	1.20
Cadmium (Cd)	3.91	3.15
Nickel (Ni)	0.94	2.50
Manganese (Mn)	2.20	2.30
Cuprum (Cu)	0.58	1.63
Chromium (Cr)	1.28	1.69
Cobalt (Co)	1.05	0.39
Contamination index	<b>1.47</b>	<b>1.81</b>

## 5 Cluster analysis and identification of correlations among pollutant concentrations

To establish correlations among the concentrations of various pollutants, as well as to identify pollutant markers, while having knowledge on the concentrations thereof for potential statistical evaluation of reliable concentrations of other substance, a multivariate statistical analysis was performed for the entire data set. Based on the analysis results, 3 clusters were identified (Figure 5).

Cluster A includes heavy metal cations, many of which are highly toxic. Compounds of these metals are widely used in the chemical industry, galvanic production, etc. Their sources of waste are electrical products, such as lamps, batteries, etc.

Cluster B is represented by ammonium cations, alkaline and alkaline earth metals, as well as by sulfate, nitrate, chloride and phosphate ions. The salts generated from these ions are mainly soluble compounds. Sources of their presence in water can be natural minerals (calcite, apatite). They are contained in the products of the chemical industry; in particular, they are part of mineral fertilizers.



**Fig. 5.** Scatter diagram with the correlations identified (solid line signifies the links with the correlation ratio  $r_{xy} > 0.6$ ; the dotted line stands for the links with the correlation ratio  $0.3 < r_{xy} < 0.6$ ).

Cluster C represents organic compounds: oil products and phenols. The sources of oil products in groundwater are the effluents from industrial enterprises, petrol stations, and the landfills – containers with waste lubricants. Phenolic compounds can stem from industrial wastewater of the chemical complex enterprises. In addition, phenolic compounds are formed in natural conditions during the transformation of native aromatic compounds, mainly macromolecular (lignin) ones.

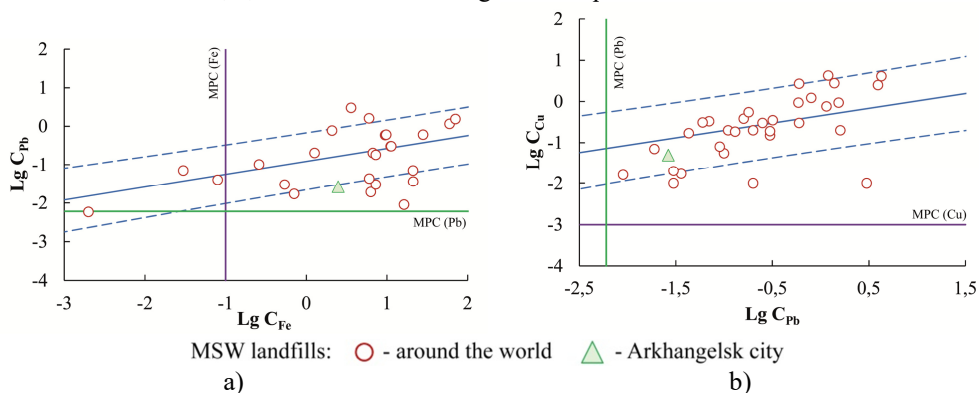
### 5.1 Correlation analysis of pollutant concentration in Cluster A

To perform the correlation analysis in Cluster A, decimal logarithms of concentrations were calculated, a correlation matrix was created, in which functional connections with the

Pearson correlation coefficient with p-level of statistical significance < 0.10 were singled out.

The matrix analysis showed that 3 statistically significant elements can be identified as marker pollutants: zinc (Zn), plumbum (Pb) and ferrum (Fe), which have the greatest number of links with other pollutants (Figure 5).

Figure 6 shows the dependences of the Pb concentration on the content of Fe (a), Cu on the content of Pb (b), and the remaining dependences are presented in Table 3, which are expressed as  $\lg C_p = a \cdot \lg C_{PM} + b$ , where  $C_p$  – pollutant concentration;  $C_{PM}$  – pollutant marker concentration; a, b - coefficients of regression equation.



**Fig. 6.** Dependence of the Pb concentration on the content of Fe (a), Cu on the content of Pb (b). The dotted lines show the confidence limits equal to  $\pm \sigma$ .

## 5.2 Correlation analysis of concentration of pollutants in Cluster B

The correlation analysis in Cluster B was performed by analogy with that of Cluster A. The analysis of the correlation matrix showed that 2 statistically significant elements could be identified as marker substances: sodium (Na) and potassium (K), which have the greatest number of connections with concentrations of the other substances (Figure 5). At the same time, chloride ion does not have stable correlations with the marker substances and should be considered as a separate marker. The coefficients of the regression equations for the concentration of marker substances with concentrations of other pollutants are presented in Table 4. Figure 7 displays the dependence of the Mg concentration on the K content (a) and the dependence the nitrate ion concentration on the chloride ion content (b).

Correlation analysis of cluster C showed that no statistically significant correlations were established between the two substances (Fig. 5).

Thus, the analysis of groundwater contamination at the base of MSW landfills allowed to identify 3 clusters of pollutants, to determine the allocation of marker substances and to establish statistically significant correlations between their concentration and the concentration of other pollutants. For verification of the obtained dependences results of monitoring the composition of groundwater at the base of landfill of Arkhangelsk city were used. Eleven pollutants (K, Na, Mg, Ca, ammonium-, chloride-, nitrate-, sulphate ion, Pb, Fe, Cu) were under monitoring. Chemical analysis of groundwater samples taken near MSW landfill in the Arkhangelsk City has shown that the pollutants concentration at the landfill in the City of Arkhangelsk corresponds to the correlation dependencies identified by the literature data.

**Table 3.** Coefficients of the regression equations Cluster A.

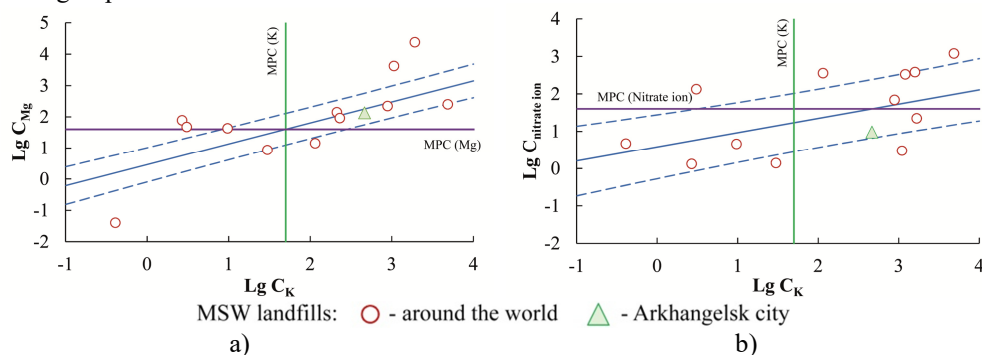
Pollutant marker (PM)	Pollutant (P)	Coefficients of the regression equations $lgC_p = a \cdot lgC_{PM} + b$	
		a	b
Zinc	Manganese	0.371	-0.247
	Nickel	0.333	-0.434
	Ferrum	0.592	1.014
	Cadmium	0.512	-1.299
	Cobalt	0.094	-0.926
	Cuprum	0.348	-0.559
	Mercury	0.182	-2.015
Plumbum	Cobalt	0.131	-0.849
	Chromium	0.263	-0.170
	Cuprum	0.359	-0.346
	Nickel	0.382	-0.209
	Ferrum	0.623	1.384
Ferrum	Cadmium	0.608	-0.940
	Zinc	0.553	-0.547
	Manganese	0.383	-0.632
	Cobalt	0.087	-1.013
	Plumbum	0.336	-0.917
	Cadmium	0.337	-1.634

**Table 4.** Coefficients of the regression equations Cluster B.

Pollutant marker (PM)	Pollutant (P)	Coefficients of the regression equations $lgC_p = a \cdot lgC_{PM} + b$	
		a	b
	Magnesium	0.810	-0.157
Sodium	Sulfate ion	0.262	-0.434
	Ammonium ion	0.278	1.688
	Magnesium	0.672	0.460
	Nitrate ion	0.381	0.583
Potassium	Phosphate ion	0.187	0.209
	Calcium	0.159	1.865

## 6 Conclusions

Analysis of scientific publications allowed determining the list of pollutants coming from MSW landfills into the groundwater. These substances were divided into three groups, the substance markers were identified, by the concentration of which the correlation dependencies make it possible to estimate the content of other substances in groundwater in each group.



**Fig. 7.** Dependence of the Mg concentration on the content of K (a) and nitrate ion dependence on the content of K (b). The dotted lines show the confidence limits equal to  $\pm \sigma$ .

## 6 Conclusions

Analysis of scientific publications allowed determining the list of pollutants coming from MSW landfills into the groundwater. These substances were divided into three groups, the substance markers were identified, by the concentration of which the correlation dependencies make it possible to estimate the content of other substances in groundwater in each group.

The study was funded by a grant from the Russian Foundation for Basic Research (Project No. 18-45-292004).

## References

1. S. B. Sosorova, M. G. Merkusheva, L. L. Ubugunov, *Geochemistry of landscapes and geography of soils: Proceedings of the All-Russian scientific conference*, 303-304 (2012)
2. N. S. Larionov, K. G. Bogolitsyn, M. V. Bogdanov, I. A. Kuznetsova, *Chem. Plant Raw Mater.* 147-152 (2008)
3. G. S. Morozov, I. P. Breus, *Geochemistry of landscapes and geography of soils: Proceedings of the All-Russian scientific conference*, 223-225 (2012)
4. I. V. Telminov, *Bull. Volgogr. St. Univ. Archit. Constr. Ser.: Constr. Archit.* **42(61)**, 25-38 (2015)