

Relation between air and soil pollution based on statistical analysis and interpolation of Nickel (Ni) and Lead (Pb): Case study of Zagreb, Croatia

Račić, Nikolina; Malvić, Tomislav

Source / Izvornik: **Mining of Mineral Deposits, 2023, 17, 112 - 120**

Journal article, Published version

Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

<https://doi.org/10.33271/mining17.02.112>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:169:669045>

Rights / Prava: [Attribution 4.0 International](#) / [Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-07-03**




Repository / Repozitorij:

[Faculty of Mining, Geology and Petroleum Engineering Repository, University of Zagreb](#)



Relation between air and soil pollution based on statistical analysis and interpolation of Nickel (Ni) and Lead (Pb): Case study of Zagreb, Croatia

Nikolina Račić^{1, 2}✉, Tomislav Malvić¹*✉ 

¹ University of Zagreb, Zagreb, Croatia

² Institute for Medical Research and Occupational Health, Zagreb, Croatia

Corresponding author: e-mail tomislav.malvic@rgn.unizg.hr

Abstract

Purpose. This paper focuses on the comparison of Ni and Pb concentrations in air and soil pollution in the Zagreb area. Due to the very limited amount of publicly available data from soil analysis samples, 2016 and 2019 were chosen as the best possible indicators of related changes in metal concentrations in soil and air.

Methods. Testing the normality of Ni and Pb concentrations in the total deposited matter (TDM) confirmed the feasibility of using two parametric statistical tools – the Pearson correlation coefficient and the *t*-test. The Inverse Distance Weighting (IDW) interpolation method was selected as the best approach for a small number of measurements.

Findings. The insufficient amount of data is the main shortcoming for urban health policy in a large area like Zagreb. The small number of air measurement stations and especially soil sampling sites cannot lead to any reliable conclusions about urban pollutants, their activity over time and direct links to soil toxic degradation based on statistical or geological methods and analyses. However, there is no doubt that urban pollution sources fill the soil with accumulated toxic elements such as Ni and Pb, especially in suburban areas located along the paths of the dominant wind directions.

Originality. This is an original research that for the first time statistically analyzes and maps publicly available air and soil pollution data for the period 2016-2019.

Practical implications. This research is a necessary step in determining the future planning of air and soil measurement stations in the Zagreb urban area.

Keywords: *pollution, Ni, Pb, statistical testing, geostatistical mapping, Zagreb, Croatia*

1. Introduction

Soil and air pollution continues to be a major problem in urban areas such as Zagreb. As the largest city in Croatia, it faces high population density and heavy traffic, which has a significant impact on air and soil quality. The total mass of pollutants transferred from the air to surfaces is called the total deposited matter (TDM). The TDM is one of the parameters for assessing environmental air quality. As part of TDM determination, TDM metal concentrations are also determined (Pb, Cd, As, Ni, Tl, etc.). Considering that air particles settle to the ground as a result of deposition in the air, it can be concluded that there is a correlation between the concentrations of metals in air and soil.

Air quality is an important environmental variable that determines the spatial accumulation of heavy metals in the soil. Some results have shown that there is a sources-receptor relationship mechanism between air quality and heavy metals in soil. Industry releases heavy metal particles into the atmosphere, which then enter the soil through atmospheric deposition and precipitation [1].

However, soil pollution in urban or rural areas can be affected by both air and water streams. Even more, both me-

chanisms are generally the most active in mining areas, increasing the concentrations of heavy metals, for example, in industrial discharged water or gases. So Kadriu et al. in [2] proved for several sampling sites that the industrial activity (mine and flotation) of the Trepča Mine is the main cause of environmental pollution. Their analysis of discharged water from the mine and the flotation showed concentrations of heavy metals in the surrounding environment.

Moreover, Polishchuk et al. in [3] defined the problem of assessing the guaranteed atmospheric air quality for a point source. Such point(s) can be defined in various fields such as mining, metallurgy, chemistry etc., but are always described as a vector random field and a variable with multidimensional distribution density. Consequently, there is a clear link between air quality and industrial sources.

Skrobala et al. in [4] presented mining as one of the most dangerous sources of environmental pollution using the case study of the Nadiya mine rock dumps in Ukraine. Pollution depends on the relief and slope exposure, which affect the distribution of chemical elements and their increase compared to the background values. Dispersion (according to the Principal Component Analysis) influenced by anthropogenic

Received: 9 May 2023. Accepted: 15 June 2023. Available online: 30 June 2023

© 2023, N. Račić, T. Malvić

Mining of Mineral Deposits. ISSN 2415-3443 (Online) | ISSN 2415-3435 (Print)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

activity is the most intense for Mg, Pb, Sn, Fe, Al, Cu, P, Ni, Zn, affecting vegetation near dumps.

However, in Croatia, mining activities do not have a direct influence on the environment due to the closure of all active coal mines several decades ago (last in 1999). Only a few relatively small quarries are still active, but their influence is not chemical, like the accumulation of heavy metals, dissolved ions etc., but in air quality due to the large quantities and volumes of mining dust. Gravel and clay pits do not have any influence on local inhabited areas, except for aquifers (unless they are properly planned).

Consequently, the biggest air pollutant in the city of Zagreb is traffic [5], in addition to several industrial plants in the eastern part of the city, a heating plants, a waste disposal site and an airport near the city. Of course, pollution can also occur when a pollutant enters the soil directly. However, this data is not considered in this work, since the main goal is to monitor where most wind-blown pollutants (mostly caused by traffic) settle, so that we can better adapt the measurement stations in further research. Polluting substances are carried by the wind and, depending on the particle size, direction and speed of the wind, settle on the soil surface near the source of pollution or are carried to more distant areas (primarily if they are small particles $PM_{2.5}$ and smaller, ultrafine particles). Many papers study the relationship between pollutants and the influence of meteorology on the behaviour of pollutants, the direction of their trajectories and, ultimately, deposition [6]-[8]. The general conclusion of all the papers is that wind direction plays a key role in pollutant migration from its source to deposition. The mechanisms that explain this relationship are complex, since many other phenomena influence the path of particles from the source to the soil surface, such as the size of the particles themselves, wind direction and speed, rainfall, vegetation, tall buildings, etc. [9].

Heavy metals are very dangerous for the environment and organisms. Through the food chain, humans suffer the consequences of soil pollution. If soil is contaminated with heavy metals, it is difficult to rehabilitate it [10]. These heavy metals are persistent and bioaccumulative, do not biodegrade or metabolize in the environment [11].

Within the framework of this paper, Ni and Pb were selected as potentially toxic elements in high concentrations. These elements were chosen because of the same elements measured in soil near air measurement stations during two years of observations.

Nickel (Ni) is a toxic and carcinogenic heavy metal. It enters the soil with diesel fuel, smelters, city effluent and mining [12]. The average concentration of Ni in soils is 20-30 mg/kg [13]. Spatial distribution of Ni in central Croatia area ranges within 12 and 427 mg/kg with an average value of 33 mg/kg, which is lower than the average value for the entire Croatia (48 mg/kg) (Fig. 1) [14].

Lead (Pb) is a heavy metal mostly presented in the top layer of soil due to the deposition from air containing smoke from vehicles. In uncontaminated soils, Pb concentrations are generally below 50 mg/kg [15]. Also, industrial activities have a large impact on heavy metal pollution, ecological and health risks. When the trace element concentration exceeds the critical threshold, it may have a potentially toxic effect on human health and the ecosystem [16]. According to the International Agency for Research on Cancer (IARC), high concentrations or extreme exposure to Pb can cause severe neurological and hematological disorders in the exposed population, mainly children.

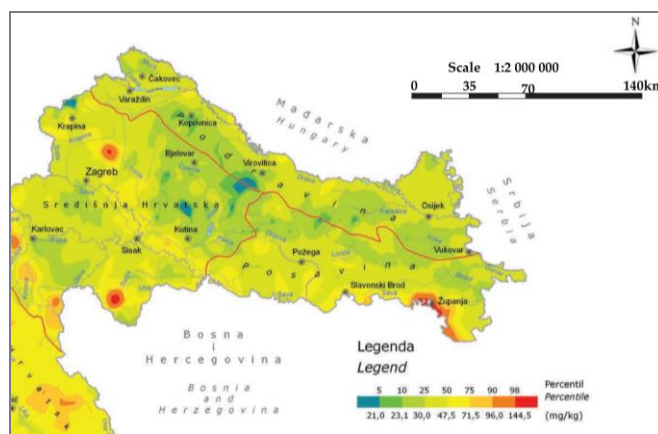


Figure 1. Regional spatial distribution of Ni [10]

On the other hand, short-term exposure to a high Ni concentration via inhalation can cause kidney and lung disorders. Exposure to high levels via ingestion can lead to neurological disorders and gastrointestinal discomfort. Moreover, long-term exposure through dermal contact may result in dermatitis, while inhalation can cause many respiratory diseases, such as nasal and lung cancer [17]. Lead content in this region varies between 14 and 217 mg/kg, with an average value of 27 mg/kg, which significantly exceeds the state average content of this element in soils (Fig. 2) [14].

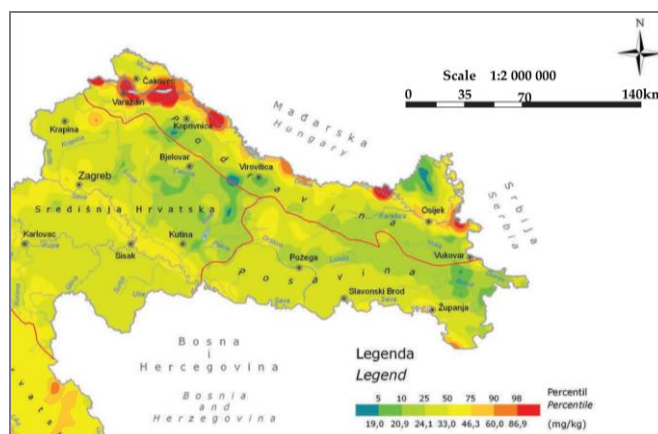


Figure 2. Regional spatial distribution of Pb [14]

Soil enrichment with Cd, Zn, As, and Pb is attributed to high-temperature coal combustion [18]. In the areas of large cities, there are mixed sources of pollution from industrial emissions, the burning of fossil fuels, emissions from vehicles, irrigation with wastewater, as well as the disposal of solid waste and natural sources of parent soil materials in urban soils [19].

The major methods for the prevention and control of heavy metals in soils are focused on the identification of heavy metals in the soil from different pollution sources. The average contribution rates of the soil-forming process in Cd, Hg, As, and Pb are around 80-90%, and those of anthropogenic activities are 10-20% [20]. Certain papers come to the conclusion that the source material plays a very important role in the distribution and quantity of metals in soil horizons. Based on the example of Cd, it was concluded that the parent material is the decisive factor determining the magnitude and depth of exogenous Cd accumulation in the soil profile. These findings suggest that the parent material-induced Cd fraction distribution and accumulation should be

considered in order to effectively explore remediation strategies for Cd pollution [21].

However, how pollutants behave in an unsaturated zone is a very complex question. Their carrier is affected by many physical, chemical and microbiological processes, such as advection, dispersion, diffusion, sorption, cation exchange, etc. [22]. Sorption processes of potentially toxic metals in soils differ at different soil pH values, and the soil ability to retain them depends on its resistance to any soil pH changes [23].

A total of 420 surface soil layer samples (0-15 cm) were sampled in Zagreb and surrounding areas using systematic sampling on a 2 km grid [24]. The conclusion of this work is that the influence of the anthropogenic urban environment on the distribution of heavy metals in the soils of Zagreb is noticeable only for Hg, Pb and Zn, while pollution caused by Cr, Ni and Mn has not revealed. When investigating the movement of pollutants in the unsaturated zone of the Zagreb aquifer [25], the highest concentrations of potentially toxic elements (Pb, Cd, Zn) were determined in the shallowest part of the soil profile. Considering the obtained results of the predicted transport model, in this case, the initial concentration of potentially toxic elements on the soil profile surface is about 10% of the elements carried into the aquifer. Therefore, the soil surface layer can be considered the most significantly polluted with potentially toxic metals from the air.

An interesting research was published by Saeiti et al. [26], where the authors pointed out mixed models as a useful way of longitudinal discriminant data analyses. Although the model was applied to specific patient deviations from the general trend over time (i.e., in biostatistics), they introduced a so called “joint distribution” to include all model random effects and assumed as multivariate normal distribution. Two sample sets are tested ($n = 250$ and 2500), where misspecification of the random effect distributions is found. Authors [26] stated that “when there is severe departure from normality, more flexible mixture distributions can give better classification accuracy. However, in many cases, the incorrect assumption of a single multivariate normal distribution has little impact on the classification accuracy”. Although analysis considered time-shifted medical data, an important general conclusion can be drawn. If there is a slight deviation (for random effects) from a normal distribution, then assuming a single Gaussian distribution will not lead to significantly less accuracy. However, if the deviation is larger, more accurate classifications will be reached with “a more flexible random effects distributions”, but it is only possible for larger datasets.

In a similar study, also in the field of medicine, presented by Sheng et al. [27], a correlation is analysed in clustered data (formed by two ears of the same individual). They also propose a linear mixed-effects model for analysing the multi-level data structures. Generally, they concluded that more efficient models are mixed, multivariate, with a larger number of inputs (in this case, method for both ears, including participants, audiologists and confounder).

Also often in geology there is an approach to increase artificially the data using bootstrap as a nonparametric statistical method. Ivšinović and Litvić [28] show that resampling an input dataset can provide a chance to reach a new set which is characterised by normal distribution. However, this method is based on a large number of resampling. The authors show that at least 1000 (new) subsurface porosity artificial data were necessary to obtain a normal distribution.

All these approaches deal with distribution of data and can be tested for both small and large datasets in different fields of science and engineering. They all have their limitations, especially bootstrap when resampled datasets may not be statistical representative, but they all can help to make decision how to spatially represent variability.

2. Methods

2.1. Study area description

The study area includes six measurement stations in Zagreb for observing air quality (A1-A6). The behaviour of air pollutants in the environment, as well as human exposure to air pollutants in Croatia, are investigated and monitored by the Institute for Medical Research and Occupational Health (IMI). During the two-year period, 2016 and 2019, daily samples of Ni and Pb concentrations in the total deposited matter were taken, and monthly and annual averages are considered for mapping and statistical analyses. Additionally, soil samples were taken from two locations (S1 and S2), and values in 2016 and 2019 were compared with air pollutants. These 2 years are the only ones when mutual air and soil measurements exist. Moreover, locations S1 and S2 were chosen due to their proximity to possible pollution sources. S1 is the location of the Zagreb wastewater facility and S2 is close to the city landfill. The entire observed area (air measurement and soil sampling stations) covers approximately 300 km². The sampled locations are shown in Figure 3.

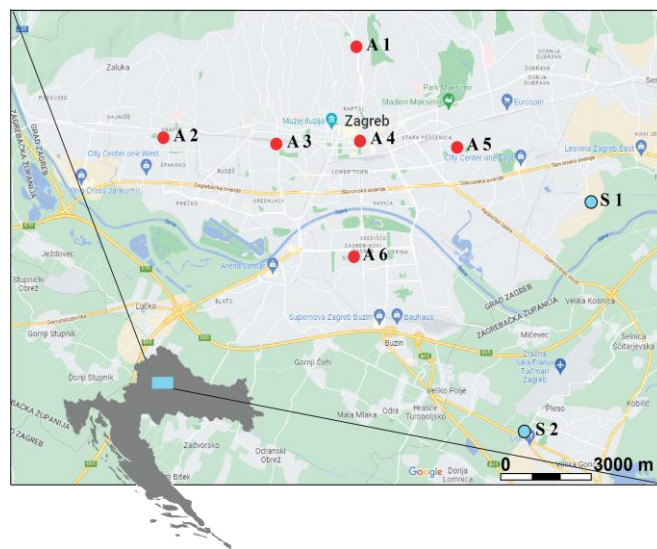


Figure 3. Study areas with measurement stations (A – air measurement stations, S – soil sampling stations)

Soil samples were taken at a depth of 20 cm and subsequently subjected to drying, sieving and preparation for sequential extraction analysis (BCR) [29]. The BCR method was used to determine the overall distribution of heavy metals in sediments.

2.2. Geological settings

The study area consists primarily of Quaternary deposits composed of Pleistocene non-carbonate loess and Holocene alluvium, flooded and splay sediments [30]. It is represented by lowlands located at the very southwestern margin of the Pannonian Basin System, below Medvednica Mt. These Holocene sediments are carried by the Sava River in numerous redeposition cycles originating from the Alpine regions [31].

Apart from alluvial deposition, a significant effect of tectonics results in pronounced heterogeneity and anisotropy [32].

Quaternary sediments surrounding Zagreb were firstly described at the beginning of the century [33]. Based on all available geologic-geophysical data, Pleistocene and Holocene sediments have been completed to a depth of approximately 100 m. At this depth, sediments are composed of loose and weakly cemented clasts, the sizes of which correspond to gravels, sands, silts and clays in variable proportions (Fig. 4).

The upper part (upper 50 m) is dominated by gravels. They are underlain by a thin layer of silty clay overlying the following sequence: gravels and sands, thin silty clayed sediments and sand, silts and clays forming the lowermost 10 meters [34]. Analysed surface soil samples were taken from 0-20 cm depth in locations S1 and S2 [25], [29], [35], while metal concentrations in the total deposited matter were measured monthly in 2016 and 2019 at six locations, namely A1-A6 (Table 1-3).

2.3. Statistical analysis

The Inverse Distance Weighting (IDW) interpolation method was selected, which will give the best data for a small number of measurements [36]. The IDW method is based on the distance ponder with an exponent, mostly the second power [37]-[39], between the measured data and the location where an unknown value is estimated. Such values are in grid nodes.

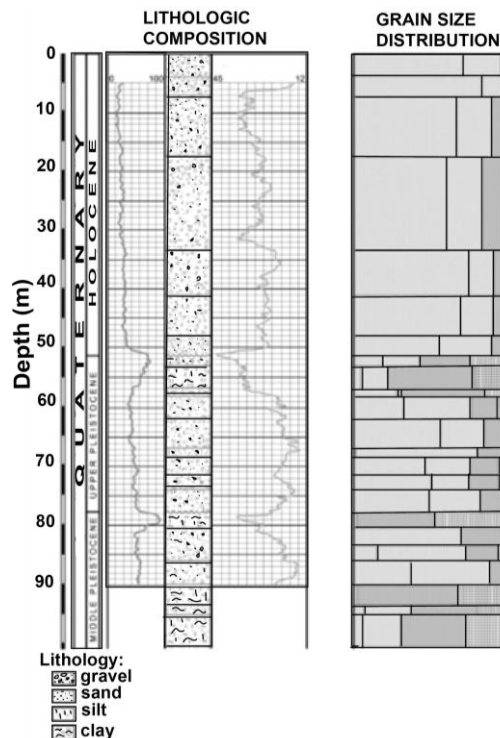


Figure 4. Lithologic column of the Jakuševac well in the Zagreb waste depository [34]

Table 1. Annual Pb concentrations in the TDM (mg/m²d)

A1		A2		A3		A4		A5		A6	
2016	2019	2016	2019	2016	2019	2016	2019	2016	2019	2016	2019
1.89	0.55	8.03	10.9	1.54	0.95	1.99	1.34	1.20	0.76	1.37	0.63
3.27	1.27	10.5	5.89	4.80	2.02	4.25	3.14	3.29	1.01	3.28	1.31
2.46	1.19	4.53	7.92	2.59	3.37	4.50	4.04	2.58	1.32	5.80	2.03
3.66	5.11	4.10	15.5	1.57	4.17	7.42	4.97	1.59	3.12	3.27	3.35
3.42	1.22	8.76	2.85	3.99	0.47	4.62	1.72	2.38	0.95	3.21	0.81
2.47	2.42	13.1	8.49	3.47	3.91	5.06	3.64	3.70	2.40	3.39	0.84
1.85	1.65	3.86	6.38	/	/	4.07	2.70	1.96	1.28	2.13	1.78
5.42	1.06	4.30	5.97	2.06	1.79	2.19	2.92	2.29	2.36	1.66	1.88
1.05	1.65	4.41	3.73	1.50	1.99	1.73	2.81	1.32	1.83	1.37	3.23
5.94	0.53	7.70	1.66	2.55	0.93	3.38	1.29	2.76	0.73	2.25	0.77
/	/	4.92	1.84	2.52	1.16	4.52	1.46	/	/	2.68	1.11
0.90	0.46	2.74	7.41	2.09	1.05	4.54	1.07	2.09	0.74	1.25	1.04

Table 2. Annual Ni concentrations in the TDM (mg/m²d)

A1		A2		A3		A4		A5		A6	
2016	2019	2016	2019	2016	2019	2016	2019	2016	2019	2016	2019
9.67	0.33	3.05	2.15	2.82	0.66	2.85	0.54	2.74	0.64	5.18	0.65
7.24	0.26	4.25	2.35	1.24	0.97	3.00	1.39	2.17	0.83	3.14	0.89
1.83	0.65	2.81	5.30	1.30	1.32	1.55	1.56	1.06	0.64	1.87	1.27
3.60	2.08	3.25	5.37	1.28	2.26	3.38	2.06	1.52	1.32	2.56	1.72
5.84	0.86	22.0	2.05	2.86	0.32	2.41	0.72	1.59	0.61	2.10	0.61
6.17	1.44	3.49	3.26	2.27	2.32	4.88	1.90	2.05	1.22	1.97	0.64
1.56	0.82	1.71	1.90	/	/	1.22	0.86	0.90	0.69	1.16	0.97
5.96	0.69	0.62	2.41	0.94	1.02	0.89	1.22	1.16	1.46	0.89	1.02
1.29	0.89	2.47	1.93	1.08	1.06	1.42	1.21	0.84	1.00	0.93	1.89
3.45	0.30	3.35	1.22	2.29	0.67	2.45	4.78	2.08	0.38	1.64	0.87
/	/	3.97	1.33	1.84	1.37	2.51	1.18	/	/	3.06	1.05
1.42	0.27	0.38	0.94	1.33	1.11	1.34	0.83	2.54	0.44	1.10	0.64

The measurement data included in the calculations are the values within the search radius/radii, i.e. inside the circle/ellipsoid of spatial dependency.

The mathematical expression for inverse distance estimation is:

$$Z_{UI} = \frac{Z_1 + Z_2 + \dots + Z_n}{\frac{1}{d_1^p} + \frac{1}{d_2^p} + \dots + \frac{1}{d_n^p}}, \tag{1}$$

where:

- Z_{UI} – estimated value; di-distance to the “i-th” location;
- p – power of distance;
- Z_i – measured values at “i-th” location ($i = 1, \dots, n$) [40].

Table 3. Average annual metal concentrations in the TDM and soil samples

Metal concentrations in the total deposited matter (mg/m ² d)				
	2016		2019	
	Pb	Ni	Pb	Ni
A1	2.94	4.37	1.5	0.8
A2	6.42	4.28	6.6	2.5
A3	2.61	1.75	2.1	1.2
A4	4.02	2.33	2.6	1.5
A5	2.29	1.7	1.5	0.8
A6	2.64	2.13	1.6	1
Metal concentrations in soil samples (mg/kg)				
	2016		2019	
	Pb	Ni	Pb	Ni
S1	18.24	205.07		
S2	18.26		27.1	38.3

The result of IDW interpolation depends on the value of the exponent p , which is obtained experimentally and has a different value in different fields of science [38]. Specifically, in this work, the value of 2 was used as the exponent p , which in previous studies in similar areas was shown to be optimal [38], [39]. This means that the weight given to each point is inversely proportional to the square of the distance between the unknown point and the known point. This gives more weight to nearby points and less weight to distant points, resulting in a smoother and more accurate interpolation surface. However, the choice of power exponent depends on the nature of the data and the desired accuracy of the interpolation. For example, a power exponent of 1 will give equal weight to all neighbouring points, while a higher power exponent will give more weight to the closest points.

To investigate the correlation between variables (in this case, metal concentrations), it is useful to check the normality of the TDM annual data distribution. In the limited datasets, a normality check is quite an unreliable analysis. However, the amount of data for normality analyses, t -test and interpolation was different. Statistical data were checked based on monthly averages for two compared years (t -test) or single years (normality check). It included 24 or 12 values, respectively. Oppositely, for interpolation, the datasets included 6 points and 2 auxiliary (for interpretation) data. So, 10+ data points are considered the minimum for the logical selection of formal normality tests, where also one was “stricter” (Shapiro-Wilk) and another “lighter” (Kolmogorov-Smirnov). One of the most frequently used and reliable (strong) tests for confirming the hypothesis of a normal distribution, the Shapiro-Wilk (SW) test is expressed by the Equation:

$$W = \frac{\left(\sum_{i=1}^n a_i \cdot x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \tag{2}$$

where:

- $x_{(i)}$ – the i -th smallest number in the sample;
- \bar{x} – a sample mean;
- a_i – a constant.

This is a hypothesis test that is applied to a sample and whose null hypothesis is that the sample was generated from

a normal distribution. If the p -value is low, we can reject such a null hypothesis and say that the sample was not generated from a normal distribution [41].

Due to the small number of data (monthly averages), we additionally check the normality with another test called the Kolmogorov-Smirnov (KS) test [42]. The test is based on the empirical distribution function (ECDF). Given N ordered data points Y_1, Y_2, \dots, Y_N , the ECDF is defined as:

$$E_N = \frac{n(i)}{N}, \tag{3}$$

where:

- $n(i)$ – the number of points less than Y_i ;
- Y_i – ordered from the smallest to the largest value.

This is a step function that increases by $1/N$ at the value of each ordered data point.

For groups where normality is confirmed by at least one formal test, the parametric t -test was made to compare the means of the two groups of monthly values between 2016 and 2019. The paired sample t -test was chosen to determine whether the mean difference between two sets of observations is zero [43]. In the t -test, we used the p -value as a measure of the statistical significance of the observed result. If the p -value is small (less than 0.05), it suggests that the observed result is unlikely to have occurred by chance, and we reject the null hypothesis. On the other hand, if the p -value is large, this suggests that the observed result could have been obtained by chance (the concentration of individual metals in different years is statistically significant). It is important to note that the p -value does not provide information on the magnitude of the effect or the size of the difference between groups; it only shows us whether the observed effect is statistically significant. Two sets of observations are paired if each observation in one set has a special correspondence of relationship with exactly one observation in the other data set [44]. The formula for estimating the paired t -test is:

$$t = \frac{m}{\frac{\sigma}{\sqrt{n}}}, \tag{4}$$

where:

- m – the mean of the group;
- σ – the standard deviation of the group;
- n – the sample size.

Also, for samples with a normal distribution, a correlation was made using Pearson correlation coefficient [45]. The Pearson correlation coefficient is a measure of the strength of a linear relationship between two variables [46]. It is commonly denoted by the Greek letter ρ (rho) and may be referred to as the population correlation coefficient. Given a pair of random variables (X, Y), the formula for ρ [44] is:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}, \tag{5}$$

where:

- cov – the covariance;
- σ_X – the standard deviation of X ;
- σ_Y – the standard deviation of Y [47].

When applied to a sample, it is commonly represented as r and may be referred to as the sample correlation coefficient or the sample Pearson correlation coefficient.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}, \tag{6}$$

where:

- r – correlation coefficient;
- x_i – values of the x -variable in a sample;
- \bar{x} – a mean of the values of the x -variable;
- y_i – values of the y -variable in a sample;
- \bar{y} – a mean of the values of the y -variable.

3. Results and discussion

In order to check the normality of the distribution of monthly values of Pb and Ni concentrations in the air during 2016 and 2019, Shapiro-Wilks and Kolmogorov-Smirnov formal tests were performed. The results are presented in Table 4.

Table 4. Results of the statistical tests of Pb, Ni concentration in the TDM (green – normal distribution is confirmed, red – is not confirmed)

		A1		A2		A3		A4		A5		A6	
		Pb	Ni	Pb	Ni	Pb	Ni	Pb	Ni	Pb	Ni	Pb	Ni
2016	KS test	Green	Green	Green	Red	Green	Green	Green	Green	Green	Green	Green	Green
	SW test	Green	Green	Green	Red	Green	Green	Green	Green	Green	Green	Green	Green
2019	KS test	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
	SW test	Red	Green	Red	Green	Green	Green	Red	Green	Green	Green	Red	Green

Most monthly concentrations have a normal distribution, especially when analysed using the KS test. It was expected because the KS test is not as strict as the SW test. The monthly Ni concentrations at the A2 measurement site in 2016 are an exception, since there was one extreme value, which was probably real and not a measurement error. Based on this, we can compare the monthly concentrations of individual measurement stations in 2016 and 2019 (except for Ni at A2) using the t -test and Person correlation coefficient (Table 5).

Table 5. Results of statistical paired t -test (p -value) and Pearson correlation coefficient (r) between concentrations in 2016 and 2019

		A1		A2		A3		A4		A5		A6	
		Pb	Ni	Pb	Ni	Pb	Ni	Pb	Ni	Pb	Ni	Pb	Ni
r		0.04	-0.12	-0.07	-0.05	-0.06	-0.32	0.49	0.24	-0.02	-0.37	0.09	-0.26
p		0.05	0.002	0.93	0.33	0.26	0.12	0.01	0.07	0.04	0.01	0.03	0.02

In the t -test, we used the p -value as a measure of the statistical significance of the observed result. If the p -value is less than 0.05, it suggests that the observed result is unlikely to have occurred by chance, and we reject the null hypothesis (the concentration of individual metals in different years is not statistically significant). When normality is not satisfied, this means that the data distribution is not Gaussian, and therefore statistical methods based on this assumption may not be valid. In the context of air pollution data, this may happen because air pollution sources are not evenly distributed in space and time, or because some pollutants are more persistent in the atmosphere than others. In this case it is important to use non-parametric methods, such as the Wilcoxon rank-sum test, the Kruskal-Wallis test, Spearman’s rank correlation or Robust regression, which do not assume a specific data distribution. Alternatively, one can

transform the data using a suitable transformation, such as logarithmic or square root transformation, to make the distribution more normal before applying parametric methods.

Pearson correlation coefficient does not significantly differ from zero. According to this correlation coefficient, there is not significant linear relationship between monthly concentrations of Ni in 2016 and 2019, as well as between respective Pb concentrations. A paired t -test is used because it can be assumed that, without additional external pollutants, the sample means will be similar, i.e., statistically related. The null hypothesis assuming similarity between 2016 and 2019 monthly means is rejected based on the p value at locations A1, A5 and A6 for both Pb and Ni concentrations. At measurement station A4, the null hypothesis is not rejected for Ni concentrations, and it is rejected for Pb. At measurement station A2 and measurement station A3, null hypothesis is not rejected for both Pb and Ni concentrations.

In short, using an additional t -test, the hypothesis of equality of their means is mostly rejected ($H1 : \mu1 \neq \mu2$), which places the time-shifted data of majority of air samples in different weather and soil conditions. This can be compared to negative correlations often calculated between the same datasets, where both values mask any relationship between stations, which probably cannot be the prevailing statement. However, finding the critical value of the rock petrophysical values when applying t -test in geological research could be a tentative process [48], and an increase in the number of data can help reach the representativeness of the statistics, i.e., belonging to the so called “parent population”. This means that increasing the network of stations probably improve the relationship between different stations, expressed by statistical values such as correlation coefficient or t -test values.

Significant differences in the sample means are also visible geographically on maps of Ni and Pb concentrations in both analysed years (Figs. 5 and 6). Measurement stations A3 and A2, for which null hypothesis is not rejected, are adjacent and geographically located from the central to the western side of the mapped area. Mapped Pb concentrations in 2016 and 2019 show the highest values in the central area (A3, A4) as well as at the western margin (A2). The average annual Pb concentration reaches 4.02 $\mu\text{g}/\text{m}^2\text{d}$ at the measurement A4 station in 2016 and 2.59 $\mu\text{g}/\text{m}^2\text{d}$ at the same station in 2019. At A2 station, an average annual concentration of Pb in 2016 is 6.42 $\mu\text{g}/\text{m}^2\text{d}$ and 6.55 $\mu\text{g}/\text{m}^2\text{d}$ in 2019. Pb concentrations of soil samples from S1 and S2 locations are 18.24 mg/kg and 18.26 mg/kg in 2016, respectively. Although S1 data for 2019 is not available, S2 reaches a high of 27.1 mg/kg in 2019.

Unlike Pb, the average annual concentrations of Ni are maximal only at the margin of the mapped area in 2016. The A1 and A2 stations show 4.37 and 4.28 $\mu\text{g}/\text{m}^2\text{d}$, respectively. This changed in 2019, with station A2 retained the highest mapped concentration, while Ni concentration at station A1 became the lowest mapped concentration. Soil samples from location S1 show a value of 205.07 mg/kg in 2016, and no data is available for 2019. Similarly, location S2 shows a sample concentration of 38.3 mg/kg in 2019, while data at S1 for 2016 is not available.

Since the spatial distribution of these potentially toxic elements (Pb and Ni) can be significantly affected by wind direction and speed, both factors were observed using an annual wind rose valid for Zagreb (meteorological station located in the uptown).

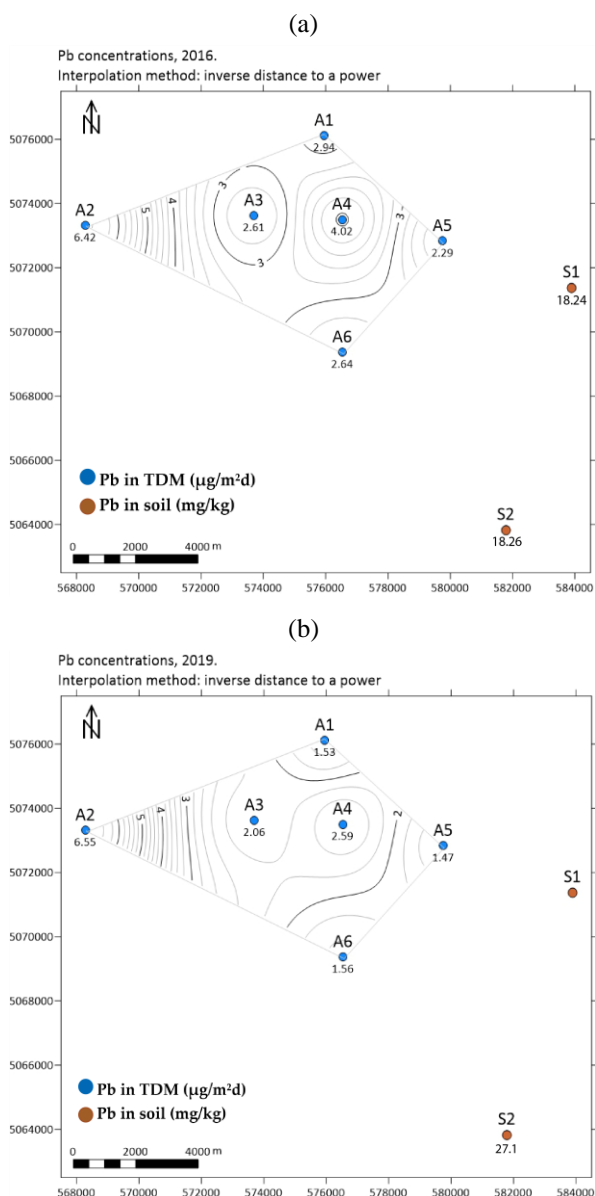


Figure 5. Map of Pb concentration in the TDM and soil (Gauss-Krüger projection): (a) 2016; (b) 2019

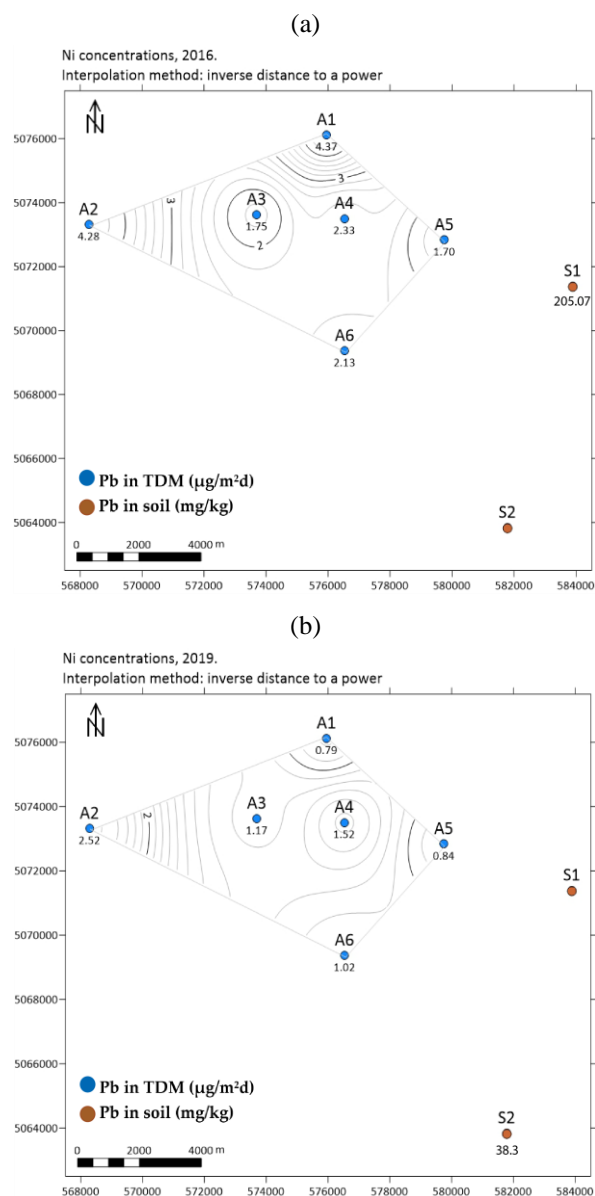


Figure 6. Map of Ni concentration in the TDM and soil (Gauss-Krüger projection): (a) 2016; (b) 2019

The data for measured wind properties reflect the same time frames (2016 and 2019) as the Pb and Ni concentration measurements from the air. The length of each wind direction spike is the cardinal direction frequency in which the wind blows. It is represented as a percentage (0, 10, 20, 30...).

In both years (Fig. 7), the wind direction was predominantly towards the NE with speed of 0-2 m/s. In addition, the second spike was directed towards the SE with speed of mostly 0-1 m/s, rarely higher. Wind roses for 2016 and 2019 are very comparable.

Moreover, we used points S1 and S2 for a quantitative comparison of mapping results (A1-A6) and dominant wind directions (Figs. 9 and 10). Two soil samples were taken as an example to see if metal concentrations in soils increase or decrease over the years, as metals from the TDM settle and accumulate in the topsoil over the years. Of course, this is an assumption that cannot be precisely confirmed with such a small amount of data, and this paper is precisely a guide for further research into the relationship between air and soil metal pollution.

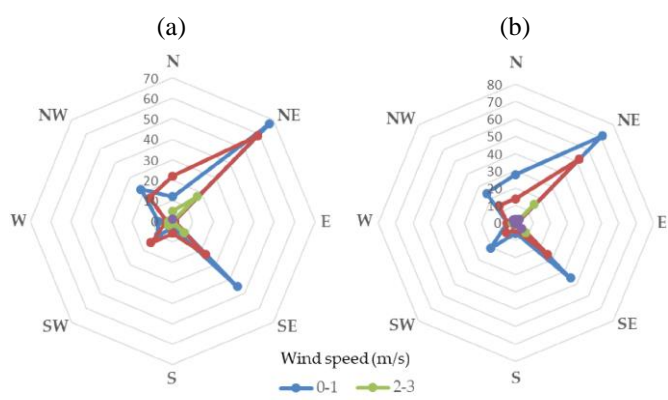


Figure 7. The wind rose: (a) 2016; (b) 2019

4. Conclusions

This paper is focused on the comparison of Ni and Pb concentrations in air and soil pollution in the Zagreb area. Due to the very limited amount of publicly available data from soil analysis samples, 2016 and 2019 were chosen as

the best possible indicators of related changes in metal concentrations in soil and air.

The main problem was that the input data contained only six measurement stations for air quality and, more importantly, just two analysed soil sampling sites for the same metals (measured in total deposited matter). As can be seen on the maps, the only area defined by the air measurement stations can be mapped, due to the inadequate number of soil sampling sites. Moreover, those two subareas cannot be mapped simultaneously, but each by itself due to different physical values, but they are interpreted as possibly dependent.

Moreover, a larger dataset of air concentrations was statistically analysed by itself, based on monthly averages for 2016 and 2019 used to calculate the (mapped) annual averages. This included formal normal distribution testing, linear correlation calculations, and *t*-test.

Testing the normality of Ni and Pb concentrations in the TDM confirmed the feasibility of using two parametric statistical tools – the Pearson correlation coefficient and the *t*-test. Interestingly, the Pearson correlation coefficient does not significantly differ from zero, i.e., the “*r*” is “not significant” and there is no significant linear relationship between the monthly concentrations of Ni and Pb in 2016 and 2019.

Furthermore, the *t*-test for most of the data shows that the null hypothesis is rejected. Thus, it can be concluded that the environment, i.e., the pollutant sources operating in the observed area, have changed over the three years. This is also confirmed by the absence of a linear correlation. In the locations where the null hypothesis was not rejected (A2 and A3), it is concluded that the area is still potentially polluted by the same sources in similar proportions.

Since the mean values of Pb and Ni concentrations in the air in 2019 are lower compared to 2016, while the Pb concentration in soil, on the contrary, increased in 2019, it can be concluded that, as expected, pollution from the air is accumulated in the surface soil layers. From the graphic display of wind directions and speeds, it is evident that the most dominant wind directions were in the northeast and southeast directions. The influence of the NE direction cannot be confirmed due to the absence of soil samplings, but the SE is clearly reflected in the increase in Pb in S1 and S2, although the air concentration in the Zagreb urban area decreased over the analysed years.

The deposition of particles from the air, their accumulation and behaviour within soil horizons is a very complex issue. Some analytical methods on how to deal with them are shown in this work. However, the most substantial conclusion is probably that the insufficient amount of data is the main shortcoming for urban health policy in a large area like Zagreb. The small number of air measurement stations and especially soil sampling sites cannot lead to any reliable conclusions about urban pollutants, their activity over time and direct links to soil toxic degradation based on statistical or geological methods and analyses. There is no doubt that urban pollution sources fill the soil with accumulated toxic elements such as Ni and Pb, especially in suburban areas located along the paths of the dominant wind directions.

However, with a rather restricted dataset and no meaningful spatial network of stations, normality and *t*-test could be used as some of the rare available statistical tools that could help to explain meaningless correlation values (all were in the range of $r = \pm 0.25$, which was very strange depending on the number of calculations, where at least one pair could

randomly reach a high correlation). Such an approach could be much more effective if the sampling network were expanded in the future. This is considered a necessary condition for any proper protection of urban soils and waters, and especially citizens, and is crucial for any reliable health policy that people can trust based on qualitative and quantitative statistical and geological tools, methods, and facts.

Acknowledgements

This research was partly carried out within the framework of the projects “Mathematical Research in Geology VII/2022 and VIII/2023” (led by T. Malvić) and for doctoral study exploration (N. Račić).

References

- [1] Song, Y., Kang, L., Lin, F., Sun, N., Aizezi, A., Yang, Z., & Wu, X. (2022). Estimating the spatial distribution of soil heavy metals in oil mining area using air quality data. *Atmosphere Environment*, (287), 119274. <https://doi.org/10.1016/j.atmosenv.2022.119274>
- [2] Kadriu, S., Sadiku, M., Kelmend, M., & Sadriu, E. (2020). Studying the heavy metals concentration in discharged water from the Trepça Mine and flotation, Kosovo. *Mining of Mineral Deposits*, 14(4), 47-52. <https://doi.org/10.33271/mining14.04.047>
- [3] Polishchuk, S., Falko, V., Polishchuk, A., & Demydenko, A. (2019). Assurance of guaranteed atmosphere air quality for a point emission source. *Mining of Mineral Deposits*, 13(2), 103-110. <https://doi.org/10.33271/mining13.02.103>
- [4] Skrobala, V., Popovych, V., Tyndyk, O., & Voloshchynshyn, A. (2022). Chemical pollution peculiarities of the Nadiya mine rock dumps in the Chervonohrad Mining District, Ukraine. *Mining of Mineral Deposits*, 16(4), 71-79. <https://doi.org/10.33271/mining16.04.071>
- [5] Golubić, J., Vogrin, Z., & Lovrić, I. (2014). Air pollution generated by road traffic in the city of Zagreb and measures proposed. *Transactions on the Built Environment*, (138), 13. <https://doi.org/10.2495/UT140271>
- [6] Kozawa, K.H., Winer, A.M., & Fruin, S.A. (2012). Ultrafine particle size distributions near freeways: Effects of differing wind directions on exposure. *Atmosphere Environment*, (63), 250-260. <https://doi.org/10.1016/j.atmosenv.2012.09.045>
- [7] Janhäll, S. (2015). Review on urban vegetation and particle air pollution – Deposition and dispersion. *Atmospheric Environment*, (105), 130-137. <https://doi.org/10.1016/j.atmosenv.2015.01.052>
- [8] Leifu, C., Shaolin, P., Jingang, L., & Qianqian, H. (2012). Dry deposition velocity of total suspended particles and meteorological influence in four locations in Guangzhou, China. *Journal of Environmental Sciences*, 24(4), 632-639. [https://doi.org/10.1016/S1001-0742\(11\)60805-X](https://doi.org/10.1016/S1001-0742(11)60805-X)
- [9] Magne, A., & Hobæk Haff, I. (2005). Generalised additive modelling of air pollution, traffic volume and meteorology. *Atmospheric Environment*, 39(11), 2145-2155. <https://doi.org/10.1016/j.atmosenv.2004.12.020>
- [10] Chao, S., Li Qin, J., & Wen Jin, Z. (2014). A review on heavy metal contamination in the soil worldwide: Situation, impact and remediation techniques. *Environmental Skeptics and Critics*, 3(2), 24-38.
- [11] Evert, N., & Richardson, D.H.S. (1980). The replacement of the non-descript term “heavy metals” by a biologically and chemically significant classification of metal ions. *Environmental Pollution Series B, Chemical and Physical*, 1(1), 3-26. [https://doi.org/10.1016/0143-148X\(80\)90017-8](https://doi.org/10.1016/0143-148X(80)90017-8)
- [12] Alloway, B.J. (1990). *Heavy metals in soils*. London, United Kingdom: Blackie and Academic Professionals, 613 p.
- [13] Echavarria, G. (2006). Assessment and control of the bioavailability of nickel in soils. *Environmental Toxicology and Chemistry*, 25(3), 643-651. <https://doi.org/10.1897/05-051r.1>
- [14] Halamić, J., & Miko, S. (2009). *Geochemical atlas of the Republic of Croatia*. Zagreb, Croatia: Croatian Geological Survey, 87 p.
- [15] Reimann, C., & De Caritat, P.K. (2005). *Chemical elements in the environment. Factsheets for the geochemist and environmental scientists*. Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong: Springer-Verlag, 598 p. <https://doi.org/10.1017/S0016756800264613>
- [16] Sridevi, J., & Gurdeep, S. (2017). Human health risk assessment of airborne trace elements in Dhanbad, India. *Atmospheric Pollution Research*, 8(3), 490-502. <https://doi.org/10.1016/j.apr.2016.12.003>
- [17] Manalis, N., Grivas, G., Protonotarios, V., Moutsatsou, A., Samara, C. & Chaloulakou, A. (2005). Toxic metal content of particulate matter (PM10), within the Greater Area of Athens. *Chemosphere*, 60(4), 557-566. <https://doi.org/10.1016/j.chemosphere.2005.01.003>

- [18] Zhijiea, L., Yic, H., Weic, Z., Yu, D., Chen, Y., Chaoc, L., & Wang, R. (2005). Effect of different industrial activities on soil heavy metal pollution, ecological risk and health risk. *Environmental Monitoring and Assessment*, (193), 20. <https://doi.org/10.1007/s10661-020-08807-z>
- [19] Ghani, J., Nawab, J., Eshaq Faiq, M., Ullah, S., Alam, A., Ahmad, I., Weqas Ali, S., Khan, S., Ahmad, I., Muhammad, A., Aziz Ur Rahman, S., Abbas, M., Rashid, A., Hasan, S.Z., & Hamza, A. (2022). Multi-geostatistical analyses of the spatial distribution and source apportionment of potentially toxic elements in urban children's park soils in Pakistan: A risk assessment study. *Environmental Pollution*, (311), 119961. <https://doi.org/10.1016/j.envpol.2022.119961>
- [20] Boa, W., Guo, L., Li, S., & Yangb, L. (2023). Source apportionment of heavy metals in the soil at the regional scale based on soil-forming processes. *Journal of Hazardous Materials*, (448), 130910. <https://doi.org/10.1016/j.jhazmat.2023.130910>
- [21] Li, Q., Deng, Q., Fang, H., Yu, X., Fan, Z., Du, Z., Li, M., Tao, Q., Song, W., Zhao, B., Chen, C., Huang, R., Yuan, D., Gao, X., Li, B., Wang, C., & Wilson, J.P. (2022). Factors affecting cadmium accumulation in the soil profiles in an urban agricultural area. *Science of The Total Environment*, 807(3), 151027. <https://doi.org/10.1016/j.scitotenv.2021.151027>
- [22] Šimůnek, J., van Genuchten, M.Th., & Šejna, M. (2012). *The HYDRUS software package for simulating the two- and three-dimensional movement of water, heat, and multiple solutes in variably-saturated porous media*. Technical Manual, Version 2.0.
- [23] Elzahabi, M., & Yong, R.N. (2001). pH influence on sorption characteristics of heavy metal in the vadose zone. *Engineering Geology*, 60(1-4), 61-68. [https://doi.org/10.1016/S0013-7952\(00\)00089-2](https://doi.org/10.1016/S0013-7952(00)00089-2)
- [24] Namjesnik, K. (1994). *Distribution of heavy metals in the soils of the city of Zagreb and its surroundings*. Master's Thesis. Zagreb, Croatia: University of Zagreb, 71 p.
- [25] Ružičić, S. (2013). *Transport model of potentially toxic elements through unsaturated zone at regional wellfield Kosnica*. Doctoral Thesis. Zagreb, Croatia: University of Zagreb, 149 p.
- [26] El Saeiti, G., Garcia-Finana, M., & Hughes, D.M. (2021). The effect of random-effects misspecification on classification accuracy. *The International Journal of Biostatistics*, 18(1), 279-292. <https://doi.org/10.1515/ijb-2019-0159>
- [27] Sheng, Y., Yang, C., Curhan, S., Curhan, G., & Wang, M. (2022). Analytical methods for correlation data arising from multicenter hearing studies. *Statistics in Medicine*, 41(26), 5335-5348. <https://doi.org/10.1002/sim.9572>
- [28] Ivšinić, J., & Litvić, N. (2022). Application of the bootstrap method on a large input data set – Case study western part of the Sava Depression. *Rudarsko-Geološko-Naftni Zbornik*, 36(5), 13-19. <https://doi.org/10.17794/rgn.2021.5.2>
- [29] Trebeš, J. (2017). *Mineralogical and geochemical characteristics of soil from the industrial zone in the eastern part of Zagreb*. Master's Thesis. Zagreb, Croatia: University of Zagreb.
- [30] Bach, O. (1981). *Basic geological map of SFRY 1:100000, Ivanić Grad Sheet*. Zagreb, Croatia: Federal Geological Survey, Beograd.
- [31] Velić, J., & Durn, G. (1993). Alternating lacustrine-marsh sedimentation and subaerial exposure phases during Quaternary: Prečko, Zagreb, Croatia. *Geologia Croatica*, 46(1), 70-91.
- [32] Brkić, Ž. (2017). The relationship of the geological framework to the quaternary aquifer system in the Sava River valley (Croatia). *Geologia Croatica*, 70(3), 201-213.
- [33] Gorjanović-Kramberger, D. (1904). *Geological overview map of the Kingdom of Croatia-Slavonia. Interpreter of the geological map Zagreb (Zone 22, Col. XIII)*ю Zagreb, Croatia: Department of Internal Affairs, 76 p.
- [34] Velić, J., Saftić, B., & Malvić, T. (1999). Lithologic composition and stratigraphy of quaternary sediments in the area of the "Jakuševac" Waste Depository, Zagreb, Northern Croatia. *Geologia Croatica*, 52(2), 119-130. <https://doi.org/10.4154/GC.1999.10>
- [35] Majhen, Lj. (2020). *Geochemical characteristics of soil in the exploration pit from Velika Gorica well field area*. Master's Thesis. Zagreb, Croatia: University of Zagreb.
- [36] De Mesnard, L. (2013). Pollution models and inverse distance weighting: Some critical remarks. *Computers and Geosciences*, (52), 459-469. <https://doi.org/10.2139/ssrn.1931636>
- [37] Malvić, T. (2008). *Application of geostatistics in the analysis of geological data*. Zagreb, Croatia: University of Zagreb.
- [38] Ivšinić, J., & Malvić, T. (2022). Comparison of mapping efficiency for small datasets using inverse distance weighting vs. moving average, Northern Croatia Miocene Hydrocarbon Reservoir. *Geologija*, 65(1), 47-57. <https://doi.org/10.5474/geologija.2022.003>
- [39] Ivšinić, J., & Malvić, T. (2020). Application of the radial basis function interpolation method in selected reservoirs of the Croatian part of the Pannonian Basin System. *Mining of Mineral Deposits*, 14(3), 37-42. <https://doi.org/10.33271/mining14.03.037>
- [40] Setianto, A., & Triandini, T.J. (2013). Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. *Journal of Applied Geology*, (5), 21-29. <https://doi.org/10.22146/jag.7204>
- [41] Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, (52), 591-611. <https://doi.org/10.2307/2333709>
- [42] Chakravarti, I.M., Laha, R.G., & Roy, J. (1967). *Handbook of methods of applied statistics*. New York, United States: Wiley, 350-355 p.
- [43] Wilkerson, S. (2008). Application of the paired t-test. *Xavier University of Louisiana's Undergraduate Research Journal*, 5(1), 7.
- [44] Diez, D.M., Cetinkaya-Rundel, M., & Barr, C.D. (2019). *OpenIntro statistics*. Fourth Edition, 422 p.
- [45] Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, (58), 240-242. <https://doi.org/10.1098/rspl.1895.0041>
- [46] Medunić, G. (2022). Basic methods of (geo)statistical analysis of data from the environment. *Manualia Universitatis Studiorum Zagrabensis*, 1-54.
- [47] Weisstein, E.W. (2022). *Statistical correlation*. Retrieved from: <https://mathworld.wolfram.com/StatisticalCorrelation.html>
- [48] Malvić, T., Velić, J., & Peh, Z. (2005). Qualitative-quantitative analyses of the influence of depth and lithological composition on lower Pontian sandstone porosity in the Central Part of Bjelovar Sag (Croatia). *Geologia Croatica*, 58(1), 73-85.

Зв'язок між забрудненням повітря та ґрунту на основі статистичного аналізу та інтерполяції нікелю (Ni) і свинцю (Pb): тематичне дослідження на прикладі Загребу, Хорватія

Н. Рачич, Т. Мальвіч

Мета. Порівняння концентрацій Ni та Pb у забрудненні повітря та ґрунту для встановлення їх взаємозв'язку у районі Загреба.

Методика. Тестування нормальності концентрацій Ni та Pb у загальному обсязі відкладеної речовини (TDM) підтвердило доцільність використання двох параметричних статистичних інструментів – коефіцієнта кореляції Пірсона та t-тесту. Метод інтерполяції зі зважуванням зворотної відстані (IDW) був обраний як найкращий підхід для невеликої кількості вимірювань. Через дуже обмежену кількість загальнодоступних даних з аналізу зразків ґрунту 2016 та 2019 роки були обрані як найкращі можливі індикатори відповідних змін концентрацій металів у ґрунті та повітрі.

Результати. Визначено, що недостатній обсяг даних є головним недоліком міської політики в галузі охорони здоров'я на такій великій території як Загреб. Зазначено, що невелика кількість станцій вимірювання повітря та особливо місць відбору проб ґрунту не може призвести до будь-яких достовірних висновків щодо міських забруднювачів, їхньої активності в часі та прямого зв'язку з токсичною деградацією ґрунту на основі статистичних чи геологічних методів та аналізів. Встановлено, що міські джерела забруднення заповнили ґрунт накопиченими токсичними елементами, такими як Ni та Pb, особливо в приміських районах, розташованих уздовж шляхів панівних напрямків вітру.

Наукова новизна. Це оригінальне дослідження, в якому вперше проводиться статистичний аналіз та надаються загальнодоступні дані про забруднення повітря та ґрунту за період 2016-2019 років.

Практична значимість. Дане дослідження є необхідним кроком у визначенні майбутнього планування станцій для вимірювання повітря та ґрунту у міській зоні Загреба.

Ключові слова: забруднення, Ni, Pb, статистичне тестування, геостатистичне картографування, Загреб, Хорватія