Assessment of the stability of mercury concentrations in municipal waste using data science tools

Radosław Jędrusiak[1], Monika Chuchro[2], Barbara Bielowicz[3], Agnieszka Gielar[4]

[1] AGH University of Science and Technology, Faculty of Geology, Geophysics and Environment Protection, Krakow, Poland,
e-mail: jedrusia@agh.edu.pl, ORCID ID: 0000-0003-0008-4075
[2] AGH University of Science and Technology, Faculty of Geology, Geophysics and Environment Protection, Krakow, Poland,
e-mail: chuchro@agh.edu.pl, ORCID ID: 0000-0002-0381-4697
[3] AGH University of Science and Technology, Faculty of Geology, Geophysics and Environment Protection, Krakow, Poland,
e-mail: bbiel@agh.edu.pl (corresponding author), ORCID ID: 0000-0002-8742-5890
[4] AGH University of Science and Technology, Faculty of Geology, Geophysics and Environment Protection, Krakow, Poland,
e-mail: agnieszkagielar@geol.agh.edu.pl

Abstract: Mercury and its compounds are among the most dangerous and toxic substances in the environment. As part of the study, several exploratory analyses and statistical tests were conducted to demonstrate how low and stable mercury content is in municipal waste. A statistical analysis of the mercury content in waste (waste codes 19 12 12 and 20 03 01) was carried out using advanced IT tools. Based on 32 results for each waste, the maximum mercury concentration was 0.062 mg/kg dry weight (EWC code 19 12 12) and 0.052 mg/kg dry weight (EWC code 20 03 01). The analysis, data inference, and modeling were performed according to the CRISP-dm methodology. The results obtained were compared with the maximum allowable mercury concentrations for agricultural soils (2 mg/kg dry weight) and the provisions of the Minamata Convention (1 mg/kg). The average, median, and maximum observed mercury concentrations in waste are significantly lower than the assumed levels of 2 mg/kg (permissible concentrations for II-1 soils) and 1 mg/kg (Minamata Convention). The stability of mercury content in waste was examined. Descriptive statistics, statistical tests, and regression modeling were used. The tests and analyses performed showed an insignificant variation in the mercury content of the wastes with codes 19 12 12 and 20 03 01. No trend or seasonality was observed. The analyses and tests performed confirmed that the data are stable, and the values are low.

Keywords: mercury, municipal waste, trend analysis, regression, seasonality, stability

INTRODUCTION

General information

Mercury belongs to transition metals; it is the only metal in a liquid state. It constitutes only 0.5 ppm of the Earth's crust (Kabata-Pendias 2011). In the atmosphere, it occurs mainly in three forms: 1) elemental mercury – Hg^0, the most stable form of Hg, accounting for about 90% of total atmospheric mercury; 2) oxidized mercury – Hg^{2+}, characterized by high solubility in water; 3) mercury bound to particulate matter in the atmosphere – Hg^p (Fu et al. 2012). It has no biological role but is highly toxic to all living organisms causing damage to the nervous, cardiovascular, digestive, excretory, and immune systems (WHO 2017, Song et al. 2018). Mercury and its compounds are used in thermometers, barometers, manometers, pressure gauges, float valves, mercury switches, mercury relays, fluorescent lamps, and other devices. Although concerns

about the toxicity of the element have led to the withdrawal of these devices in favor of alternatives, mercury remains in use in scientific applications and as an amalgam for dental restoration.

Mercury is released into the environment through both natural and anthropogenic processes. Natural sources of mercury include volcanic eruptions, volatilization from water, soil, and vegetation, rock erosion, and geothermal activity in the Earth's crust. According to UNEP (United Nations Environmental Programme) data published in 2018, in 2015 the global mercury emissions from natural sources amounted to about 220 Mg per year and were mostly related to volcanic emissions (UNEP 2018).

This element is most commonly found in metallic ore deposits – both pure mercury and various mercury mixtures. It is located primarily in young orogenic regions and volcanic zones. Mercury deposits are distributed all over the world and are mainly associated with minerals such as cinnabar (HgS), calomel (Hg_2Cl_2), and native mercury. There are silica-carbonate, hot springs, and Almaden-type mercury deposits. Mercury is also produced as a by-product of gold, silver, and sulfide mining, which account for 5% of global production.

Mercury occurs in minerals worldwide and does not pose much of a threat; that being said, mercury deposits have been mined since the 16th century. It is the interference with the Earth's crust that is inextricably linked to the increased release of mercury-containing compounds into the biosphere but, unfortunately, they tend to accumulate in water. In addition to natural sources, mercury enters the environment through human activities. Historically, mercury emissions in Europe were high but over the last 40 years, significant legislative action has significantly reduced its use and release into the environment. In the rest of the world, the use of mercury and its emissions has increased over time, which is associated with continuous economic development and industrialization. The global emissions from anthropogenic sources in 2015 were estimated at about 2,220 Mg per year (UNEP 2018). Most mercury emissions come from the combustion of solid fuels such as coal, peat, lignite, and wood. However, it should be noted that small-scale gold mining is the world's largest source of mercury pollution (UNEP 2018). National emissions of mercury reported to the Convention on Long-range Transboundary Air Pollution by EU countries decreased in the reported period by 94% in agriculture sector, energy production, by 82% in distribution sector, by 82% in industrial processes and product use sector and by 78% in waste industry sector respectively (EEA 2019). European countries restrict the use of mercury in industrial activities, with the ultimate goal of ending its use in the industry. Mercury is no longer used in EU gold mining process and its use in the production of vinyl chloride is limited to one plant in Slovakia, which stopped using mercury in 2022. Until 2015, mercury was used for the production in chloralkali industry, where 85 tons were used, and in dental applications, where 56 tons were utilized. The use of mercury in industrial chlorine processes was banned in 2017. Estimated mercury consumption in the European Union in 2005–2015 and in 2021 is presented in Figure 1.
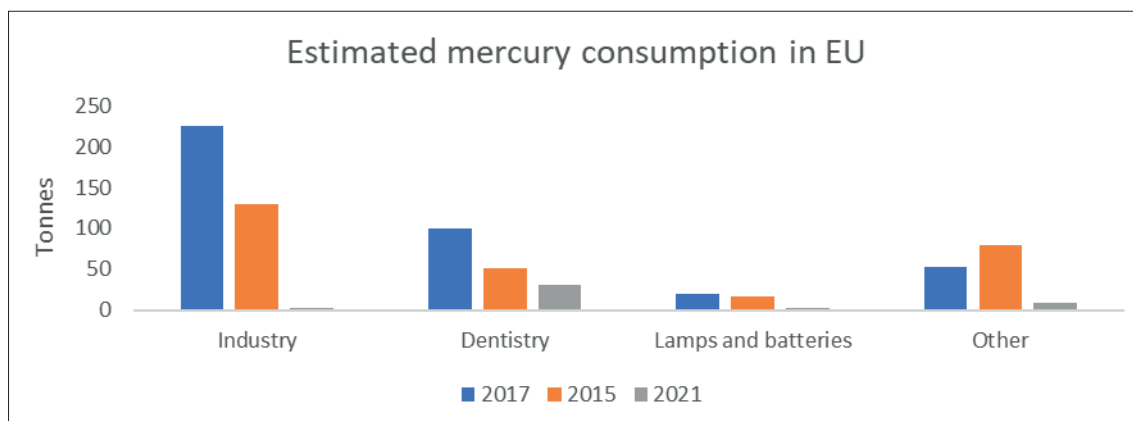

Fig. 1. Estimated mercury consumption in EU countries by the consumption sector (modified from Marnane 2018)

There is no denying that mercury can still be found in municipal waste. Separate collection of such waste can help to reduce the amount of mercury in a total load of mixed municipal solid waste load, but in practice comprehensive separate collection is not achieved.

Legislation on mercury

Several regulations aimed at reducing emissions and counteracting mercury-related environmental pollution have been adopted. To address the mercury problem, the first global international agreement, the Minamata Convention, was adopted in October 2013. The Convention was ratified by 98 parties and entered into force in 2017. The Minamata Convention on Mercury, adopted at Kumamoto, on 11 October 2013, is a new international instrument of a global scope aimed at protecting human health and the environment from anthropogenic emissions and releases of mercury and mercury compounds (UNEP 2013).

The provisions of the Convention regulate issues related to mercury mining, trade of mercury-added products, emissions of mercury into the atmosphere, the release of mercury into the water and soil, as well as the use of mercury in industrial products and processes in a comprehensive manner. The convention also covers the management of mercury-containing waste and the use of environmentally friendly mercury storage methods, as well as measures for the remediation of mercury-contaminated sites.

Poland signed the Convention in New York on September 24, 2014. It was ratified on August 21, 2021 and entered into force on December 29, 2021.

The European Union has also provided a list of instruments in matters governed by the Minamata Convention. The Union is competent for the performance of those obligations from the Minamata Convention on Mercury regarding which the provisions of Union legal instruments, in particular those listed below, establish common rules and insofar as these common rules are affected or altered in scope by the provisions of the Minamata Convention or an act adopted in implementation thereof.

The main EU documents on mercury are the Regulation (EU) 2017/852 of the European Parliament and of the Council of 17 May 2017 on mer-cury, and repealing Regulation (EC) No 1102/2008 (EP and CEU 2017), Directive 2011/65/EU of the European Parliament and of the Council of 8 June 2011 on the restriction of the use of certain hazardous substances in electrical and electronic equipment (EP and CEU 2011), and the Directive 2006/66/EC of the European Parliament and of the Council of 6 September 2006 on batteries and accumulators and waste batteries and accumulators and repealing Directive 91/157/EEC (EP and CEU 2018).

In Poland, there are several regulations concerning mercury, including sources of the supply and trade of mercury, mercury-added products, emissions of mercury into the atmosphere or mercury waste, and their temporary storage, which result from EU and national law. Regulation (EP and CEU 2017) implemented provisions of the Convention not yet regulated by EU law. Thanks to this, all the provisions of the Convention are reflected in EU law, which means that they are applicable in Poland (*Ustawa…* 2021).

PERMISSIBLE MERCURY CONCENTRATIONS

According to CLP Regulation (EP and CEU 2008a), inorganic mercury compounds are classified as Acute Tox. 2 – H300, and Annex III to the Waste Directive (EP and CEU 2008b) defines hazardous waste as waste that displays HP (hazardous properties) 6 – Acute Tox. 1 Oral – H300, at the level of $\geq$0.1% (1000 mg/kg); due to the many times lower concentrations in the tested wastes, this limit value is omitted in the following. However, the national legislation on acceptable levels of pollutants in soils and the Minamata Convention were adopted as the benchmark. To demonstrate low mercury concentrations in waste, the Regulation of the Minister of the Environment of 1 September 2016 on the method of conducting the assessment of soil surface pollution, Journal of Laws 2016 item 1395 (*Rozporządzenie…* 2016), was adopted as a benchmark.

The study compares the mercury concentration in waste (waste codes 19 12 12 and 20 03 01) to the most restrictive limit value for agricultural soil (class II-1). This value is 2 mg/kg dry weight and indicates the acceptable limit for mercury in

soil and subsurface, taking into account the impact of this substance on human health and the environment. Class II soils include arable land I, family allotment gardens (R1), orchards (S), permanent meadows (Ł), permanent pastures (Ps), land under ponds (Wsr), land under ditches (W) and family allotment gardens (Bz).

The second document on maximum mercury levels is the Minamata Convention. According to the Minamata Convention report of 16 August 2021, the mercury content of waste should be in the range of 1–25 mg/kg, and the exact limit should be set at the national level. Above 25 mg/kg, waste can be considered to be contaminated with mercury. However, wastes with a mercury concentration below 1 mg/kg should not be considered category C (potentially mercury-contaminated) (UNEP 2021). The present study compares the value of 1 mg/kg dry weight with wastes (19 12 12 and 20 03 01).

SUBJECT OF THE STUDY

Based on the established regulations and activities conducted around the world, it should be stated that the mercury emission management system is clear and contributes to reducing pollution by mercury and its compounds. Therefore, it is important to continuously monitor the mercury content of municipal waste that is processed or stored. The mercury content in waste samples should be assessed and compared with regulations. What is more, the volatility of mercury values over time should be analyzed. One way to prove the stability and low mercury concentration in waste is to use advanced statistical tools and data science (Murtagh & Devlin 2018, Schmidt et al. 2019, Godyń et al. 2020). This is the approach taken in this study, where the stability and mercury content of waste were investigated based on the results of chemical analyses of waste. We use the term stability as second-order stationarity, lack of trend, and seasonality. What is more, it is the first step in complex municipal waste analysis. Carrying out comprehensive tests and analyses will allow for the future use of other wastes such fly ash or flue gas cleaning residues from waste incineration without environmental hazards (Giro-Paloma

et al. 2017, Cho et al. 2020, Fabricius et al. 2020, Godyń & Dutka 2023).

The aim of the article is also to present the methodology developed for assessing the mercury concentration in waste and the stability over time of the mercury concentration in waste. The analyses and modeling performed included a comprehensive approach to data in accordance with the Cross-Industry Standard Process for Data Mining (CRISP-DM) (Wirth & Hipp 2000, Schröer et al. 2021,) methodology. It is an open standard process model that describes common approaches used by data mining experts and is the most widely-used analytic model. CRISP-dm consists of six phases covering understanding business conditions, understanding data, preparation, modeling, evaluation, and implementation. The advantage of this methodology is the possibility of flexible and free transition between phases and returning to earlier phases in case of unsatisfactory results. CRISP-dm encourages best practices and allows projects to replicate. What is more, this methodology provides a uniform framework for planning and managing a project (Wirth & Hipp 2000). The first phase of the methodology is presented in "Introduction" section. Phase one focused on the occurrence of mercury in the environment and waste and on legislation on mercury content in waste and soil. Understanding the data, i.e., the second phase of the CRISP-dm methodology, is presented in "Permissible mercury concentrations" section, and the subsequent phases including data preparation for analysis, modeling, and evaluation are presented in "Research methodology" section. Assessing the feasibility of implementation is discussed in the last two sections.

In this paper, the authors wanted to present the benefits of using the Cross Industry Standard Process for Data Mining methodology together with advanced statistical and data science approaches in assessing the content and stability of mercury in municipal waste ashes. The application of the presented methodology together with chosen methods allows for complex information to be obtained about the nature of the data in the context of mercury content and stability. What is more, the obtained results are reproducible.

RESEARCH METHODOLOGY

Sampling

Within an 8-month period from April to November 2021, sampling was carried out for two types of municipal waste. The first was residue from the mechanical treatment of municipal waste with EWC code 19 12 12 – other waste (including mixed substances and objects) from the mechanical processing of waste other than those mentioned in 19 12 11. The second was non-recyclable waste collected directly from residents as part of the selective municipal waste collection system with EWC code 20 03 01 – non-segregated (mixed) municipal waste. The sampling was conducted in accordance with the following standards: *Solid secondary fuels – Sampling methods* (CEN 2011a) and *Solid secondary fuels – Methods for preparing a laboratory sample* (CEN 2011b). Samples were collected from randomly selected waste delivery vehicles, during the delivery day (waste 20 03 01) and within a 500 Mg batch (waste 19 12 12). Ten 5 kg primary samples were collected three times per day: during the initial unloading phase, the middle, and the final phase (randomly) – in this way, initial bulk samples with a given code were obtained. The samples collected in this way were deposited on a waste dump site. At the end of the cycle (day/patch), the conical waste dump obtained from the primary samples was averaged by mixing, scattering a cone and quartering with a steel plate on a previously prepared clean surface – until a laboratory sample of approximately 10 kg was obtained. The prepared samples, after appropriate labeling and description, were stored in tightly closed containers in a refrigerator at 4°C and transported to an analytical laboratory within a maximum of 48 hours. Three laboratory samples (daily or corresponding to a batch of 500 Mg) were averaged based on 10 primary samples taken three times during the waste discharge into the bunker (randomly – initial, middle, and final phase). After collection, the containers were sealed with clingfilm. The collected material was clearly labeled and transported to the laboratory. The analytical sample was obtained by mixing, shredding, and dividing three laboratory samples for each waste code.

Mercury concentration analysis

The analysis of the samples was conducted in accordance with the EPA Method 7473 (SW-846) (U.S. EPA 1998), ASA method, CVAAS (cold vapor absorption atomic spectrometry) method, and the amalgamation technique with a limit of quantification within the scope of accreditation not higher than 0.01 mg/kg dry matter for total (organic and inorganic) mercury (U.S. EPA 1998). The analysis used an MA-3000 mercury analyzer (Nippon Instruments Corporation). The calibration curve was performed using a liquid CRM compliant with the ISO 17034 standard catalog number MSHG-100PPM (Inorganic Ventures). The analyzer was checked using the following standards/ reference materials: the BCR-143R sewage sludge amended soil (Institute for Reference Materials and Measurements, IRMM) and a bituminous coal standard.

Data

The data used in the analysis consist of two-time series with equal intervals. Each time series has 32 observations recorded over 8 months in 2021, with one observation per week. There is one gap in the dataset for the 35th week of the year. Both time series contain information on the mercury concentration expressed in mg/kg dry weight for waste codes 19 12 12 and 20 03 01. In the dataset 19 12 12 there are 7 non-detects with values lower than 0.01 mg/kg dry weight, which is 21.9% of observations. In the dataset 20 03 01, there are 14 non-detects, which constitute 43.7% of observations (Tab. 1).

Preprocessing and data analysis

This section covers two steps of the CRISP-dm methodology: preprocessing and data modeling. Proper data preparation is crucial in order to obtain optimal results for analysis. The preprocessing phase involved replacing the non-detects, also known as left-censored results. Non-detects are usually reported as being less than a given threshold and such observations should never be deleted from the dataset to avoid a strong upward bias in estimating location measures since such observations could be replaced. In the imputation method, a different value is assigned to each non-detect

observation. In datasets with more than 15% of non-detects, it is advisable to use extrapolation methods, robust regression on order statistics (ROS), Kaplan–Meier method, Cohen's method, Weibull regression, or maximum likelihood estimation (MLE) (U.S. EPA 2000, Helsen 2005).

Non-detects in the dataset were replaced by the robust ROS method. This method is commonly applied in many environmental data sets. Robust ROS is a semi-parametric method that substitutes non-detects based on least-squares regression on a probability plot (Helsen 2005). This method assumes that there must be at least three detected values and a detection frequency greater than 50%. It also assumes that the detected data can be fitted to a known distribution on a probability plot, from which imputations are made for the non-detects. The estimated summary statistics are computed from a combination of the known and imputed measurements, rather than from the parameters of the fitted model (Helsen 2005, 2012).

Robust regression on order statistics can be presented as follows (Helsen 2005): for $n = n_0 + n_1$ independent normally distributed data, with mean μ and variance μ_2, n_0 – observations are non-detects, n_1 – observations are larger than detection limit.

$$y_i = \mu + \sigma \Phi^{-1}(P_i),$$

where $P_i = P(Y_i \leq y_i)$ and $\Phi^{-1}(.)$ denotes the inverse cdf of a $N(0.1)$ distribution.

The procedure replaces probabilities with adjusted ranks; the regression equation becomes:

$$y_i = \hat{\mu} + \hat{\sigma}\, \Phi^{-1}\!\left(\frac{i - 3/8}{n + 1/4} \right) + \varepsilon_i,$$

where:

$$i \;=\; n_0 + 1,\, n_0 + 2\,,\ldots,\, n_0 + n_1,$$
$$\hat{\mu}, \hat{\sigma} \;-\; \text{least } \mu, \sigma \text{ squares estimates of } \mu,$$
$$\Phi^{-1}\!\left(\frac{i - 3/8}{n + 1/4} \right) - \text{ normal score calculated by Altman's formula,}$$
$$\varepsilon_i \;-\; \text{residual errors.}$$

The modeling phase was divided into two parts. The first included analysis aimed at proving statistically significantly lower values than the assumed threshold value. For this purpose, descriptive statistics, visualizations, and a *t*-test for a single sample were performed. The second part

included analysis and modeling aimed at assessing the stability of the data. In this part, the autocorrelation analysis, DF tests, and regression models were performed.

In the first part of the modeling, a one-sample *t*-test was performed for large samples (more than 30 observations). The *t*-test is highly sensitive to non-detects (U.S. EPA 2000); however, it is the most popular test for assessing whether the mean value in the sample is lower than the value of comparison (Mishra et al. 2019).

One-sample *t*-test can be presented as (Mishra et al. 2019):

$$t = \sqrt{n}\, \frac{\bar{x} - \mu}{s},$$

where the test has the *t*-distribution with $n - 1$ degrees of freedom for n ($n \geq 30$), $p = 0.05$, and:

    $\bar{x}$ – mean value in sample,
    μ – value of comparison,
    s – standard deviation in the sample.

The null hypothesis (H0) is that the true difference between the sample mean and the value of the comparison is zero.

The alternate hypothesis (Ha) is that the true difference is different from zero.

The second part of the modeling phase covered the issue of data stability. Stability is a process where inputs and conditions are consistent over time. When the process is stable, it means that the sources of variation are consistent over time and the process does not show unpredictable variability (Chow 2007). We can also link data stability with stationarity. The stationary process is a stochastic process where all moments and total moments are constant. The data that is devoid of any trend or periodicity can be considered stable (Manuca & Savit 1996). To assess the stability, the autocorrelation function (ACF), partial autocorrelation function (PACF), stationarity tests, and regression models were performed.

The stability of the time series, considered as no trend or periodicity, can be assessed by ACF and PACF. The autocorrelation function is the correlation between a given time series and a lagged version of itself over successive time intervals. This function is intended to measure the relationship between a variable's present value and

any past values. The autocorrelation function for lag k is the linear Pearson's correlation coefficient between a given time series and the same time series separated by k intervals (lags). The partial autocorrelation function for lag k is the linear Pearson's correlation coefficient between a given time series and the same time series separated by k intervals, after removing the effect of any correlations due to the terms at shorter lags (Dégerine & Lambert-Lacroix 2003).

As mentioned above, data stability can also be linked to stationarity. When the mean value, variance, and the autocorrelation function variate with the change of time, the random process $x(t)$ is called non-stationary. In the special case, when the mean value and the autocorrelation function do not depend on time t_1, the random process $x(t)$ is called weak-sense stationary or wide-sense stationary (WSS). In such cases, the mean value and covariance are constant (Manuca & Savit 1996). The unit root test (an augmented Dickey–Fuller test) was chosen to assess the stationarity of the data. It is the most commonly used test to assess the stationarity of a time series. The ADF test expands the Dickey–Fuller test equation to include high order regressive process in the model (Fuller 1995).

The null hypothesis (H0) is that a unit root is present in a time series sample.

The alternative hypothesis (Ha) is different depending on which version of the test is used (usually stationarity or trend-stationarity).

ADF statistics can be presented as:

$$\Delta y_t = \mu + \gamma y_{t-1} + \sum_{i=1}^{P} \alpha_s \Delta y_{t-s} + \varepsilon_t,$$

$$\text{ADF} = \frac{\hat{\gamma}}{SE(\hat{\gamma})},$$

where:

μ – mean value,
γy_{t-1} – lagged values of the dependent variable,
ε_t – noise.

Maximum lag value is expressed as follows (Fuller 1995):

$$n\text{lag} = \left(\frac{4 \cdot n}{100}\right)^{\frac{2}{9}},$$

where n – number of observations.

The computed ADF value can be compared to the relevant critical values for the Dickey–Fuller test. This test is asymmetrical, so concerned with negative values of test statistics. If the calculated test statistic is more negative than the critical value, then the null hypothesis is rejected and no unit root is present (Fuller 1995).

Additionally, both simple and multiple regression models were created. Regression models assume that there are dependencies between the dependent variable and the explanatory variable or variables. We can write the multiple regression model as (Montgomery 2012):

$$Y = a + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n + \varepsilon,$$

where:

Y – dependent variable,
a – y-intercept (constant term),
b_n – slope coefficients for each explanatory variable,
X_n – explanatory variables,
ε – the model error term (residuals).

The parameters of the regression model were estimated using the least squares method (Ross 2005). The assessment of the model fit to empirical data was based on the determination coefficient R^2. The determination coefficient shows what part of the total variability of the explained variable is the variability of theoretical values. The closer R^2 is to 1, the better the model fit (Montgomery 2012). Additionally, the evaluation of the model is performed based on the F statistics. It is the global statistics returned for the F-test. The statistics explain the prediction ability of the regression model by confirming that all regression coefficients in the model are significantly different from 0. The F-test analyzes the cumulative effect of the explanatory variables but does not test the individual explanatory variables. p-value, which calculates the probability that the relationships in the data are random, is associated with the F-test. Since the p-values are based on probabilities, they range from 0.0 to 1.0. A low p-value, typically 0.05 or less, is required to conclude that the relationships in the model are true (not random) and to reject the null hypothesis. In this case, the probability that the relationships in the model are random is 0.05. On the other hand, the probability

that the relationships in the model are true is 0.95 (Montgomery 2012).

Software

Analyses, statistical tests, modeling, and visualizations were performed using the R language (version 4.1.2) in the integrated programming environment RStudio version 1.4.1106, and the Windows 10 – 64 bit operating system. Packages used in this project included: tidyverse version 1.3.1 (including the following packages: dplyr, ggplot2, readr), ggthemes version 4.2.4, NADA version 1.6-1.1, stats version 3.6, and TSA version 3.1.2 (R packages 2022).

RESULTS AND DISCUSSION

Non-detects were replaced with robust regression on order statistics (ROS). Mercury values below the measurable quantification level were <0.01 mg/kg dry weight of waste. The observations replaced with the ROS method are presented in Table 1 in bold. In the case of the 19 12 12, the modification covered 7 observations, and for the 20 03 01 dataset the modification covered 14 observations out of 32 observations in the studied sample.

To assess the low mercury concentration in waste with codes 19 12 12 and 20 03 01, descriptive statistics (Tab. 2), visualizations, and the *t*-test for a single sample were performed. The mean value of mercury concentration in waste code 19 12 12 is 0.018 mg/kg DM, for code 20 03 01 it is 0.013 mg/kg DM, which respectively constitutes less than 1% and 0.65% of the permissible value for II-1 land according to the Regulation of the Minister of the Environment of 1 September 2016 on the method of conducting the assessment of soil surface pollution (*Rozporządzenie…* 2016). The maximum values are 0.062 mg/kg dry weight (3.1% of permissible value) and 0.052 mg/kg dry weight (2.6% of permissible value) for 19 12 12 and 20 03 01 wastes. 50% of the samples have a mercury content of less than 0.014 mg/kg dry weight for 19 12 12 and 0.010 mg/kg dry weight for 20 03 01 waste (Tabs. 1, 2). Graphical presentation of mercury content in wastes 19 12 12 and 20 03 01 compared to the Regulation of the Minister of the Environment of 1 September 2016 on the method of conducting the assessment of soil surface pollution (*Rozporządzenie…* 2016) and the Minamata Convention for wastes that should not be considered contaminated with mercury is shown in Figure 2.

Table 1
Mercury in waste [mg/kg dry weight] with non-detects replaced with ROS (bold)

Week in year	Values		Week in year	Values	
	19 12 12	20 03 01		19 12 12	20 03 01
13	0.015	**0.001**	29	0.012	0.011
14	**0.005**	**0.007**	30	0.062	0.003
15	0.028	0.017	31	0.013	0.015
16	0.058	0.033	32	0.015	**0.002**
17	0.019	0.052	33	0.017	**0.005**
18	0.013	**0.006**	34	0.018	0.035
19	0.011	0.010	36	0.011	0.012
20	0.019	0.015	37	0.014	**0.004**
21	0.039	0.019	38	0.010	**0.003**
22	0.014	0.035	39	**0.004**	**0.007**
23	**0.003**	0.012	40	**0.006**	**0.004**
24	**0.007**	0.013	41	0.012	**0.005**
25	**0.006**	**0.006**	42	0.024	0.014
26	0.012	**0.002**	43	0.032	0.030
27	**0.007**	0.024	44	0.010	**0.003**
28	0.022	0.013	45	0.028	0.010

Table 2
Descriptive statistics and t-test results

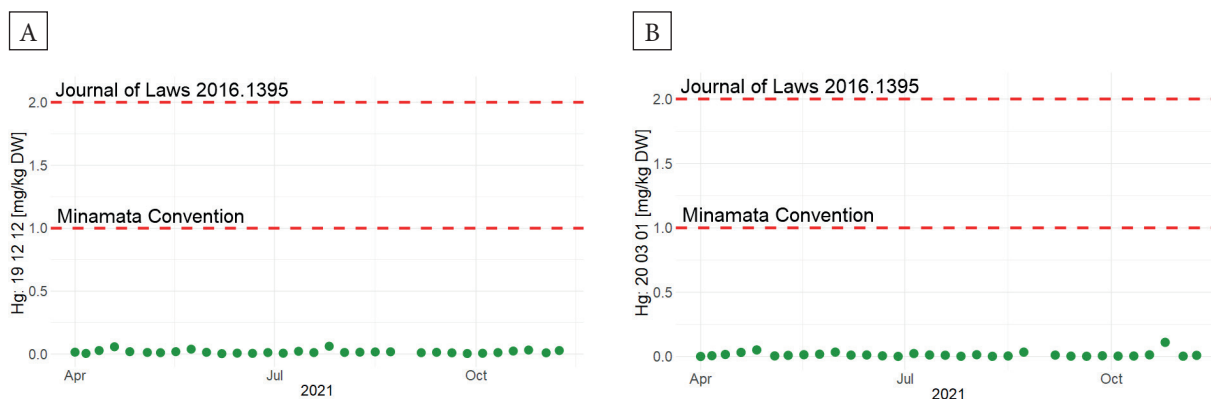| Code	Descriptive statistics					t-test		p-value t-test
	mean	median	min	max	SD	less than 2	less than 1	
19 12 12	0.018	0.014	0.003	0.062	0.014	−809	−401	<2.2e-16
20 03 01	0.013	0.010	0.001	0.052	0.012	−537	−266	<2.2e-16

Fig. 2. Comparison of the mercury content in wastes with codes 19 12 12 (A) and 20 03 01 with the permissible levels (B)

The mercury values presented are significantly below the permitted value for both the abovementioned regulations (UNEP 2021). The results of one-sample *t*-test, for 31 degrees of freedom, and $p = 0.05$ show that mercury concentrations in both waste codes are statistically significantly lower than both permissible values (see Tab. 2).

As stated in studies conducted in Italy in the1990s, the average mercury concentration in the municipal waste stream was about 4 mg/kg dry weight. Concentrations have decreased significantly over the last decade thanks to a significant reduction in the use of mercury and the introduction of efficient electrical and electronic waste return systems.

In 1997, the mercury concentration in municipal solid waste was about 2 mg/kg (van Velzen et al. 2002). Also, 2001 data from Germany indicates that typical municipal waste contains 1–5 mg/kg DM of mercury (Neuwahl et al. 2019). Comparing the mercury content in the analyzed data to the typical mercury content in municipal waste in Germany, a much lower level of this element can be noticed in the analyzed waste with codes 19 12 12 and 20 03 01 in Poland 2021.

The stability of the time series considered as no trend or periodicity in waste with codes 19 12 12 and 20 03 01 was assessed using ACF (Fig. 3A, B),

and PACF (Fig. 3C, D). If there would be a trend in the time series, then the Pea'son's linear correlation coefficient values would be statistically significant in the ACF plots for the first *n* lags, and its values would be high and decrease slowly with the increasing lags. In the case of seasonality, in the PACF plot for further lags, the value of Pearson's linear correlation would be statistically significant and appear cyclically for each *m* lag (Dégerine & Lambert-Lacroix 2003). For any of the lags, the correlation values are not statistically significant, they do not exceed the confidence levels (blue dashed line in the graphs), both for the autocorrelation function and the partial autocorrelation (Fig. 3). This means that the mercury concentrations in waste with codes 19 12 12 and 20 03 01 do not show a trend or seasonality.

The unit root test (an augmented Dickey–Fuller test) was chosen to assess the stationarity of the data. The test was computed for lag = 3. The empirical significance level (*p*-value) is the probability of obtaining the computed value of the test statistic, assuming that the null hypothesis is true. If this probability is small (lower than 0.05), as in Table 3, then the null hypothesis should be rejected. Based on the test, the mercury concentration in waste with codes 19 12 12 and 20 03 01 can be considered stationary.
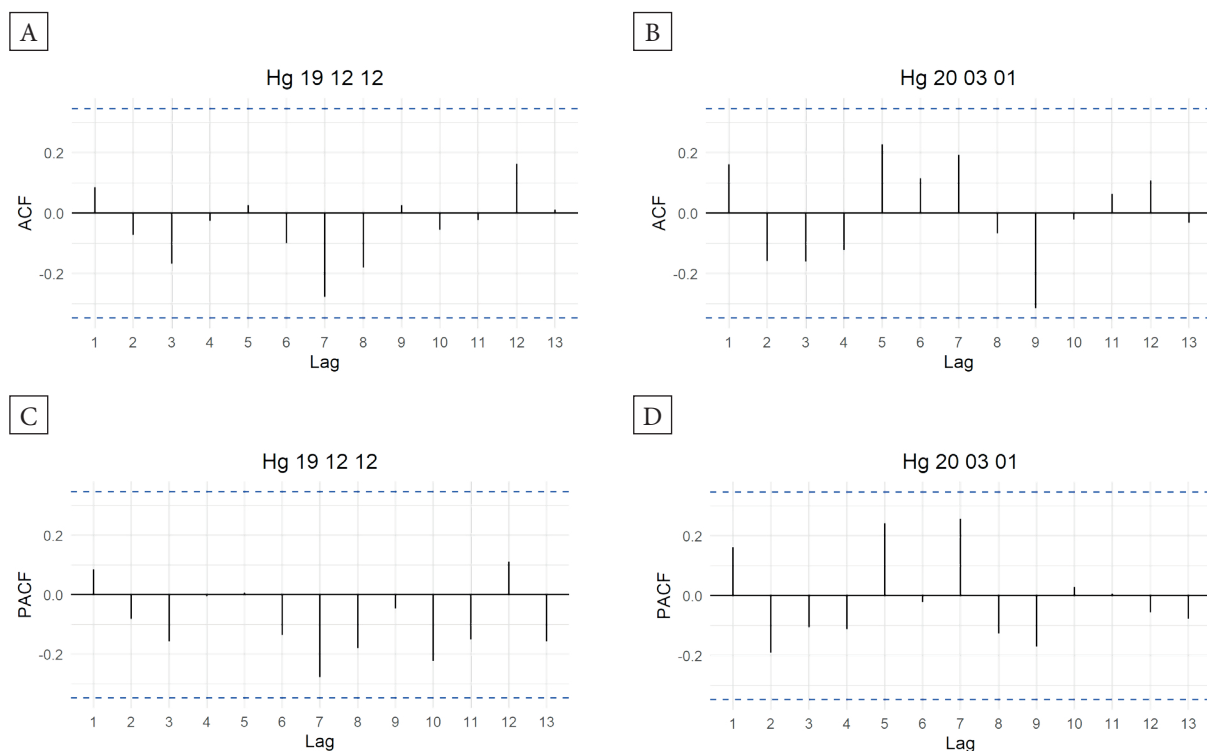
Fig. 3. *ACF (A, B) and PACF (C, D) for mercury concentration in wastes with codes 1 9 19 12 and 20 03 01*

Table 3
ADF test of mercury concentration in waste with codes 19 12 12 and 20 03 01

Lag order	Hg 19 12 12		Hg 20 03 01	
	ADF stat	*p*-value	ADF stat	*p*-value
3	−2.38	0.02	−2.54	0.01

To assess the absence of a trend in the data, regression models were created. Several linear and nonlinear parameter models were tested. Table 4 presents selected models with the best R^2 results including the linear model, the second-order polynomial (parabolic), and the logarithmic model. The linear model included the intercept (*a*) and the trend line (*b*1). The parabolic model also contained the intercept (*a*) and two components of the trend *b*1 and *b*2. The logarithmic model included the intercept (*a*) and a logarithmic trend (*b*1). Presented models are not statistically significant and are inaccurate. Only the intercept is statistically significant, with intercept coefficients close to 0 for each presented model. Other model coefficients are close to 0 (Tab. 4 – coefficients section) too, and statistically insignificant. The quality of the models also indicates a lack of fit. The values of the R^2 determination coefficient are close to 0, i.e., the created models do not explain the variability of the original data. The highest R^2 value is 0.062 for the parabolic mercury model of 20 03 01 waste, which indicates that this model explains only 6% of the original data variability. The values of the *F* statistics and the *p*-value, which indicate the ability to predict, also confirm an extremely low fit of the presented models.

Analyzing results: the mercury concentration in waste with codes 19 12 12 and 20 03 01 can be considered stationary (ADF test); they do not have a trend (ACF, PACF, regression models) and periodicity (PACF). Therefore, they can be considered stable.

Table 4
Selected regression models for mercury content in waste with codes 19 12 12 and 20 03 01

Data	Model parameters and quality		Model		
			linear	parabolic	logarithmic
Hg 19 12 12	Coefficients	a	0.019***	2.43e-02**	0.022*
		b1	0.000	−9.07e-04	−0.001
		b2	–	2.35e-05	–
	Quality	F-statistic	0.241	0.375	0.268
		p-value	0.627	0.691	0.608
		R^2	0.008	0.025	0.009
Hg 20 03 01	Coefficients	a	0.018***	1.97e-02**	0.019
		b1	−0.000	−5.05e-04	−0.002
		b2	–	5.73e-06	–
	Quality	F-statistic	1.936	0.958	0.825
		p-value	0.174	0.395	0.371
		R^2	0.060	0.062	0.027
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

CONCLUSIONS

Based on the obtained results it can be concluded that it is possible to analyze trends, the level of occurrence, and the stability of mercury concentrations in waste based on statistical tools and data mining, even in the case of an incomplete data set. The cross-industry standard process for data mining methodology allows the mercury content and stability assessment to be performed both transparently and seamlessly. The use of CRISP-dm makes it easy to trace the flow of data, analysis, and results. Its use will make it possible to obtain the repeatability of the results.

The mercury content in waste (code 19 12 12 – Other waste (including mixed substances and objects – from mechanical treatment of waste other than those mentioned in 19 12 11) and 20 03 01 – Unsorted (mixed) municipal waste) is significantly lower than the most stringent values given in the Regulation of the Minister of the Environment of 1 September 2016 on the method of conducting the assessment of soil surface pollution (*Rozporządzenie…* 2016) of 2 mg/kg s.m. for II-1 soils and the Minamata Convention Report of 1 mg/kg dry weight for waste.

Student's *t*-tests confirmed (see Tab. 2) that the mercury concentrations in wastes with codes 19 12 12 and 20 03 01 are lower than the most stringent values given above. Thus, due to the exceptionally low mercury content of municipal waste, it should not be considered potentially contaminated with mercury. Both mean values for the two waste groups and maximum and average values were compared. Based on the analysis, it can be concluded that waste with codes 19 12 12 and 20 03 01 have exceptionally low mercury concentrations. The variation in Mercury content in wastes with codes 19 12 12 and 20 03 01 shown in Figure 2 is low and the presented values are close to 0. The graph shows no time variation or periodicity. The graph confirms the hypothesis that low Mercury content in waste and data stability are related to the lack of unpredictable volatility, trend, and seasonality.

The correlations are not statistically significant for any of the lags, as they do not exceed the designated confidence levels for both the autocorrelation and partial autocorrelation functions (Fig. 3). Thus, there is no trend or seasonality when it comes to the mercury concentration of wastes (19 12 12 and 20 03 01). Based on the ACF and PACF charts, the data can be considered stable (with no unpredictable variability).

Based on ADF test statistics, it was shown that the data for both codes: 19 12 12 and 20 03 01 are stationary (Tab. 3). The mercury concentration of the tested wastes can be considered stable (considering stationarity as stability).

The simple linear regression, multiple regression, and non-linear regression models performed proved that there is no developmental trend in the data that can be described by a linear equation, a second-order polynomial, or a logarithm. The values of the coefficient of determination R^2 are close to 0 for simple linear regression models and non-linear models due to explanatory variables (parabolic, logarithmic). Such a low quality of both models indicates the lack of linear and non-linear relationships in the data in relation to explanatory variables.

Based on the analysis of the results from the above descriptive statistics, i.e., average and maximum values, data visualization, and the student's t-test, it can be concluded that the mercury content in waste codes 19 12 12 and 20 03 01 is low.

Based on the ACF and PACF plots of the regression models, it can be concluded that there is no linear or non-linear trend in the data, i.e., no variability of the data over time. The ACF and ADF test and data visualization confirmed that there is no non-stationarity in the data, and the mercury concentration can be considered stable. The restriction of the marketing of mercury (Fig. 1)-containing products in Europe and the separate collection of municipal waste have led to a significant decrease in the content of this element in municipal waste, from 4 mg/kg dry weight in 1990 (Italy), by 2 mg/kg dry weight in 1997 (Europe) to 1–5 mg/kg dry weight in 2001 (Germany) and 0.062 mg /kg dry weight in 2021 (Poland).

REFERENCES

CEN (European Committee for Standardization), 2011a. *EN 15442:2011: Solid recovered fuels – Methods for sampling.* https://standards.iteh.ai/catalog/standards/cen/1cce-c9d1-cd27-493a-ba71-f927ca8e57ca/en-15442-2011 [access: 19.06.2022].

CEN (European Committee for Standardization), 2011b. *EN 15443:2011*: *Solid recovered fuels – Methods for the preparation of the laboratory sample.* https://standards.iteh.ai/catalog/standards/cen/b049702f-7c9c-450e-abf7-ca06800c99cd/en-15443-2011 [access: 19.06.2022].

Cho B.H., Nam B.H., An J. & Youn H., 2020. Municipal solid waste incineration (MSWI) ashes as construction materials – A review. *Materials (Basel, Switzerland)*, 13(14), 3143. https://doi.org/10.3390/ma13143143.

Chow S.-Ch., 2007. *Statistical Design and Analysis of Stability Studies*. Chapman and Hall/CRC, New York. https://doi.org/https://doi.org/10.1201/9781584889069.

Dégerine S. & Lambert-Lacroix S., 2003. Characterization of the partial autocorrelation function of nonstationary time series. *Journal of Multivariate Analysis*, 87(1), 46–59. https://doi.org/https://doi.org/10.1016/S0047-259X(03)00025-3.

EEA (European Environment Agency), 2019. *Changes in cadmium, mercury and lead emissions for each sector (EEA-33).* https://www.eea.europa.eu/data-and-maps/daviz/change-in-cadmium-mercury-and-5#tab-chart_1 [access: 19.06.2022].

EP and CEU (European Parliament and the Council of the European Union), 2008a. Consolidated text: *Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006.* Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008R1272&from=PL [access: 19.06.2022].

EP and CEU (European Parliament and the Council of the European Union), 2008b. Consolidated text: *Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain Directives.* Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02008L0098-20180705 [access: 19.06.2022].

EP and CEU (European Parliament and the Council of the European Union), 2011. Consolidated text: *Directive 2011/65/EU of the European Parliament and of the Council of 8 June 2011 on the restriction of the use of certain hazardous substances in electrical and electronic equipment.* Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02011L0065-20211101 [access: 19.06.2022].

EP and CEU (European Parliament and the Council of the European Union), 2017. *Regulation (EU) 2017/852 of the European Parliament and of the Council of 17 May 2017 on mercury, and repealing Regulation (EC) No 1102/2008.* Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0852 [accessed: 19.06.2022].

EP and CEU (European Parliament and the Council of the European Union), 2018. Consolidated text: *Directive 2006/66/EC of the European Parliament and of the Council of 6 September 2006 on batteries and accumulators and waste batteries and accumulators and repealing Directive 91/157/EEC.* Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02006L0066-20180704 [accessed: 19.06.2022].

Fabricius A.-L., Renner M., Voss M., Funk M., Perfoll A., Gehring F. et al., 2020. Municipal waste incineration fly ashes: from a multi-element approach to market potential evaluation. *Environmental Sciences Europe*, 32(1), 88. https://doi.org/10.1186/s12302-020-00365-y.

Fu X., Feng X., Sommar J. & Wang S., 2012. A review of studies on atmospheric mercury in China. *The Science of the Total Environment*, 421–422, 73–81. https://doi.org/10.1016/J.SCITOTENV.2011.09.089.

Fuller W.A., 1995. *Introduction to Statistical Time Series*. 2ⁿᵈ ed. Wiley Series in Probability and Statistics, John Wiley & Sons, New York.

Giro-Paloma J., Ribas-Manero V., Maldonado-Alameda A., Formosa J. & Chimenos J.M., 2017. Use of municipal solid waste incineration bottom ash and crop by-product for producing lightweight aggregate. *IOP Conference Series: Materials Science and Engineering*, 251(1), 12126. https://doi.org/10.1088/1757-899X/251/1/012126.

Godyń K. & Dutka B., 2023. Preliminary studies of slag and ash from incinerated municipal waste for prospective applications. *Energies*, 16(1). 117. https://doi.org/10.3390/en16010117.

Godyń K., Dutka B., Chuchro M. & Młynarczuk M., 2020. Synergy of parameters determining the optimal properties of coal as a natural sorbent. *Energies*, 13(8). https://doi.org/10.3390/en13081967

Helsen D.R., 2005. *Nondetects and Data Analysis : Statistics for Censored Environmental Data*. Wiley-Interscience, Hoboken, N.J.

Helsen D.R., 2012. *Statistics for Censored Environmental Data Using Minitab and R*. 2ⁿᵈ ed. John Wiley & Sons, Hoboken, N.J.

Kabata-Pendias A., 2011. *Trace Elements in Soils and Plants*. 4ᵗʰ ed. CRC Press.

Manuca R. & Savit R., 1996. Stationarity and nonstationarity in time series analysis. *Physica D: Nonlinear Phenomena*, 99(2), 134–161. https://doi.org/https://doi.org/10.1016/S0167-2789(96)00139-X.

Marnane I., 2018. *Mercury in Europe's environment: A priority for European and global action*. EEA Report, 11/2018, EEA, Copenhagen.

Mishra P., Singh U., Pandey C.M., Mishra P. & Pandey G., 2019. Application of student's t-test, analysis of variance, and covariance. *Annals of Cardiac Anaesthesia*, 22(4), 407–411. https://doi.org/10.4103/aca.ACA_94_19.

Montgomery D.C., 2012. *Introduction to Linear Regression Analysis*. John Wiley & Sons, Hoboken, N.J.

Murtagh F. & Devlin K., 2018. The development of data science: Implications for education, employment, research, and the data revolution for sustainable development. *Big Data and Cognitive Computing*, 2(2), 14. https://doi.org/10.3390/bdcc2020014.

Neuwahl F., Cusano G., Benavides J.G., Holbrook S. & Roudier S., 2019. *Best Available Techniques (BAT) reference document for waste incineration: Industrial Emissions Directive 2010/75/EU (Integrated Pollution Prevention and Control)*. JRC Science for Policy Report, Publications Office of the European Union, Luxembourg. https://op.europa.eu/pl/publication-detail/-/publication/075477b7-329a-11ea-ba6e-01aa75ed71a1/language-en [access: 8.12.2021].

R packages, 2022. *R packages list with full documentation*. https://cran.r-project.org [access: 25.06.2022].

Ross S.M., 2005. Linear Regression. [in:] Ross S.M., *Introductory Statistics*, 2ⁿᵈ ed., Academic Press, 519–584.

Rozporządzenie Ministra Środowiska z dnia 1 września 2016 r. w sprawie sposobu prowadzenia oceny zanieczyszczenia powierzchni ziemi. Dz.U. 2016 poz. 1395. http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=wdu20160001395 [access: 19.06.2022].

Schmidt J., Marques M.R.G., Botti S. & Marques M.A.L., 2019. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials*, 5(1), 83. https://doi.org/10.1038/s41524-019-0221-0.

Schröer C., Kruse F. & Gómez J.M., 2021. A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534. https://doi.org/https://doi.org/10.1016/j.procs.2021.01.199.

Song Y., Jiang T., Liem-Nguyen V., Sparrman T., Björn E. & Skyllberg U., 2018. Thermodynamics of Hg(II) bonding to thiol groups in Suwannee River natural organic matter resolved by competitive ligand exchange, Hg LIII-Edge EXAFS and 1H NMR spectroscopy. *Environmental Science and Technology*, 52(15), 8292–8301. https://doi.org/10.1021/acs.est.8b00919.

UNEP (United Nations Environment Programme), 2013. *Final Act of the Conference of Plenipotentiaries on the Minamata Convention on Mercury | Minamata Convention on Mercury*. https://www.mercuryconvention.org/en/documents/final-act-conference-plenipotentiaries-minamata-convention-mercury [access: 19.06.2022].

UNEP (United Nations Environment Programme), 2018. *Global Mercury Assessment*. https://www.unep.org/resources/publication/global-mercury-assessment-2018 [access: 29.11.2021].

UNEP (United Nations Environment Programme), 2021. *Minamata Convention Progress Report 2020*. https://www.unep.org/resources/report/minamata-convention-progress-report-2020 [access: 12.06.2022].

U.S. EPA (U.S. Environmental Protection Agency), 1998. *Method 7473 (SW-846): Mercury in Solids and Solutions by Thermal Decomposition, Amalgamation, and Atomic Absorption Spectrophotometry*. Revision 0. Washington, DC. https://www.epa.gov/esam/epa-method-7473-sw-846-mercury-solids-and-solutions-thermal-decomposition-amalgamation-and [access: 8.06.2022].

U.S. EPA (U.S. Environmental Protection Agency), 2000. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis*. https://www.epa.gov/sites/default/files/2015-06/documents/g9-final.pdf [access: 12.06.2022].

Ustawa z dnia 20 maja 2021 r. o ratyfikacji Konwencji z Minamaty w sprawie rtęci, sporządzonej w Kumamoto dnia 10 października 2013 roku. Dz.U. 2021 poz. 1201. https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20210001201 [access: 19.06.2022].

van Velzen D., Langenkamp H. & Herb G., 2002. Review: Mercury in waste incineration. *Waste Management and Research*, 20(6), 556–568. https://doi.org/10.1177/0734242X0202000610.

WHO (World Health Organization), 2017. *Mercury and health*. https://www.who.int/news-room/fact-sheets/detail/mercury-and-health [access: 29.11.2021].

Wirth R. & Hipp J., 2000. CRISP-DM: Towards a standard process model for data mining. [in:] *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, Practical Application Company, 29–39.