IOŚ-PIB OCHRONA ŚRODOWISKA I ZASOBÓW NATURALNYCH
ENVIRONMENTAL PROTECTION AND NATURAL RESOURCES 2013
VERSITA
VOL. 24, NO 3(57): 45–51
DOI 10.2478/oszn-2013-0020

Damian Zasina*, Jarosław Zawadzki**

# Statistical analysis of data set on national reporting of emission of air pollutants. Part I: investigation of outliers

# Analiza statystyczna zbioru danych pochodzącego z krajowej sprawozdawczości emisji zanieczyszczeń do powietrza, Cz. I: wykrywanie wartości odstających

** Mgr inż. Damian Zasina - National Centre for Emission Management, Institute of Environmental Protection- National Research Institute, Chmielna 132/134 St, 00-805 Warsaw, e-mail: damian.zasina@gmail.com*

*** Dr hab. inż. Jarosław Zawadzki, prof. nadzw. - Faculty of Environmental Engineering, Warsaw University of Technology, Nowowiejska 20 St, 00-653 Warsaw e-mail: j.j.zawadzki@gmail.com*

**Keywords:** Poland, emission inventory, LRTAP, $NO_X$, CO, $SO_X$, TSP, outliers, CRAN-R
**Słowa kluczowe:** Polska, inwentaryzacja emisji, LRTAP, $NO_X$, CO, $SO_X$, TSP, wartości odstające, CRAN-R

**Abstract**

The Polish emission reporting system – "Krajowa baza o emisjach gazów cieplarnianych i innych substancji" (or National Emission Database (NED)) – was established at the end of 2010. Initially (data submitted for 2010), the database contained reported emission data for greenhouse gases and air pollutants from plants that have had proper Integrated Pollution Prevention and Control permissions (i.e., integrated permission for the release of gases and dusts into the air). The emissions reported to the NED are recognised as the emissions from local sources and partly as the emissions from point sources, with the possibility of including them into a national emission inventory as point source data (in the case of air pollutants). In the near future, it is planned that the database will be perceived as an integrated system for national air emission management (and the emission data from all sources will be required to pay a "tax for the use of the environment", which will be regulated by national Polish law).

This paper is a part of the work related to the analysis of reported emission data. Additional research on the data collected in the national database might be used to develop a National Emission Inventory, in addition to evolution of country-specific emission factors (e.g. from combustion and industrial processes). The analysed data (emission of $NO_X$, CO, $SO_X$ and TSP) were taken from the data for point sources submitted for 2011 primarily with the aim of improving the quality of data submitted previously – for 2010. This paper is the first study in the research to investigate outliers among the reported data using some basic statistical methods.

**Streszczenie**

„Krajowa baza o emisjach gazów cieplarnianych i innych substancji'' (zw. dalej „Krajową bazą"), powstała pod koniec 2010 roku ma docelowo stanowić zintegrowany, polski system zbierania informacji o emisjach gazów cieplarnianych i innych substancji na potrzeby zarządzania emisjami gazów cieplarnianych i zanieczyszczeń powietrza na szczeblu krajowym. Dane o emisjach pochodzące z Krajowej bazy sprzyjają rozwojowi metodyki krajowej w zakresie zarówno inwentaryzacji emisji zanieczyszczeń powietrza (włączenie emisji pochodzącej ze źródeł punktowych – dane indywidualne) jak i opracowania krajowych wskaźników emisji (ang. *country specific*). Dane za rok 2010 obejmowały podmioty posiadające jedno z właściwych pozwoleń: pozwolenie zintegrowane albo pozwolenie na wprowadzanie gazów i pyłów do powietrza. Docelowo w Krajowej bazie mają być zawarte dane o emisjach ze wszystkich podmiotów wprowadzających opłaty za korzystanie ze środowiska.

Niniejszy artykuł ma stanowić wstęp do pracy poświęconej analizie danych (o emisjach zanieczyszczeń powietrza), które pochodzą z Krajowej bazy (emisje of $NO_X$, CO, $SO_X$ oraz TSP). Do analizy wstępnej wybrano dane za rok 2011 przede wszystkim w aspekcie poprawy jakości raportowania w stosunku do roku 2010. W pierwszym kroku prowadzenia analizy opracowanie poświęcono analizie wartości odstających (ang. outliers) z wykorzystaniem podstawowych technik statystycznych.



## 1. INTRODUCTION

The Polish emission reporting system – the National Emission Database (NED) – was established in December 2010. Initially, this system is meant to be an integrated system for air pollutant emission management at the country level. Moreover, the data submitted to the NED are meant to be a substitute for the current system of paying "environmental tax", which is required by Poland [LAW2]. The data collection submitted for 2011 includes emissions for air pollutants ($NO_X$, CO, $SO_X$ and TSP) from plants that had received integrated permission for the release of gases and dusts into ambient air[1]. Emissions from plants that had submitted the proper questionnaire are classified according to the classification system in the Polish regulations [LAW3, LAW1], which is established in multi-level system.

The system of classification [LAW1; IEP–NRI, NCEM, 2012] consists of four basic classes: sector[2], installation[3], emission source[4] and emitter[5]. In addition, it also established an auxiliary class called

1 For details, see: http://www.ippc.mos.gov.pl/ippc or Directive 96/61/EC (Council Directive 96/61/EC of 24 September 1996 concerning Integrated Pollution Prevention and Control).

2 Distinguished nine main sectors of the economy (and the auxiliary sector: "other"), i.e.. *Heat and power* (energy sector), *Production and treatment of metals, Mineral industry, Chemical industry*, etc. Heat and power is shared among the subsectors that are typical of the Polish economy (the subsector is meant to be an indicator of the type of emitting plant, e.g. public power, combined heat and power, industrial power plant).

conduit, which is an intermediary between installation, (emission) source and emitter.

Among the various types of collected data, numerous types of errors could be observed, primarily related to the outlying observations (outliers). An outlying observation was described by Hawkins (1980) as "an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism"; a number of papers focussing on outlying observations have been published by Barnett and Lewis (1994), Johnson (1992), Acuna and Rodriguez (2004), Ilango, Subramanian and Vasudevan (2012) and others.

A significant part of the work focussing on outlying observations and sources of uncertainties about compiling emission inventories was published by Frey (2007a, 2007b).

The issue of outlying data also concerns the emission data submitted to the NED. The majority of observed outliers occur by the "human factor"[6], which creates the probability of occurrence of an error almost everywhere among the data. Checking the values row-by-row every time is a tedious and ineffective way of doing data analysis[9].

This article is a short revision of several simple techniques based on applied statistics that might be very useful in case of "first step" validation of submitted data.

## 2. DATA PREPARATION

The basic data sets on emissions contained two separated subsets:

– CO and $NO_X$ from 3804 plants.
– $SO_X$ and TSP from 2449 plants.

All the data were adapted from submissions reported for 2011 by various plants, primarily emission sources connected to fuel combustion (also LCP[8] sources).

Separated data set on LCP also included similarly:

– CO and $NO_X$ from 97 plants.
– $SO_X$ and TSP from 97 plants.

Submitted emissions of air pollutants were logarithmised before. Main statistical characteristics are presented in Table 1.

The main cause of arbitrary selection of the year 2011 was because of the huge improvement in submitted data compared with the year 2010. The order of data preparation is enumerated below (see the approach applied by Frey 2007a, 2007b, in the context of errors in emission inventories):

1 Correction of significant errors.
  Significant errors were distinguished arbitrarily. There was a considered situation in which the occurrence of that kind of errors tends to significantly change the results (in the context of emission structure).
  1.1 Corrected *allocation errors* (wrong geographical coordinates of emission source or emitter; among the data submitted for 2011, the number of data wrongly coordinated decreased from over 10% to below 1%).
  1.2 Corrected *errors in estimation* (the result of submission of mistaken values or emission units, sometimes missing values), e.g. plant 6673[9] (small brickyard), submitted emission of $NO_X$ is comparable with emission from large power plant or plant 2734 (dairy processing plant), submitted emission of $SO_X$ and TSP is comparable with emission from large power plant.
  1.3 Corrected *classification errors* (refer to emission source, installation or emitter, primarily improper assignment of emission source to wrong sector, impossible to investigate using statistical tools), e.g. plant 2076 (confectionery), assignment of confectionery oven to energy sector (should be included in food production and processing sector) or plant 104 (power plant in coal mine), assignment of power plant to public power and energy sector (according to used nomenclature included in mining sector).

2 Correction of less significant errors.
  Less significant errors let carrying out sectoral analysis without committing mistakes that would change the structure of emission.
  2.1 Corrected name errors (Myatt 2007) (occurred as an assignment of various names to the same or comparable emission source; usually a negligible type of error, but it might be significant in case of carrying out a detailed analysis, among over than 10,000 installations; the number of name errors could be up to 1.5–2%), e.g. Hoffman kiln (there are five distinguished independent Polish synonyms for the same installation).
  2.2 Corrected error of assignment (assignment of emission source to auxiliary category in case of existence proper class).
  2.3 Other corrected errors connected to improper assignments (e.g. misuse of the auxiliary category type or assignment of improper statistical code NACE or PKD[10]).

A part of data correction is done by using simple tools such as row-by-row investigation or the analysis of extreme values (it refers to significant errors, primarily).

## 3. METHODOLOGY

Initially prepared data set was used as an input data to analysis in CRAN-R statistical package. A large number of statistical tools such as histograms, scatterplots and Grubbs T-test were used.

### Histograms

Detailed description of examples for conducting analysis by histograms was published by Myatt (2007) and Kriegel, Kröger and Zimek (2010). As outliers usually identify items/pieces of data where there

a) are observable extreme values – in case of emission submissions, each submission is removed so that its value of logarithm of submitted emission is not greater than 0; the main cause of removing that kind of data was the impossibility of estimation of the value of annual emission $NO_X$, CO, $SO_2$ or TSP from all plants with an accuracy of 1 kg (or less); the data from LCP sources are better in the context of installation and using specialised equipment for continuous emission measurement;

3 Within each type of sector, almost 30 types of installations are distinguised: *Wood production and processing* – 7 types of installations (including the auxiliary); *Chemical industry* – 27 types (including the auxiliary type).
4 Basic unit of classification, distinguished based on the specificity of the emission inventory and the submitting system. Any device that is able to emit air pollutants directly or indirectly (e.g. utility boiler, oven, stationary engine, conveyor belt) is considered an emission source;, all emission sources are geographically located: a group of emission sources connected to each other by pipes is considered an installation.
5 Every unit is able to release pollutants into ambient air directly (e.g. stack); emitters are connected to emission sources via conduits (e.g. pipes); all emitters are geographically located.
6 Outlying observations occur often as a result of unintended mistakes and errors connected to both plant submitting data and controlling system.
7 Based on own experience.
8 Large combustion plants.
9 Identification number in National Emission Database.
10 NACE – Statistical classification of economic activities in the European Community , PKD – National Polish system of statistical classification; for details, see: http://ec.europa.eu/eurostat/ramon/ (NACE) or www.stat.gov.pl/klasyfikacje/ (PKD).

Table 1. Summary statistics of basic data sets (parameters calculated only for positive values are in parentheses)

| Statistics | Logarithmised data on emission | | | |
|---|---|---|---|---|
| | CO | $NO_x$ | $SO_x$ | TSP |
| N | 3804 (3666) | 3804 (3689) | 2449 (2271) | 2449 (2294) |
| Minimum | 4.605 (0.010) | 4.605 (0.058) | 4.605 (0.058) | 4.605 (0.020) |
| First quartile | 3.844 (4.129) | 4.305 (4.489) | 3.199 (3.905) | 4.225 (4.920) |
| Median | 6.398 (6.572) | 6.354 (6.478) | 5.964 (6.356) | 7.009 (7.313) |
| Mean (average value) | 6.370 (6.667) | 6.415 (6.665) | 6.243 (6.834) | 6.443 (7.018) |
| Third quartile | 9.157 (9.220) | 8.688 (8.788) | 9.809 (10.100) | 9.185 (9.333) |
| Maximum | 18.290 | 17.520 | 18.180 | 14.580 |
| Shapiro-Wilk normality test result (α = 0.05) | – | – | – | – |

b) are observable atypical clusters of value classified as separate group (e.g. several dozens of plants that reported an annual emission of $SO_2$ between 500 and 1000 kg and also lack of plants that submitted greater values).

In case of conducting the analysis of values connected with emissions, the existence of big uncertainty of estimation should be taken into consideration. Only the data submitted by LCP can be considered as sufficiently reliable.

### Scatterplots

Scatterplots are one of the most popular tools for the investigation of outliers (Ben-Gal 2005; Ilango, Subramanian & Vasudevan 2012; Kriegel, Kröger & Zimek 2010; Myatt 2007), which give the possibility to carry out exploratory data analysis in a quite fast and effective way. Also, using scatterplots let distinguish clusters, considered as occurrence of characteristic dependences in a certain range of values of investigated variables. Analysis of scatterplots can be complementary to analysis of technological structure of industrial plants.

### Statistical tests for the investigation of outliers

Statistical testing can be a useful tool for the investigation of outliers. The most popular statistical tests for investigating outliers are: Grubbs T test (Engineering Statistics Handbook n.d.; Grubbs 1950; Komsta 2011), Dixon's Q test (Dixon 1950, 1951; Komsta 2011; Rorabacher 1991) and $\chi^2$ test for outliers (Dixon 1950). Statistical tests (as above) are useful for investigating single outlying value that causes the situation, where necessary testing in an iterative way ("step by step") (Iwaniec 2008). The highest effectiveness of statistical testing is possible where at most several outliers among data are expected. For the investigation of outliers, Grubbs T test was used.

## 4. RESULTS

### Histograms

In Figures 1–4, histograms are presented for the case of logarithm of submitted emission of $NO_x$ for four different situations: basic data set, set of positive values, set of positive values <15 and set of values >9.

The value of logarithm of submitted emission >9 corresponds to plants, from which submitted emission of $NO_x$ was >8103 kg.

### Scatterplots

Considering dependency between emissions of CO and $NO_x$, also $SO_2$ and TSP (Gallardo et al. 2011; Kumari et al. 2011; Likens, Buso & Butler 2005), scatterplots for two variables (value of logarithm of submitted emission) are presented. Scatterplots for selected LCP are presented in Figures 5–8. Most of the emission sources (90% of emission sources in installations of public power sector) are equipped with a device for continuous measuring of emission, the rest (10% of sources) of the sources do not have that kind of devices installed.

### Statistical testing

Below, there are presented results of statistical testing with using Grubbs T test. As the tested samples were used the data set of logarithm of submitted emissions of CO and $NO_x$ from LCP sources (primarily power or CHP[11] plants). As the tool for conducting Grubbs T test , the "outliers" package for CRAN-R (Komsta 2011) was used. The samples were tested in iterative way ("step by step"). The result of the testing is presented in Table 2.

In the fourth step of iteration – the outlying observations detected with Grubbs T test were different for both samples. As two different observations, values submitted by two various plants were considered.

Table 2. Results of using Grubbs T test – four iterations (submitted emission of CO and $NO_x$)

| Step | Sample | | Status |
|---|---|---|---|
| | CO (ID of detected element) | $NO_x$ (ID of detected element) | |
| 1 | 1013* | 1013 | Detected, removed from sample |
| 2 | 405 | 405 | Detected, removed from sample |
| 3 | 341 | 341 | Detected, removed from sample |
| 4 | 1176 | 998 | Detected as an outlying value |

*Identification key in National database.

11 Combined heat and power.

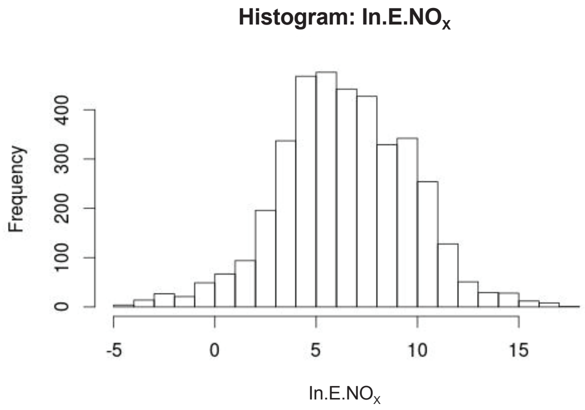


Figure 1. Histogram of logarithm of reported emission of $NO_X$ – basic data set (own analysis).

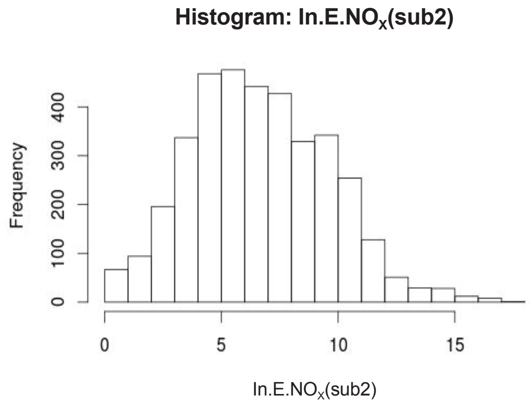


Figure 2. Histogram of logarithm of reported emission of $NO_X$ – positive values (own analysis).

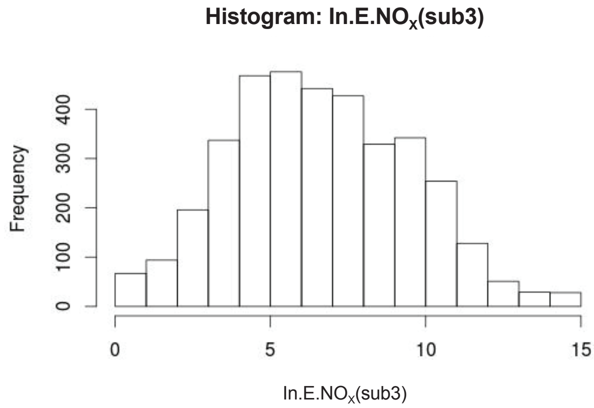


Figure 3. Histogram of logarithm of reported emission of $NO_X$ – positive values <15 (own analysis).

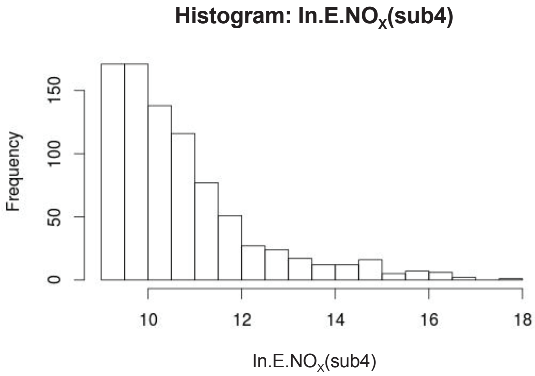


Figure 4. Histogram of logarithm of reported emission of $NO_X$ – values >9 (own analysis).

After the third step of iteration, the sample was split into two independent subsamples: based on CO and $NO_X$. Further analysis was carried out separately for both subsamples.

Testing of samples with Grubbs T test was stopped after the following:

– Four iterations (CO); in the fifth iteration, the highest value was detected as an outlier (highest value detected referred to power plant equipped with measurement device) (Figure 9).
– Six iterations ($NO_X$); in the seventh iteration, the lowest value was detected as an outlier (iteration loop was stopped due to the shape of the finished histogram) (Figure 10).

The results of the analysis are presented in Table 3.

## 5. CONCLUSIONS

In this paper, they tried to analyse the usefulness of simple statistical techniques for the development of emission inventory and partly for emission data management in the context of investigation of outlying data. For the purpose of conduction of analysis, both simple selected techniques of exploratory statistical data analysis and simple statistical testing for detecting outliers were used. According to Frey's approach (2007a, 2007b) in the context of sources of data uncertainties in emission inventories, own approach considering errors in the NED was applied (by analogy). As an indirect result of analysis, there were observed dependencies among the investigated data:

– Existence of positive correlation between logarithms of submitted emission of CO and $NO_X$ (based on scatterplots: Figures 5 and 7), because of the existence of large (not estimated) uncertainty of data – calculation of correlation factor was omitted.
– Existence of positive correlation between logarithms of submitted emission of $SO_2$ and TSP (based on scatterplots: Figures 6 and 8), because of the existence of large (not estimated) uncertainty of data – calculation of correlation factor was omitted.

As a direct result, the following were observed:

– The possibility of using simple statistical techniques (especially techniques based on histograms and scatterplots) as an auxiliary device for the investigation of outlying observations.
– The same or analogous sources of errors in National Emission Database than those described by Frey (2007a, 2007b).
– Specificity of submitted data (or its logarithm) in terms of statistical properties (uncertainty), despite the large uncertainties of data – there were observable (scatterplots) significant dependencies (Gallardo et al. 2011; Kumari et al. 2011; Likens, Buso & Butler 2005).
– In both the cases of scatterplot data analysis (for all plants and for LCP), the number of outlying values could be up to dozen (LCP, Figures 7 and 8) or even several hundreds (all plants, Figures 5 and 6).

- Further, the correlation between data on $SO_X$ and TSP seems to be less than correlation of data on CO and $NO_X$ (based on observations of scatterplots); this fact might have caused complexity of combustion process or selection of wrong emission factors, inconvenient for the type of conducted process (Frey 2007a, 2007b).
- Apart from significant drawback (Iwaniec 2008), Grubbs T test could be a very useful method for the investigation of outlying values, especially it is worth to support statistical testing with histograms (Table 3, Figures 9 and 10).
- Presented statistical analysis could be effectively used in emission reporting, using Grubbs T test let simplify analysis of particular type data sets (LCP are considered the particular class of emission sources); for the main purposes, using statistical testing could be useful for detecting normal distributions among reported data; for carrying out the analysis of complete data sets, there is a necessity to conduct independent analysis of outlying values.
- That kind of support analysis might also be used for analysing uncertainties in emission reporting.

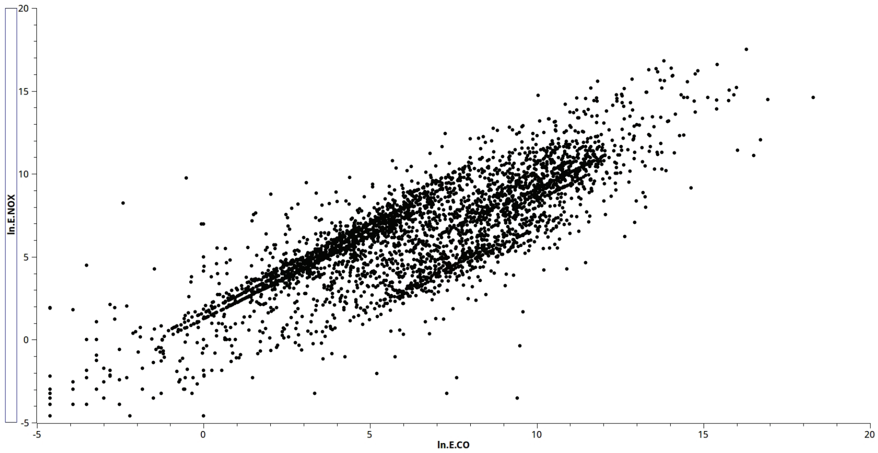


Figure 5. Scatterplot of logarithm of submitted emission of $NO_X$ (independent variable) and logarithm of submitted emission of CO (own analysis).

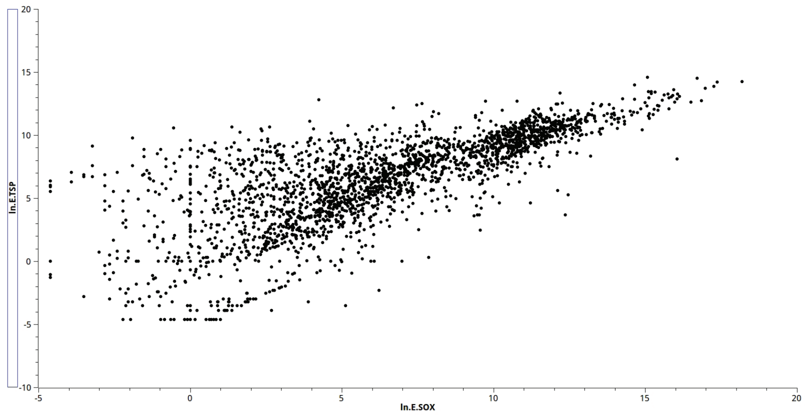


Figure 6. Scatterplot of logarithm of submitted emission of TSP (independent variable) and logarithm of submitted emission of $SO_X$ (own analysis).

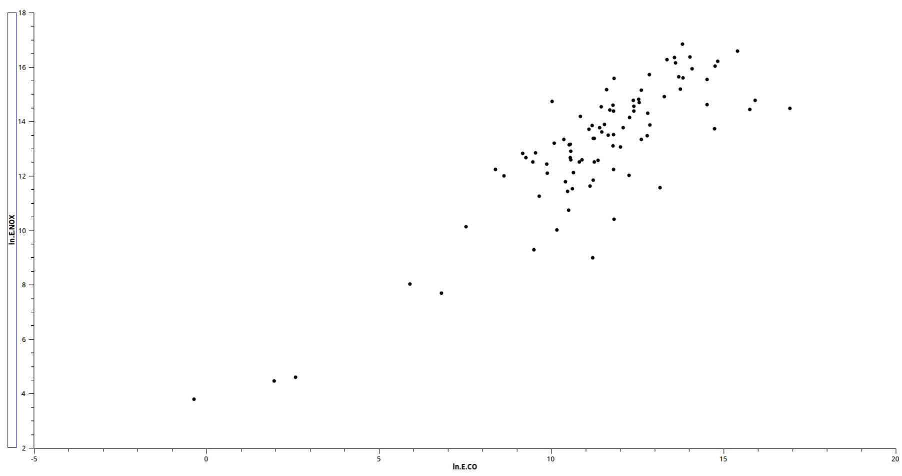


Figure 7. Scatterplot of logarithm of submitted emission of $NO_X$ (independent variable) and logarithm of submitted emission of CO (dependent variable) – large combustion plants (own analysis).

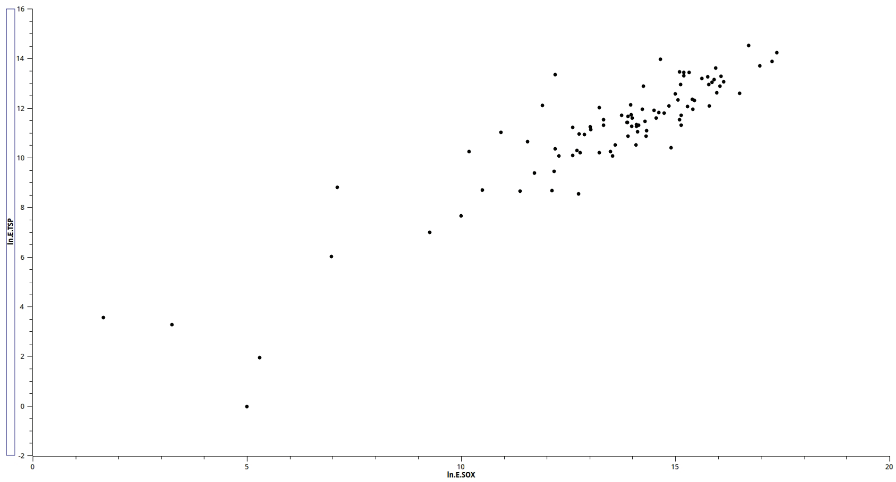


Figure 8. Scatterplot of logarithm of submitted emission of TSP (independent variable) and logarithm of submitted emission of $SO_X$ (dependent variable) – large combustion plants (own analysis).

Table 3. Rejection of outliers, Grubbs T test: summary

| Statistics | CO | | $NO_X$ | |
|---|---|---|---|---|
| | Before | After (four iterations) | Before | After (six iterations) |
| N | 97 | 93 | 97 | 91 |
| Minimum | 0.357 | 6.812 | 3.784 | 9.282 |
| First quartile | 10.510 | 10.560 | 12.230 | 12.510 |
| Median | 11.470 | 11.610 | 13.380 | 13.520 |
| Mean (average value) | 11.400 | 11.780 | 13.140 | 13.600 |
| Third quartile | 12.800 | 12.860 | 14.690 | 14.750 |
| Maximum | 16.930 | 16.930 | 16.830 | 16.830 |
| Shapiro-Wilk normality test result ($\alpha$ = 0.05) | (p-value = $1.54 \times 10^{-7}$) | + (p-value = 0.612) | (p-value = $1.75 \times 10^{-7}$) | + (p-value = 0.554) |

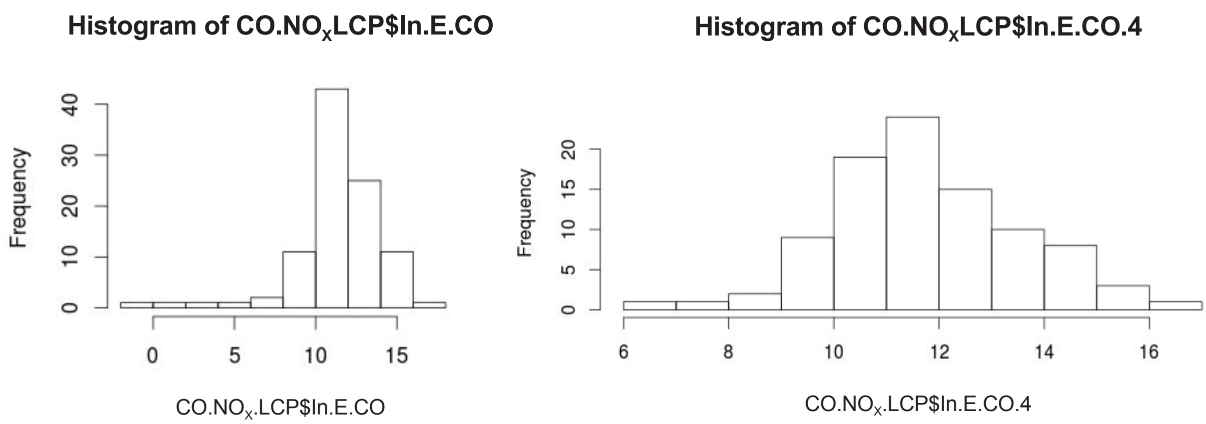


Figure 9. Rejection of outliers, submitted emission of CO, LCP (on the left: before rejection, on the right: after four iterations using Grubbs T test, own analysis).

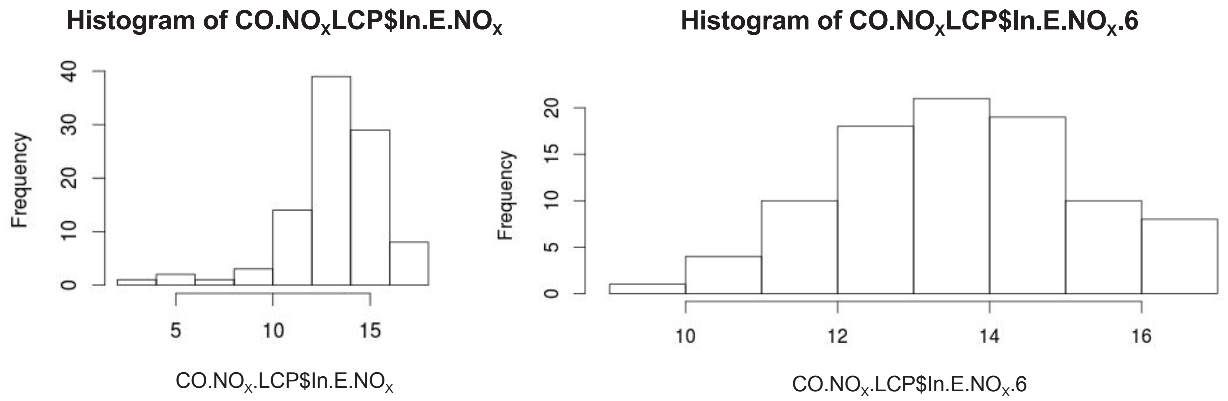


Figure 10. Rejection of outliers, submitted emission of $NO_X$, LCP (on the left: before rejection, on the right: after six iterations using Grubbs T test, own analysis).

## REFERENCES AND LEGAL ACTS

ACUNA, E & RODRIGUEZ, CA 2004, *Meta analysis study of outlier detection methods in classification*, Technical Paper, Department of Mathematics, University of Puerto Rico at Mayaguez, Venice.

BARNETT, V & LEWIS, T 1994, *Outliers in statistical data*, John Wiley.

BEN-GAL, I 2005, 'Outlier detection' in *Data mining and knowledge discovery handbook: a complete guide for practitioners and researchers*, eds O Maimon & L Rockach, Kluwer Academic Publishers, ISBN 0-387-24435-2.

DIXON, WJ 1950, 'Analysis of extreme values', *Annals of Mathematical Statistics*, no. 21, pp. 488–506.

DIXON, WJ 1951, 'Ratios involving extreme values', *Annals of Mathematical Statistics*, no. 22, pp. 68–78.

*Engineering statistics handbook (ESH), Grubbs' test for outliers* n.d. Available from: <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35h1.htm>.

FREY, HCH 2007a, Quantification of uncertainty in emission factors and inventories.

FREY, HCH 2007b, Quantification of uncertainty in air pollutant emissions inventories.

GALLARDO, L, ESCRIBANO, J, DAWIDOWSKI, L, ROJAS, N, DE FÁTIMA ANDRADE, M & OSSES, M 2011, 'Evaluation of vehicle emission inventories for carbon monoxide and nitrogen oxides for Bogotá, Buenos Aires, Santiago, and São Paulo', *Atmospheric Environment*, doi:10.1016/j.atmosenv.2011.11.051.

GRUBBS, F 1950, 'Sample criteria for testing outlying observations', *Annals of Mathematical Statistics*, no. 21, pp. 27–58.

HAWKINS, D 1980, *Identification of outliers*, Chapman and Hall.

IEP-NRI, NCEM, n.d. Zespół Ochrony Powietrza, 2012 Materiały szkoleniowe dla operatorów instalacji dotyczące wypełniania wniosku do Krajowej bazy o emisjach gazów cieplarnianych i innych substancji za rok 2012, Warszawa, edt. 2012 and 2013 [in Polish].

ILANGO, V, SUBRAMANIAN, R & VASUDEVAN, V 2012, 'A five step procedure for outlier analysis in data mining', *European Journal of Scientific Research*, no.3, pp. 327-339, ISSN 1450-216X.

IWANIEC, M 2008, *Walidacja metod pomiarowych*, StatSoft Polska [in Polish].

JOHNSON, R 1992, *Applied multivariate statistical analysis*, Prentice Hall.

KOMSTA, Ł 2011, *"Outliers" package for CRAN-R*. Available from: <http://cran.r-project.org/web/packages/outliers/index.html>.

KRIEGEL, H-P, KRÖGER, P & ZIMEK, A 2010, 'Outlier detection techniques', *The 2010 SIAM International Conference on Data Mining*.

KUMARI, H, JOON, V, CHANDRA, A & KAUSHIK, SC 2011, 'Carbon monoxide and nitrogen oxide emissions from traditional and improved biomass cook stoves used in India', in *International Conference on Chemistry and Chemical Process*, IPCBEE, vol. 10 (2011) © (2011), IACSIT Press, Singapore.

**[LAW1] Regulation of Polish Minister of Environment concerning the report questionnaire and way of its input to National Emission Database (pl: Rozporządzenie Ministra Środowiska z dnia 28 grudnia 2010 r. w sprawie formularza raportu oraz sposobu jego wprowadzania do Krajowej bazy o emisjach gazów cieplarnianych i innych substancji),** Dz. U. Nr 3/2011, Poz. 4 [in Polish].

**[LAW2] The Environmental Protection Law Act (pl: Prawo Ochrony Środowiska),** Dz.U. Nr 62/2001, Poz. 627 with changes entered into force after 27th of April, 2001 [in Polish].

**[LAW3] The law act on system of management of emissions of GHGs and other pollutants (pl: Ustawa z dnia 17 lipca 2009 r. o systemie zarządzania emisjami gazów cieplarnianych i innych substancji),** Dz.U. Nr 130/2009, Poz. 1070 [in Polish].

LIKENS, GE, BUSO, DC & BUTLER, TJ 2005, Long-term relationships between $SO_2$ and $NO_X$ emissions and $SO_4^{2-}$ and $NO_3$ concentration in bulk deposition at the Hubbard Brook Experimental Forest, NH.

MYATT, GJ 2007, *Making sense of data (a practical guide to exploratory data analysis and data mining)*, Wiley.

RORABACHER, D 1991, 'Statistical treatment for rejection of deviant values: critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level', *Analytical Chemistry*, no. 63, pp. 139–146.