Pol. J. Environ. Stud. Vol. 29, No. 1 (2020), 853-860

DOI: *10.15244/pjoes/112620* ONLINE PUBLICATION DATE: 2019-12-04

Original Research

Outlier Identification of Concentrations of Pollutants in Environmental Data Using Modern Statistical Methods

Petr Veselík[1]*, Marie Sejkorová[2], Aleksander Nieoczym[3], Jacek Caban[4]

[1]University of Defence, Department of Quantitative Methods, Brno, Czech Republic
[2]University of Pardubice, Faculty of Transport Engineering, Pardubice, Czech Republic
[3]Lublin University of Technology, Faculty of Mechanical Engineering, Department of Machine Design and Mechatronics, Lublin, Poland
[4]University of Life Sciences in Lublin, Faculty of Production Engineering, Lublin, Poland

Received: 12 June 2018
Accepted: 8 September 2019

Abstract

The article is focused on identification of outlier measurements in environmental data which may significantly affect the future results of the analysis and interpretation of results. For this reason, their identification forms an integral part of data analysis. The aim of this article is to perform statistical analysis that automatically identifies segments of outlier measurements. The results were demonstrated on real concentration data. The methodological procedure was used to evaluate particulate matter of the PM_{10} fraction size from two monitoring stations located in Brno, Czech Republic.

Keywords: capability indices, control charts, particulate matter PM_{10}, kernel regression, outlier measurements

Introduction

Sustainable development is perhaps the most important idea of our present time. This concept anticipates major civilizational change on the ecological, social and economic level [1]. Despite the imposition of European Union limit values for particulate matter (PM), frequent exceedances of the limit for PM_{10} concentrations have been observed widely across Western Europe, particularly in Switzerland, Belgium, Germany, Italy, Norway and the Czech Republic [2]. Poland also suffers from particulate pollution, for example PM_{10} particulate matter concentration is a substantial and as yet unresolved problem in the area of Warsaw, and other big cities [3, 4]. Airborne particulates are significant conveyors of metals, some of which are toxic and typically exist naturally and excessively in the environment [5]. Recent studies on the health problems associated with exposure to PM with an aerodynamic diameter of 10 μm or less (PM_{10}) have identified a variety of health-related problems that include deterioration in lung function, chronic pulmonary disease, heart disease, and premature death, along with a rise in mortality [6]. Consequently, heavy metals in association with PM definitely affect the biological and physiological functioning of the human body [5].

*e-mail: petr.veselik@unob.cz

To assess the value of the operating conditions we used a number of approaches, due to the fact that in practice there are different circumstances allowing (or not) the application of certain methods [6]. Studies have made some valuable attempts at environmental efficiency research and laid a theoretical foundation for our research, but there is still room for improvement in this field [7]. In the context of automatic continuous monitoring of environmental measurements there sometimes occur outliers – observations that appear inconsistent with the rest of the data. The outliers may be caused by, e.g., measurement error, natural variability of analysed pollutants in the air or by the presence of a new factor affecting the observed variable. If the outlier is caused by measurement error, then it is an invalid value and must be removed from the dataset. In this case, the fact that the measurement significantly deviated from the rest is correct and is likely to represent an extreme phenomenon, and its exclusion from the dataset would result in the loss of significant information.

Understanding the behaviour of outliers plays an important role in the quality of measurement and in the research of air pollution. Therefore, there is a growing requirement for validation of measured data due to the presence of outliers in the environmental area. In recently published articles relatively little attention has been paid to methods for identifying outliers that do not require prior knowledge of the distribution of the analysed variable. However, measurements from environmental areas are often dependent on accompanying climatic factors and therefore it is very difficult to estimate the distribution from which these data come.

In the problematic identification of outliers, most attention is given to mainly practical tasks in the field of economics and hydrology. The traditional areas are recently joined by applications in other fields. From currently addressed issues we can mention, e.g., analysis of large fires [8], forecast of precipitation extremes [9] and also air pollution [10-12]. An overview of the methods for outlier detection on temporal data is given in [13], outlier detection techniques for time series are described in, for example, [14] or [15].

Material and Methods

The performed analysis is based on hourly PM_{10} concentration measurements. PM_{10} mass concentrations were automatically measured at two monitoring stations (namely Arboretum and Zvonařka) in the city of Brno, Czech Republic. These locations are characterized by increased air pollution. The data were provided by the Council of the City of Brno and the time period of monitoring was from 1 November 2007 until 16 November 2015 from Zvonařka station, and from 1 November 2006 until 16 November 2015 from Arboretum station.

Monitoring station Zvonařka is situated on Opuštěná street in an area with heavy traffic. North of the station there is the Vaňkovka Gallery Shopping Centre. South and east of the station there is a parking lot and the Zvonařka bus station. Nearby is the train station and also many heavy-traffic roads. The station operates in continuous full automatic operation and is classified as traffic.

Arboretum is a station located in the Arboretum of Mendel University of Agriculture and Forestry oriented toward the building complex at Třída Generála Píky 3. The station operates in continuous full automatic operation and is the only representative station of the statutory city of Brno, serving as a source of data for announcements concerning the smog situation and PM_{10} concentrations in the air.

The presented analysis is based on the method described in the article [16] and consists of two steps. In the first step the original data are smoothed by using kernel regression with local bandwidth [17]. Smoothing the data was carried out using lokerns function available in the *R* program library [18]. Estimate of regression function based on local smoothing bandwidth compensates the absence of accompanying variables such as speed and wind direction, which affect the measured PM_{10} concentrations. Therefore, it is not possible to work with the original data and it is assumed, that newly obtained residuals are not influenced by unknown accompanying variables. In the second step automatic identification of outlying segments of residuals is subsequently performed using control charts [19] and Six sigma methodology [20, 21].

Control charts

The principle of control charts is based on graphical illustration of the changes of residual process in time and assessment of its statistical stability. In charts partial characteristics calculated from *n*-size subgroups are plotted. Although the choice of the parameter *n* is usually based on the measurement time interval, in case hourly PM_{10} concentrations, the best results were obtained using subgroups of small sizes ($n = 2, 3$). Control charts visualise characteristics from disjoint subgroups, and two horizontal lines – upper control limit (UCL) and lower control limit (LCL). These visualisation allows identify segments where the residual process is out of statistical control.

To detect segments of observations where the outliers responsible for the change in expectation and variance of the residuals occur $\bar{x}$ and *R* charts constructed from characteristics from disjoint subgroups and graphically representing sample means $\bar{x}_1, ..., \bar{x}_k$ and sample ranges $R_1, ..., R_k$ are used, respectively. The control limits based on Chebyshev´s inequality states that independently of the distribution type of the residuals maximally $(1/L^2)\%$ of sample means $\bar{x}_1, ..., \bar{x}_k$ falls outside the limits $\mu_{\bar{X}} \pm L\sigma\mu_{\bar{X}}$, where $\mu_{\bar{X}} = \mu$ denote expectation and $\sigma_{\bar{X}}$ standard deviation of $\bar{x}_1, ..., \bar{x}_k$.

For the normal distribution the probability that the sample mean falls within 3-sigma limit is 99,7% and the limits μ ± 4σ, μ ± 5σ, μ ± 6σ then 99,994%, 99,99994%, 99,99999%. Therefore the control limits of $\overline{x}$ chart are defined as $\mu_{\overline{X}} \pm L\sigma\mu_{\overline{X}}$. The parameter $\sigma_{\overline{X}}$ is standard deviation of the sample mean, i.e. $\sigma_{\overline{X}} = \sigma/\sqrt{n}$, the $\overline{x}$ chart control limits can be derived in the form [16]:

$$LCL = \overline{\overline{x}} - L\frac{\widehat{\sigma}_s}{\sqrt{n}}, \quad UCL = \overline{\overline{x}} + L\frac{\widehat{\sigma}_s}{\sqrt{n}}. \tag{1}$$

More information on the construction of the control limits of *R* and *s* charts can be found in [16, 19].

Six Sigma

Six sigma methodology was used to divide the residual process into groups of size 24 (each group corresponds to the daily time interval) and further each group was divided into subgroups of size $n = 3$. Upper specification limit (USL) and lower specification limit (LSL) were set using $\overline{x}$ chart with preselected parameters *n* and *L* based on empirical experience with historical data.

The assessment of statistical capability of the residual process was done using capability index C_p given by:

$$C_p = \frac{USL - LSL}{6\sigma}, \tag{2}$$

...where: σ - value of the standard deviation of the residual process.

If $C_p > 1.33$, the residual process is considered to be highly capable, when $1 \leq C_p \leq 1.33$, it is moderately capable, and in case $C_p \leq 1$, it is statistically incapable.

For further analysis the estimate of capability index C_p given by:

$$\hat{C}_p = \frac{USL - LSL}{\widehat{\sigma}_s}, \tag{3}$$

...where: $\widehat{\sigma}_s$ given by: $\widehat{\sigma}_s = \frac{\overline{s}}{c_4(n)}$.

Based on the obtained estimate $\hat{C}_p$ of capability index C_p was then completed with 100(1–*α*)% confidence interval for C_p given by [22]. One-sided 100(1–*α*)% confidence interval for C_p with left-sided interval estimate $\hat{C}_{p,L}$ and right-sided interval estimate $\hat{C}_{p,U}$ are obtained analogous given by [16].

Since statistical incapability is indicative of outliers present in the residual process the point and interval estimates of capability index can be used for direct detection of segments, when the outlier residuals occur. Considering the significance level *α*, for testing the null hypothesis $C_p = 1.33$ against the alternative hypothesis $C_p < 1.33$ the right-sided confidence interval (-∞; $\hat{C}_{p,U}$) is used [23]. Rejecting the null hypothesis indicates the presence of outliers in the corresponding time intervals.

Results and Discussion

The analysis of the datasets of PM_{10} concentrations on which the statistical analysis was performed were measured in 2015. The monthly means of PM_{10} concentration were 28,29 μg m^{-3} from 1 January 2015 until 16 November 2015 for Arboretum station. The lowest (0,40 μg m^{-3}) and the highest (167,20 μg m^{-3}) concentrations were both measured in November (Table 1). The monthly means of PM_{10} concentration were 33,04 μg m^{-3} from 1st January 2015 until 16th November 2015 for Zvonařka station. In this period, the lowest monthly concentration (1,20 μg m^{-3}) was measured in January, while the highest concentration (266,50 μg m^{-3}) was in October (Table 1).

For Arboretum station we focus on concentrations measured in January 2015. The results obtained via the methods described in the Material and Methods section are shown in Fig. 1. In Fig. 1a) and b) one can see the PM_{10} concentrations. Outliers obtained using $\overline{x}$ and *R* charts are denoted by red colour. The values of entrance parameters *n* and *L* are given in the description of Fig. 1. Point and right-sided interval estimates (for α = 0,05) of capability index C_p is visualised in a logarithmic scale together with the horizontal line at level 1.33 in Fig. 1c).

In Fig. 1 several identified outliers based on control charts are shown. Using $\overline{x}$ chart outliers on 13st January were detected using *R* chart outliers on 1st and 13st January. Measurements that were marked as outliers using *R* chart consist of similar values as those identified by the $\overline{x}$ chart and several others, especially in the period of 1st January (Fig. 1a and b). From Fig. 1c) we can see that by using the Six sigma-based method outliers on 1st and 13st January were identified. In these days the null hypothesis ($C_p = 1.33$) was rejected and alternative hypothesis ($C_p < 1.33$) was accepted at a significance level of 5%.

Fig. 2 shows PM_{10} concentrations corresponding to sample means that exceed the limits of $\overline{x}$ chart ($n = 3$) with different choices of parameter *L* for Arboretum station in September 2015 are highlighted by red colour. From Fig. 2a) it is clear that by using control limits suggested by Shewhart (constructed with parameter ($L = 3$), a large number of outliers was identified. Fig. 2(b-d) identified outliers using control limits with values of parameter $L = 4$, 5 and 6 highlighted by red colour. It follows from the above, that the choice of parameter *L* is a major problem in the construction of control limits and its choice plays a key role in identifying outliers. Although the usual setting of parameter is $L = 3$, for hourly PM_{10} concentration measurements better results were obtained using $L = 5$ and $L = 6$. Parameter *L* must be chosen based on the character of the data. For different types of data, a different value of parameter *L* may be optimal. If the distribution of the data is known, the choice of parameter *L* can be based on the quantiles of this distribution.

Table 1. Descriptive statistics of the measured PM$_{10}$ concentrations in ng m^{-3} from Brno Arboretum and Brno Zvonařka stations in 2015.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Arboretum station											
Mean	28.05	37.57	34.86	22.55	18.76	20.7	20.57	31.7	18.14	29.14	49.15
SD	21.31	18.7	19.93	12.95	7.87	10.17	10.01	14.37	9.88	13.24	38.86
Median	22.6	36.5	32.35	20.9	17.8	18.8	19.6	29.9	16.6	28.1	40.5
Min	1.6	2.4	1.4	1.4	3.7	2.7	2.5	1.2	1.5	3.2	0.4
Max	143.9	112.3	96.2	88.9	50.4	53.1	55.1	81.8	59.5	73.7	167.2
Zvonařka station											
Mean	32.23	44.45	39.47	25.13	20.05	18.85	25.93	29.71	24.18	41.93	61.5
SD	23.49	22.99	23.39	14.77	9.35	9.87	18.75	14.34	16.96	22.66	45.96
Median	28.2	42.35	36.7	22.4	17.4	17.35	21.65	29.7	20.3	37.8	53.8
Min	1.2	1.9	1.5	1.7	4.3	3.8	3.5	2.5	2.8	10.7	4.1
Max	153.3	154.4	168.3	89.6	55.7	74.5	124.5	79.5	221.9	266.5	197.1

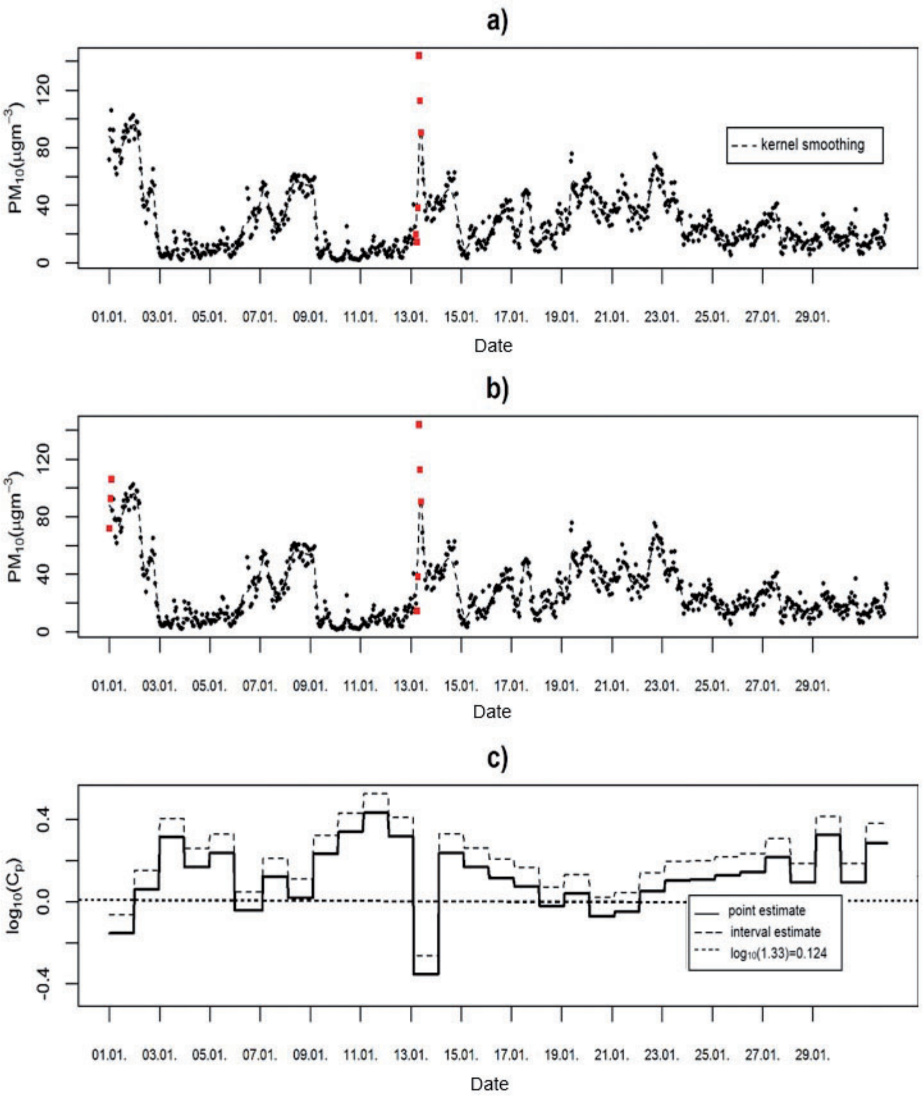

Fig. 1. Arboretum (January 2015): a) segments detected based on $\bar{x}$ chart ($n = 3$, $L = 5$), b) segments detected based on R chart ($n = 3$, $L = 6$), c) point estimate $\hat{C}_p$ and right-sided interval estimate $\hat{C}_{p,U}$ for capability index, horizontal line at the level $\log_{10}(1.33) = 0.124$.

For the sake of clarity, the different choice of parameter L was also applied to the concentrations measured in September for Zvonařka station. From Fig. 3 we can see outliers identified using $\bar{x}$ chart ($n = 3$) highlighted by red colour; Fig. 3c) and d) confirm that better results were obtained by setting $L = 5, 6$.

The points highlighted by red colour indicate the measurement segments that need attention. The user of the measurement process can choose from "suspicious" values to those that are apparent outliers. This is a particular problem according to various measured pollutants in the air and type of measuring instrument.

From Figs 2 and 3 it is clear that some measurements that visually appear as outliers were not identified. The true reason for the presence of outliers in the dataset cannot be specified automatically. Therefore, automatic identification must be followed by manual validation in order to reduce the number of outliers. It is also appropriate to add information about the measuring process and any unusual events happening at the monitoring stations to the instrument logs of the relevant analysers.

Detection of outliers is a problem that is constantly discussed in technical literature, and many methods have been proposed for its solution. Methods and algorithms for detection of outlier measurements in environmental data can be found in [24, 25]. Methods for detecting outliers in spatial environmental data were proposed in [26, 27].

This article proposes a method for the automatic detection of segments with outliers based on the smoothing of the original data and the subsequent analysis of the residuals which are no longer affected by the accompanying variables. The smoothing techniques used in chemical applications are based, for example, on moving averages or the Savitzky-Golay method [28], kernel regression with the global width of the smoothing window [29] or regression splines [30]. Another approach to smoothing is based on robust methods [31]. Here the smoothing was performed using non-parametric kernel regression, which estimates the regression function at each point as the weighted average of the surrounding observations, and by which use it is possible to eliminate the trend from the original data.

The amount of observations used for averaging is determined by a parameter called bandwidth, which can be determined locally or globally. Many methods and algorithms have been proposed for its optimal choice, has a significant effect on the resulting estimate of the regression function. An overview and comparison of

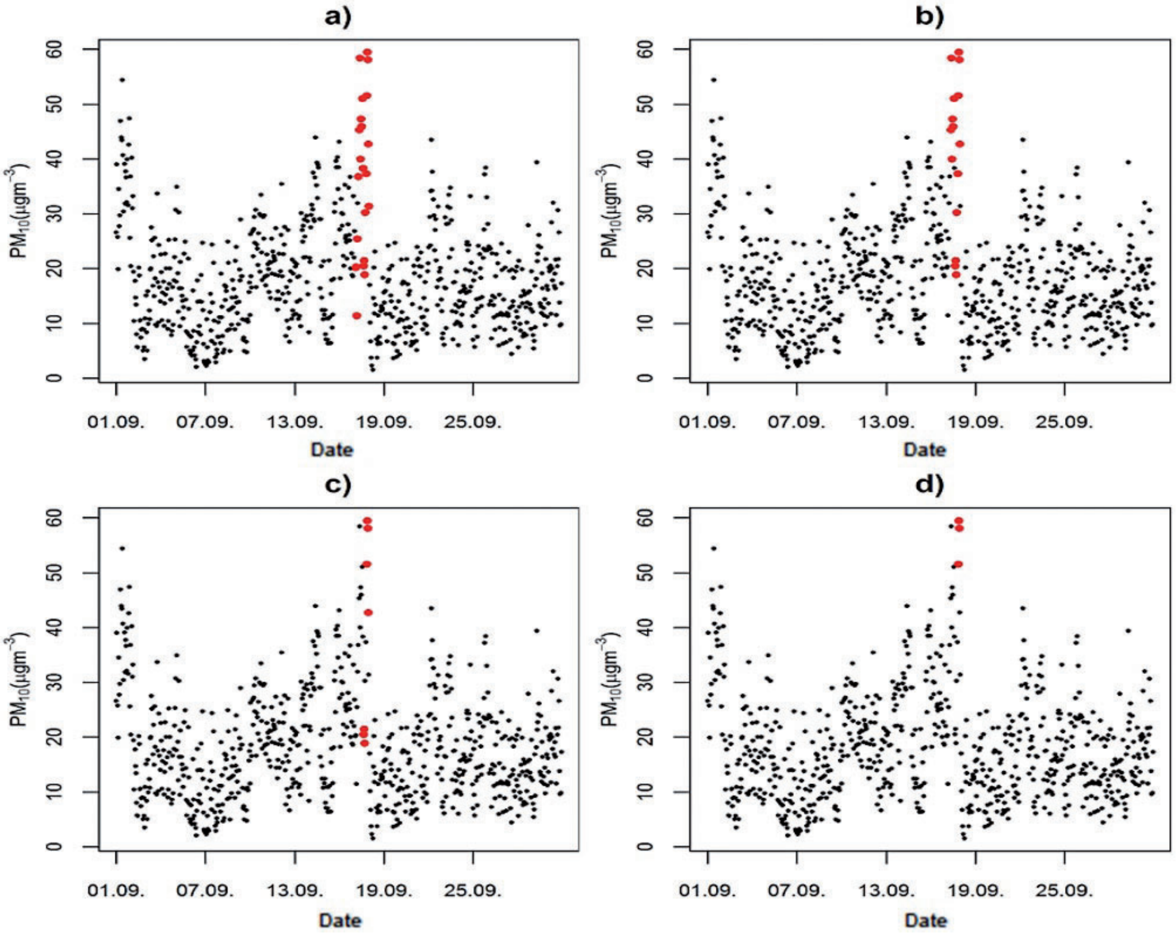

Fig. 2. Arboretum (September 2015): segments of outliers detected based on $\bar{x}$ chart ($n = 3$) using different values of parameter L: a) $L = 3$, b) $L = 4$, c) $L = 5$, d) $L = 6$.

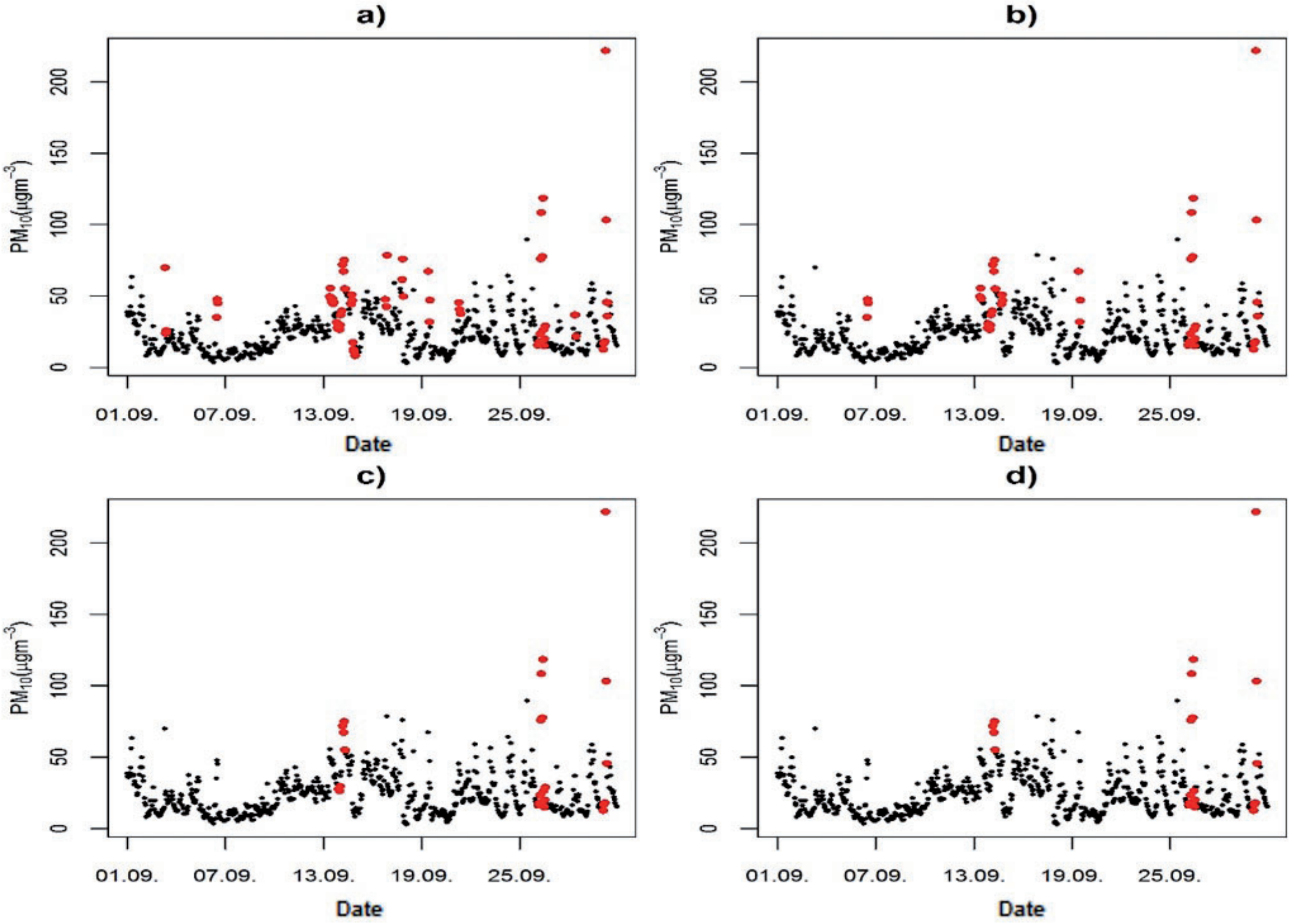

Fig. 3. Zvonařka (September 2015): segments of outliers detected based on $\bar{x}$ chart (n = 3) using different values of parameter L: a) L = 3, b) L = 4, c) L = 5, d) L = 6.

some of these methods is provided in [32]. Since the regression curve based on the local bandwidth is more adapted to the local data structure, the proposed method for detecting outliers is based on a kernel regression with local bandwidth estimated using a local plug-in algorithm.

Since the outlier appears in the original observation as an outlier value in the corresponding residual, the kernel regression is followed by residual analysis. With the use of control charts and the Six sigma methodology, the segments where the residuals are present are subsequently detected.

More strict or milder criterion for the detection of outliers can be achieved by a different choice of parameters n and L. The value of these parameters affects the amount of detected outliers, due to fact that when using control charts the outliers correspond to the respective sample means or sample ranges that are exceeded and also because the Six sigma-based method automatically denotes segments of size 24 (1-day intervals), the parameters n and L have to be chosen with respect to the number of observations. For this reason and also because we focus on outliers exceeding 3σ, smaller values of n (n = 2, 3, 4) are recommended. The parameter L must be chosen with regards to the character of the data, because each monitoring station is site specific. Therefore, the value of parameter L was determined based on the analysis of historical data that has already been manually validated.

In the case of setting the value of parameter L in such a way that the obtained residuals deviated less than 3 sigma from other residuals it was after subsequent manual control that it identified too many observations. Therefore, the Six sigma-based method cannot be used to detect outliers that deviate less than three standard deviations from the mean value of the neighbouring observations.

Conclusions

The automatic detection of segments in environmental data, where the outliers occur, has been performed in the article. The segments, where the outliers occur, were detected by using control charts and Six sigma methodology. The results of the analysis were demonstrated on real concentration data, which has been obtained by measuring PM$_{10}$ concentrations from two monitoring stations located in Brno (Czech Republic).

It is advisable that the detected segments of outlier measurements are further evaluated and interpreted by a specialized operator who assesses their quality. The main benefit of the presented analysis is to reduce the number of observations that must be assessed manually. Although the analysis was applied on environmental data, the procedure is general and can be adapted to data from various areas (e.g. econometrics or areas of risk prediction). The statistical analysis was performed using software *R*.

Acknowledgements

This work was supported by the Ministry of Defence of the Czech Republic: project ROZVOLOG (DZRO K-109).

Conflict of Interest

The authors declare no conflict of interest.

References

1. PAWŁOWSKI A. The role of environmental engineering in introducing sustainable development. Ecol. Chem. Eng. S. **17** (3), 263, **2010**.
2. HARRISON R.M., STEDMAN J., DERWENT D. New directions: why are PM_{10} concentrations in Europe not falling? Atmos. Environ., **42**, 603, **2008**.
3. MAJEWSKI G., PRZEWOŹNICZUK W. Study of particulate matter pollution in Warsaw area. Pol. J. Environ. Stud. **18** (2), 293, **2009**.
4. ĆWIKLAK K., PASTUSZKA J.S., ROGULA-KOZŁOWSKA W. Influence of traffic on particulate matter polycyclic aromatic hydrocarbons in urban atmosphere of Zabrze, Poland. Pol. J. Environ. Stud., **18** (4), 579, **2009**.
5. ELHADI R.E., ABDULLAH A.M., ABDULLAH A.H., ASH'AARI Z.H., KURA N.U., D.Y. G., ADAMU A. Source identification of heavy metals in particulate matter (PM_{10}) in a Malaysian traffic area using multivariate techniques. Pol. J. Environ. Stud., **26** (6), 2523, **2017**.
6. MARCZUK A., CABAN J., SAVINYKH P., TURUBANOV N., ZYRYANOV D. Maintenance research of a horizontal ribbon mixer. Eksploatacja i Niezawodnosc – Maintenance and Reliability, **19**, (1), 121, **2017**.
7. MA D., HE F., LI G., CHEN L. Estimation and comparative analysis of environmental efficiency in China, with and without consideration of haze. Pol. J. Environ. Stud., **27** (1), 201, **2018**.
8. JIANG Y., ZHUANG Q. Extreme value analysis of wildfires in Canadian boreal forest ecosystems, Can. J. Forest Res., **41** (9), 1836, **2011**.
9. FUKUTOME S., LINIGER M.A., SÜVEGES M. Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in Switzerland. Theor. Appl. Climatol., **120** (3), 403, **2014**.
10. BOBBIA M., MISITI M., MISITI Y., POGGI J.M., PORTIER B. Spatial outlier detection in the PM_{10} monitoring network of Normandy (France). Atmos Pollut Res, **6** (3), 476, **2015**.
11. SHAADAN N., JEMAIN A.A., LATIF M.T., DENI, S.M. Anomaly detection and assessment of PM10 functional data at several locations in the Klang Valley, Malaysia. Atmos Pollut Res, **6** (2), 365, **2015**.
12. KOKALJ M., RIHTARIČ M., KREFT S. Commonly applied smoothing of IR spectra showed unappropriate for the identification of plant leaf samples. Chemometr. Intell. Lab., **108** (2), 154, **2011**.
13. GUPTA M., GAO J., AGGARWAL C.C., HAN J. Outlier detection for temporal data: A survey. IEEE T. Knowl. Data En., **26**, (9), 2250, **2014**.
14. FOX A.J. Outliers in Time Series. J. R. Stat. Soc. Series B, **34** (3), 350, **1972**.
15. BURMAN J., OTTO M. Outliers in time series. Statistical Research Division Report Series CENSUS/SRD/RR-88/14, Bureau of the Census, 44, **1988**.
16. ČAMPULOVÁ M., VESELÍK P., MICHÁLEK J. Control chart and Six sigma based algorithms for identification of outliers in experimental data, with an application to particulate matter PM10. Atmos Pollut Res, **8** (4), 700, **2017**.
17. ČAMPULOVÁ M., MICHÁLEK J. The Application of Kernel Regression with Local Bandwidth to Identification of Outlier Values in Environmental Data, Forum Statisticum Slovacum, **12** (1), 1, **2016**.
18. R CORE TEAM. A language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing. http://www.Rproject.org/ (20.10.2017). **2013**.
19. ČSN ISO 8258 - Shewhart control charts, Czech Standards Institute, 36, **1994**.
20. MICHÁLEK J. Capability and performance indices of manufacturing process, CRQ. Prague, **2009**.
21. MONTGOMERY D.C. Introduction to Statistical Quality Control, John Wiley & Sons. New York, **2009**.
22. VESELÍK P., DVORSKÁ A., MICHÁLEK J. Half a year of co- located gaseous elemental mercury measurements: investigation of temporal changes in measurement differences. Fresen. Environ. Bull., **26** (5), 3128, **2017**.
23. VESELÍK P. Testing of environmental process capability and performance on the example of measurement gaseous elemental mercury. Forum Statisticum Slovacum, **6**, 184, **2015**.
24. ČAMPULOVÁ M., MICHÁLEK J., MIKUŠKA P., BOKAL D. Nonparametric algorithm for identification of outliers in environmental data. J. Chemom., **32** (5), 17, **2018**.
25. HOLEŠOVSKÝ J., ČAMPULOVÁ M., MICHÁLEK J. Semiparametric outlier detection in nonstationary times series: Case study for atmospheric pollution in Brno, Czech Republic. Atmos Pollut Res, **9** (1), 27, **2018**.
26. MARTÍNEZ J., SAAVEDRA Á., GARCÍA-NIETO P.J., PIÑEIRO J.I., IGLESIAS C., TABOADA J., SANCHO J., PASTOR J. Air quality parameters outliers detection using functional data analysis in the Langreo urban area (Northern Spain). Appl. Math. Comput., **241**, 1, **2014**.
27. GUPTA M., GAO J., AGGARWAL C. Outlier detection for temporal data: A survey. IEEE Trans Knowl Data Eng, **26** (9), 2250, **2014**.
28. KANEKO H., FUNATSU K. Smoothing-combined soft sensors for noise reduction and improvement of predictive ability. Ind Eng Chem Res, **54** (50), 12630, **2015**.
29. HENRY R., NORRIS G., VEDANTHAM R., TURNER R. Source region identification using kernel smoothing. Environ. Sci. Technol., **43** (11), 4090, **2009**.

30. HE S., FANG S., LIU X., ZHANG W., XIE W., ZHANG H., WEI D., FU W., PEI D. Investigation of a genetic algorithm based cubic spline smoothing for baseline correction of raman spectra. Chemometr Intell Lab Syst, **152**, 1, **2016**.

31. BOROWSKI M., FRIED R. Online signal extraction by robust regression in moving windows with data-adaptive width selection. J Stat Comput Simul, **24** (4), 597, **2014**.

32. KÖHLER M., SCHINDLER A., SPERLICH S. A review and comparison of bandwidth selection methods for kernel regression. Int Stat Rev, **82** (2), 243, **2014**.