*Pol. J. Environ. Stud. Vol. 28, No. 1 (2019), 197-203*
DOI: *10.15244/pjoes/84837* ONLINE PUBLICATION DATE: 2018-07-31

*Original Research*

# Calculating Environmental Background Value: A Comparative Study of Statistical Versus Spatial Analyses

**Sun Linhua***

School of Resources and Civil Engineering, Suzhou University, Anhui, China
National Engineering Research Center of Coal Mine Water Hazard Control, Anhui, China

*Received: 28 November 2017*
*Accepted: 27 January 2018*

## Abstract

Local environmental background is important for environmental management. In this study, lead concentrations of shallow groundwater samples from the urban area of Suzhou in Anhui Province, China were measured and analyzed by statistical and spatial analyses for calculating the environmental background value. The results show that the lead concentrations in the groundwater range from 4.16-11.5 μg/L, and all of the samples were classified to be Class III or better according to the groundwater quality standard of China. The samples have medium coefficient of variation and low p-values of normal distribution test, suggesting that it may have been influenced by anthropogenic activities, which was further demonstrated by the consistency of the distribution of the samples with high lead concentrations and the areas with high density of transportation, as well as the high-low cluster of the spatial autocorrelation analysis. The environmental background values have been calculated to be 3.74-8.62 and 3.48-10.3 μg/L with box plot and spatial autocorrelation analyses, respectively. The study demonstrated that for calculating the environmental background value, the statistical and spatial methods should be chosen according to the current state – especially pre-consideration about the distribution of the elements or pollutants.

**Keywords**: environmental background, heavy metals, groundwater, statistical analysis, spatial autocorrelation analysis

## Introduction

Heavy metals, which are defined as metals with relatively high densities, atomic weights, or atomic numbers (e.g., copper, zinc, lead, cobalt, iron, etc.), have attracted large number of studies because of their toxicity and worldwide distribution [1]. Among these metals, chromium, cadmium, mercury, and lead have the greatest potential to cause harm on account of their extensive use, the toxicity of some of their combined or elemental forms, and their widespread distribution in the environment [2].

Pollution of groundwater in the current world is a major concern of the government and environmental scientists, because of the importance of the groundwater

---

*e-mail: sunlinh@126.com

for the survival and development of human society [3]. And therefore, monitoring and assessing groundwater pollution is becoming more and more necessary in the current world. Moreover, because of their special characteristics (e.g., toxicity, persistent nature, non-biodegradability, and the ability to bio-accumulate in the food chain [4]), the pollution of groundwater by heavy metals has long been a concern of scientists, and the research includes monitoring and assessing, source identification, approximation, and remediation [5-8].

Monitoring concentrations of heavy metals in groundwater is essential for maintaining groundwater quality as doing so can provide basic information for contamination management. Therefore, many studies have tried to determine the extent of contamination by comparison with the background concentrations or the universal standards (e.g., the water quality standard of the World Health Organization and the National Groundwater Quality Standard of China (GB/T 14848-93)) [9-12]. However, if the purpose of the study is to identify the anthropogenic contribution, the use of universal background is inappropriate, because different areas have different environmental backgrounds: e.g., the groundwater in the area with Pb-Zn mineral deposits must have higher Pb and Zn concentrations relative to the groundwater in other areas.

Northern China is an area with a serious water shortage, and most of the water used for industrial, agricultural, and domestic purposes in the area is obtained from underground. To be a typical city in northern China, the city of Suzhou has undergone a long period of groundwater utilization [13-15]. However, the local environmental background of the groundwater is still lacking, which restricts not only the groundwater quality monitoring, but also the identification of the anthropogenic influence on the groundwater environment.

Taking into account the lack of the environmental background value and the importance of lead in environmental research, as well as the importance of groundwater in the city, a total of 62 groundwater samples from shallow wells in the urban area have been collected, and two kinds of methods (including statistical and spatial) have been applied to their lead concentrations for calculating the environmental background value. The study can provide some new information about establishing the local environmental background value relative to previous studies.

## Materials and Methods

### Study Area

Suzhou is a city located in northern Anhui Province in China. It is located south of the Huang-Huai plain, with annual precipitation of 857 mm, and an average temperature of 14.4 degrees (centigrade). There are many rivers in the area, including the Kui, Sui, Tuo, and Hui. However, groundwater is the main water source for industrial production, agricultural activities, and urban living.

### Sampling and Analysis

A total of 62 shallow groundwater samples were collected from shallow wells (<30 m) in the urban area of the city, and all of the wells are used for living purposes (Fig. 1). The sampling procedures were as follows:

1) Sampling. Before sampling, the polyethylene bottle cleaned in the laboratory was rinsed three times in water, and then the samples were acidified to be pH<2 by $HNO_3$ to prevent the adsorption of the elements by

Fig. 1. Locations of groundwater samples.

the bottle, and finally labeled and sent to the laboratory for processing. Meanwhile, the GPS location and site conditions (including the living conditions, flow of people, transportation, and garbage, etc.) were recorded.

2) Laboratory. In the laboratory, all of the samples were filtered by 0.22 μm film before analysis for removing debris. The concentrations of lead were analyzed by atomic absorption spectrometry, and quality control was carried out by standard sample (the correlation coefficient between actual and measured concentrations was higher than 0.99). All of the analyses were conducted in the Engineering Research Center of Coal Mine Exploration, Anhui province, China.

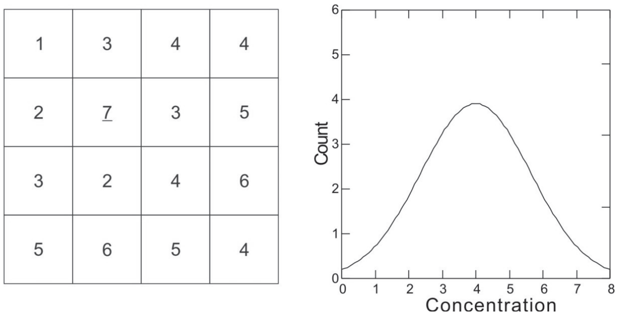
### Data Treatment

All of the data were first processed for statistical analysis using Mystat 12 software, and the minimum, maximum, mean, standard deviation, coefficient of variation, and p-value of the normal distribution test were obtained. Then the contour map of the lead concentrations was plotted by Surfer 11 software (with the natural neighbor grid method). Finally, the spatial autocorrelation analysis was processed by Geoda 1.8.3 software for obtaining the significant map and the spatial clustering map.

### Concepts and Methods

Previous studies revealed that the environmental background value is the concentrations or range of concentrations of the elements in the relatively clean area (with low or no contribution from anthropogenic activities), and is the basis for determining the pollution degree of the regional environment [16-17].

Currently, there are two kinds of methods for determining environmental background values, including the direct (geochemical) and indirect (statistical) methods [18]. Among these two kinds of methods, the direct method is often criticized as having subjective sample selection criteria, high costs, and heavy laboratory workload, and therefore the application of the direct method is limited. Comparatively, the indirect (statistical) methods have been used more frequently, which are used not only for assessing background concentrations, but also for the separation of geochemical anomalies from geochemical background. The methods include regression analysis always applied for soil environmental background calculation [19], the fractal method (C-A and S-A) for identification of anomaly during mineral exploration [20], and the probability plots applied for environmental background calculation of the nitrate in the groundwater [21] et al.



| | | | |
|---|---|---|---|
| 1 | 3 | 4 | 4 |
| 2 | 7 | 3 | 5 |
| 3 | 2 | 4 | 6 |
| 5 | 6 | 5 | 4 |

Fig. 2. Distribution of an element's concentrations (the number in the grid).

### Consideration in this Study

Although some of the studies prefer to use normal or lognormal distribution as the basis for calculating the environmental background value [16-17, 21], it has long been demonstrated to be not true [22]. Such information can also be achieved from Fig. 2. As can be seen from the figure, the numbers of the samples with concentrations 1-7 are 1, 2, 3, 4, 3, 2, and 1, respectively. Obviously, these concentrations are normally distributed. However, this does not mean that all of these samples are environmental background samples. Because concentration No. 7 is obviously different from its nearby samples with concentrations ranging 1-4, it can be considered an anomaly that should be removed from the background samples [20].

In this study, two kinds of methods were applied for determining the environment background value:

1) Traditionally: the box plot by Mystat 12, based on the assumption that the environmental background value is in line with normal distribution. After processing, the samples outside the lower and upper hinges of the box plot were removed (repeated until no abnormal samples), and then we calculated the mean and standard deviation of the rest of the samples. The environmental background value was then calculated to be mean ±2*standard deviation.
2) Spatially: based on the spatial autocorrelation analysis, the basis is that there is no significant change of the concentrations in the sample relative to its nearby ones, or namely "no mutation." Therefore, only the "non-significant" samples (not belonging to the high-high, low-low, high-low, and low-high clusters) [23] after spatial autocorrelation analysis were considered to be environmental background samples, and the environmental background value was also calculated by the mean ±2*standard deviation of them.

## Results and Discussion

### Concentrations

The National Groundwater Quality Standards of China (GB/T 14848-93) classify groundwater quality into five degrees (from good to bad) with lead concentrations: Class I (≤0.005 mg/L) and Class II (≤0.01 mg/L), which

Table 1. Statistical description of lead in this study (μg/L).

| | N | Min | Max | Mean | SD | CV | p-value |
|---|---|---|---|---|---|---|---|
| Pb | 62 | 4.16 | 11.5 | 6.58 | 1.67 | 0.253 | <0.01 |

represent natural background; Class III (≤0.05 mg/L), suitable for drinking, industrial, and agricultural use; Class IV (≤0.10 mg/L), suitable for industrial and agricultural use (can also be used for domestic purposes after treatment); and Class V (>0.10 mg/L), not suitable for any purpose. In this study, the lead concentrations of the groundwater are 4.16-11.5 μg/L (mean = 6.58 μg/L; Table 1). In comparison with the above standard, 8, 51, and 3 samples were classified as Class I, II, and III, which suggests that the groundwater in the study area is of good quality when considering only their lead concentrations.

Coefficient of variation (CV = standard deviation / mean) is an index that can be used for identifying the anthropogenic contribution degree for pollution in environmental studies [24-25], previous studies revealed that when CV<0.10 and >0.90 mean low and high anthropogenic contributions, respectively. In this study, the CV of the lead concentrations of the groundwater samples is 0.253, which indicates that the groundwater has probably been influenced by anthropogenic activities. Moreover, the p-value of the normal distribution test is <0.01, implying that the lead concentrations in this study cannot pass the normal distribution test (p>0.05), which also suggests the anthropogenic contribution [16-17].

## Spatial Distribution

Contour maps have long been used for environmental studies because of the visualization of pollution [26]. As can be seen from the contour map of lead concentrations in Fig. 3, two areas with high lead concentrations can be found in the west and east-central parts of the study area. In comparison with the sample locations in Fig. 1, it can be found that the east-central area is located between the train station, bus stations, and the business area. Our observation during sampling confirmed that this area is characterized by a high density of people, business, and transportation. And therefore, the high lead in this area might be contributed to by natural and anthropogenic sources simultaneously, because except for the weathering of lead-bearing minerals (e.g., galena), the burning of gasoline is the main source for releasing lead into the urban environment [24].

## Spatial Autocorrelation Analysis

According to the classification of Moran'I index in the local spatial autocorrelation (LISA) [23], all of the samples can be subdivided into two major categories after calculation: "not significant" and "significant." Samples classified into the former are considered as no "mutation" relative to its nearby samples, and the samples classified into the latter can be divided into four sub-categories: high-high, low-low, low-high, and high-low, which represent the relationship between a sample and its surrounding ones. The first two are called "hot spot" and "freezing spot," respectively, which reflects the regional anomaly such as the surface pollution related to sewage irrigation, whereas the low-high and high-low samples are abnormal ones, which may be related to the influence of other factors (e.g., anthropogenic point pollution) [27-29]. The results of spatial autocorrelation analysis are shown in Fig. 4.

As can be seen from the figure, 41 samples are classified as "non-significant" samples, whereas the sample numbers classified as high-high, low-low, low-high, and high-low clusters are 2, 4, 0, and 15, respectively. It can also be noticed from the figure that the samples divided into the high-low cluster are mainly concentrated in the area near the railway station (north) and the bus station

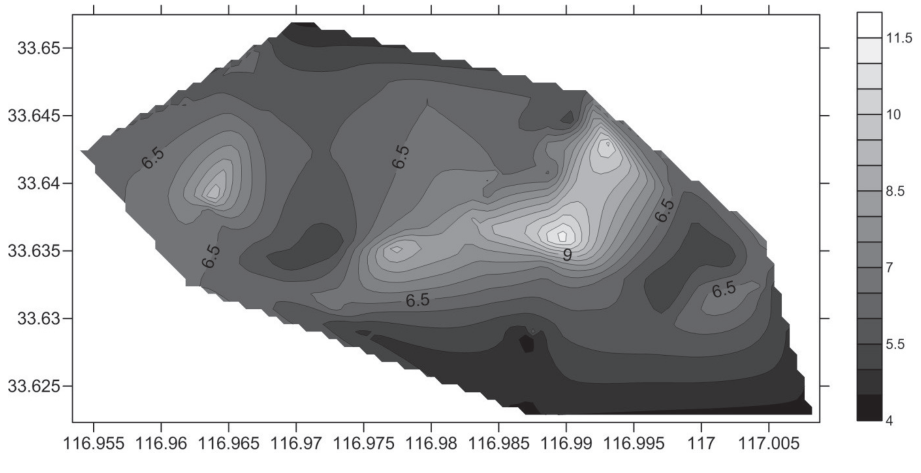


Fig. 3. Contour map of lead concentrations (μg/L).

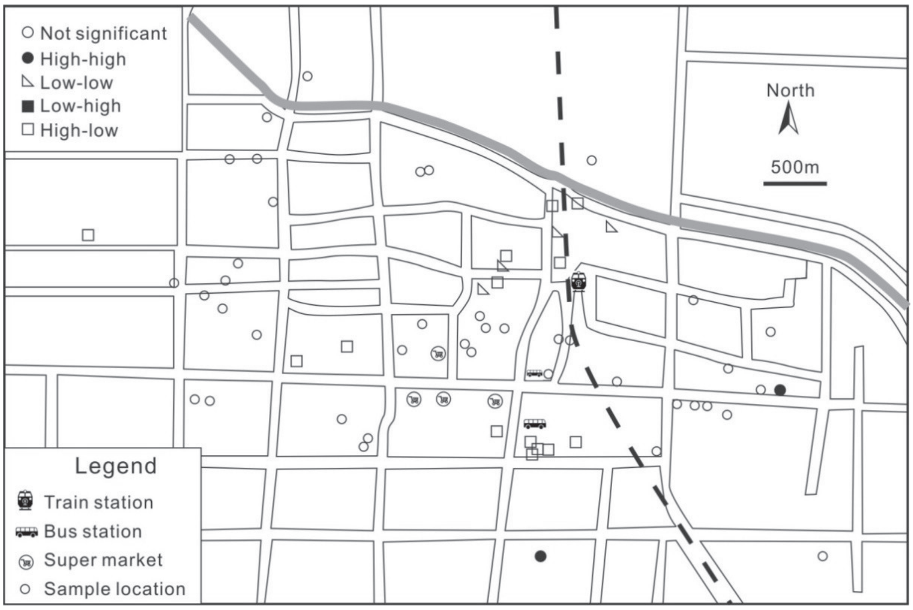


Fig. 4. Results of spatial autocorrelation analysis.

(south). This area is consistent with the distribution of the samples with high lead concentrations in Fig. 3, which imply that the release of lead from transportation is the main influencing factor.

## Environmental Background Value Calculation

Fig. 5 is the density plot of lead concentrations. It can be seen from the figure that the concentrations of lead have at least two peaks, which probably have two kinds of information: one is that the lead in the groundwater might have been influenced by anthropogenic activities (the second peak), and another is that the lead in the groundwater without anthropogenic influence might follow the normal distribution (the first peak). Moreover, based on the box plot of the lead concentrations, there are 4 samples with concentrations near or higher than 10 μg/L, which were identified as outliers (Pb-1 in Fig. 6), and another 2 samples can be identified as outliers with the second application of the box plot (Pb-2 in Fig. 6). After the removal of the six, the remaining 56 samples were calculated and the mean value was 6.18 μg/L (with standard deviation = 1.22 μg/L, Table 2). Therefore, the background value calculated by the statistical method was 3.74-8.62 μg/L.

As to the spatial autocorrelation analysis, the above-mentioned 41 samples belonging to the "non-significant" category were calculated, and their mean value was 6.88 μg/L with the standard deviation of 1.70 μg/L (Table 2). Therefore, the background value calculated by the spatial autocorrelation method was 3.48-10.3 μg/L.

For comparison, the iterative standard deviation, distribution function method [30] and QQ plot [21] have also been applied for calculating the environmental background. The results show that the environmental background values of lead in the groundwater are 4.1-7.8 μg/L (iterative standard deviation) and 4.2-8.4 μg/L (distribution function), respectively. As can be seen from the QQ diagram

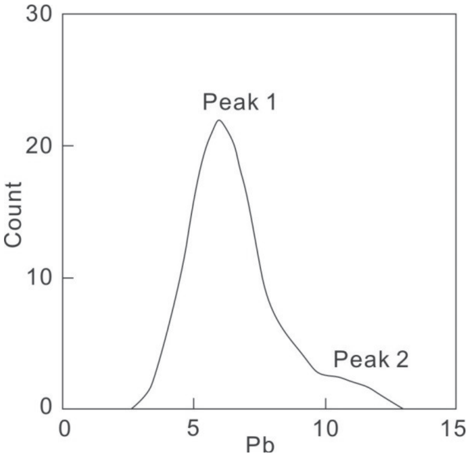


Fig. 5. Density plot of lead concentrations.

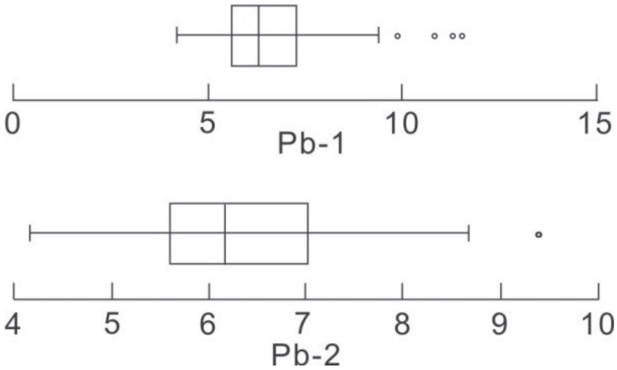


Fig. 6. Box plot of lead concentrations.

Table 2. Statistical description of lead samples after outlier removal (μg/L).

| | N | Min | Max | Mean | SD | CV | p-value | Background value |
|---|---|---|---|---|---|---|---|---|
| $Pb^1$ | 56 | 4.16 | 8.67 | 6.18 | 1.22 | 0.182 | >0.15 | 3.74-8.62 |
| $Pb^2$ | 41 | 4.16 | 11.5 | 6.88 | 1.70 | 0.247 | <0.01 | 3.48-10.3 |

Note: 1 and 2 are the samples after outlier removal based on the box plot and spatial autocorrelation analysis, respectively.

(Fig. 7), it is hard to find an inflection point, and therefore it is hard to determine the environmental background through the method.

According to the results obtained by the above methods, it can be seen that the calculated environmental background values are different from each other, and the main reason for this difference is their different basis: for the statistical methods (including box plot, QQ plot, and iterative standard deviation, distribution function), the basis is that the background samples follow the normal distribution and are calculated by the environmental background to be 3.74-8.62 μg/L, whereas the spatial analysis focused on the "non-mutation" feature (no significant spatial autocorrelation), the environmental background calculated by it is 3.48-10.3 μg/L.

During the application, if the distribution of the element/pollutant can be determined or demonstrated (most commonly normal or lognormal), the statistical method (or the method with same principle, such as the iterative standard deviation and the probability distribution plot [16-19, 21]) can be considered. However, if the distribution of data is uncertain, the method of spatial analysis should be chosen. If considering the lead in this study, it is hard to determine which distribution to follow, because it is hard to find a straight line in the QQ plot even before or after lognormal transformation (Table 1 and Fig. 6). And therefore, the spatial autocorrelation method should be chosen for calculating the environmental background value.

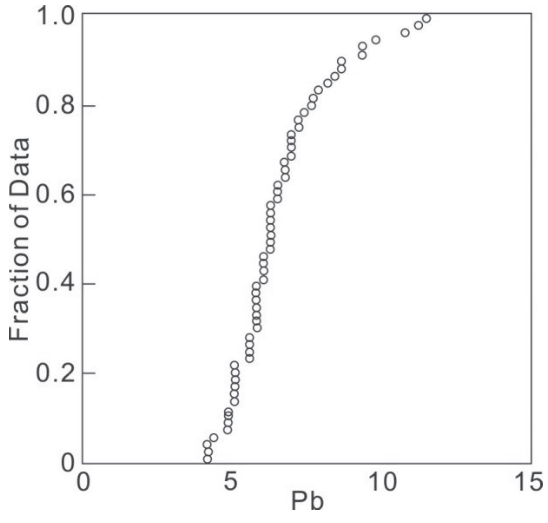


Fig. 7. QQ plot of lead concentrations.

## Conclusions

Based on the statistical and spatial analyses of the lead concentrations of the shallow groundwater in Suzhou, Anhui province, China, the following conclusions have been made:

1) All of the samples were classified as Class I, II, and III according to the groundwater quality standard of China, which means that the groundwater can be used for drinking according to their lead concentrations. The samples have medium coefficient of variation and low p-values of normal distribution test, implying that it may have been influenced by anthropogenic activities.
2) The spatial distribution of the lead concentration suggests that samples with high concentrations of lead are located in an area consistent with the high-low anomaly cluster identified by spatial autocorrelation analysis, indicating that the area has been influenced by anthropogenic activities, most probably transportation.
3) The environmental background values have been calculated to be 3.74-8.62 and 3.48-10.3 μg/L with box plot and spatial autocorrelation analyses, respectively.
4) For calculating the environmental background value, the statistical and spatial methods should be chosen according to the current state, especially the pre-consideration about the distribution of the elements or pollutants.

## Acknowledgements

Thanks to Feng Songbao and Cheng Chen at Suzhou University for their help in sampling and analysis. This work was financially supported by the Academic Funding for Top-talents in Disciplines of Universities in Anhui Province (gxbjZD48), the Foundation of Scholarship Leaders (Reserve) in Suzhou University (2018XJHB08) and the Key Research Program of Suzhou University (2016yzd08).

## Conflict of Interest

The authors declare no conflict of interest.

## References

1. AGARWAL S.K. Heavy metal pollution. APH publishing, **2009**.
2. BAIRD C., CANN M. Environmental chemistry. Macmillan Higher Education, **2012**.
3. SUN N.Z., SUN A. Mathematical modeling of groundwater pollution. Springer Science & Business Media, **2013**.
4. ACHPARAKI M., THESSALONIKEOS E., TSOUKALI H., MASTROGIANNI O., ZAGGELIDOU E., CHATZINIKOLAOU F., RAIKOS N. Heavy metals toxicity. Aristotle University Medical Journal, **39** (1), 29, **2012**.
5. MUHAMMAD S., SHAH M.T., KHAN S. Health risk assessment of heavy metals and their source apportionment in drinking water of Kohistan region, northern Pakistan. Microchemical Journal, **98** (2), 334, **2011**.
6. MOMODU M.A., ANYAKORA C.A. Heavy metal contamination of groundwater: the Surulere case study. Research Journal of Environmental Earth Sciences, **2** (1), 39, **2010**.
7. ZHANG X., WANG H., HE L., LU K., SARMAH A., LI J., HUANG H. Using biochar for remediation of soils contaminated with heavy metals and organic pollutants. Environmental Science and Pollution Research, **20** (12), 8472, **2013**.
8. HASHIM M.A., MUKHOPADHYAY S., SAHU J.N., SENGUPTA B. Remediation technologies for heavy metal contaminated groundwater. Journal of Environmental Management, **92** (10), 2355, **2011**.
9. SIRAJUDEEN J., VAHITH A.R. Applications of water quality index for groundwater quality assessment on Tamil Nadu and Pondicherry, India. Journal of Environmental Research and Development, **8** (3), 443, **2014**.
10. LONGE E.O., BALOGUN M.R. Groundwater quality assessment near a municipal landfill, Lagos, Nigeria. Research Journal of Applied Sciences, Engineering and Technology, **2** (1), 39, **2010**.
11. ZHANG B., SONG X., ZHANG Y., HAN D., TANG C., YU Y., MA Y. Hydrochemical characteristics and water quality assessment of surface water and groundwater in Songnen plain, Northeast China. Water Research, **46** (8), 2737, **2012**.
12. SUN L., GUI H., PENG W. Heavy metals in groundwater from the Wolonghu coal mine, northern Anhui Province, China and their hydrological implications. Water Practice and Technology, **9** (1), 79, **2014**.
13. LIU Q., LIANG H. The application of electrometric method and numerical simulation to evaluate groundwater resources in the Suzhou city, Anhui province. Chinese Journal of Engineering Geophysics, **11** (5), 688, **2014**.
14. LIN M.L., PENG W.H. Concentrations and health risk assessment of heavy metals in groundwater in rural areas of Suzhou City. Water Resources Protection, **32** (6), 114, **2016**.
15. YU B., ZHI Y.H. Analysis on the outcomes of drinking water monitoring in urban areas in Suzhou of Anhui province from 2015 to 2016. Journal of Tropical Diseases and Parasitology, **15** (2), 102, **2017**.
16. REIMANN C., FILZMOSER P., GARRETT R.G. Background and threshold: critical comparison of methods of determination. Science of the Total Environment, **346** (1), 1, **2005**.
17. REIMANN C., GARRETT R.G. Geochemical background-concept and reality. Science of the Total Environment, **350** (1), 12, **2005**.
18. GAŁUSZKA A., MIGASZEWSKI Z. Geochemical background-an environmental perspective. Mineralogia, **42** (1), 7, **2011**.
19. JIANG J., WANG J., LIU S., LIN C., HE M., LIU X. Background, baseline, normalization, and contamination of heavy metals in the Liao River Watershed sediments of China. Journal of Asian Earth Sciences, **73**, 87, **2013**.
20. GUILLÉN M.T., DELGADO J., ALBANESE S., NIETO J.M., LIMA A., DE VIVO B. Environmental geochemical mapping of Huelva municipality soils (SW Spain) as a tool to determine background and baseline values. Journal of Geochemical Exploration, **109** (1), 59, **2011**.
21. PANNO S.V., KELLY W.R., MARTINSEK A.T., HACKLEY K.C. Estimating background and threshold nitrate concentrations using probability graphs. Ground Water, **44** (5), 697, **2006**.
22. REIMANN C., FILZMOSER P. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. Environmental Geology, **39** (9), 1001, **2000**.
23. ANSELIN L. Local indicators of spatial association-LISA. Geographical Analysis, **27** (2), 93, **1995**.
24. SARKAR D., DATTA R., HANNIGAN R. Concepts and applications in environmental geochemistry (Vol. **5**). Elsevier, **2011**.
25. SUN L.H., CHENG B.X., CHEN S.H. Trace metal concentrations in the river water near the urban area: a case study in Suzhou, northern Anhui province, China. Fresenius Environmental Bulletin, **26** (6), 4017, **2017**.
26. ROHDE R.A., MULLER R.A. Air pollution in China: mapping of concentrations and sources. PloS one, **10** (8), e0135749, **2015**.
27. TANG R., MA K., ZHANG Y., MAO Q. The spatial characteristics and pollution levels of metals in urban street dust of Beijing, China. Applied Geochemistry, **35**, 88, **2013**.
28. MAAS S., SCHEIFLER R., BENSLAMA M., CRINI N., LUCOT E., BRAHMIA Z., GIRAUDOUX P. Spatial distribution of heavy metal concentrations in urban, suburban and agricultural soils in a Mediterranean city of Algeria. Environmental Pollution, **158** (6), 2294, **2010**.
29. LI W., XU B., SONG Q., LIU X., XU J., BROOKES P.C. The identification of ‘hotspots’ of heavy metal pollution in soil-rice systems at a regional scale in eastern China. Science of the Total Environment, **472**, 407, **2014**.
30. NAKIĆ Z., POSAVEC K., BAČANI A. A visual basic spreadsheet macro for geochemical background analysis. Ground Water, **45** (5), 642, **2007**.