GIS IN THE STUDIES OF SOIL AND WATER ENVIRONMENT

Rafał Rozpondek¹, Katarzyna Wancisiewicz¹, Małgorzata Kacprzak¹

¹ Faculty of Infrastructure and Environment, Czestochowa University of Technology, Dąbrowskiego 69 Str., 42-201 Czestochowa, Poland, e-mail: raf.rozpo@gmail.com, katarzynawancisiewicz@gmail.com, mkacprzak@is.pcz.czest.pl

ABSTRACT

GIS with the geostatistics are more frequently-used technique in environmental studies. Providing spatial information creates a possibility to place the results in space. Spatial information significantly affects the credibility and quality of research projects. It can be useful in determining certain relationships between monitoring and reclamation of degraded areas and to minimize and evaluate the costs related to environmental pollution. On the basis of practical and scientific examples, potential methods of applications of GIS and geostatistical methods in the study of water and soil environment was presented.

Keywords: Geographic Information System, geospatial analyze, kriging, land degradation, reclamation.

INTRODUCTION

The problem of soil contamination is a serious problem related to the developed industry, progressive urbanization, agriculture or mining activities. Due to a significant solubility of heavy metals compounds (eg. Pb, Cu, Zn) they may pose not only a threat to the top soils, but also for surface water and groundwater.

With the rapidly developing technology, it is possible to use modern computer software and geographic information system (GIS) as an additional help for activities related to monitoring and reclamation of degraded lands. GIS allows supplementing the conducted research with important information - spatial information. It allows placing the results in space, which can be beneficial both from the point of view of proper remediation planning, detecting sources of pollution or locating and detecting certain relationships between different phenomena (eg. between the rainfall intensity, slope and the level of erosion).

Environmental studies are often held in areas with significant surface. In order to interpret the results correctly, a useful tool seems to be GIS in conjunction with geostatistics, and special issue of prediction. It allows the design of maps of pollutants distribution for the whole studied area - it is a complex, multi-step process. It requires careful planning in every phase of activities, starting from the corresponding assumptions of measurement network, through adequate extraction and analysis of environmental samples and ending with the use of the correct interpolation method. There are many interpolation methods: as simple as nearest-neighbor interpolation (also known as proximal interpolation), average, maximum, minimum, and more complex like kriging methods. Notable kriging, beyond the expected value, also provides information about the accuracy of its designation. Such maps, in combination with the knowledge of environmental pollution, are based for planning of monitoring or reclamation of degraded areas. The obtained results can also be important for scientists.

Geographic Information System (GIS)

The main component of GIS is geographic information. It concerns the object or phenomenon, which has its spatial location. Additionally, it provides details on the location, shape, size and

nature of the specified object. Geographic Information System is used to manage, archive, enter, visualize and analyze geographic information.

GIS data can be divided into two main groups – descriptive data and spatial data. Spatial data relates to the location and shape of the object, while descriptive data contains information about what is in a particular area. Sets of information can come from many sources, it can be obtained, from [Drzewiecki, 2015]:

1. Laboratory analysis – samples of material taken from the site are tested in a laboratory, then the results are subjected to positioning in space. Example - Map of cadmium content in the soil (Figure 1)

2. Aerial photographs vectorisation – one of the important results can be a land use map (Figure 2). It arises as an outcome of the insertion of closed figures (areas) around specific land cover.

3. Satellite images – satellite images use mainly the dependence of reflectance of wavelength for specific land covers. As a result, it is possible to obtain some helpful features. The good example can be Normalized Differential Vegetation Index NDVI [Shisanya, 2011], which uses the dependence of reflectance by vegetation in the red and near-infrared channel (re-

flection of vegetation in the red channel tends to zero, and in near-infrared it is very strong). Hence, it is possible to determine not only the location of biomass, but also its quality. An important advantage of satellite images is their relatively easy availability – some satellite systems publish images on the internet web pages– they can be downloaded free of charge (e.g. The Landsat system).

4. Surveying – beyond the traditional use of geodetic measurements (staking out buildings, updating records of land and buildings etc.), it can be used as the main center for insertion results in space. By using geodetic instruments (eg. GPS transmitter), the user is able to determine the location of the sampling or restore previously planned measurement points in the field. This provides the ability to take samples from the same places at certain intervals of time [Wysocki, 2010].

5. Airborne laser scanning – the principle of the airborne laser scanning can be compared to a laser rangefinder, which emits a high frequency laser beam (Figure 3). Based on the time difference between the moment of sending and returning of the beam the distance between the scanner and terrain is determined. By using GPS-INS (GPS – Global Positioning System,
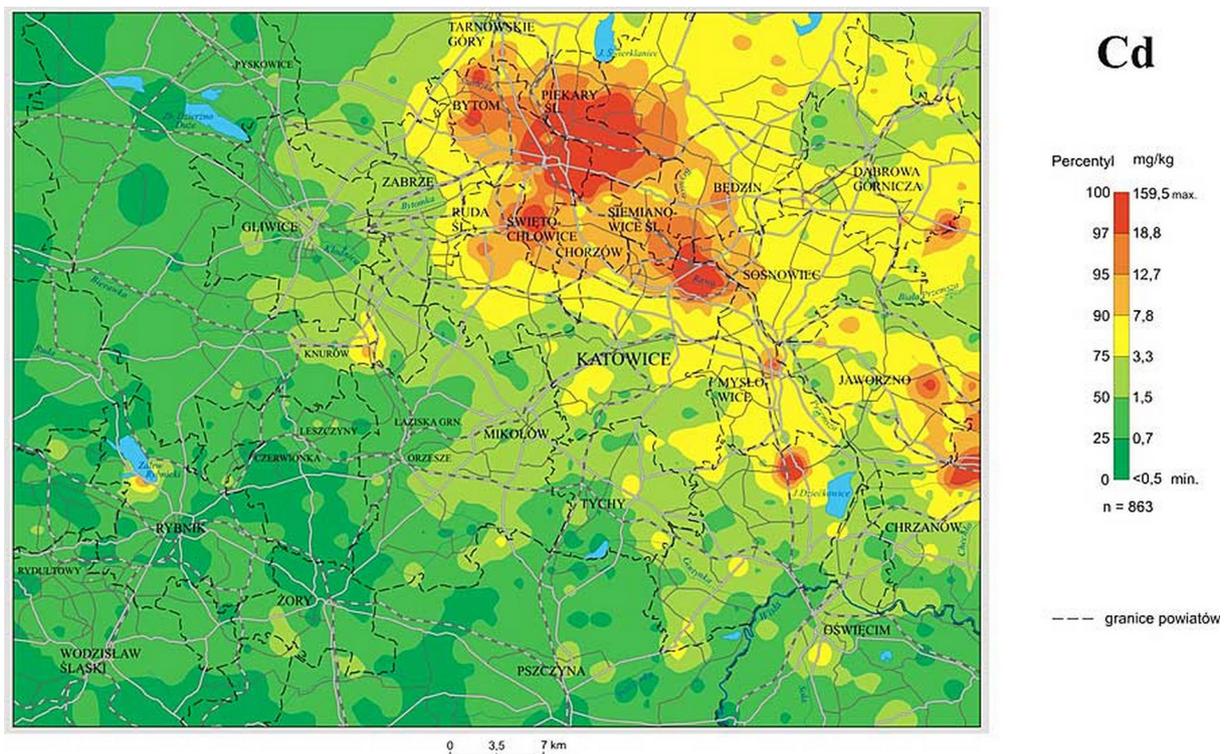

Figure 1. Cd content in the soil
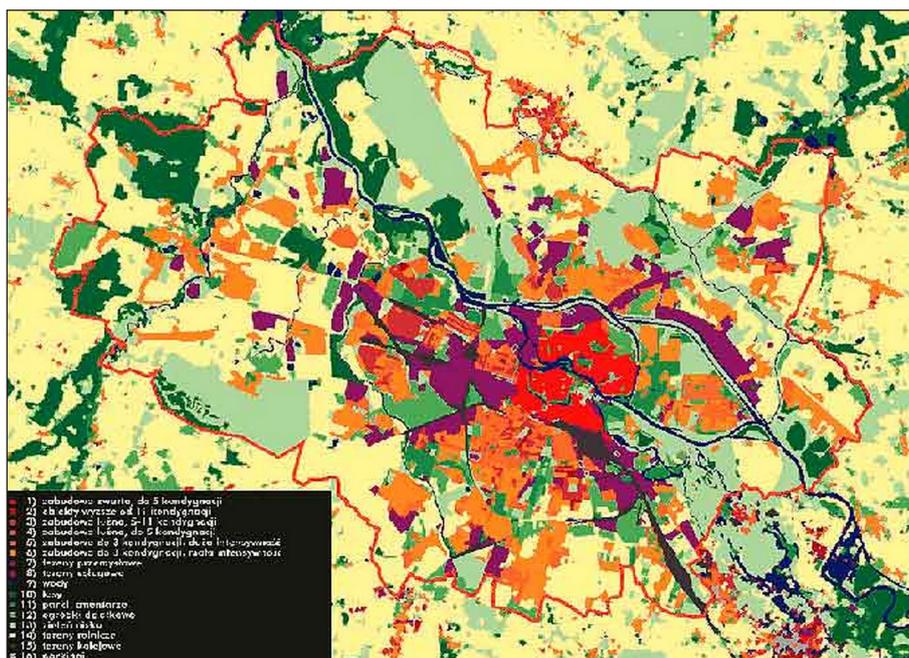[source:*http://www2.pgi.gov.pl/images/stories/artykuly/mapy_geochem/cdsl_4.jpg*]

Figure 2. Land use map [source: *http://www.eko.org.pl/wroclaw/srodowisko/img/k-rys6.jpg*]

INS- Inertial Navigation System), which allows to determinate the current angular position and the position of the scanning head (from which the measurement is made), it is possible to calculate the XYZ coordinates of points on the field. The recorded points form a point clouds. From which it is possible to generate digital terrain models, containing information on the shape and form of surface coverage. TLaser pulse can be reflected from the target more than once. An example of this is a wooded area. The first reflection occurs from the crown of the tree and the last from the surface. Between these two reflections indirect reflections are recorded. Registering them provides more accurate information on the condition of vegetation in the studied area [Kurczyński Z., 2014]
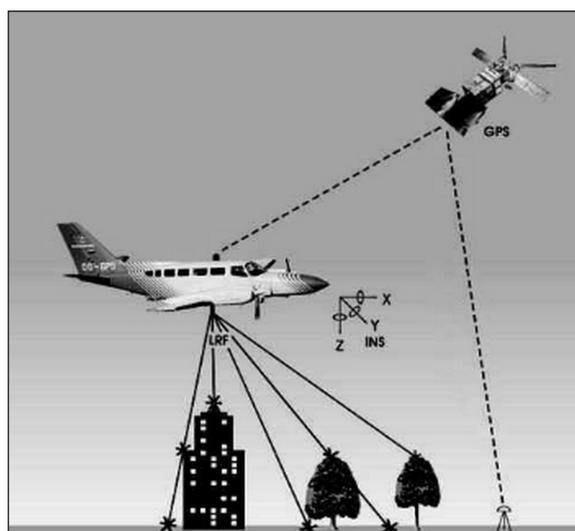
The obtained results, formed during the processing the input data are the basis for further analysis. Modern GIS software allows to perform both relatively simple and more advanced activities aimed at enriching and imaging the conducted research. GIS system allows to get answers to some important questions e.g.:

- Where the conditions are met?
- What has changed in a certain time slot?
- How phenomena are interdependent?

The answer to the question "what has changed since…?" is crucial in the issue related to the monitoring of the environment. It provides possibility to take appropriate steps to minimize

Figure 3. The principle of airborne laser scanning [Kurczyński, 1999]

the damage caused by the presence of contaminants. An additional advantage is observation of the progress of the used method of contamination remediation [Chrisman N., 1999]

Interpolation methods

The GPS system allows placing the results of performed tests in space. Unfortunately, they can be represented only as a point (a sample of material is usually collected in one particular place). For more information on the distribution of the phenomenon studied on a certain area, it is a suitable to use some method of interpolation. These

can be divided into simpler group (e.g. the nearest neighbour, inverse distance weighted), and more complex methods (e.g. kriging methods).

In the nearest neighbour method, any estimated point is set to a value of the nearest source point [Olivier, 2012]. The main drawback of this model is the formation of abrupt changes in value. In the inverse distance weighted method (IDW), which can be assigned to a group of kriging methods, estimated points are determined on the basis of source points, found in its surroundings. The result is affected by several parameters such as range searches, the number of points involved in the analysis and power factor. The process of IDW interpolation can be divided into the following steps:

1. Searching for points that meet the criterion of neighborhood (the amount or the distance).
2. Allocating weights to each typed point. At this step, it is possible to determine the power factor (p), the bigger it is, the points which are further, will have a greater impact on the result (Equation 1).
3. Calculating the value of estimating points [Lu G, 2008].

$$w_n = [(x_n - x_0)^2 + (y_n - y_0)^2]^{(-\frac{p}{2})}$$
$$z = \sum (w_n * z_n / \sum w_n) \quad (1)$$

where: w_n – weight of the points used to interpolate
z_n – value of the point used in the interpolation,
x_o, y_o – the coordinates of estimating point
p – power factor,
z – value of the estimated point.

Kriging-draft

Kriging is a group of advanced geostatistical methods. Among them we can distinguish: simple kriging, ordinary kriging, universal kriging, disjunctive kriging, co-kriging, indicator kriging and lognormal kriging. All these methods are based on two main principles: unbiased estimator and minimization of the mean square error of estimation. In the linear methods, estimation of unknown value is expressed by a linear combination of the neighbouring values. This may mean that in the estimation process of a single point, all source points can be involved. However, increasing the distance from the source is directly proportional to the weight reduction. For this reason, it is advisable to limit the search in the dependance on the location of the point. This allows

to minimize the number of equations necessary for obtaining reliable results. The main problem of geostatistical analysis is to determine the spatial variability. This phenomenon is to determine how to decrease the similarity research elements, when increases the distance between them. The analysis of the spatial variability involves the determination of the empirical values as a function of the distance between the analyzed data points, and then modelling of the obtained dependencies. A plot of the semivariance function and the distance between the points is called a semivariogram [Ligas, 2015].

Due to the complexity of kriging methods, it is recommended to perform preliminary analyzes at the stage of designing measurement network. Proper spatial sampling should provide as much information as possible on the spatial distribution of the studied phenomenon in a specific area with minimization of cost and time. The most important problems in measurement network planning:

- Shape of the network – simple and systematic random and systematic deterministic network. In the case of a simple random network, the location of all points is entirely coincidental. Systematic networks are represented by a set of points in a form of regular geometric figures (e.g. triangular, rectangular or hexagonal). Systemic random network differs from deterministic only on the choice of starting point, from which other points are calculated (systematic random – random position of starting point, systematic deterministic - the position of the first point is specified). The shape of the networks should be appropriate to the studied phenomenon. The most favourable, from the point of view of convenience of calculation, is systematic form (although not always possible to achieve).
- The distance between the points – this parameter essentially determine the number of points in the network. It should be set based on analysis of spatial correlation. Selection of this property determines the reliability of measurements. Too rare a network can result in inaccurate results, while very dense – in high costs of research.
- The geostatistical analysis is a very complex problem. It starts at the planning measurements stage and ends on the correct interpretation of the results. Geostatistics is a relatively new, rapidly growing field. There are many computer programs that allow its conduct.

Probably their number and range of opportunities will increase in near future. However, even the best software will not be able to replace the knowledge about the environment. Understanding the phenomenon that is being examined (e.g. heavy metals content and distribution in top soils) is the basis for the proper planning and resolving the full research process [Zawadzki, 2011].

Practical and scientific application of GIS in search of ground-water environment – examples

Recent years have seen a considerable increase in search of GIS, geostatistics and their applications in environmental studies. There are several research examples based on scientific articles connected with described area of study:

1. Spatial distribution of pH and organic matter in urban soils and its implications on site specific land uses in Xuzhou China [Mao Y., 2014]. By using of GIS and geostatistics (linear kriging), the authors of the article determined the relationship between the content of pH, organic matter SOM (Soil Organic Matter) and the form of land use. Based on 172 collected soil samples, maps of the spatial distribution of pH and SOM were generated (Fig. 4). It is noted that in wooded areas the pH and SOM achieve high values, while lower values were observed in urban areas related to industry. Generated maps can be useful for planning and environmental management.

2. Multivariate statistical and GIS-based approach to evaluate heavy metals behavior in mine sites for future reclamation [Acosta J.A., 2011]. The subject of the study was the area of the closed Pb-Zn mine. Not only the content

of heavy metals, but also the other phenomena that affect the distribution of heavy metals were analysed (e.g. pH, soil organic matter, conductivity). The authors emphasize that soluble compounds of heavy metals can pose a threat, not only for soil, but also for surface water and groundwater. Knowledge of the spatial distribution can be crucial for planning the reclamation, both in the selection of appropriate technology and the estimation areas with the highest level of degradation.

3. Analysis of the spatial relationship between heavy metals in soil and human activities based on landscape geochemical interpretation [Yu H., 2014]. The development of chemical elements in the soil is caused not only by the parent material, climate and biology, but also anthropogenically. The main source of heavy metal content are the anthropogenic activities. The subject of the study was to determine the spatial impact of human activity and landforms on the content of heavy metals in soils. It has determined several significant correlations e.g.: Zn, Cr, Cu, Pb, Hg, Cd, Ni and Ag were associated with an urban area, the areas of the river showed increased content of Ni, Pb, Hg, Ag and Cd. It was also presented that highest value of Zn, Cr, Cu, Pb, Hg, Cd, Ni and Ag were found near roads. The results of the research revealed the possibility of using GIS and geostatistics to analyze the chemical properties of soil, realizing at the same effective combination of macroscopic spatial information and microscopic migration mechanism of soils.

4. The modeling of spatial distribution of chlorine ion concentration in groundwater from post-flotation reservoir "Żelazny Most". The aim of the study was modelling the distribu-
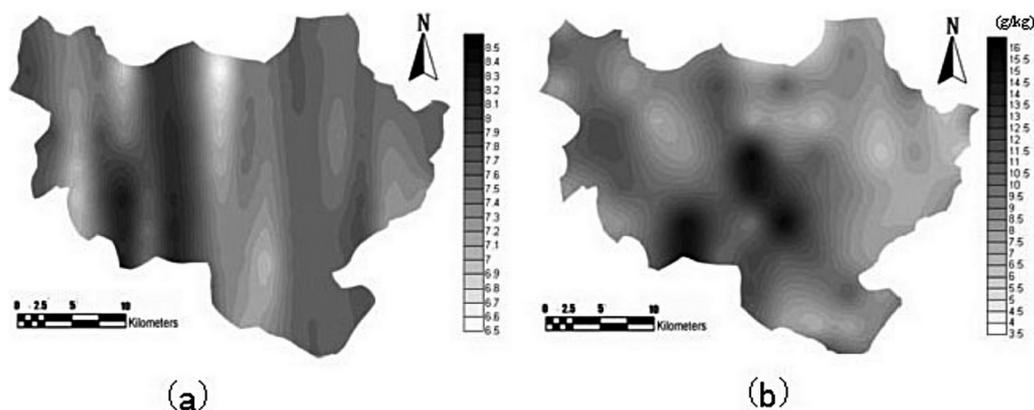

Figure 4. Spatial distribution of pH (a) and SOM (b) in soil of Xuzhou, Chine [Mao Y, 2014]

tion and migration of pollutants in the post-flo-tation reservoir "Żelazny Most". In the study different interpolation methods were used (inverse distance weighted, indicator kriging and point kriging). The main result of the research was correct assessment of the range and concentration of groundwater pollution. This is important information exerting a major impact on environmental studies [Zawadzki 2004].

5. Metal contamination in urban, suburban and country park soils of Hong Kong. A study based on GIS and multivariate statistics studies [Celine Sio-Jan, 2005]. Due to the constantly progressive urbanization, heavy metals are continuously emitted to the terrestrial environment and pose a significant threat to human health. Soil samples were collected in a systematic manner – 5 samples per km^2 in urban areas and 2 samples per km^2 in suburban areas. The results of the analysis showed that the soil in both urban and suburban contain Cu, Pb and Zn. The content of Pb in urban soils exceeded the limit value. Statistical analysis using PCA transformation (Principal component analysis) and cluster analysis CA showed various links between heavy metals and the main components of soils in urban, suburban and country parks. Using GIS, maps of soil contamination was generated – and it was possible to identify places where heavy metal content was highest.

6. Heavy metal contamination of topsoils around a lead and zinc smelter in the Republic of Macedonia [Stafilov T., 2010]. The study used GIS technologies and multivariate statistics to identify areas most susceptible to contamination. The main focus was on the content of As, Au, Cd, Cu, Hg, In, Pb, Sb, Se, Zn.

7. Groundwater arsenic contamination risk prediction using GIS and classification tree method [Hossain M., 2013]. Contamination of groundwater by arsenic is a serious problem for the health of both the environment and people. Information on the probability of contamination may create a possibility to take countermeasures, before the damage. The study included six factors that may affect the contamination of arsenic – geology, soil pH, elevation, SOM, the Fe content in the soil and distance from the river. Prediction model was generated using MATLAB software. The results showed a correlation between the measured factors and arsenic content in ground-

water. The best model created by the decision tree obtained 87.9% accuracy. In addition, the GIS was used to generate risk maps [Figure 5].

8. Comparison of spatial interpolation methods for estimating heavy metals in sediments of Caspian Sea [Kazemi S., 2011]. The studies estimated distribution of six heavy metals: As, Cd, Cu, Hg, Pb, Zn. It was collected a total of 80 samples – 18 in Iran, 16 – Azerbaijan, 21 – Turkmenistan, 13 – Kazakhstan and 21 – Russia. Interpolation was carried out by four methods: ordinary kriging (OK), conditional simulation (CS), adaptive network fuzzy inference system (ANFIS), genetic algorithm based on artificial neural network (GA-ANN). Among those methods, GA-ANN proved to be the most advantageous.

9. Assessing land cover change and degradation in the Central Asian deserts using satellite image processing and geostatistical methods [Karnieli A., 2008]. Analyzing the satellite images – it was observed as fading light strips directly proportional to the increasing distance from the well. The aim of the study was to determine the spatial and temporal distribution of degradation. On the basis of selected two areas of desert, using the brightness index Tasseled Cap's semi-variogram of the stud-
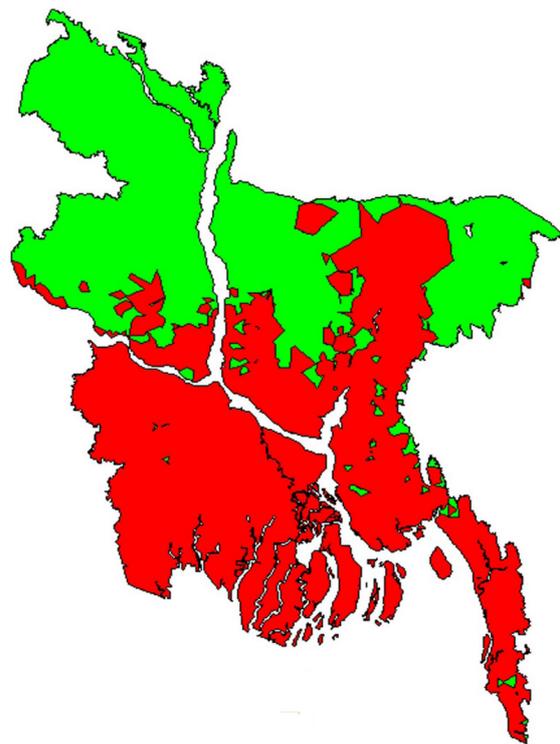

Figure 5. Groundwater arsenic contamination risk prediction map (green – safe zone, red – danger zone) [Hossain M., 2013]

ied phenomenon was established – it allowed the use of the kriging method to obtain the spatial distribution.

10. Assessing heavy metal contamination in soils of the Zagreb region Northwest Croatia using multivariate geostatistics [Solitto D., 2010]. Using multivariate geostatistical analysis, the chemical composition of soils was estimated. Strong analysis kriging (FKA – factorial kriging analysis) was used to investigate the correlation structure of some variables, modelling ten chemical variables and basic components of co-kriging. By analyzing FKA two main factors were established – the first of shorter range for the Zn, Pb, Cd, Cu and Ni indicates the various sources of pollution of anthropogenic origin, and Ca (mainly for larger areas) was associated with lithology and composition of parent metal. The methodology was useful for the separation of geological and anthropogenic factors influencing the content of heavy metals in the soil and to determine the source of contamination.

Spatial distribution of pH in soil of zinc smelter "Miasteczko Śląskie"

Zinc smelter "Miasteczko Śląskie" has been operating since the early 1970s. The plant produces zinc, crude, refined lead and sulfuric acid. As a result of long term emissions into the atmosphere, the areas around the plant is strongly contaminated by heavy metals (Pb, Zn, Cd), other trace element (Al, As) [Table 1] and gases (SO_2, CO). In the immediate vicinity of the zinc smelter, areas totally or largely devoid of vegetation were created.

In 2015 it was decided to re-conduct research on the zinc smelter "Miasteczko Śląskie" area. On the basis of aerial photographs and reconnaissance the study area were selected. During the selection process, distribution of vegetation was taken mainly into account - lack of vegetation may indicate a contaminated area. Network of 29 measuring points (figure 6), with particular emphasis on the area with the least amount of vegetation, was planned. Extra points were located around this area. Measurement grid has a form of a grid of different edge lengths – for the central region the length is 175 meters, and 350 meters for external. In March 2016 on each of the points two samples of soil were collected from the two levels of depth: 0–20 cm and 40–60 cm.

Samples were analyzed in terms of pH. It was determined in a solution of KCl by potentiometric method according to the standard ISO 10390:1997. It should be noted that the pH is one of the main parameters affecting the migration of heavy metals. Using the pH results and interpolation by inverse distance weighted method [Lu, 2008] the spatial distributions of the pH on examined levels of depth were generated (figure 7).

Analyzing the obtained results it was found that the studied area is dominated by acidic soil with pH value in range from 3 to 5. The area is characterized by a semi constant distribution, the soil with the lowest pH values are mainly in the central part of studied area. A significant relationship between the pH value and studied levels of depth was noted – in 40–60 cm the pH value is higher (about 0.5) than in the layer of 0–20 cm. The obtained results are the basis for further detailed research focusing on heavy metal and organic matter distribution.

Table 1. Trace elements content in soils (0–20 cm) in the area of zinc smelter "Miasteczko Śląskie" in 2004 – 2010 [Kacprzak, 2007; not published data]

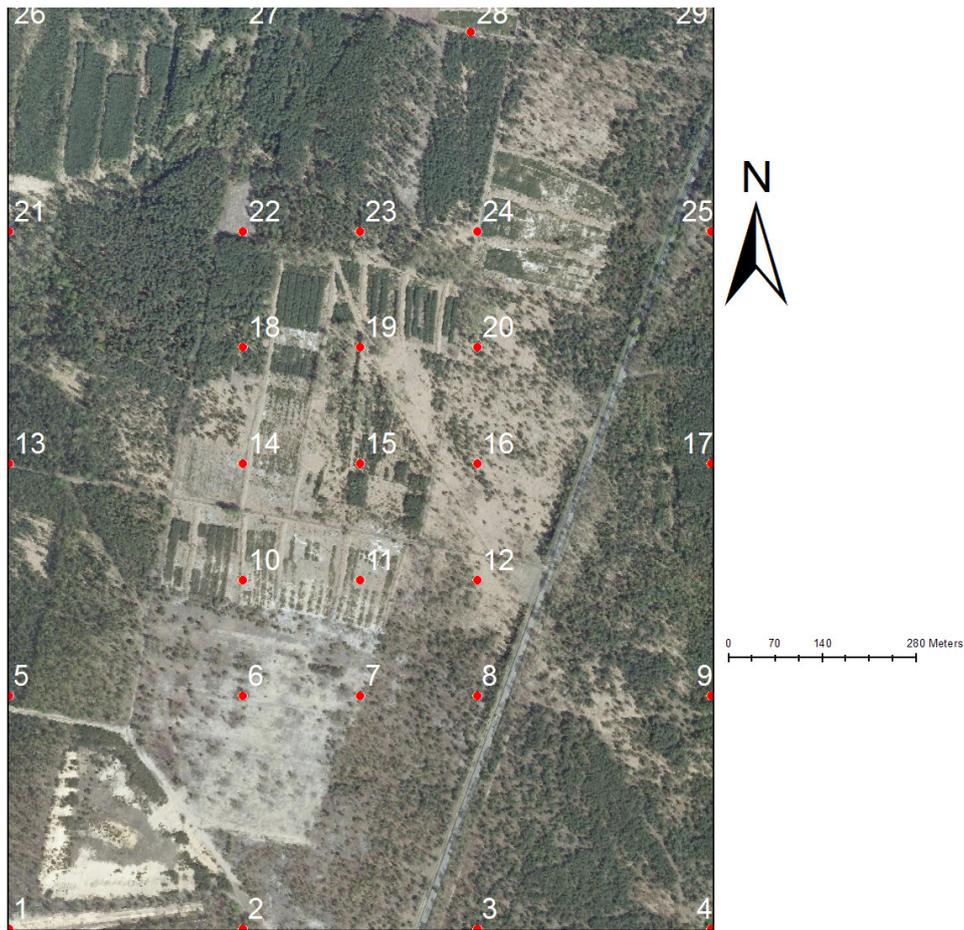| trace element	content of heavy metals in soils of „Miasteczko Śląskie" [mg/kg]	average content of heavy metals in soils of Poland [mg/kg]
Hg	5.6 – 20.0	0.02 – 0.2
Ag	3.5 – 56.0	0.03 – 0.1
As	4.5 – 67.0	2 – 13
Al	4 533.0 – 8 494.0	10 000 – 35 000
Pb	540.0 – 1 1237	13 – 25
Cd	4.0 – 10.9	0.05 – 0.3
Ni	5.0 – 13.0	5 – 23
Zn	428.0 – 1 220.0	35 – 80
Cu	6.0 – 74.3	5 – 19
Cr	10.4 – 16.6	15 – 740
Fe	7 820.0 – 14 100.0	3 000 – 10 000

Figure 6. Soil sampling locations in the area of zinc smelter "Miasteczko Śląskie"
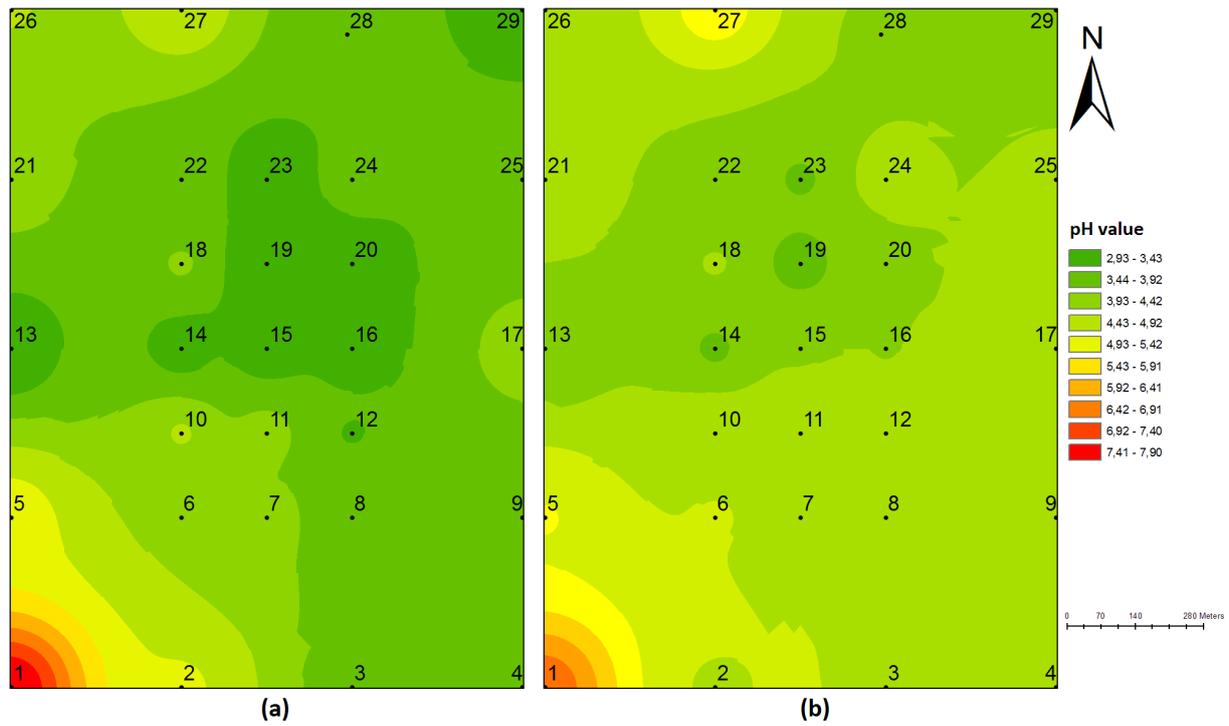(Silesian Region, south Poland)

Figure 7. Spatial distribution of pH in soil of "Miasteczko Śląskie" on two levels of depths: 0 – 20 cm (a),
40 – 60 cm (b).

SUMMARY

GIS and geostatistics can be useful tools in the environment studies. They can be applicate in many areas of research – mainly where spatial distribution can be significant. An important advantage of advanced geostatistical methods is credible assessment of the accuracy of the model, accompanying the process of prediction. In addition to the product in a form of maps, GIS allows observation of the impact of certain factors on the studied phenomenon (e.g. the impact of anthropogenic and geological content of heavy metals or migration of organic contaminants in the soil). However, appropriate use of GIS requires knowledge both in terms of geostatistics and environmental engineering. Rapid development of GIS facilitates its use, but it is important to remark that lack of appropriate theoretical preparation can lead to erroneous conclusions – the software will never replace the knowledge of the phenomenon studied.

Acknowledgements

REFERENCES

1. Acosta J.A., Faz. A., Martinez-Martinez S., Zroznoza R., Carmona D.M., Kabas S., 2011. Multivariate statistical and GIS based approach to evaluate heavy metals behavior in mine sites for future reclamation, Journal of Geochemical Exploration.

2. Celine Siu-lan L., Xiangdong L., Wenzhong S., Sharon Ching-nga C, Iain T., 2005. Metal contamination in urban, suburban and country park soils of Hong Kong: A study based on GIS and multivariate statistics, Science of the Totel Environment 356, 45–61.

3. Chrisman N, 1999. What Does 'GIS' Mean?, Transcation in GIS, 3(2), 175–186.

4. Drzewiecki W., 2015. GIS w skrócie, http://home.agh.edu.pl/~awrobel/resources/GIS%20w%20skrocie.pdf, from 20.11.2015.

5. Hossain M., Piatanakulchai M., 2013. Groundwater arsenic contamination risk prediction using GIS and classification tree method, Engineering Geology 156.

6. Karnieli A., Gilad U., Ponzet M., Svoray T., Mirzadinov R., Fedorina O., 2008. Assessing land-cover change and degradation in the Central Asian deserts using satellite image processing and geostatistical methods, Journal of Arid Environments.

7. Kazemi S.M., Hosseini S.M., 2011. Comparison of spatial interpolation methods for estimating heavy metals in sediments of Caspian Sea, Expert systems with Applications, 38.

8. Kurczyński Z. 2014. Fotogrametria, Wydawnictwo naukowe PWN, 241–267.

9. Kurczyński Z. 1999. DTM inaczej, Geodeta – Magazyn Geoinformacyjny.

10. Ligas M., 2015. Zaawansowane Metody Analiz Przestrzennych Geostatystyka – wstęp, http://home.agh.edu.pl/~ligas/konspekty/geo_intro.pdf, from 06.12.2015.

11. Lu G, Wong D., 2008. An adaptive inverse-distance weighting spatial interpolation technique, Computer&Geosciences, 1044–1055.

12. Mao Y., Shuxun S., Shiqi L., Jinlong J., 2014. Spatial distribution of pH and organic matter in urban soils and its implications on site specific land uses in Xuzhou China, Comptes Rendus Biologies.

13. Olivier R, Hangian C., 2012. Nearest Neighbor Value Interpolation, International Journal of Advanced Computer Science and Application, Vol. 3.

14. Stafilov T., Sajn R., Pancevski Z., Boev B., Frontasyeva M., Strelkova L., 2010. Heavy metal contamination of topsoils around a lead and zinc smelter in the Republic of Macedonia, Journal of Hazardous Materials 175.

15. Shisanya C., Recha C., Anyamba A., 2011. Rainfall variability and its impact on Normalized Difference Vegetation Index in arid and semi-arid lands of Kenya. International Journal of Geosciences, 2, 36–41.

16. Sollitto D, Romic M., Castrignano A., Romic D., Bakic H., 2010. Assesing heavy metal contamination in soils of the Zagreb region (Northwest Croatia) using multivariate geostatistics, Catena.

17. Wysocki J., 2010. Podstawowa rola geodezji, fotogrametrii, teledetekcji oraz geomatyki w tworzeniu instrastruktury informacji przestrzennej w zakresie opisu ukształtowania powierzchni terenu, Roczniki geomatyki, tom VIII, Zeszyt 1(37).

18. Yu H., Ni S., He Z., Zhang C., Nan X., Kong B. Weng Z., 2014. Analysis of the spatial relationship between heavy metals in soil and human activities based on landscape geochemical interpretation, Journal of Geochemical Exloration.

19. Zawadzki J., 2011. Metody geostatystyczne dla kierunków przyrodniczych i technicznych, OPW, Warszawa.

20. Zawadzki J., Kucharek M., Trichel W. 2004. The modeling of spatial distribution of chlorine ion concentration in groundwater from post-flotation reservoir "Żelazny Most", Engineering and Environmental Protection, 7(3-4), 381–392.

21. Kacprzak M., 2007. Wspomaganie procesów remediacji gleb zdegradowanych, seria Monografie nr 128, Wydawnictwo Politechniki Częstochowskiej, Częstochowa.