Proceedings of ECOpole

Vol. 3, No. 2 2009

Jarosław ZAWADZKI[1] and Adam TARGOWSKI[1]

EVALUATING THRESHOLD RADIAL AND ADAPTIVE FILL SECONDARY SAMPLING METHODS FOR CONTAMINATED SOIL SAMPLING EFFICIENCY

OCENA SKUTECZNOŚCI OPRÓBKOWANIA WTÓRNEGO METODAMI UZUPEŁNIANIA PROMIENIOWEGO I WYPEŁNIANIA ADAPTACYJNEGO W BADANIACH ZANIECZYSZCZENIA GLEB

Abstract: The purpose of the work was to compare effectiveness of common secondary sampling methods for assessing the distribution of soil pollution. The study case is based on an example of assessing the spatial distribution of soil contamination with lead in Slawkow area (Upper Silesian Industrial Region). This comparison was made in regard to both precision of the spatial estimation and minimization the cost of measuring campaign. The special attention was given to the often applied secondary sampling designs such as threshold radial (also known as adaptive cluster sampling) and adaptive fill sampling. These two methods were tested in typical municipal and suburban environment in Slawkow area. The work contains also detailed statistical and geostatistical analysis of the above-mentioned contamination, and elaboration of series of its spatial distributions using numerous alternative sampling designs. The determined sampling plans make it possible to find compromise between ecological and financial aspects. A combination of the obtained results with the legal regulations in force concerning concentrations of heavy metals in soils are the basis for reliably estimation the ecological hazard arising from the soil contamination with lead in the Slawkow area. The results of performed analyses show that better efficiency in terms of cost and precision of measuring campaign gives rather coarser preliminary sampling design followed by appropriate secondary sampling then use the one-stage very dense measuring grid.

Keywords: geostatistics, heavy metals, soils, ecological risk, secondary sampling designs

The choice of appropriate sampling design is essential in different soil related surveys. This arises from the fact, that our knowledge on soil in their natural state is never fully known and collecting samples as well as laboratory analysis is expensive and time consuming, especially when investigations are performed on large areas. Secondary sampling is very important stage of many environmental studies, which can significantly improve the analysis by relatively low cost. Secondary sampling design can refine a model, get a deeper insight into studied phenomenon, clarify situation and thus make right decision.

The goal of the work was a study-case based evaluation of effectiveness of commonly used secondary spatial sampling designs such threshold radial design (also known as adaptive cluster sampling) and adaptive fill design for delineation of the extent of the area polluted with heavy metals.

Site description and data collection

Study area was located in Slawkow city and its vicinity (Upper Silesian Industrial Region). The studies were performed using soil samples from the archives of the Polish Geological Institute [1]). A satellite view of study area with sample point locations (exhaustive data set) were shown in Figure 1.

---

[1] Environmental Engineering Department, Warsaw University of Technology, ul. Nowowiejska 20, 00-661 Warszawa, email: j.j.zawadzki@gmail.com, adam@targowski.pl

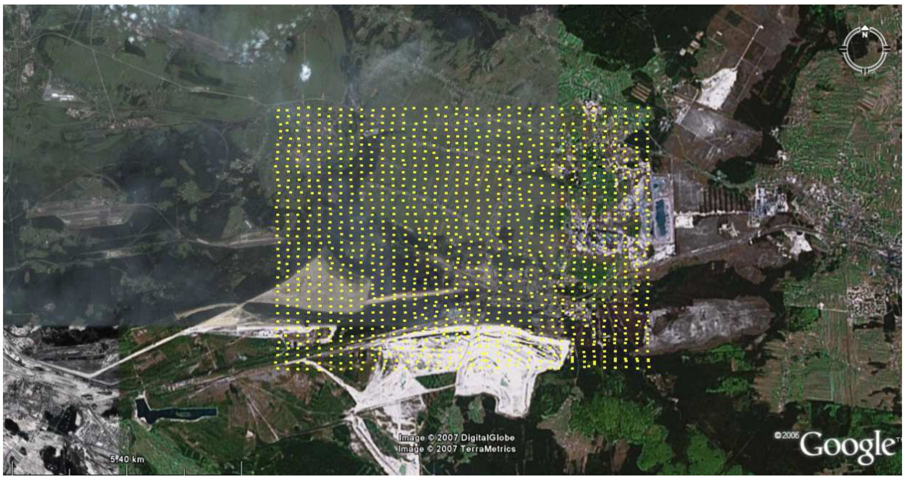

Fig. 1. Satellite view of Slawkow with sample point locations (exhaustive data set)

Secondary sampling strategies and geostatistical methods

In order to limit the time consuming analyses we focused in our work on two commonly used secondary sample designs [2, 3]:

1. Threshold Radial (also known as Adaptive Cluster Sampling) which is a straightforward secondary sample design that places samples in a radial pattern around existing data points that exceed a decision threshold. Threshold radial can be useful in situations where one have a lot of very low or undetected samples and one or two very high measurements.

2. Adaptive Fill Design

In this sample design, samples are placed in the largest spatial gaps among data points. Unlike Threshold Radial, this design gives no regard to the measured values, only their relative positions. A set of new sample candidates is defined by a grid (much like a spatial model) that overlays the data points and acknowledges site boundaries, polygons, and whether layers are active or not. From this set of N candidates the first winning location is simply that value which has the maximum distance to its closest neighbor. The design searches for the second location among the remaining candidates by comparing with the N+1 locations. If there are ties among the two locations, then the tie breaker method is used. The process repeats until one of the following becomes true:

- The total number of samples has been located
- There are no remaining candidates
- No remaining candidate satisfies the minimum distance constraint.

To study soil contamination with lead, it was necessary to obtain the spatial distribution of lead concentrations in soils, as well as the spatial distribution of estimation errors. This was done using geostatistical methods. The ordinary kriging was selected as the most appropriate technique for our analysis. Ordinary kriging is the most effective linear estimator as it assumes that the average value of the estimation error equals zero, and thus minimises the variance of the estimation error [4-10]. The series of spatial distribution of lead concentrations in soils were produced. In the first step preliminary coarse sampling designs (systematic or random) were chosen from the very dense exhaustive data set shown in Figure 1. These designs were treated as pre-information for subsequent sampling. Then above-described spatial distributions were created by careful variogram modeling and kriging technique. At this stage all modeled spatial distributions of lead concentrations in soils were validated using cross-validation methods [11]. Each of the data points was individually removed from the data set, and after that, its value was modeled and subsequently compared with the measured one. Next, the scatter plots of estimated values versus the measured ones were calculated. Using these scatter plots several estimation errors were carefully calculated. Furthermore, modeled spatial distributions were validated using the true values taken from exhaustive data set. Another important measure of sampling design was the total cost of measuring campaign calculated on the basis of the cost of single measurement. By comparing the quality of the spatial distribution with the total cost of measurements it was possible to evaluate the efficiency of the sampling grid under investigation. Then, it was decided what type of secondary sample design should be applied, and how many additional measurements should be used. This allowed for significant reduction of uncertainty by relatively low cost. Then all above-described analyses were repeated once or twice. For clarity multistage sampling process was abbreviated. For instance, the abbreviation "radial_100_100_rand" means, that preliminary sample grid was random, and furthermore 100 additional points were added using threshold radial sampling technique.

Analyses were performed using Arc Gis software (namely, FIELDS, Geostatistical Analyst and Spatial Analyst components) [12].

Results and discussion

Below, in Figures 2 and 3, as well as Figures 6 and 7, exemplary sampling designs are presented (eg systematic and random preliminary sampling design followed by threshold radial and adaptive fill secondary samplings) with the spatial distribution of prediction errors (root mean square prediction error-RMSE).

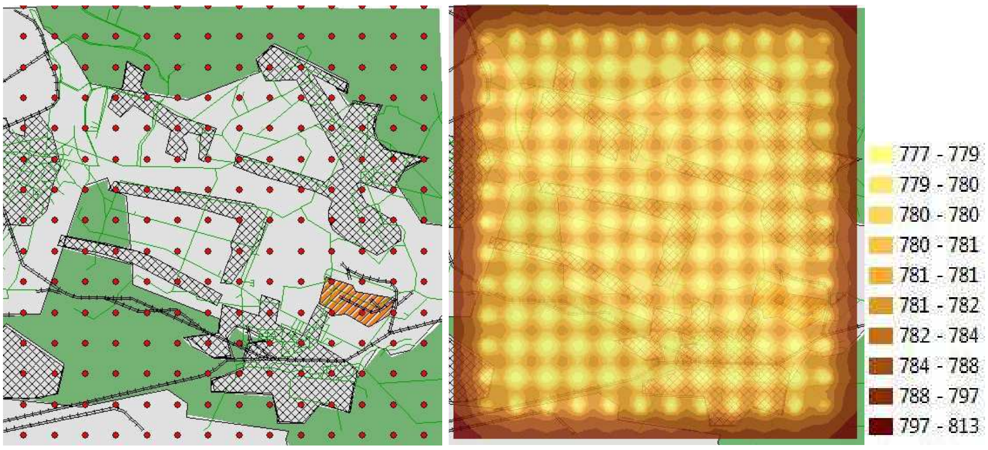

Fig. 2. Preliminary systematic sampling design of 200 observations (left) with the spatial distribution of RMSE (right)

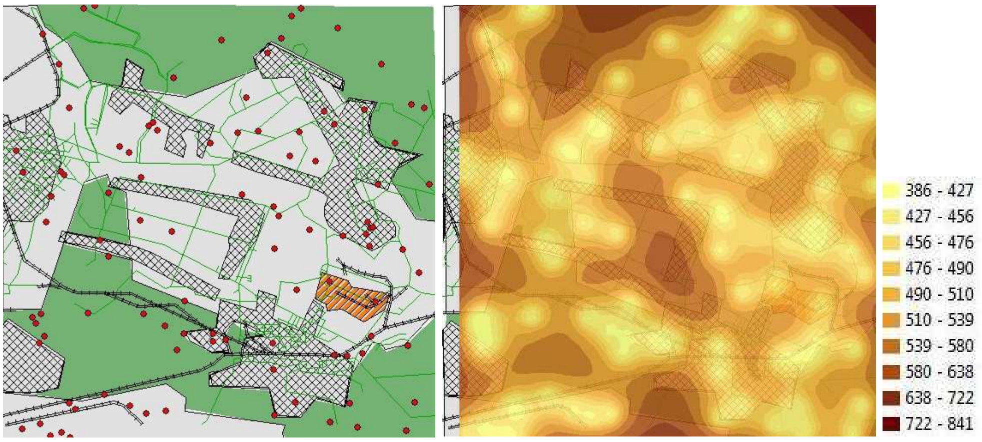

Fig. 3. Preliminary simple random sampling design of 100 observations (left) with the spatial distribution of RMSE (right)

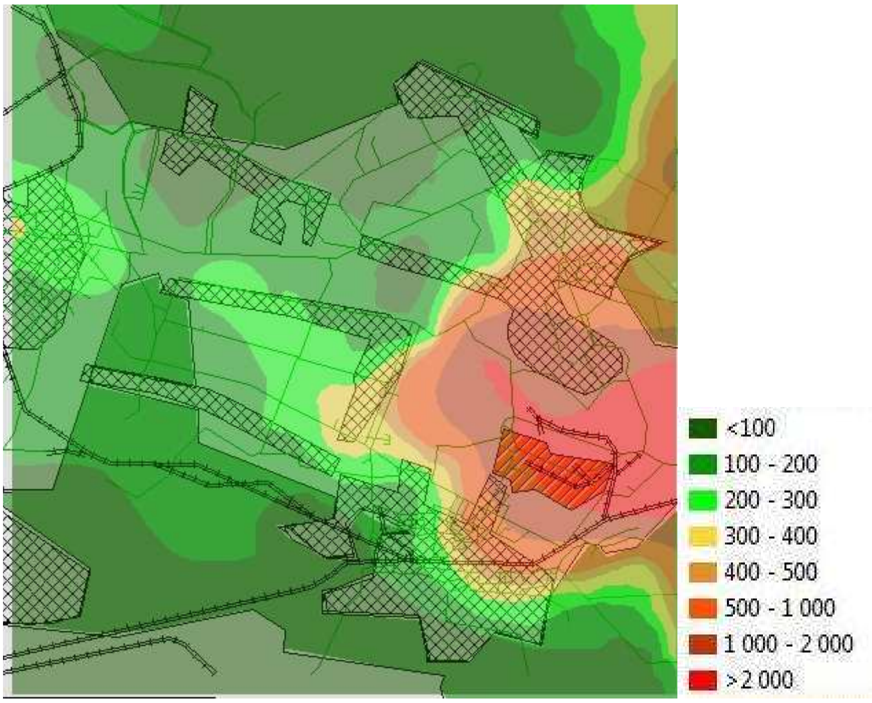

Fig. 4. Spatial distribution of lead contamination in [mg/kg] made on basis sampling design from Figure 2

Figures 4 and 5 show spatial distributions of lead contamination obtained on basis sampling from the Figures 2 and 3, respectively. In Figure 8 root mean square standardized prediction error as a function of cost of measuring campaign is presented for numerous different preliminary and secondary sampling designs.

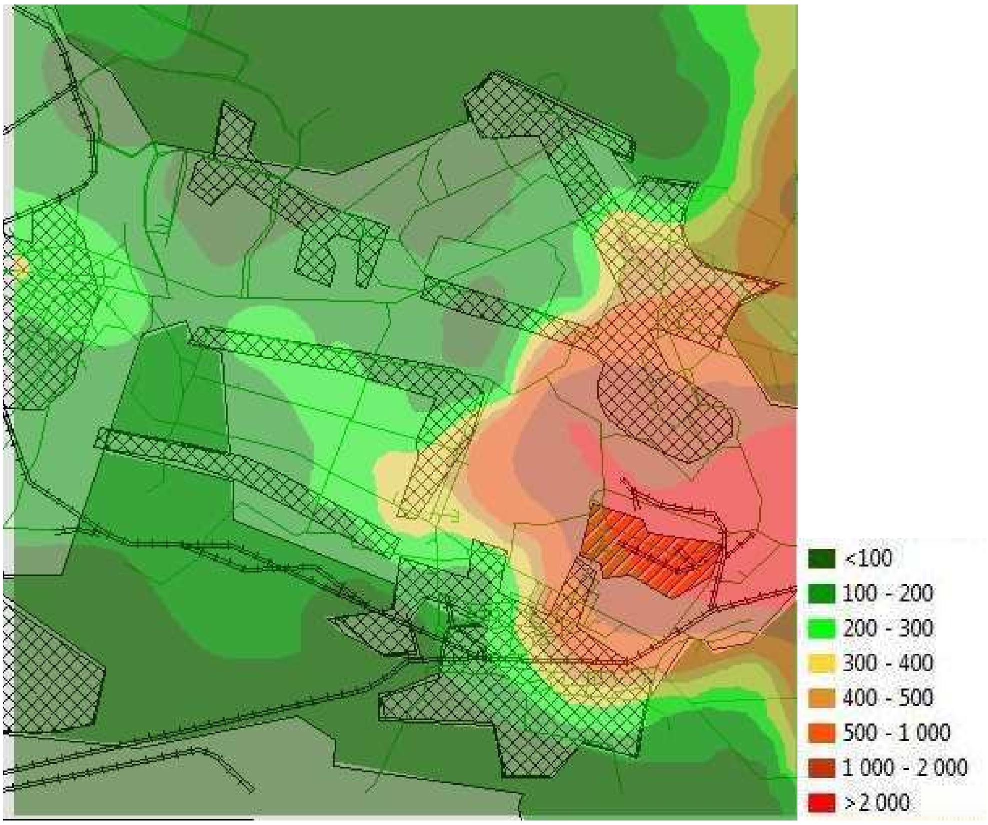

Fig. 5. Spatial distribution of lead contamination in [mg/kg] made on basis sampling design from Figure 3

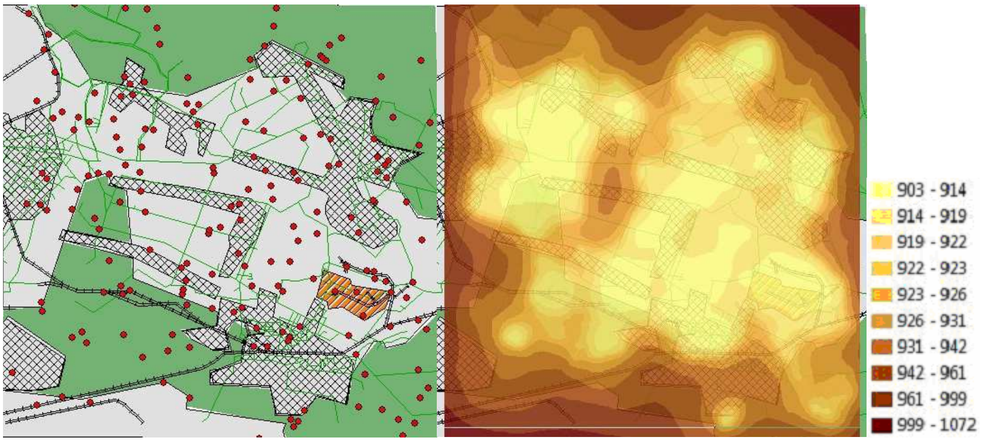

Fig. 6. Optimisation of sampling design from Figure 3 (random_100) with 100 additional sampling points (left) with with the spatial distribution of RMSE (right)

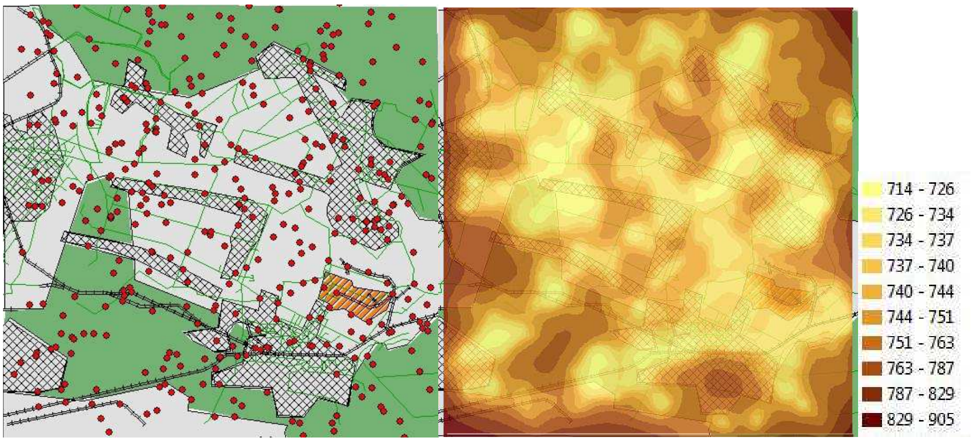

Fig. 7. Optimisation of sampling design from Figure 3 (random_100) with 200 additional sampling points (left) with the spatial distribution of RMSE (right)

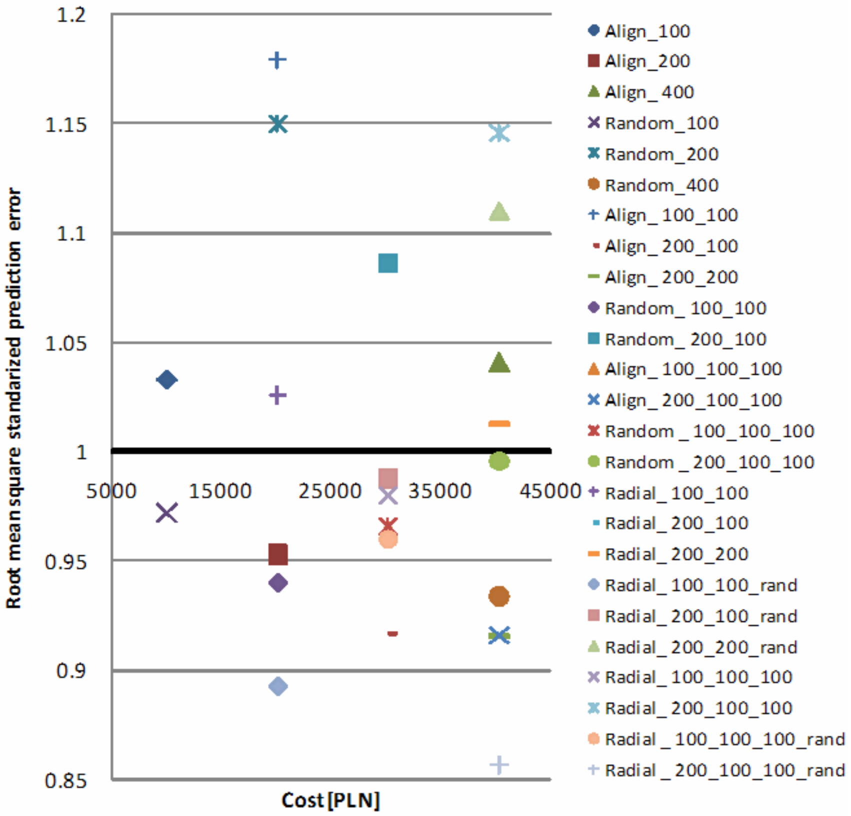

Fig. 8. Root mean square standardized error as a function of cost of measuring campaign calculated for different preliminary and secondary sampling designs

At first glance, the above-shown results seem to be rather ambiguous. However, detailed analysis of estimation errors for many multistages sampling designs showed that much better results in terms of cost and precision give coarser preliminary sampling designs followed by appropriate chosen secondary sampling than use of the one-stage very dense measuring grid. This result arises from the fact that although mostly used dense regular sampling grids, are relatively precise, but in the same time they are very costly. While preliminary random sampling designs give unbiased results, but in general, are less precise than systematic sampling designs. Multi-stage sampling allows for treating the intermediate results as the pre-information for subsequent sampling. This make it possible to better control and tune the whole sampling process according to circumstances occurring during sampling campaign.

Acknowledgements

The authors would like to thank dr J. Lis and dr A. Pasieczna from Polish Geological Institute, for giving access to the soil samples.

References

[1] Lis J. and Pasieczna A.: Szczegółowa mapa geochemiczna Górnego Śląska. Państwowy Instytut Geologiczny, Warszawa 1999.
[2] Stewart R. and Purucker T.: SADA Technical Support: Using Secondary Sampling Strategies. University of Tennessee, Knoxville 2000.
[3] Guidance on Choosing a Sampling Design for Environmental Data Collection. United States Environmental Protection Agency, Office of Environmetnal Information, Washington DC 20460, 2002.
[4] Journel A.G. and Huibregts C.J.: Mining Geostatistics. Academic Press, London 1978.
[5] David M.: Handbook of Applied Advanced Geostatistical Ore Reserve Estimation. Elsevier Scientific Publishing 1988.
[6] Isaaks E.H. and Srivastava R.M.: An Introduction to Applied Geostatistics, Oxford University Press, New York 1989.
[7] Goovaerts P.: Geostatistics for Natural Resources Evaluation. Oxford University Press, New York 1997.
[8] Demougeot-Renard H., Foquet C. and Renard P.: *Forecasting the number of soil samples required to reduce remediation cost uncertainty*. J. Environ. Qual., 2004, **33**, 1694-1702.
[9] Zawadzki J.: *Geostatistical methods for continuity correlation evaluation: Studies of Fe, Pb and Zn concentrations in soil*. Inż. Ochr. Środow., 2002, (3-4), 369-391.
[10] Carlen C, Critto A, Nathanail P. and Marcomini A.: *Risk based characterisation of a contaminated industrial site using multivariate and geostatistical tools*. Environ. Pollut., 2000, **111**(3), 417-27.
[11] Kohavi R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. Proc. 14th Int. Joint Conf. on Artificial Intelligence, IJCAI, Montreal, CA, 1995, 1137-1143.
[12] ArcGIS 9; What is ArcGIS 9.1, ESRI 2005.

OCENA SKUTECZNOŚCI OPRÓBKOWANIA WTÓRNEGO METODAMI UZUPEŁNIANIA PROMIENIOWEGO I WYPEŁNIANIA ADAPTACYJNEGO W BADANIACH ZANIECZYSZCZENIA GLEB

Wydział Inżynierii Środowiska, Politechnika Warszawska

Abstrakt: Celem pracy było porównanie efektywności wybranych metod opróbkowania dodatkowego, wykonywanego w celu wyznaczenia rozkładu zanieczyszczenia gleby. Studium przypadku dotyczyło wyznaczania rozkładu przestrzennego zanieczyszczenia gleby ołowiem w okolicach Sławkowa (Górnośląski Okręg Przemysłowy). Głównymi kryteriami efektywności metod opróbkowania dodatkowego były oceny dokładności rozkładu przestrzennego oraz koszty kampanii pomiarowej. Szczególną uwagę zwrócono na często stosowane opróbkowanie dodatkowe metodami uzupełniania promieniowego (zwanego również adaptacyjnym opróbkowaniem klastrowym) oraz wypełniania adaptacyjnego. Te dwie metody były przetestowane w typowym miejskim i podmiejskim środowisku, na terenie Sławkowa i w jego okolicach. Praca zawiera również statystyczną i geostatystyczną analizę omawianego zanieczyszczenia gleby, określenie jego ciągłości przestrzennej, a także wyznaczenie serii rozkładów przestrzennych zanieczyszczenia gleby ołowiem z wykorzystaniu wyżej wymienionych metod opróbkowania dodatkowego. Wyznaczone sieci pomiarowe pozwoliły na znalezienie kompromisu pomiędzy aspektem ekologicznym a finansowym. Rezultaty analizy statystycznej i geostatystycznej wraz z obowiązującymi uregulowaniami prawnymi dotyczącymi zawartości metali ciężkich w glebie są podstawą do rzetelnego określenia potencjalnego ryzyka ekologicznego wynikającego z zanieczyszczenia gleb ołowiem w okolicach Sławkowa. Rezultaty wykonanych analiz pokazują, że lepszą skuteczność określaną kosztem i precyzją kampanii pomiarowej daje rzadsze opróbkowanie wstępne uzupełnione odpowiednim opróbowaniem dodatkowym niż jednoetapowa kampania pomiarowa z gęstą siecią pomiarową.

Słowa kluczowe: geostatystyka, metale ciężkie, ryzyko ekologiczne, opróbkowanie dodatkowe