J. Jpn. Soc. Soil Phys.
土壌の物理性
No. 128, p.5 ～ 10 (2014)

JpGU 特集
論　文
Original Paper

Geostatistical modeling of the spatial distribution of surface soil arsenic around a smelter

Pierre GOOVAERTS[1] and Gregory GLASS[2]

Abstract: For almost 100 years, the Asarco Company operated a copper smelter in Tacoma, Washington State. Air pollution from the smelter settled on the surface soil over more than 2,600 square kilometers of the Puget Sound basin. As part of the Tacoma Smelter Plume project, there have been a number of studies looking at soil arsenic contamination, leading to the collection of more than 5,000 surface soil samples over the years. The present study aimed to pool all of the sampling results to create a model that can be used to further prioritize areas for additional sampling and remediation since not enough money is available to clean all residential parcels. This paper describes a geostatistical simulation-based approach to compute for each block-group the expected number of residential parcels where a given arsenic concentration threshold is exceeded with a minimum probability. This information is being currently used to select widely contaminated block-groups where all residential parcels are systematically sampled and the ones exceeding a target threshold of 100 ppm are remediated.

Key Words : stochastic spatial simulation, remediation, airbone contamination, soil sampling, wind direction

1. Introduction

Etymologically, the term “geostatistics” designates the statistical study of natural phenomena. The early developments of geostatistics in the 1950’s and 1960’s aimed to improve the evaluation of recoverable reserves in mining deposits (Journel and Huijbregts, 1978). Since then its field of application has expanded considerably to include most fields of geoscience (e.g. geology, geochemistry, geohydrology, soil science) as well as a vast array of disciplines that all deal with the analysis of spatial data, such as oceanography, hydrogeology, remote sensing, agriculture, and environmental sciences. One of the main characteristics of the aforementioned data is their structured distribution in space which reflects the impact of various factors (e.g. geology, weather, human activities, land cover) operating at different spatial scales. Geostatistics (Goovaerts, 1998) provides a set of tools to model this spatial structure and interpolate scattered measurements to create spatially exhaustive layers of information and assess the corresponding accuracy and precision.

The cost of collecting field data is often prohibitive, leading to targeted sampling and areas that might be sparsely sampled. It is thus critical to incorporate any source of information available in the interpolation of environmental attributes. For example, the information available for site characterization impacted by incinerators typically includes air and soil measurements collected near the source, complemented by data about the source and its environment — such as meteorological data, wind speed and direction, the characteristics of the stack, and exhaust emissions. Geostatistics allows the combination of detailed process-based modeling of atmospheric deposition and monitoring data within a probabilistic framework in order to incorporate and propagate the uncertainty attached to the different pieces of information (Goovaerts et al., 2008).

This paper illustrates the application of geostatistics to the characterization of a large area including 583,291 residential and undeveloped parcels whose soil has been potentially contaminated with arsenic as the result of the 100-yr operation of a copper smelter. Parcels are used for tax assessment with an average size of 728 m^2 in urban environment. Field data are combined with secondary information such as proximity to smelter and wind rose data to model the uncertainty about soil arsenic concentrations at the parcel and block-group levels.

2. Material and methods

2.1 Study area and data available for spatial modeling

The study area covers five counties of Washington State centered on the city of Tacoma where the Asarco Company operated a copper smelter for over 100 years. Air pollution

[1]BioMedware, Inc., 121 W. Washington St., 4th Floor-TBC, Ann Arbor, MI 48104, U.S.A. Corresponding authour: Pierre Goovaerts.
[2]Gregory Glass Environmental Consulting, 7558 Brooklyn Avenue NE, Seattle, WA 98115, U.S.A.
2014 年 2 月 21 日受稿　2014 年 7 月 12 日受理

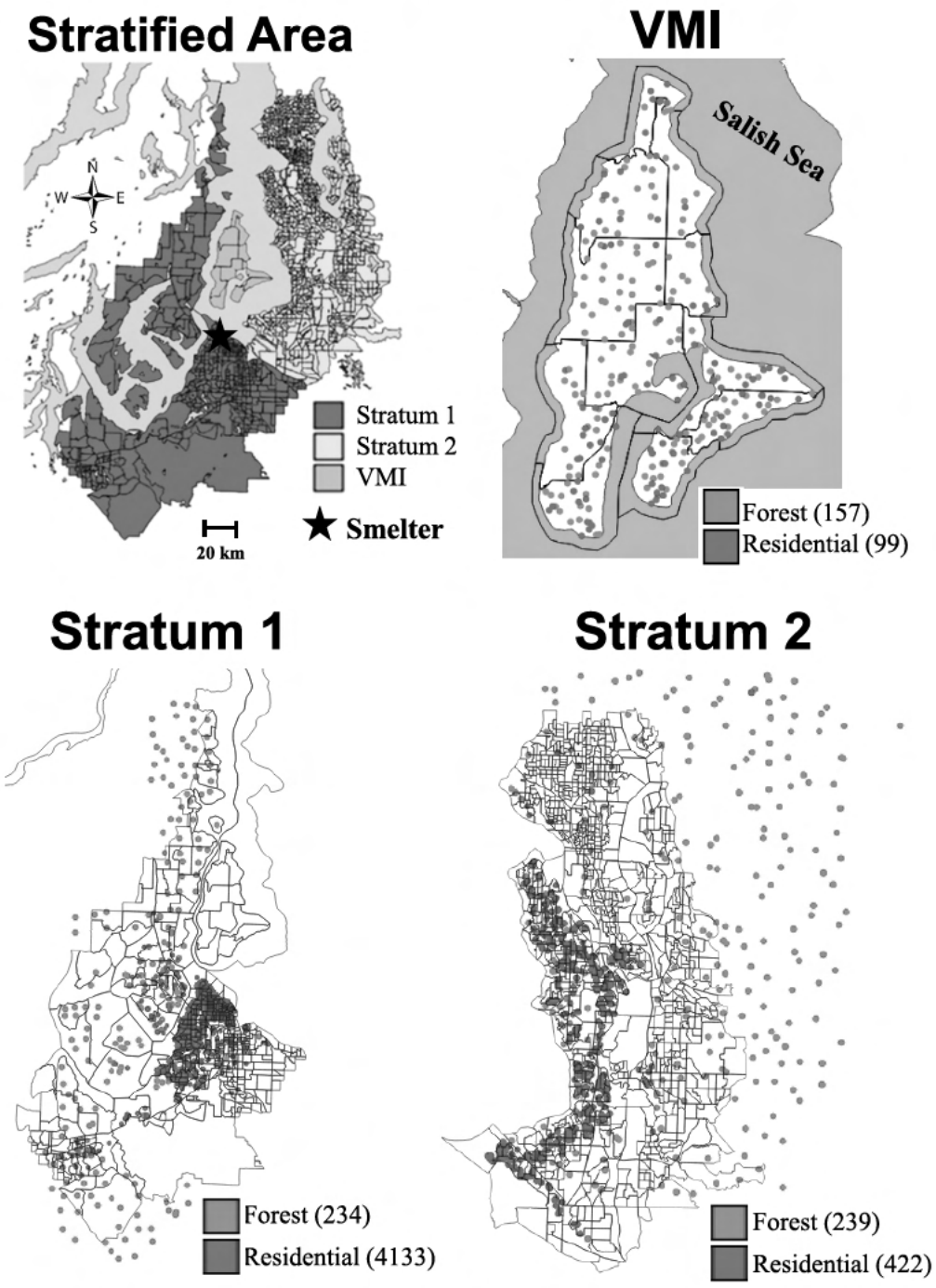


Fig. 1 Stratification of the study area into three zones (Stratum 1 & 2, Vashon Maury Island) that were modeled separately using geostatistical analysis of residential and forest samples. The polygons in each map correspond to block-groups.

from the smelter settled on the surface soil over more than 2,600 square kilometers of the Puget Sound basin. The geospatial characterization of the area was based on the following data:

1. Average arsenic concentrations measured in 4,654 residential and 630 undeveloped/forested parcels (Fig. 1). All measurements were standardized to a 0 to 15.24 cm depth for uniformity.
2. Separation distance and azimuth of each parcel with respect to the location of the smelter. This information was used as a proxy to capture the impact of atmospheric deposition on arsenic concentration. A more rigorous approach would have required the use of an air dispersion model to characterize the different types of atmospheric deposition (e.g. wet and dry depositions); yet such model was not available here. The reference point for the smelter was the tall stack. This modeling effort did not attempt to make any distinctions based on different locations for low-level fugitive emissions (e.g., at the converter building) versus tall stack emissions.

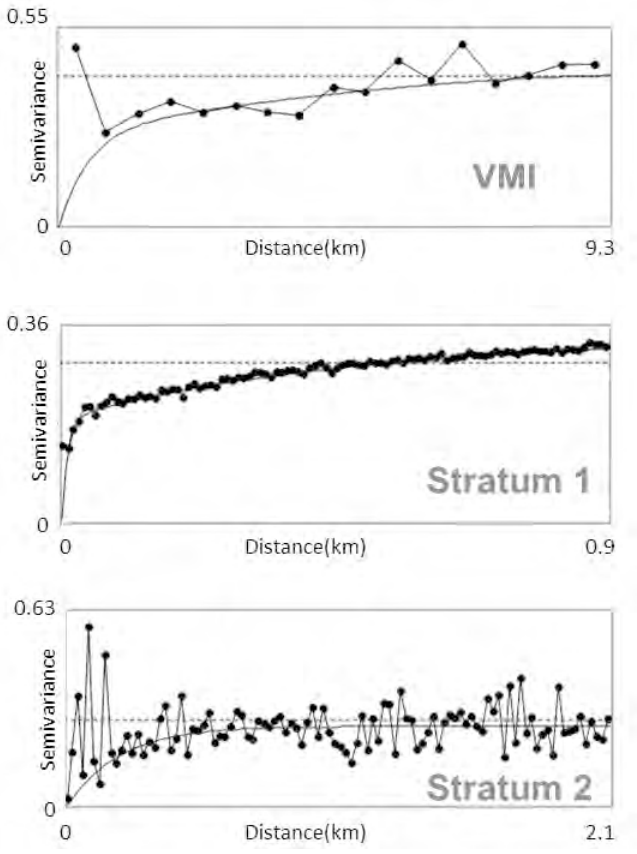


Fig. 2 Experimental semivariogram of regression residuals with the model fitted for each of the three zones (Stratum 1 & 2, Vashon Maury Island).

3. Digital Elevation Model (DEM) attributes, such as elevation, slope and aspect.

2.2 Geostatistical analysis

The study area was stratified into three zones (Fig. 1) that were deemed more homogeneous than the entire study area for modeling purposes: Stratum 1, Stratum 2, and Vashon Maury Island (VMI). On VMI, the average single family parcel is much larger than in urban neighborhood (8,094 m^2 versus 728 m^2) and these parcels generally are more predominately forested. The wind rose for the Tacoma Smelter has two dominant (seasonal) downwind directions: north-northeast and south-southwest axis, which controlled to a large extent the distribution of the contamination. The area usually gets south winds during cloudy or rainy weather, while north winds usually come with the sunny, dry weather. The division of the study area into zones is consistent with a separation of primary downwind areas according to the wind rose. The following methodology developed in Goovaerts et al. (2008) was implemented for each of the three modeling zones:

1. All arsenic concentrations (forest and residential values) were first normal score transformed (Goovaerts, 1997) to correct for the strongly positively skewed sample histogram.
2. The transformed arsenic values were regressed against covariates (explanatory variables) that were based on separation distance and azimuth with respect to the location of the smelter, as well as DEM-related attributes. This spatial trend model took the form:

$$\begin{aligned} Trend = & Intercept + b1 \times Cos(Azimuth - \theta_1) \\ & + b2 \times Log(Distance\ \ to\ \ smelter) \\ & \times |Cos\{Azimuth - \theta_2)| \\ & + b3 \times Log(Distance\ \ to\ \ smelter) \\ & + b4 \times Elevation \end{aligned} \quad (1)$$

where the offset angles θ_1, θ_2 and θ_3 were selected through a sensitivity analysis that aimed at maximizing the linear correlation between the covariate and the transformed arsenic values. For all three zones and two types of land use stepwise linear regression indicated that slope and aspect consistently did not contribute significantly to the regression model, hence they were not included in the final model. For VMI, the trend model for residential parcels was fitted using the forest data because they were more numerous, covered the island uniformly (Fig. 1) and reflect the undisturbed impact of the smelter on the soil in the island. For the two other zones, the trend model for residential parcels was fitted using the residential data. The R^2 of the regression model, which represents the percentage of the variance in arsenic values explained by the covariates, ranged from 4 % for Stratum 2 to 63 % (VMI) and 66 % (Stratum 1). The low R^2 for Stratum 2 means that the secondary information explains little of the spatial variability of arsenic values and the interpolation will be based almost essentially on the field samples. This regression model was used to predict the arsenic concentration and standard error at each sampled location and at the geographical centroid of each residential parcel.
3. The spatial variability of regression residuals (i.e. field data minus trend model) was characterized using the experimental semivariogram which describes how the average half squared difference between pairs of observations changes as a function of the separation distance. Fig. 2 shows the omnidirectional residual semivariogram with the isotropic model fitted for each of the three zones. The semivariogram is well structured for Stratum 1 which has been sampled exhaustively (4,133 soil samples). Only 99 residential samples were available for VMI, which explains the more erratic semivariogram; the value for the first lag was ignored during

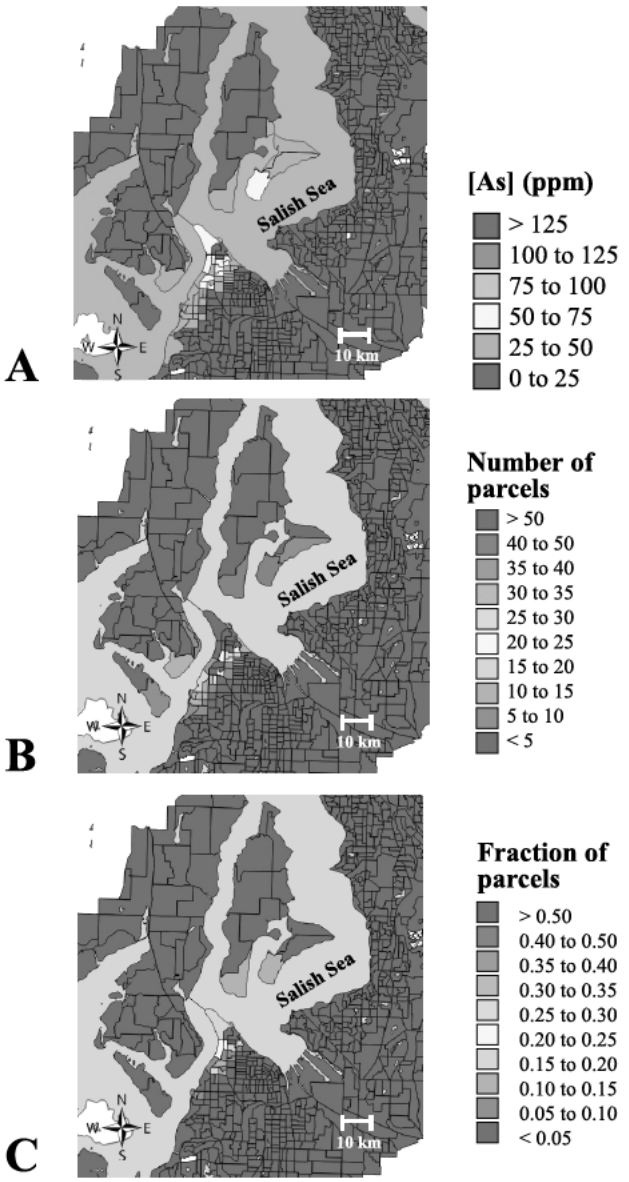


Fig. 3 Results of the geostatistical simulation approach: (A) Map of average block-group arsenic concentration, (B) Map of the average number (expected value) of residential parcels per block group that exceed an arsenic concentration of 100 ppm, and (C) Map of the estimated fraction of residential parcels per block group that exceed an arsenic concentration of 100 ppm with probability above 0.50.

the modeling because if was based on only a few data pairs.

4. Sequential Gaussian simulation (Goovaerts, 1997) was used to simulate the spatial distribution of arsenic values conditionally to the sampled data, the trend model inferred in step 2 and the pattern of correlation modeled in step 3. A set of 999 realizations were generated over all 501,505 residential parcels located in the study area.
5. Parcel-level simulated values were post-processed to compute for each of the 1,452 block-groups: 1) the mean arsenic concentration, 2) the average (expected) number of residential parcels that exceeds a threshold of 100 ppm, and 3) the fraction of the residential parcels for which the threshold of 100 ppm is exceeded with a minimum probability of 0.5.

3. Results and discussion

The end-result of the simulation procedure is a set of 999 maps of simulated parcel-level concentrations for each of the three modeling zones. The most straightforward way to summarize this information is to simply average the simulated arsenic values within each block-group and across all 999 simulations. These average block-group concentrations are mapped in Fig. 3A which illustrates the higher arsenic concentrations observed closer to the smelter and along the prevailing wind directions. There is however some uncertainty attached to these average concentrations which can be assessed empirically using the histogram of 999 simulated block-group values.

In 2009, the State of Washington received a settlement from Asarco, including a lump sum to pay for partial cleanup. The clean-up procedure requires the systematic sampling of residential parcels suspected of being contaminated followed by the remediation of the ones exceeding a target threshold of 100 ppm. Since remediation involves digging up contaminated soil and taking it to a landfill, it is more efficient to conduct the cleaning one neighborhood at a time to avoid moving heavy equipment around. Also, because of the high sampling and analysis costs resulting from the systematic investigation of all residential parcels within a block group, it is desirable to start with neighborhoods where a large proportion of residential parcels are likely contaminated to avoid a low "hit rate" (e.g. sampling 1,000 parcels to only find 5 contaminated) .

The simulation-based approach allowed the computation of different quantities that can guide the selection of neighborhoods where resources should be allocated first because of the extent and likelihood of the contamination. For example, Fig. 3B shows a map of the expected number of residential parcels per block group that exceed an arsenic concentration of 100 ppm. This map reflects the spatial distribution of the contamination displayed in Fig. 3A and the population density: number of parcels is higher in downtown Tacoma in the vicinity of the smelter compared to inland VMI where residential lots are larger. Still this map does not account for uncertainty in the prediction since it is based on an average. Uncertainty assessment is tackled by the map of Fig. 3C that shows for each block-group the estimated fraction of residential parcels that exceed an arsenic concentration of 100 ppm with a probability above 0.50.

4. Conclusions

Any quantitative assessment of soil contamination is incomplete as long as it ignores the spatial coordinates of observations. Geostatistics allows one to capitalize on the first law of geography which states that "*Everything is related to everything else, but near things are more related than distant things*" to estimate pollutant concentrations at unmonitored locations. Geostatistics offers a palette of algorithms to incorporate different types of information in the interpolation process, such as wind direction and location of the point source of contamination in the present case-study. Results can take the form of arsenic concentration maps or maps of the probability to exceed critical values, such as regulatory thresholds. An additional difficulty is the need to upscale the modeling results for residential parcels to the level of block-groups which are the geographical units considered for the prioritization of sampling and remediation sequence. This can be accomplished using stochastic simulation to generate many maps of residential parcel soil concentrations that all honor the data and reproduce aspects of the patterns of spatial dependence or other statistics deemed consequential for the problem at hand. This ensemble of realizations enables the derivation of concentration estimates and answer complex questions, such as "*What is the fraction of residential parcels exceeding a target threshold T with a minimum probability P?*", over block-groups of different size and shape as required by the decision-making process. The use of this geostatistical model in the ongoing clean-up phase demonstrates the applicability of the methodology for real-world environmental projects.

It is noteworthy that the model is known to be incomplete in some respects. First, the absence of air dispersion model forced us to use an empirical modeling approach where the impact of deposition was approximated on the basis of the separation distance and azimuth of each parcel with respect to the location of the smelter. However, the fact that the study area is highly urbanized and the smelter stopped operating 30 years ago means that the site has been greatly disturbed since the pollution event and the current local distribution of contaminants might bear little relationship to the original deposition pattern that would be predicted by an air dispersion model. On the other hand, the global trend of contamination level declining away from the smelter and at smaller pace in the downwind direction can be easily captured by the large number of samples that were collected throughout the study area. Second, since age of residence was not available for many sampled properties it is not included in the model, but it is very likely to be an important variable affecting soil arsenic concentration. A block group where a large fraction of residential parcels are for homes built well after smelter operations ended (1986), with few or no sampling results for those residences in the modeling data set, may be overestimated by model outputs. Some larger parcels characterized as residential are known to have relatively small cleared and developed areas within a largely forested or undeveloped area (e.g., VMI). The undeveloped portions of such properties are likely to have higher soil arsenic concentrations than shown by the residential models for the small developed portions of the parcels.

5. Acknowledgments

This research was funded by the Tacoma-Pierce County Health department (contract # 2011-00000044). Data were assembled by the Department of Ecology of the State of Washington.

References

Goovaerts, P. (1997): Geostatistics for Natural Resources Evaluation. Oxford Univ. Press, New York.

Goovaerts, P. (1998): Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Biology and Fertility of Soils 27: 315–334.

Goovaerts, P., Trinh, H.T., Demond, A.H., Franzblau, A., Garabrant, D., Gillespie, B., Lepkowski, J. and Adriaens, P. (2008): Geostatistical modeling of the spatial distribution of soil dioxin in the vicinity of an incinerator. 1. Theory and Application to Midland, Michigan. Environmental Science & Technology 42: 3648–3654.

Journel, A.G. and Huijbregts, C.J. (1978): Mining Geostatistics. Academic Press, New York.

要　　旨

米国ワシントン州タコマにおいて Asarco 社は，約 100 年にわたって銅精錬所を運転し，精錬所からの大気汚染物質が，Puget Sound 流域の 2600 平方キロメートル以上の表土に沈着した．タコマ精錬所汚染プロジェクトの一部として，多くの土壌ヒ素汚染の調査研究が行われ， 5000 以上の表層土壌試料が収集された．すべての住宅用土地に対して汚染物を除去する予算はなく，本研究は，これまでのデータから土壌試料の収集および修復が必要な土地について優先順位をつけるためのモデルを構築することを目的とした．本論文では，地球統計学的シミュレーションにより，地区ごとに平均でどれだけの土地がどのくらいの確率でヒ素濃度が閾値を超えるかを求めた．この情報は，汚染されている地区を選択し，地区内のすべての土地で土壌試料を収集し，設定された目標値の 100ppm を超える土地について修復を施すため，用いられている．

キーワード：確率論的空間シミュレーション，浄化，空中汚染物質，土壌サンプリング，風向