ThS_A8: Sampling methodologies

ADAPTIVE SAMPLING STRATEGIES TO IMPROVE SITE ASSESSMENT

Patrik von Heijne[1], Christian Maurice[2], Björn Öhlander[3]

Luleå University of Technology, Chemical Engineering and Geosciences, 971 87 LULEA, SWEDEN
[1]+46920 49 1504, patrik.von.heijne@ltu.se
[2]+46920 49 1755, christian.maurice@ltu.se
[3]+46920 491 478, bjorn.ohlander@ltu.se

Keywords
adaptive sampling, measurement uncertainty, uncertainty propagation, geostatistics, interpolation quality

Summary
In an adaptive sampling strategy based on multi-stage sampling, information gathered in earlier stages is used to choose what information to search for in the next stage. This should be the information that has the greatest probability to best improve the basis for the decisions to be made. Based on literature this paper discusses how to use this strategy to improve investigations of contaminated soil by assessing decision uncertainty.

The measurement error, the number of samples and choice of interpolation method are among the factors that have large effect on the usability of the outcome of a multi-stage sampling investigation using interpolation. In the light of increased usage of field-portable test instruments the interaction of measurement error and number of samples needs to be further investigated. Other issues that need attention are how to measure the quality of the basis for decision and how to decide when the quality is good enough.

Introduction

According to the European Environment Agency (2007) there are about 240 000 identified contaminated sites that needs remedial treatment within the countries of west central Europe and some countries in the south eastern Europe. Remediation of even a small fraction of these will be very expensive.

After a site has been classified as probably in need of remedial treatment the next step often is to delineate the contamination and better assess the amount of contaminant and contaminated soil. This can be accomplished by sampling, analysing and using interpolation methods. Several different methods exist in which different parameters are to be set. A common feature is that they all use results from sampled points to estimate results for unsampled points. Some of them can also assess the uncertainty of the estimations and can thus be used to decide where to take additional samples to best reduce the uncertainty.

The aim of this article is to discuss important aspects of how the use of an adaptive sampling strategy, i.e. stepwise based on the incoming information, together with geostatistical methods may improve the quality of the site investigations. The article focuses on:

1. measurement error and sampling density
2. methods to compare interpolation results
3. the interpolation result of interest to compare

Problem descriptions and methods

Support, sampling and analytical error

Each sample has a so called sample support which for a grab sample is the original size, shape, and orientation of the material in the sample before it was sampled. Soil samples from contaminated sites often weigh less than 1 kg and are in most cases several order of magnitude smaller than the size of the decision support - the size, shape and orientation of the soil volume for which a decision will be made. Englund (1987), studying how to conduct a phased sampling in a cost efficient way, has shown that if the volume of the support is increased then its variability will decrease. A consequence is that the statistical distribution for one support can not be used directly for another without risks for improper conclusions. The difference between the supports depends on the heterogeneity for volumes the size of the smaller support at the scale of the bigger support.

One way to address this issue by increasing the sample support is to take larger samples. Another way is to take composite samples as their support is larger than just the soil volume in the samples.

When a sample is taken, there is always a risk that the sample composition does not reflect the composition of the original soil volume before sampling. This difference is called the sampling error. The sampling error has many causes, e.g. some soil fall of, some of the compounds of interest evaporate, the sample is contaminated or is not taken in the exact right position. If the sample and decision supports are different and the sample is used to represent the decision unit the difference between the composition of the sample and the decision unit can be thought of as sampling error.

If it is not possible to measure the sampling error there is no way to know the real amount of compounds to be determined and thereby it is difficult to take good decisions. If the distribution of the error can be assessed then at least it is possible to predict a probability interval for the composition.

The effect of the sampling error can be reduced in three different ways by:

- using a sampling method that produces samples with a composition more alike that of the intended sample before it was taken.
- having sample and decision supports that are more alike or by handling the results in a way that better compensate for the difference. This can be possible if using simulation or kriging methods.
- sampling in a way that makes it possible to estimate the error distribution. This is possible by sampling and evaluating duplicates, as proposed by Ramsey (1992), where both samples could be used to represent the same volume.

No matter how a sample is analyzed there is always a risk of an analytical error that adds to the nuisance induced by the sampling error. Together they compose the measurement error by

$$\sigma_m^2 = \sigma_s^2 + \sigma_a^2 \qquad (1)$$

where σ_m^2 is the measurement error, σ_s^2 the sampling error and σ_a^2 is the analytic error all expressed as variances.

The ratio between the sampling and the analytical part of the measurement error is of no interest when using an analytical result. It is only the size of the total measurement error that matters. But when trying to predict or optimize the size of the error then the relative sizes of its components needs to be assessed. As Ramsey (1992) showed in equation (1), if the smaller of the two error parts (expressed as standard deviation) is increased from zero to half the larger error part it will only increase the total measurement error by about 12% (25% expressed as variance), see *Fig 1*.

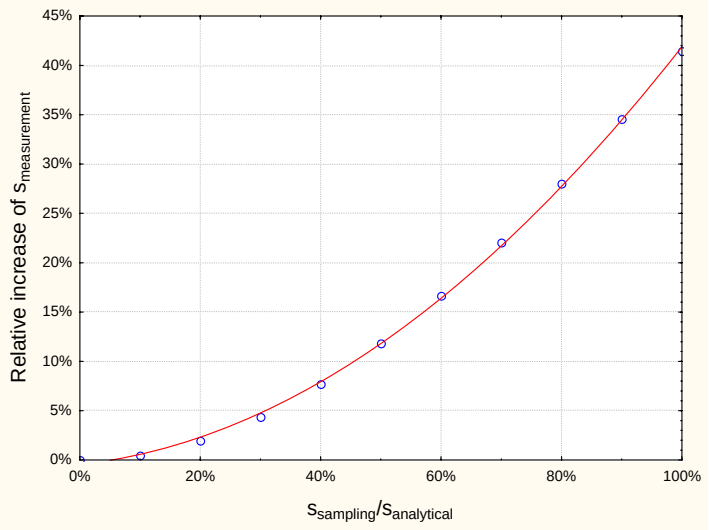

Fig 1 Effect of increased sampling error when the analytic error is constant

It is seldom cost-effective to spend energy to decrease the smaller error part to low relative levels. Instead it may pay off to look for cheaper or faster sampling or analytical techniques that may increase the smaller error part to a more adequate level. Dividing the measurement error into parts is possible with the method of sampling duplicate mentioned above. This is of special interest in the light of increased usage of field-portable instruments that, compared to sending samples to ordinary laboratories, can make faster analyses and sometimes measurements with a lower and maybe more appropriate level of error.

The quality and usability of analysis results are defined already at the planning stage, deciding the location, the size, the number of samples.

Interpolation methods and their differences

If a soil parameter has been measured for which a spatial autocorrelation exists then geostatistical interpolation can be used to estimate the level of the parameter in unsampled locations. Depending on the interpolation method that is used and the distribution of the parameter it might also be possible to assess the uncertainty of the estimation.

There are at least three different types of interpolation methods that are frequently used to estimate unsampled points of contaminated soil: deterministic, geostatistical and stochastic simulation methods. They all estimate values for unsampled points based on known values of sampled points in the adjacent surroundings called the neighbourhood. It is shaped like a circle or if the data is anisotropic, i.e. the variability of the data is different in different directions, as an ellipse. The shape of the neighbourhood as well as a minimum and maximum number of known samples to include usually has to be set. The implementations of the models generally include different possibilities to transform data, mainly to achieve a normal distribution as it is required for certain statistical analysis. After the estimations the result is back transformed before presented to the user.

The three method types differ in what input that is needed, how computer intense they are and in how and what they calculate.

- The deterministic methods use distance between sampled and unsampled points to weigh the impact nearby sampled point's values shall have when estimating those unsampled.
- The geostatistical methods, different versions of kriging, use the variability of the pair wise difference between sampled points a certain distance apart and, in anisotropic conditions, in the same direction to decide on the weight of nearby sampled point's values. In addition to calculating estimations these methods also calculate a kriging variance for each point which corresponds to the uncertainty of the estimation if the data is normal distributed, as shown by Heuvelink (2002). As both the deterministic and the geostatistical methods calculates the estimates by weighting, they smooth the variation resulting in an estimation that is less variable than the samples and thus probably the reality. Consequently these estimates should not be used for applications sensitive to variance, extreme values and continuity patterns, (Goovaerts, 1999).
- The stochastic simulation models use the known values within the neighbourhood to stochastically generate a probable estimate for the point. When this is done for all unsampled points a so called realization is created, that is a map showing one possible combination of values for all points. By repeating this procedure many times, a number of equiprobable realizations are created. As this is done stochastically they are usually all different. By evaluating the realization both an estimate and the uncertainty of that estimate can be calculated for each point or the area as a whole. The variance of the estimates is smoothed in the same way as for the geostatistical models whereas the variance for each realization will still match the one of the original samples, (Lin, 2001). Therefore, the realizations may be used in applications sensitive to variance.

Berthet (2009) studied the impact of different interpolation methods and parameter settings for two sites and found that for Inverse Distance Weighting (IDW), a deterministic method, changes of the power or neighbourhood parameters could result in differences of contaminated soil volume or contaminant mass with up to 20%. The largest difference was observed for a marshalling yard where the contamination was following the rail roads in an anisotropic way, see Fig 2.

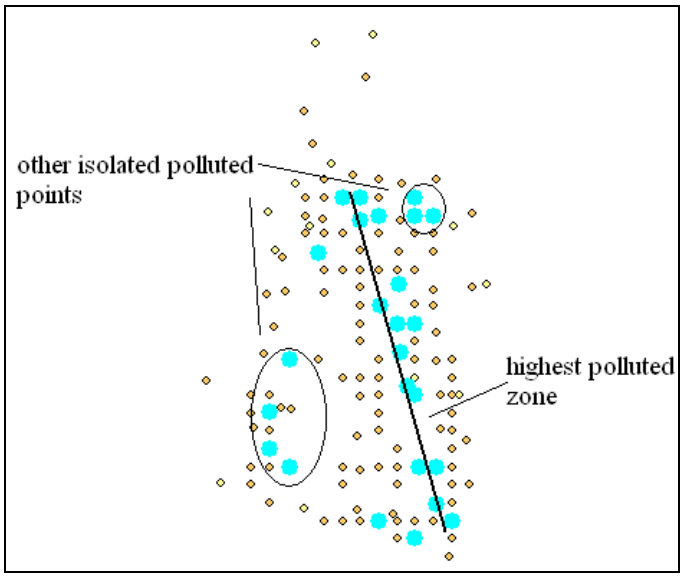

Fig 2 Visualization of the soil arsenic concentration above 40mg/kg (larger dots), (Berthet, 2009)

Comparisons of none transformation and log transformation showed differences of up to 45% indicating a possible erroneous back-transformation. The IDW is not robust to outliers, something that existed in the used datasets and might be one reason for the big differences. The greatest differences were in sparely sampled areas.

For Ordinary kriging (OK) the result of the same changes of neighbourhood settings was negligible, probably due to the method's ability to take into account the spatial structure and thereby being less sensitive to different neighbourhood settings.

As stated by Goovaerts (1999), "Provided the soil variable averages linearly, the average value over the large support can be estimated by block kriging or, indirectly, by the arithmetical mean of kriging estimates", it is possible to calculate estimates for larger supports than the sample support for all three types of models. By use of the realizations the simulation models can also calculate the uncertainty and thereby create probability maps for those lager supports.

Based on the difference in sensitivity and the type of results the methods can produce, the choice of method should depend both on the specificity of the site as well as what kind of decisions the interpolation results will be used for.

Multi-phased and multi-staged sampling

Lack of knowledge at the initial phase of a site investigation makes it difficult to know what to look for, where to look, how many samples that is needed and even what the final investigation result needs to contain. This makes it complicated to plan the investigation and is one of the reasons why most investigations are run in phases. The idea is to use conclusions from earlier phases to decide if further investigation is needed and if so to better plan the next phase.

The phases can be divided into stages where data from earlier stages are used to select the locations for a number of new samples that are most probable to give the information needed for a better assessment. This is iterated until the result is good enough for the need or optimal until the cost of continuing sampling is bigger than the return of more adequate information.

With interpolation the gathered data can be processed to assess where to increase sample density to decrease the uncertainty of the area of interest. Depending on the distribution of the parameter of interest as well as what kind of result that is needed both geostatistical and stochastic simulating methods can be used. Three crucial questions to be answered are which methods are to prefer, how to measure the quality of the result and how to decide when to stop sampling.

Measure quality of an interpolation result

How well an interpolation result resembles the truth is of interest when:

- evaluating the quality of an interpolation result
- comparing the results of different interpolation methods or parameter settings to decide which is most appropriate
- deciding when to stop iterating in a multi-stage sampling

The most straight forward comparison is the difference between the estimated and the true values. In real investigations the true value is unknown in all unsampled areas. This makes those evaluations impossible and therefore other ways of measuring the quality are needed.

Cross-validation

By removing one measurement at a time, making the interpolation and comparing the estimation at the removed point with the remove value the method's difference in that point can be calculated. The method is called cross-validation. By doing this for all measurements the overall indicator mean error (ME) (2) and root mean square error (RMSE) (3) can be calculated to express the bias and the error of the estimations.

$$ME = \frac{1}{n}\sum_{i=1}^{n} Z(x_i) - z(x_i); \quad (2)$$

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n} (Z(x_i) - z(x_i))^2} \quad (3)$$

where $Z(x_i)$ is the true value and $z(x_i)$ is the estimated value for sample i. n is the number of samples.

As stated by Armstrong (1998) and Berthet (2009) these indicators are strongly affected by outliers.

Uncertainty estimates

For more advanced assessments, e.g. in cost-benefit or risk of missing contamination analysis, the use of uncertainty estimation is needed. In Goovaerts' (2001) study about modelling uncertainty two different plots based on cross-validation are used to compare the overall uncertainty result for different simulation models.

- The Exceedence probability plot: the y-axis shows the amount of cross validation points that fall within each point's calculated probability interval. The x-axis shows the probability levels. The graph should be as near a straight line with the slope gradient as close to one as possible, e.g. for the 60% probability level about 60% of the cross-validation points should be within their probability intervals.
- The Narrowness of probability interval plot: on the y-axis it shows the average width of the probability intervals. The x-axis shows the probability levels. The slope of the graph should be positive and as flat as possible.

Stop iterating a multi-stage sampling

If the sampling is done in stages then the size of the change of results from one stage to the other can be used as an indication of when to stop iterating. This was studied by Verstraete (2008) using data

from an exhaustively investigated site - before remediation 240 plus 1005 samples where taken in two stages from a 5 ha site. He revaluated the data as a multi-stage sampling by using the first 240 samples as stage one. Of the remaining 1005 samples 200 were excluded as a control set and the other 805 were used in sets of 25 for each new stage. For each stage a sequential Gaussian simulation (SGS) was used to calculated two overall indicators of uncertainty, the median conditional variance ($me(\sigma^2)$) and the average coefficient of variation ($\overline{CV_C}$). Both indicators would have been possible to calculate if this had been a real investigation. He also used the control set to calculate the RMSE for each stage. In a real investigation this probably would not have been possible as there usually is no control set of 200 samples. The two indicators had different patterns from stage to stage and also different local minimums. By comparing the patterns against that of the RMSE the conclusion was for this site:

- The changes of $\overline{CV_C}$ was a better measure of remaining accuracy improvements than that of $me(\sigma^2)$.
- A multi-stage sampling stopping at the first local minimum of the $\overline{CV_C}$ would have needed only 365 samples to delineate the contamination with an accuracy that could not be improved significantly by further sampling.

Cost of uncertainty
These three different ways of assessing the quality of the result mentioned above all reflect the overall quality of the estimation which can be of interest e.g. when to decide if to remediate. Sometimes the quality is needed more in some parts of the site than in others, e.g. when it comes to decide which parts to remediate. Englund (1994) studied how many stages and how many samples that would be optimal to take in remediation of a specific modelled site. He used OK to calculate estimates and their uncertainties and after each sampling stage a cost function was used to assess the cost of taking the remediation decisions. The function was fed with cost information about sampling, remediation and erroneous non-remediation for the grid of remediation units. New samples were taken in units with the highest expected losses.

One conclusion that probably is valid for all sites is “The success of a two-phase design depends on the estimates from the first phase being at least approximately correct”. In the case of the used interpolation and site models about 75% of the samples were optimal to be taken in the first stage.

In cases where the impact on decision of a certain level of uncertainty conspires with the estimation this kind of approach is more adequate than the overall indicators presented earlier. Its main drawback is the cost function. In the case of remediation it needs someone to value the return of remediating and the costs of erroneously not remediating a contaminated volume of soil, e.g. what is the return of remediating if the contamination level is less than the guideline value and what is the loss of not remediating even though the actual contamination level is higher than the guideline value. The cost and return will be functions of the true concentration of the contaminant and probably difficult to set.

Discussion and conclusions
Many factors affect the quality of the result of an interpolation. Sample type, number and coordinates, analysis type and interpolation method and parameter settings are among the important factors that are controllable during an investigation and that should be used to attaining an adequate quality. Depending on the investigation phase, different types of decisions are to be made and thereby different types of interpolation results are of interest. When assessing the quality of an interpolation result to decide if further sampling is needed or to compare the results of different interpolation

methods or parameter settings, the final use of the interpolation results should be taken into consideration. If the decision is based on economic terms the usage of cost functions is one way to evaluate the correct aspect of the interpolation result.

The usage of cost functions needs relevant input which is not always available or easy to determine. Therefore there is a need for an alternative and more simple way of measuring the quality of the part of interpolation results that have effect on decisions. In the same way as the cost function it needs to increase the weighting of those parts of the interpolation results that are most critical to the decisions.

Field-portable instruments are being used in investigations of contaminated soil and as they still get better their usage will probably increase. Their main advantage to ordinary laboratory analysis is their quickness in returning result and cost of analysing. Their drawback is the higher error levels. For some of them, e.g. the XRF, the analysing time can be changed which then affect the error level. How the measurement error propagate through interpolations and how it together with the sampling density conspire to the quality of the interpolation results has not been covered in the research referenced in this paper. As the total measurement cost is depending on the number of samples and sample and analysis type those aspects are related. To make a better use, especially of field-portable instruments, it is essential with a deeper understanding of how those factors interact.

Acknowledgements

This work was supported by the European Union Structural Funds, Northern Sweden Soil Remediation Centre, EDF Objective 2, Contract MCN IO No 43173 which are greatly acknowledged.

Thanks to Ramboll Sweden for its financial support and to Loraine Berthet, Master's student at Luleå University of Technology.

References

Armstrong, M. (1998). *Basic Linear Geostatistics*: Springer-Verlag Berlin Heidelberg.

Berthet, L. (2009). Polluted site assessment using Inverse Distance Weighted and Ordinary Kriging. 59.

EEA, E. E. A. (2007). Environment and health and the quality of life. *State of the environment report No 1/2007. Europe's environment - The fourth assessment, chapter 2.*

Englund, E. J. (1987). Spatial Autocorrelation: Implications for Sampling and Estimation. *ASA/EPA Conferences on Interpretation of Environmental Data*.

Englund, E. J. (1994). Phased sampling for soil remediation. *Environmental and ecological statistics, 1*(3), 247-263.

Goovaerts, P. (1999). Geostatistics in soil science: State-of-the-art and perspectives. *Geoderma, 89*(1-2), 1-45.

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma, 103*(1-2), 3-26.

Heuvelink, G. (2002). Is the ordinary kriging variance a proper measure of interpolation error? . *Proceedings of: 5th international symposium on spatial accuracy assessment in natural resources and environmental sciences*.

Lin, Y. P. (2001). Characterization of soil lead by comparing sequential Gaussian simulation, simulated annealing simulation and kriging methods. *Environmental geology, 41*(1-2), 189-199.

Ramsey, M. H. (1992). Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance. *Journal of geochemical exploration, 44*(1-3), 23-36.

Verstraete, S. (2008). A multi-stage sampling strategy for the delineation of soil pollution in a contaminated brownfield. *Environmental pollution, 154*(2), 184-191.