Geostatistical sampling optimization of contaminated premises

Nicolas JEANNEE[1], Yvon DESNOYERS[1], Fabrice LAMADIE[2], Bertrand IOOSS[3]

[1] *GEOVARIANCES, 49bis Av. Franklin Roosevelt, BP91, 77212 AVON, FRANCE. Phone: +33 (0)1.60.74.91.00 – Email: jeannee@geovariances.com/desnoyers@geovariances.com*

[2] *CEA DEN/VRH/DTEC/SDTC/LTM, Centre de VALRHO, BP 1717 – 30207 BAGNOLS SUR CEZE, FRANCE. Phone: +33 (0)4.66.79.65.97 – Email: fabrice.lamadie@cea.fr*

[3] *CEA DEN/CAD/DER/SESI/LCFR, Centre de CADARACHE, 13108 SAINT PAUL LEZ DURANCE, France. Phone: +33 (0)4.42.25.72.73 – Email: bertrand.iooss@cea.fr*

ABSTRACT

This paper provides a methodological study of geostatistical approaches suitable to radiological evaluation in nuclear premises. By modelling the spatial continuity of activities, geostatistics provide sound methods to estimate and map surface activities, together with their uncertainty. The paper first recalls the main geostatistical principles, then investigate how it is possible to optimise the sampling strategy and avoid an exhaustive sampling, which usually leads to prohibitive costs. The kriging results of the sampling scenarios are compared to the exhaustive dataset in order to evaluate the sampling efficiency in terms of activity estimation and hot-spot detection.

Keywords: sampling strategy, nuclear premises, radiological evaluation, geostatistics, spatial structure, kriging, uncertainty.

INTRODUCTION

Decontamination of nuclear premises requires a radiological assessment of residual activity levels. From this point of view, the set up of an appropriate sampling strategy is of crucial importance. Nowadays, the radiological assessment is divided into two steps; first, a systematic (exhaustive) control of the surface activity is performed using in situ measurements methods such as in situ gamma spectrometry. Then, in order to assess the contamination depth, concrete samples are collected and analysed from drill-holes performed at several locations within the premises. Combined with historical information, such data allow to confirm a preliminary waste zoning.

In situ sampling therefore aims at answering several issues:

- Identifying surface hot spots, e.g. areas that present a surface activity greater than a target level;
- Quantifying the average residual contamination level or the residual amount of radionuclide present in the premises;
- Assessing the spatial continuity of the radionuclides within the premise, to provide a reliable estimate of activity levels.

The final goal is to set up a sampling strategy that guarantees at least the detection of given percentage of hot spots, for instance 95%. Classical statistics allow the computation of such confidence intervals and the determination of the number of samples that should be collected to reach this goal [7]. However, these techniques assume that the contamination is randomly distributed within the premises. This is the case after the decontamination process, but commonly not the case before, during the initial contamination characterization step. Actually, the initial activity usually presents a spatial continuity within the premises, with preferential contamination of specific areas or the existence of activity gradients. Taking into account this spatial continuity is essential to avoid bias while setting up the sampling plan.

In such a case, geostatistics extend the statistical framework by providing methods that integrate the spatial structure of the contamination. Once characterized this spatial structure, most probable estimates of the surface activity at unsampled locations can be derived using kriging-like techniques. Variants of these techniques also give access to estimates of the uncertainty associated to the spatial prediction, or to the probability to exceed a given decontamination threshold [2].

The ability of geostatistics to provide alternative sampling strategies to the systematic

control has been evaluated on several premises located in former nuclear installations of the CEA in France. Focusing on one of these premises, the strategy presented in this paper consists in regularly sampling only a given percentage of the surface to be characterized. For a given percentage, for instance 25%, random and regular sampling patterns are compared. Also, the added value of "sampling crosses" (several very close samples), that allow a precise identification of the contamination spatial variability, is illustrated.

Then, starting from this first sampling, it is possible to identify clean areas, hot spots and uncertain areas that require additional sampling in order to reach a target level of precision for the activity estimates. Once the additional samples are collected, the estimates and the classification of clean/contaminated areas are updated.

The efficiency of these sampling strategies is evaluated on several bases: statistical criteria (bias, mean squared error, etc), empirical comparison of the estimated activity map with the systematic control case, ratio of number of measurement points over the proportion of identified hot spots.

MATERIAL AND METHODS

Investigated Area

In 1996, the "9095est" wall of the AT1 facility (CEA, La Hague) was evaluated using an in situ detection system. That measurement system provides α and β average activities for a 300 cm^2 surface (10 cm x 30 cm). The detection limit (DL) is about 0,1 Bq/cm^2, which correspond to 30 seconds measurement time. This paper is only based on α contamination (Fig. 1).

We focus here on surface activities, not considering a possible migration of the contamination in the concrete. The contamination depth is the objective of the second step with concrete samples.

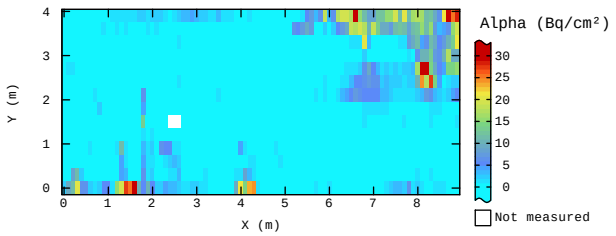

Fig. 1: α contamination: surface activity in Bq/cm^2 for each mesh of 300 cm^2.

So as to evaluate the different α activity levels, the activity statistical distribution is classically presented trough an histogram (Fig. 2), that shows a strong dissymmetry partly due to a large number of measures below the detection limit.

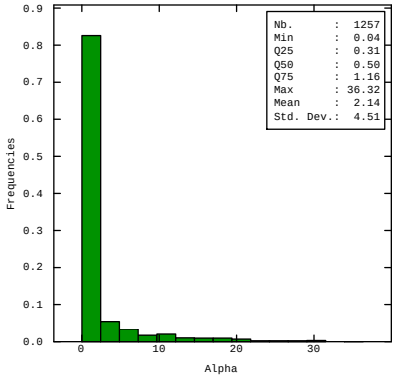

Fig. 2: Histogram and elementary statistics for α activity in Bq/cm^2 for each mesh of 300 cm^2.

Geostatistics

Spatial Variability and Variogram

The whole interest of the geostatistical methodology consists in taking into account the spatial continuity of the phenomenon to predict it at unsampled locations, together with an estimate of the prediction uncertainty. The characterization of this spatial continuity, or spatial variability, is obtained through the calculation and the model fitting of the variogram, a step called "variographic analysis" [6][1].

The experimental variogram is calculated by averaging, in distance classes, the variability contribution of each measure couple; this contribution is usually quantified by the half squared difference of the measured activities. Generally, in the study of a structured phenomenon, spatial variability increases with distance and tends to stabilise from a distance named "distance of maximum correlation" or "range". Measures separated by a larger distance are no longer spatially correlated.

The experimental variogram for α contamination is shown on Fig. 3, the calculation lag is 0.3 m, equals to the vertical sampling mesh. The evolution of the spatial variability is clearly identified: the spatial variability quickly increases up to 1 m and then more gradually, which signifies the existence of important changes of activity levels for a distance of 1 m.

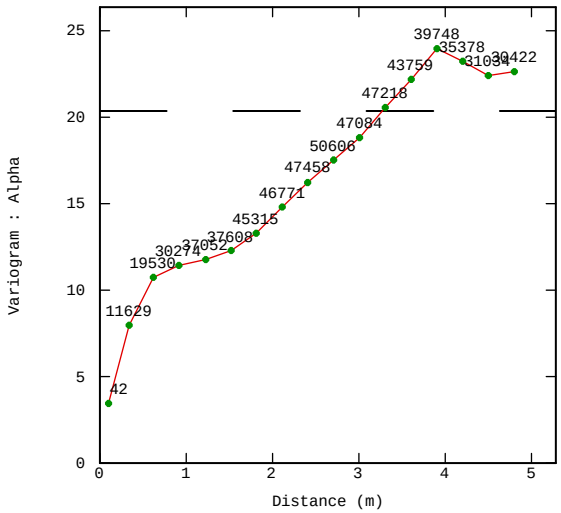

Fig. 3: Experimental variogram for α activity. Indication of the number of pairs used for each experimental point calculation and the statistical variance in dash.

The existence of a spatial structure makes the classical statistical approaches inadequate as they assume spatial randomness of the measured activities.

Such a spatial structure can be masked due to the presence of a few very high values (which is usually the case for nuclear contamination as shown by the strong dissymmetry of the histograms). The variability produced by these values is preponderant and modifies the average spatial variability described in the experimental variogram. For example, on another nuclear facility (Rapsodie, CEA Cadarache), the influence of the removal of the two highest values is shown on the next figures (Fig. 4 and Fig. 5). However, even if the removal of these high values decreases the global variability of the phenomenon (the statistical variance decreases), the spatial structure remains unchanged.

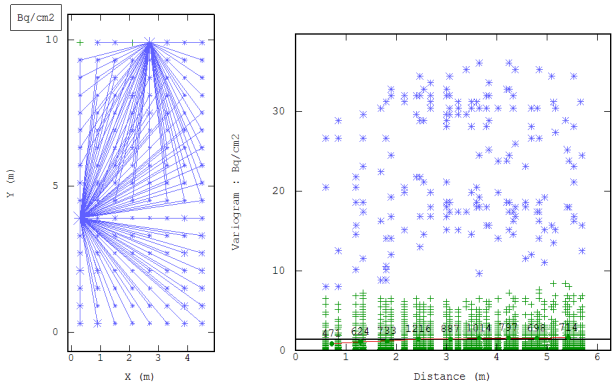

Fig. 4: Surface activity base map (left) and variographic cloud (right) in the Rapsodie reactor. In blue, pairs of points with a high variability.

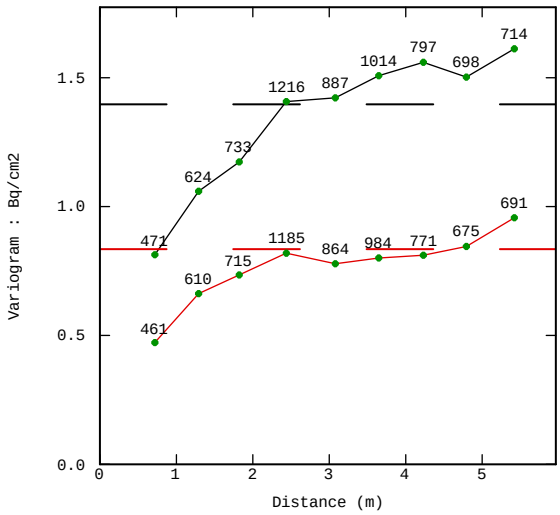

Fig. 5: Experimental variograms in the Rapsodie reactor: with (black) and without (red) the two highest values. Indication of the corresponding statistical variances in dash.

In addition to the temporary removal of the highest values, an other exploring approach of the spatial structure consists in transforming the raw data so as to decrease the influence of high values. Two transformation are commonly employed : a logarithm transformation or a Gaussian transformation (intuitively, the raw histogram is deformed to become a Gaussian one, see below). The resulting data is usually more spatially structured than the raw data and the analysis of the transformed data makes easier the spatial structure determination.

Variogram Model Fitting

The kriging procedure requires a model fitting of the experimental variogram. Indeed, the spatial variability should be known whatever the distance and should integrate the a priori information on the phenomenon, which is not always illustrated by the measures. Of course, different but consistent variogram fittings are possible. One example is presented in Fig. 6. It correctly adjusts the experimental variogram until a distance of 2 m. This is perfectly satisfying due to the high density of measures: for any point to be estimated, there will be no need to take into account measurement points more than 2 m away.

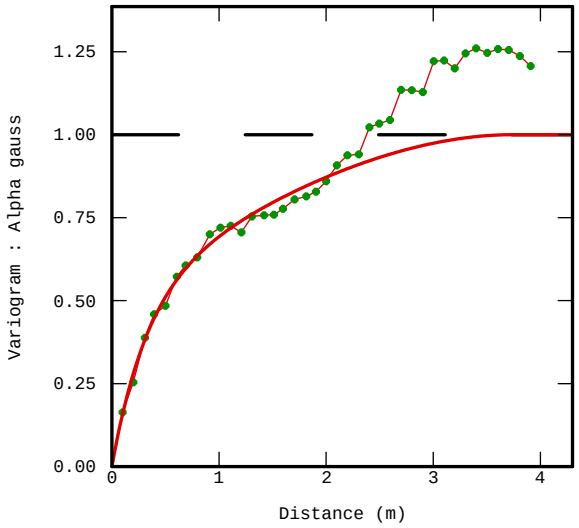

Fig. 6: Experimental variogram and its model fitting of the Gaussian transformation of α activity on "9095est" wall.

Estimation using Kriging

The data and the previously fitted variogram model lead to the surface activity estimation through the kriging procedure. As for classical interpolations, the kriging estimation of the surface activity at a point x_0, written $\hat{z}(x_0)$, is calculated using a linear combination of the *n* known activities at measurement points: $\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i)$.

The kriging estimation differs from other interpolators in the choice of the λ_i coefficients. Named kriging weights, they depend on:

- the distances between the data and the point to be estimated (as for classical interpolators)
- the distances between the data (clusters…),
- the spatial structure of the studied phenomenon. Indeed, for equal geometric data configuration, the cartography should integrate by construction the phenomenon spatial continuity (for example, very smooth or heterogeneous).

Knowing the data configuration, the position of the point to be estimated and the variogram model, the kriging procedure consists in determining the kriging weights, only unknown values of a linear system to be solved. These weights assure:

- an unbiased estimation,
- the minimization of the estimation error variance $\mathrm{Var}[Z\text{-}\hat{Z}]$, which intuitively corresponds to the minimization of the error risk.

From this point of view, the kriging technique is the best linear unbiased estimator [5].

Practically, the estimation for a given point can be realized using the whole set of data (unique neighbourhood) or using only the closest points of the target point (moving neighbourhood).

In the case of the "9095est" wall, the exhaustive data set makes the surface activity interpolation useless. However, the spatial prediction is necessary in the case of a non-exhaustive sampling of the area to be estimated. Fig. 7 illustrates the kriging results realized with the variogram model on Fig. 6, on the hypothesis that only 1 measure out of 10 is known, regularly located on the area.

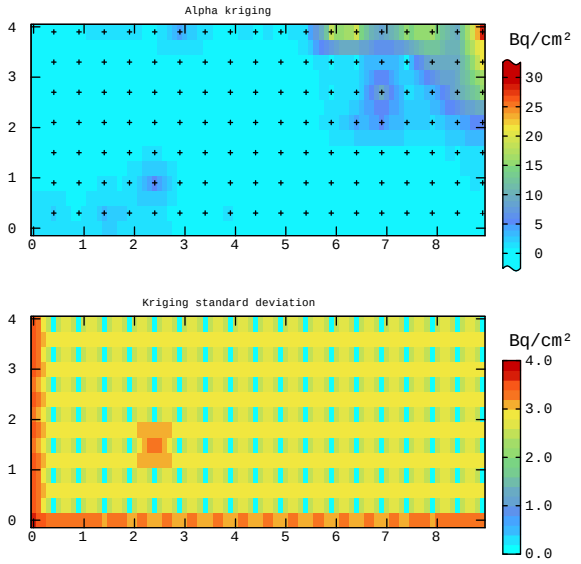

Fig. 7: Maps of the α activity kriging estimates from 1 out of 10 measures (top) and associated kriging standard deviation (bottom).

The whole study of the error analysis due to partial sampling is the object of the next part of this article.

Kriging Standard Deviation

In comparison to classical interpolators, geostatistical estimation techniques provide the quantification of the related estimation uncertainty. This quantification is possible due to the spatial variability modelling.

Uncertainty is usually illustrated with the kriging (error) standard deviation map, as shown on Fig. 7 for α contamination:

- it takes minimal values close to data points, the estimation confidence is high;
- it increases with the distance between the target point and the data points, in function of the chosen variogram model.

Kriging standard deviation depends exclusively on the data geometric configuration and the variogram model. This is its main advantage as it can be a priori calculated knowing the contamination spatial structure. This is also its limit because it does not consider the uncertainty variation related to activity levels.

Uncertainty Quantification

Generally, the kriging standard deviation map is a good indicator of the estimation quality. The Gaussian framework enables a more quantitative use of kriging standard deviation so as to get a real standard deviation or a confidence interval for the radionuclide activity.

Minimizing by construction the error risk, the kriging techniques smooth the extreme values, unlikely to be estimated, and tend to bring them to the local contamination average. This is the kriging smoothing property: the contamination real spatial variability is not reproduced through the interpolation step. Consequently, the resulting kriging map cannot be directly used to estimate the probability to exceed a cut-off and the corresponding contaminated surface.

The kriging procedure does not need any hypothesis about the statistical distribution, Gaussian or not, for the data to be estimated. However, non linear techniques and quantitative calculations of the estimation uncertainty require multi-normal distribution [8]. So the raw data should be Gaussian, which is generally not the case. The raw statistical distribution is then transformed into a Gaussian one via a transformation function called “anamorphosis” (See Fig. 8 for an example).

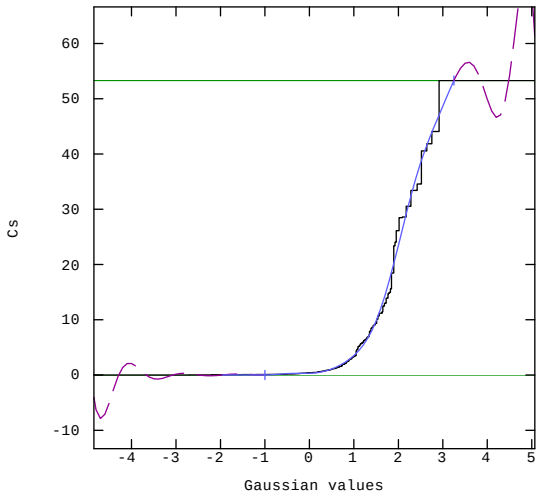

Fig. 8: Gaussian anamorphosis example: experimental (black) and modelled (blue) functions.

Some risk analysis results can be directly drawn from the theoretical properties of the Gaussian distribution. In particular, this is the case for the probability to be above cut-off calculations or for a confidence interval.

Considering the gaussian transformation function φ (anamorphosis), the involved cut-off *s*, the simple kriging (with known mean) of the Y^K radionuclide gaussian transformation and the related kriging standard deviation σ_K, the probability for the radionuclide *Z* to be above *s* is given by:

$$\mathrm{P}\left[Z(x) \geq s\right] = \mathrm{P}\left[Y(x) \geq \varphi^{-1}(s)\right] = 1 - G\left(\frac{\varphi^{-1}(s) - Y^K(x)}{\sigma_K(x)}\right),$$

where G is the density function of the normal distribution.

In the same way, with the gaussian transformation kriging estimate and the related standard deviation, the confidence interval can be directly calculated in a statistical meaning. For example, the two bounds for a 95% confidence interval are calculated using the following formula:

$$\mathrm{CI}_{95\%} = \left[\varphi\left(Y^K(x) - 1.96\sigma_K(x)\right), \varphi\left(Y^K(x) + 1.96\sigma_K(x)\right)\right]$$

Application for α activity

The 95% confidence interval bounds are calculated using the anamorphosis transformation, as mentioned above. It is then possible to derive a qualitative uncertainty map by computing the difference between the confidence interval bounds or a standard deviation (difference between the $CI_{95\%}$ bounds divided by the 3.92). The figure below (Fig. 9) illustrates such a map, for the sampling presented in Fig. 7.

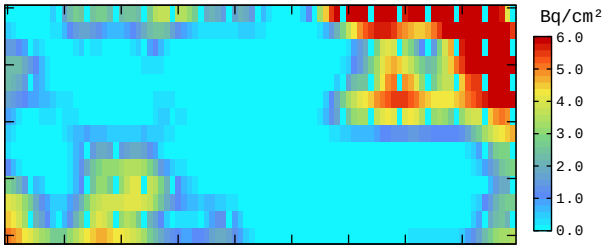

Fig. 9: Uncertainty map for α contamination (1/10 sampling strategy).

For α contamination, uncertainty is logically higher in high activities zones due to the anamorphosis shape.

Complementary investigations can be based on this kind of uncertainty maps so as to reduce the high uncertainty zones (see below).

RESULTS AND DISCUSSION

In order to evaluate a non exhaustive control of the investigated area, numerous partial sampling strategies are considered, supposing known only 1 mesh out of 2, then 1/4, 1/6 and 1/10. These measurement points are supposed to be the only available data and they are used in order to estimate the contamination level and the related uncertainty. In a second step, it will be possible to sample the locations where the estimation uncertainty exceeds a given threshold and then to update the contamination estimation.

Several scenarios are presented and then evaluated using several criteria, by comparison to the exhaustive dataset: statistics and variogram reproduction, computation of a 95% confidence interval and of the percentage of detected hot-spots.

Regular Sampling Sets

Statistical Distributions

The first evaluation of the sampling sets quality consists in comparing the statistics differences for α contamination between each scenario and the exhaustive dataset (see Fig. 2 for the corresponding statistics).

Due to the strong dissymmetry of the statistical distribution, comparisons are based on robust statistics like median (Q50) and inter-quartiles difference (Q75-Q25). Classical statistics (mean, standard deviation) are presented in the same manner in Table 1. Only the average results are presented, grouped by sampling density.

Table 1: α contamination – Average relative statistic differences between a sampling strategy and the exhaustive dataset.

Sampling strategy	Q50	Q75-Q25	Mean	St. dev.
1 out of 2	2%	2%	3%	2%
1 out of 4	3%	10%	3%	5%
1 out of 6	9%	12%	7%	5%
1 out of 10	7%	14%	8%	6%

The quality of the statistics relative to the different sampling strategies is globally good. This can be explained by the large number of data even for the 1 out of 10 strategy where more than 100 points are still available.

Some of the 1/6 or 1/10 scenarios show very close results with the complete set. However, the global statistics are worse than for the 1/2 and 1/4 strategies ones.

Structure Identification

As well as global statistics, it is essential to correctly capture the contamination spatial structure, in particular for short distances, so as to perform a correct estimation using the geostatistical methodology. As usual, this spatial structure is analysed through the experimental variogram. The variograms, grouped by sampling density, are presented in Fig. 10.

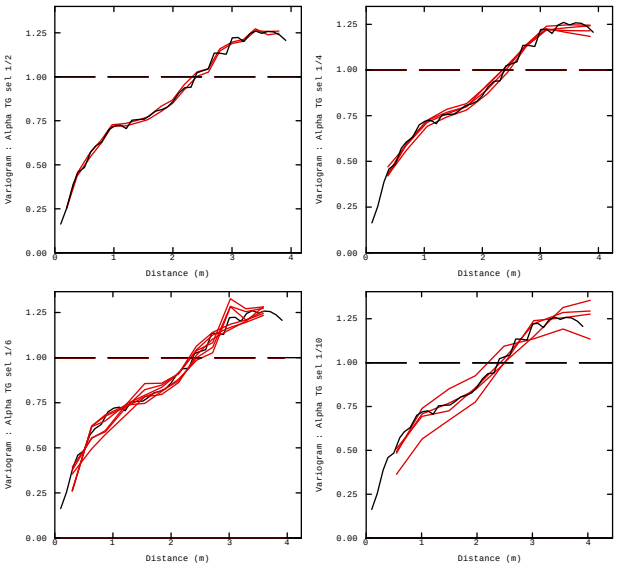

Fig. 10: Experimental variograms for α contamination gaussian transformation for regular sampling sets.

The denser the sampling strategy, the better the spatial structure reproduction. As previously, the 1/2 and 1/4 strategies present good results. The 1/6 and 1/10 strategies are less precise, in particular for short distances. This lack of information has significant consequences on further kriging estimation.

Sampling crosses, presented in a next paragraph, will constitute a possible solution to improve the estimation of the spatial structure at short distances.

Bias and Precision of Estimates

With the objective of evaluating the different sampling strategies, only the selected points are considered so as to determine the spatial structure of the contamination. The experimental variograms are independently adjusted. Then, in relation to the exhaustive data set, it is possible to quantify the estimation error calculated at the points which are not selected (at validation points). In Table 2, the different kriging estimates $\hat{Z}$ are compared to the true value Z, using the Bias criterion and the

Root Mean Squared Error (RMSE) criterion defined by:

$$Bias = \frac{1}{n_v}\sum_{i=1}^{n_v}(Z_i - \hat{Z}_i)\ ,$$

$$RMSE = \sqrt{\frac{1}{n_v}\sum_{i=1}^{n_v}\left(Z_i - \hat{Z}_i\right)^2}\ ,$$

Where n_v is the number of validation points.

Table 2: Statistic evaluation of the different strategies at validation points, global bias and root mean squared error (RMSE).

Sampling strategy	Alpha	
	Bias	RMSE
1 out of 2	0.15	1.95
1 out of 4	0.20	2.38
1 out of 6	-0.02	2.85
1 out of 10	0.00	3.02

Firstly, the estimation bias is systematically close to 0, which confirms the unbiased estimation of activity levels.

The error dispersion, quantified by the root mean squared error, gradually increases with the sampling degradation. The results for each scenario are not presented here but the statistics are almost the same for a given degradation strategy. This confirms that the sampling resolution is far more important than the starting points for the mapping global quality.

Identifying Hot-spots

A 95% confidence interval is also associated to the estimation. For each sampling strategy, global statistics are reported for true values outside this interval of estimation, in particular if the true value is higher than 0.4 Bq/cm² (see Table 3). This activity threshold will define the points considered as contaminated.

Table 3: Evaluation of the number of points outside the estimation 95% confidence interval. Percentages in relation to the exhaustive data set.

Sampling strategy	Alpha	
	Out. CI	Out. CI > 0.4
1 out of 2	6%	3%
1 out of 4	6%	3%
1 out of 6	11%	5%
1 out of 10	10%	5%

Theoretically, in all cases, the percentage of values outside the 95% confidence interval should be close to 5%. Significant differences exist, in particular if the sampling density is low.

A more advanced analysis shows that these differences derive from an evaluation error of the spatial structure at short distances. The error risk becomes higher while the sampling density decreases. For example, we consider the second scenario of the 1 out of 10 strategy. For α contamination, the values outside CI percentage are around 1%. The corresponding variogram model for this percentage is presented in Fig. 11. A significant nugget effect is used, consistent with what the experimental variogram shows. Nevertheless the nugget effect is probably over-estimated. The variability over-estimation results in the artificial widening of the confidence interval, and leads to the low number of points outside this CI. If we do not consider any nugget effect, the confidence interval is reduced and the number of points outside the CI increases. Then the percentage reaches 5% of the validation points.

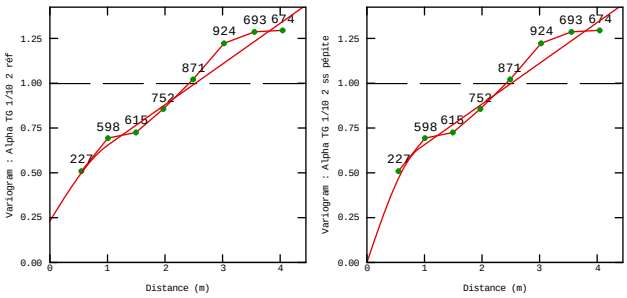

Fig. 11: Experimental variograms for α contamination for sampling set "1/10 2" and model fittings with (left) or without (right) nugget effect.

This simple example emphasises the importance of a well-characterized spatial structure, precisely at short distances, in particular if the evaluation objective is a precise confidence interval of the estimate.

To pursue the reasoning, it should be possible to use a "percentage of points outside the $CI_{95\%}$ equals to 5%" criterion so as to guide the model fitting choices. But from a practical point of view, this approach seems very fastidious to carry out.

Regular or Random Sampling?

Until now, only regular sampling strategies are considered as they provide more reliable information for the different distance classes in order to calculate the experimental variograms. However, classical statistical approaches generally recommend a random sampling. This paragraph aims at considering such sampling schemes and to compare them to the results of a same sampling effort, regularly located in the area.

The exercise is done on the 1/4 sampling effort. Four random scenarios, that constitute a partition of the exhaustive dataset, are considered and compared with the four regular ones (Fig. 12).

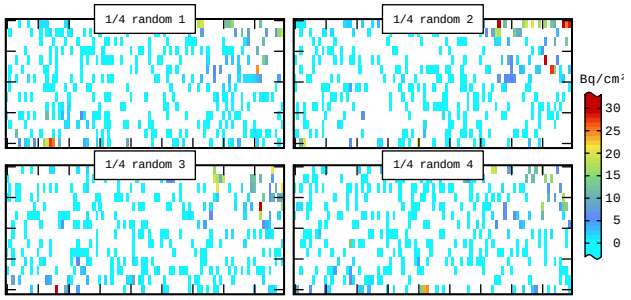

Fig. 12: 1/4 random sampling plans for α contamination.

As for statistical distribution estimation, the four random sampling plans present results close to the regular sampling ones. However the variability between the different scenarios is higher than what is observed for the regular ones (Table 4); for example with the Q75-Q25 interval, the differences with the exhaustive dataset vary between 5% and 16% for regular scenarios and between 4% and 27% for random scenarios. This is probably the consequence of the spatial structure of the contamination as, in the case of a random data, random sampling theoretically produces better results, which is not the present case.

Table 4: 1/4 random sampling statistics. Differences with the complete set for α and β contaminations.

	Q50	Q75-Q25	Mean	St. dev.
1/4 random	5%	15%	9%	11%
1/4 regular	3%	10%	3%	5%

Regarding spatial structure (experimental variogram), random sampling provides more information for very short distances as randomly chosen points can be spatially very close. This is not possible with a regular sampling. However, the variations around the experimental variogram of the exhaustive set are more significant (Fig. 13).

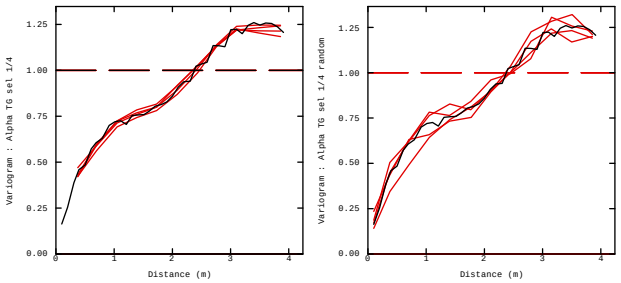

Fig. 13: α contamination - experimental variograms for 1/4 regular (left) or random (right) sampling strategies

The random strategy statistics are globally comparable to regular ones as 5% of the validation points are outside the $CI_{95\%}$ and 3% for values > 0.4 Bq/cm², for α activity.

All these considerations lead to the classical recommendation in geostatistics to realize, if possible, regular sampling.

Added Value of Sampling Crosses

The previous sections remind the importance of a well-characterized spatial structure for short distances in experimental variograms. Consequently, a classical sampling recommendation in geostatistics consists in adding a few sampling crosses so as to locally refine the regular sampling.

In order to illustrate the added value of sampling crosses, we start with the 1/4 regular sampling and then 9 and 18 sampling crosses are added. The eight corresponding sample schemes are presented in Fig. 14.

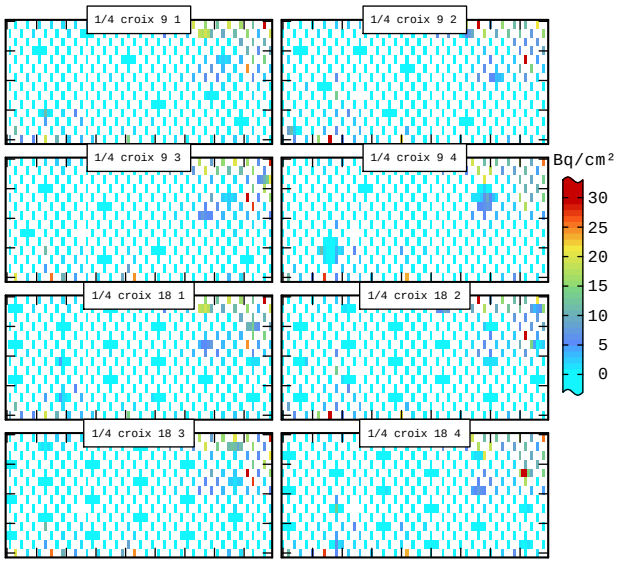

Fig. 14: α contamination – 1/4 sampling strategies completed with 9 or 18 sampling crosses.

As for the spatial structure of the contamination, Fig. 15 shows that the variograms are clearly better defined for short distances thanks to the sampling crosses.

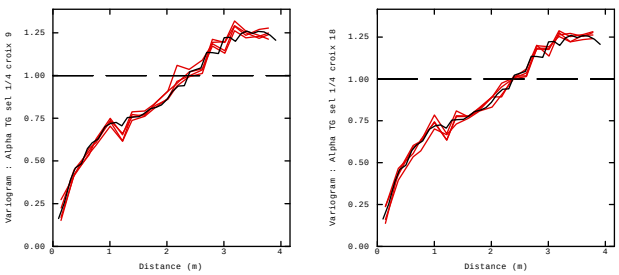

Fig. 15: α contamination – Experimental variograms for a 1/4 sampling strategy completed with 9 (left) and 18 (right) sampling crosses.

Complementary Sampling of Uncertain Areas

Whatever the sampling strategy is, it remains impossible to reduce the risk of hot spot non detection to nil, except by the use of an exhaustive sampling and if we assume no sampling error.

However, for a given sampling resolution, for example 1/4 of the area to be characterized, it is possible to map the contamination estimation with its uncertainty. Such uncertainty maps may give directions for complementary investigations.

One possible solution is to re-investigate the area where the uncertainty is high. Nevertheless, the main goal is not to reduce estimation uncertainty everywhere, but to be able to decide, with a confidence level, on the risk for activity to be over a given threshold, for example 0.4 Bq/cm².

For example, no complementary samples will be carried on a zone where the activity estimation is 30 Bq/cm² with 5 Bq/cm² kriging uncertainty for a 0.4 Bq/cm² threshold. As areas of large kriging uncertainty are located where the activity values are high, no additional information for the risk to be over the interest threshold of 0.4 Bq/cm² will be added by re-sampling these areas.

Therefore, complementary investigations based on a risk to be over 0.4 Bq/cm² map looks more appropriate. This probability map underlines three different zones (see Fig. 16):

- Zones where the activity is bound to be over the threshold (with a given confidence level);
- Zones where it is bound to be under the threshold;
- Zones where the risk is intermediate, then a reliable decision cannot be taken.

Complementary investigations are then orientated towards these intermediate zones where the risk uncertainty is high.

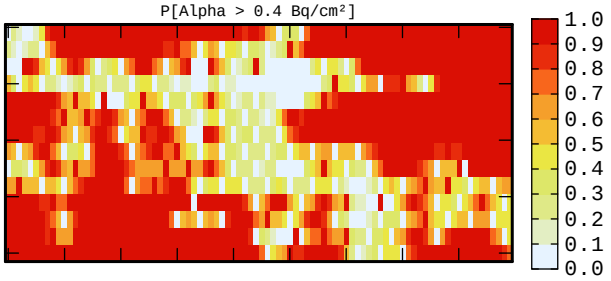

Fig. 16: Risk to be above 0.4 Bq/cm² map for α contamination. 1/4 regular sampling with 18 sampling crosses.

On Fig. 16, with only 29% of investigated area and a 95% confidence level, more than 54% of the meshes are characterized towards the 0.4 Bq/cm² threshold, meaning that the 95% confidence interval attached to the prediction is entirely above or below the threshold.

Considering the probability histogram and if we assume a maximum probability decision rule (a mesh with a probability larger than 0.5 is considered as contaminated and otherwise clean), the following results can be calculated:

- 85.7% of the area is correctly characterized,
- 11.5% of the meshes are assumed contaminated while they are actually clean,
- 2.8% of the meshes are considered clean while contaminated (4.6% of the hot spots).

Complementary investigations then depend on the different confidence levels that can be accepted. For example, if new samples are located on meshes where the risk to be above 0.4 Bq/cm² is between 5% and 50%, this means that a mesh with a 50% or more risk is considered contaminated and only low risk meshes are re-investigated. Other confidence intervals are also possible (10% and 50%, 20% and 80%). The choice only depends on the risks that remain acceptable.

With the choice of measuring all points for which the probability to be above 0.4 Bq/cm² is between 5% and 50%, 233 new points are measured, 47.7% of the area is finally investigated and the estimates and the classification of clean/contaminated areas are updated:

- 92.4% of the area is correctly characterized,
- 5.5% of the meshes are assumed contaminated while they are actually clean,
- 2% of the meshes are considered clean while contaminated (3.4% of the hot spots).

It is then possible to update the risk to be above the threshold taking into consideration the second sampling data. This leads to a new map (Fig. 17). The procedure can be iterated as long as the results are not satisfying. In our case, we stop with the second sampling.

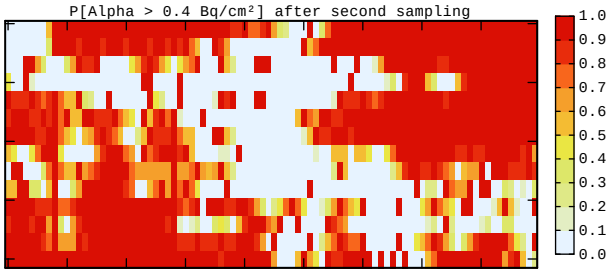

Fig. 17: Risk to be above 0.4 Bq/cm² map for α contamination taking into account the second sampling data.

Such a sequential strategy seems to be quite efficient as a high percentage of the area (50%) is not investigated with very low error results (2% and 5.5%) due to the use of the contamination spatial structure.

CONCLUSION

This paper has recalled the principles of geostatistics and demonstrated how it provides promising alternatives to the exhaustive sampling of contaminated premises.

The relevance of such approaches relies on the presence of spatial continuity for the radiological contamination. In this case, geostatistics provide pertinent activity estimates.

Besides, different sampling strategies have been analysed and compared: partial regular sampling, random and regular sampling patterns and also the added value of "sampling crosses". This work demonstrates the advantages of regular sampling strategies rather than random, coupled with short distance investigation such as sampling crosses.

Moreover, with non exhaustive data sets, geostatistics provide interesting and useful results such as uncertainty and probability maps. The risk to be above a given threshold can be calculated so as to guide new investigations on the area.

The approach has been applied on recently investigated premises, where an exhaustive sampling of the surface activity would have lead to prohibitive costs. The conclusions of this study have been employed so as to optimise the sampling plan taking into consideration the contamination spatial structure. More geostatistical treatments are currently carried on to use all available pieces of information: the historical context and preliminary investigations are taken into account. Likewise, a multivariate geostatistics approach links surface measurements and in situ gamma spectrometry measurements. Multivariate geostatistics is used to highlight structural relations between variables, improve their estimates and ensure their consistency.

The next step of our work will be the integration of destructive samples in relation to the in-situ measurements in order to assess the contamination depth for the decontamination step. Contaminated volumes and their location in the premises will possibly be calculated using conditional simulation based on the contamination spatial structure. The main objective remains the waste production minimization and the control of the decontamination workings costs.

The final aim of the ongoing research work will be to integrate these geostatistical methodologies within a software dedicated to help decision-making and sampling optimisation of contaminated premises and soils during decommissioning and dismantling of nuclear installations.

ACKNOWLEDGMENTS

All geostatistical calculations and graphics have been obtained using the ISATIS software [3]. This work was supported by the Decommissioning project (CEA/Nuclear Energy Division) . The authors would like to thank Christophe Le Goaller and the "Département de Démantèlement et de Conduite d'Opérations" (called today "Département des Projet d'Assainissement-Démantèlement") of CEA Marcoule who enable this methodological study and for the provided data sets.

REFERENCES

1. Chilès J.P., Delfiner P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, Wiley series in Probability and Statistics, New-York, 695p.

2. Géosipol (2005), *Apports pratiques de la géostatistique à la caractérisation et à la gestion des sites et sols pollués*. Rapport technique ADEME-INERIS, 139p.

3. Geovariances (2008), *Isatis technical references*, version 8.0, 148p.

4. IRSN (2001), *Gestion des sites industriels potentiellement contaminés par des substances radioactives*. Version 0, Guide DPPR-BRGM.

5. Isaaks E. H., and Srivastava R.M. (1989), *An introduction to applied geostatistics*. Oxford University Press, 561p.

6. Matheron G. (1989), *Estimating and Choosing* - An Essay on Probability in Practice, Springer Berlin, 141p.

7. Pérot N., and Iooss B. (2008), *Quelques problématiques d'échantillonnage statistique pour le démantèlement d'installations nucléaires,* Actes de la conférence $\lambda\mu$ 16, Avignon, France, octobre 2008, soumis.

8. Rivoirard (1994), *An introduction to disjunctive kriging and non linear geostatistics*. Clarendon Press, 182p.