Improved Nuclear SIte characterization for waste minimization in DD operations under constrained EnviRonment

Research and Innovation action
NFRP-2016-2017-1

WP3 – Sampling strategy State-of-the-art report

Deliverable D3.1

Version # 2

Authors: Nadia Pérot (CEA), Yvon Desnoyers (Geovariances), Grégoire Augé (ONET), Frédéric Aspe (ONET), Sven Boden (SCK•CEN), Rogiers Bart (SCK•CEN), Oleksandr Sevbo (Energorisk), Olaf Nitzsche (Brenk), Joost de Groot (Brenk)

http://www.insider-h2020.eu

This project has received funding from the Euratom research and training programme 2014-2018 under the grant agreement n°755554

Document Information

Grant Agreement #:	755554
Project Title:	Improved Nuclear SIte characterization for waste minimization in DD operations under constrained EnviRonment
Project Acronym:	INSIDER
Project Start Date:	01 June 2017
Related work package:	WP 3: Sampling strategy
Related task(s):	Task 3.1: Statistical approach status
Lead Organisation:	CEA
Submission date:	19 December 2017
Dissemination Level:	Public

History

Date	Submitted by	Reviewed by	Version (Notes)
2017/12/19	Nadia Pérot	WP3 participants	
2018/01/11	Nadia Pérot	Danielle Roudil	#1
2019/11/18	Nadia Pérot	End User Group	#2

Table of contents

List of Tables

List of Figures

Summary

The main objective of work package 3 (WP3) is to draft a sampling guide for initial nuclear site characterization in constraint environments before decommissioning, based on a statistical approach. The first task consisted in an overview of the available sampling design methods described in standards and guides followed by a brief presentation of the statistical methods that can be used in the present context according to the objectives specified for initial nuclear site characterization. That overview was done by selecting state-of-the-art techniques for sampling design optimization, using prior information and multiple iterations, testing the approach through different case studies and reviewing the feedback from overall uncertainty calculations.

This document fulfils the first-step objectives. First it presents the topic by listing the various objectives of initial site characterization, defining the expected benefits and providing an overview of the relevant existing standards and guidelines in the field of initial site characterization before decommissioning. Furthermore, specific chapters on preliminary statistical analyses and data processing describe available and currently applied methods. As regards data processing, we distinguished approaches for data with and without a spatial structure. In the context of initial site characterization in constraint environments, the very limited amount of available data is particularly challenging. That is why a chapter is entirely dedicated to this specific issue.

This document reports on the currently used methodologies and its main objective is to provide input for the next step of the WP3 process, that is: strategy development. Nevertheless, the generic sampling design techniques and state-of-the-art statistical techniques used in preliminary analyses and data processing are often considered as stand-alone methods. In many cases, an integrated and overall pre-decommissioning characterization approach, which notably consists in evaluating historical data, making on-site measurement campaigns, sampling and analysing, developing scaling factors and applying numerical codes, is missing.

This document is not aimed at providing a comprehensive overview of all the existing statistical methods or guidelines to the end user. The sampling design methods and state-of-the-art statistical techniques described in the present document are preselected methods that will be used as main inputs for the next step of the process, strategy development. In this next step, WP3 participants will develop a decision tree to guide the user throughout the radiological process and the choice regarding sampling scheme as well as the selection of statistical tools for preliminary analyses and data processing.

List of abbreviations

ALARA	As Low As Reasonably Achievable
ASN	Autorité de Sûreté Nucléaire (French Nuclear Safety Authority)
BC	Bienaymé-Chebytchev
CERN	European Organization for Nuclear Research
CM	Camp-Meidell
EPA	Environmental Protection Agency (US)
D&D	Decommissioning and Dismantling
DA	Destructive Analysis
DIN	Deutsches Institut für Normung e.V.
DTM	Difficult To Measure
DOE	Department of Energy (US)
DOD	Department of Defense (US)
ETM	Easy To Measure
FCA	Factorial Correspondence Analysis
FMCA	Factorial Multiple Correspondence Analysis
FDA	Factorial Discriminant Analysis
GLM	General Linear Model
GRS	Gesellschaft für Anlagen-und Reaktorsicherheit (German nuclear safety institute)
IAEA	International Atomic Energy Agency
INSIDER	Improved Nuclear SIte characterization for waste minimization in Decommissioning under constrained EnviRonment
IQR	InterQuartile Range
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual (US)
MOX	Mixed Oxide (plutonium/uranium nuclear fuel)
NEA/OECD	Nuclear Energy Agency/Organization for Economic Co-operation and Development
NDA	Non-Destructive Analysis
NRC	Nuclear Regulatory Commission (US)
P	Percentile
PCA	Principal Component Analysis
PLS	Partial Least Squares
Q	Quartile
QA	Quality Assurance
QC	Quality Control
QM	Quality Management
SF	Scaling Factor
VD	Van Danzig
WP	Work Package

1 Introduction

The EURATOM work program project INSIDER (Improved Nuclear SIte characterization for waste minimization in Decommissioning under constrained EnviRonment) was launched in June 2017. It aims at improving the management of contaminated materials arising from decommissioning and dismantling (D&D) operations by proposing an integrated methodology of characterization. The methodology is based on advanced statistical processing and modelling, coupled with adapted and innovative analytical and measurement methods, in line with sustainability and economic objectives.

The main objective of Work Package 3 (WP3) is to draft a sampling guide for initial nuclear site characterization in constraint environments before decommissioning, based on a statistical approach. This is done by selecting state-of-the-art techniques concerning sampling design optimization, using prior information and multiple iterations, testing the approach through different case studies and reviewing the feedback from overall uncertainty calculations. The process followed to meet the main WP3 objective consists of four steps:

1. Status: provide an overview of the available sampling design methods and state-of-the-art statistical techniques.
2. Development: develop a strategy/methodology and implement it in a software package by making use of (and possibly extending) state-of-the-art techniques.
3. Implementation: apply the methodology to the different test cases considered in order to test its adequacy.
4. Guidance: summarize all the findings in a comprehensive sampling strategy guide.

This document fulfils the first-step objectives. First it presents the topic by listing the various objectives of initial site characterization, defining the expected benefits and providing an overview of the relevant existing standards and guidelines in the field of initial site characterization before decommissioning. Furthermore, specific chapters on preliminary statistical analyses and data processing describe available and currently applied methods. As regards data processing, we distinguished approaches for data with and without a spatial structure. In the context of initial site characterization under constraint environments, the very limited amount of available data is particularly challenging. That is why a chapter is entirely dedicated to this specific issue.

This document reports on the methodologies currently being used and its main objective is to provide input for the next step of the WP3 process, that is: strategy development.

1.1 Objectives of initial site characterization

The generally accepted purpose of D&D is to allow the removal of part or all of the regulatory controls that apply to a nuclear site, whilst securing the long-term safety of the public and the environment and continuing to protect the health and safety of decommissioning workers throughout the process. Underlying this purpose, of course, are other practical objectives including the release of valuable assets such as buildings and sites for alternative use, the recycling and reuse of materials and the restoration of environmental amenity. In all cases, the basic objective is to achieve an end result that is sensible in technical, social and financial terms, that properly protects the workers, the public and the environment and that, in summary, complies with the basic principles of sustainable development. Stringent regulatory controls during D&D protect the public, the environment and the workers from the hazards associated with nuclear facilities. These hazards arise both from the radioactive inventory of the facility and from the nature of the operations carried out.

The initial site characterization upstream from the decommissioning of nuclear installations is essential in ensuring that decommissioning is accomplished in a safe and cost-effective manner. This pre-decommissioning radiological characterization is relevant for shutdown nuclear installations, the buildings containing these nuclear installations and the related (nuclear) sites.

At the various phases of a plant cycle, coordination and synergy with regard to radiological characterization is essential for an effective initial site characterization. In order to safely and cost-

effectively plan and manage a decommissioning project, the pre-decommissioning radiological characterization process should:

- Qualify the radiological status of the facility and the nature and extent of potential problem areas in order to decide on:
  - dismantling procedures: direct, semi-remote or remote working and tools required, shielding
  - safety assessment and ALARA planning (e.g. radiological protection of workers, the general public and the environment);
- Quantify the material inventory and couple it with the radiological status:
  - waste classification: waste management & transportation during and after dismantling (recycling, reuse, temporary storage, final disposal)
  - economic optimization: decontamination processes in order to declassify certain waste streams.

In order to achieve a general economic optimization, the waste management in all its aspects is of crucial importance. According to [1], it can be useful to take a waste-led approach to develop the characterization strategy, by working backwards from a waste and materials end-state perspective since the characterization requirements are most stringent at the end of the lifecycle. However, this approach needs to be considered with care as there can be very specific and stringent characterization requirements at any stage of the lifecycle, for example meeting the waste acceptance criteria for a waste treatment plant with a narrow operating envelope or meeting specific transportation requirements.

The main end products of a decommissioning project consist in several types of materials being released from regulatory control and declassified as radioactive waste. Figure 1 shows a conceptual representation of the waste classification scheme according to [2].

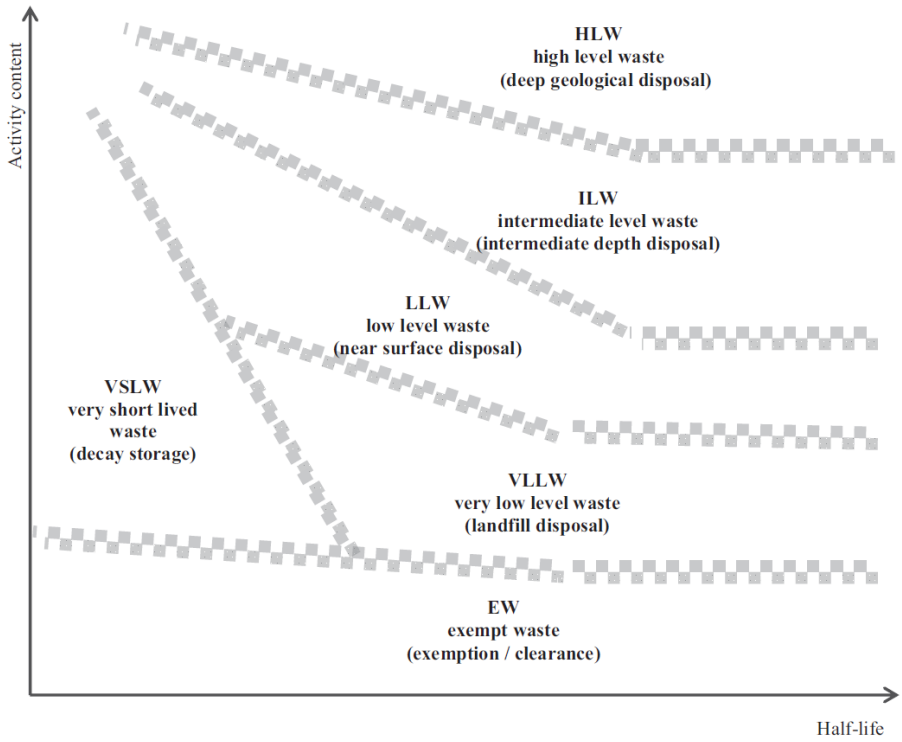


Figure 1: Conceptual representation of the waste classification scheme according to IAEA (2009)

Every European country has its own regulations and acceptance criteria for the different kinds of end products. Radiological acceptance criteria are usually expressed as activity concentration levels by

radionuclides per mass or volume, averaged out over a certain volume. Sometimes additional requirements apply, notably regarding dose rate (transport regulations) and surface activity levels.

The presence and concentration of radionuclides are therefore the main statistical parameters to be estimated, and these are based on physical quantities and their related uncertainties, i.e. activity concentration averages over a specific volume or mass.

1.2 Expected benefits

The term “radiological characterization” ([3]) usually represents the determination of the nature, location and concentration of radionuclides in a nuclear installation. It is one of the fundamentals on which a decommissioning project is based. Radiological characterization must be seen as an on-going process and it will only cease after the successful completion of the final survey and the termination of the nuclear license. It does not only consist in sampling, measuring and analysing the results, it also involves assessing the information derived from the operating history, calculations, the collection of existing data and many other sources. At the very beginning of the decommissioning planning stage, the purpose is to collect sufficient pieces of information to assess the radiological status of the facility and the nature and extent of problem areas. Data collected during this initial characterization step are generally based on the available information, notably that related to historical operations, and they are used to plan the overall decommissioning programme and to prioritize and sequence major decommissioning activities. As the planning progresses, characterization objectives move towards developing more detailed data concerning the physical, chemical and radiological conditions of a nuclear facility. This can include calculating induced activity, taking samples or conducting inspections designed to fill the gaps in the data drawn from the previous characterization step. It can be done to determine preliminary details, including cost, risk and waste generation estimates.

The main objective of a sampling process is to obtain a realistic estimate of the involved quantity of interest in order to decide whether it corresponds to compliance or non-compliance given a certain criterion or specification. It is necessary to take into account the measurement uncertainty and the confidence level.

According to INSIDER Work Package 2 [4], the following recommendations must be duly followed to guarantee the successful completion of the pre-decommissioning characterization of a given nuclear facility:

Regarding characterization objectives and sampling process:

1. It is necessary to define which radionuclides, non-radioactive elements and other properties will be measured and how they will be measured (through NDA and/or DA). It is also critical that representative samples are taken over the entire range of expected radioactive materials.
2. The sampling plan needs to detail the analytical procedures that are to be applied, including the necessary quality assurance (QA), quality control (QC) and quality management (QM) requirements. It should also ensure that the procedures meet all the applicable levels of accuracy, as well as the reliability and precision needed to control and monitor the treatment or conditioning processes adequately, without neglecting the time required to gather and analyse the samples because it will have a direct impact on the performance of such processes.
3. A minimum number of replicates has to be established in order to have sufficiently robust criteria for the statistical assessments of the proficiency test. A minimum of three replicates per sample are typically required in case of destructive analysis (with similar masses or volumes), two of them being used (mineralized or prepared for measurement) for the required analyses and the other one being kept as a reserve aliquot to cover any accidental event that may result in a loss of information.

4. Extrapolating information from in-situ measurements and samples to the whole area under study should be done very carefully since it requires planning a suitable strategy beforehand based on the analysis of the historical documentation related to the nuclear facility and of the advanced statistical approaches.
5. Representative samples depend on the physico-chemical nature of the radioactive materials being considered and on the chemical properties of the elements that need to be isolated. For example, if a volatile material is to be determined, the sample treatment should avoid exhausting the gases from the system.

Regarding in-situ measurements and laboratory analyses of samples:

6. Both in-situ measurement and radio-analytical methodologies, along with uncertainty assessment and calibration procedures, must have been previously tested using internal and/or external protocols.
7. The radioactivity distribution can be checked by measuring one or more easy-to-measure (ETM) radionuclides at different locations and all the collected samples must be as homogeneous as possible. If not, any heterogeneity should be corrected in order to ensure a good comparison of the results obtained.
8. The development and use of reference samples, either used to simulate non-radioactive or radioactive materials, is of prime importance to compare the quality of their analytical procedures using different techniques and instrumentation, to provide a means to improve the accuracy and precision of methods and techniques, to facilitate direct inter- and intra-laboratory comparison of analytical results when the samples are used as “blind samples”, and to allow for the bias correction of analytical results when they are processed alongside actual waste samples.
9. An inactive matrix with the same physico-chemical properties as those of the considered family of radioactive materials to be measured must be provided to subtract the blank contribution for each measurement and to allow a good estimation of the minimum detectable activities.
10. It is preferable to perform intercomparison exercises with real samples that are representative of radioactive contaminated concrete and/or soil. Although it is difficult to know a priori the "true values" of such samples, this kind of exercises will consolidate, harmonize and identify the actual limits of the different methodologies applied.
11. Moreover it is recommended to perform a common benchmarking exercise for both in-situ measurements and destructive analyses, for example by considering a piece of concrete that can be measured in situ before taking representative samples for laboratory analysis.

Regarding other practical considerations:

12. It is preferable to conduct all the characterization steps in cooperation with the experienced staff of the facility, since these latter are most familiar with the site history and specificities and they should have the necessary technical skills.

1.3 Overview of relevant standards, guides, etc.

The IAEA (International Atomic Energy Agency) reference document on the radiological characterization of shut down nuclear reactors for decommissioning purposes is the technical report series # 389 ([3]). Besides a description of characterization objectives, health and safety considerations, the characterization process, the radionuclide inventory, characterization methods

and techniques and quality assurance requirements, a short subchapter is dedicated to sampling and analyses. In this report, the IAEA advises to divide the sampling plan into unbiased and biased sampling schemes. Unbiased sampling schemes should be used for areas where only little or no surface contamination is expected, or for general areas that are expected to be homogeneous in terms of degree and characteristics of contamination. The facility to be characterized should be divided into discrete sampling areas and survey units for measurement purposes. These population samples can then be compared to a background population to determine whether they have been affected by the facility operations. Biased sampling is used to find or define a contamination or an activation that is known or expected to exist. The biased sampling program actively examines sample locations in areas where contamination or activation is likely to be present. To figure out whether to use an unbiased or a biased survey, the expected range of measurement values must be determined on the basis of the plant operational and historical data. If the results of all the survey measurements are expected to be uniformly distributed, then an unbiased survey should be used. On the contrary, if the results of the survey measurements are expected to be non-uniform (e.g. with hot spots present), then a biased survey should be used. In situations where the unbiased survey method is used, the characterization effort should imply a measurement and sampling regime that provides an acceptable confidence level in determining true surface contamination or activation levels within a specified error, typically 95 % confidence level. The results of the initial measurements and sample collections are used to determine whether additional data are needed to achieve the desired accuracy in determining the true mean surface contamination or activation levels.

The NEA/OECD (Nuclear Energy Agency/Organization for Economic Co-operation and Development) has just updated its report on the Radiological Characterisation for Decommissioning of Nuclear Installations towards practices and experiences ([1] and [5]). According to [5]: “In a very simplified representation, the main sampling and measurement methodologies are judgemental or probabilistic-based. Judgemental approaches can hold down sampling and measurement costs by focused sampling, but rely on good prior knowledge and the validity of their results depends on the quality of judgement used. They are good for looking for worst cases, with a limited number of points rather than estimating means or testing hypotheses due to the obvious sampling bias. Probabilistic-based approaches (random or regular mesh) assume a random or spatial distribution hypothesis for which some historical information is necessary. Their mathematical structure allows inferences (using statistical or geostatistical data analysis) to be made from the samples to the whole population. Sampling optimization (number of random points or size of the regular mesh) is therefore possible to meet the objective due to the uncertainty quantification. In complex situations (e.g. where there is a high abundance of hard-to-detect radionuclides) a mixture of judgemental and probabilistic-based approaches may be necessary. Another case is the calibration of deterministic models (activation, ground water flow, atmospheric dispersion…).” The Appendix B of [5] “lists detailed considerations when determining a sampling and measurement strategy” and “Appendix C lists considerations for in-field requirements during performance of sampling and measurement”. The document [1] discusses the practical aspects of how a high level characterization strategy can help optimize characterization in practice (e.g. waste-led characterization, holistic characterization).

One of the most commonly used references in the final phase of decommissioning projects is the U.S. MARSSIM ([6]). The MARSSIM provides information on planning, conducting, evaluating, and documenting final status radiological surveys for buildings and surface soils in order to guide the demonstration of compliance with radiation dose- or risk-based regulations or standards. The MARSSIM is a multi-agency consensus document that was developed collaboratively by four federal agencies having authority and control over radioactive materials: Department of Defense (DOD), Department of Energy (DOE), Environmental Protection Agency (EPA), and Nuclear Regulatory Commission (NRC). The MARSSIM's objective is to describe a consistent approach for final status surveys applied to buildings and surface soils in order to meet established radiation dose or risk-based release criteria, while encouraging an effective use of resources. The basic principles consists in classifying areas according to their impact and performing systematic or judgmental measurements with an appropriate coverage. The MARSSIM recommends the use of nonparametric statistical tests to evaluate data (e.g. Sign test, Wilcoxon Rank Sum test). A number of alternate statistical tests are also proposed in specific situations. The Manual focuses on the demonstration of compliance during the final status survey following scoping, characterization and any necessary

remedial actions and therefore its scope is more limited than the scope of the INSIDER project. Since many of the MARSSIM principles rely on EPA experiences, you will find below a set of EPA reference documents:

- Guidance on Choosing a Sampling Design for Environmental Data Collection (2002) ([7]).
- Guidance for Data Quality Assessment - Practical Methods for Data Analysis (2000) ([8]).
- Data Quality Assessment: Statistical Methods for Practitioners (2006a) ([9]).
- Data Quality Assessment: A Reviewer's Guide (2006b) ([10]).
- Statistical Software for Environmental Applications for Data Sets with and without Non-detect Observations (2013) ([11]).

The ISO/DIS 18557 standard ([12]) contains both guidelines and references to other documents which may be useful in combination with characterization principles for soils, buildings and infrastructures contaminated by radionuclides for remediation purposes.

Other standards also focus on the final phase of decommissioning projects, notably the German DIN25457-4 ([13]), DIN 25457-6 ([14]) and 25457-7 ([15]) which deal with the clearance of metal scraps, buildings and rubble, and ground surfaces and excavated soils. For the remediation of buildings and lands, a two-step process consisting of preliminary and decision-based measurements is usually chosen. As in other standards and guidelines (e.g. MARSSIM, Guide ASN n°14 [16]), a graded approach using a categorization based on the contamination hazard is being applied. The preliminary measurements made in buildings or sites containing nuclear installations during radiological characterization phase are used to determine whether and how statistical measurements can be applied for clearance. The German DIN categories are as follows:

- Category 1: A contamination above clearance levels is or has been present – these areas usually require measurements on the entire area.
- Category 2: Possible contamination (there are hints of contamination, but measured values do not exceed clearance levels) – these areas only require statistical measurements in grids.
- Category 3: Contamination can be excluded (no hint of contamination) – these areas only require measurements to secure evidence that no contamination is present.

Decision-based measurements in land areas are carried out to demonstrate that surface- or mass-related activities comply with the appropriate clearance levels. A prerequisite for decision-based grid measurements is that all parts of the area have a comparable contamination history and mechanism. In other words, it would not be acceptable to treat lawns and streets/transportation routes as a single statistical entity. The grid shall consist of rectangles or squares. Measurements can be carried out directly on the grid areas or on samples taken from the grid areas. The size of the grid area should correspond to the averaging area that is allowed for surface-related measurements (e.g. 100 m² in case of clearance of land areas). Depending on the measurement procedure, the measurement area and the grid area / averaging area can obey the following relations (see Figure 2):

- Case 1: The measurement area is much smaller than the grid area. A certain number of measurements have to be carried out inside the grid area. The percentage of the grid area which is covered by these single measurements must be appropriately determined (see [17]).
- Case 2: The measurement area is of a similar size as the grid area. The measurements have to cover the entire grid area.
- Case 3: The measurement area is equal to or larger than the grid area. The measured activity has to be related to the grid (averaging) area.

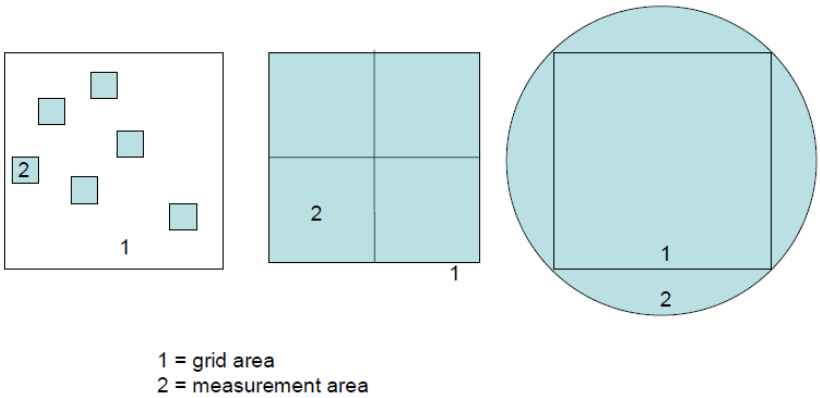


Figure 2: Relations between the measurement area and the grid area (cases 1, 2, 3: measurement area much smaller, smaller and larger than the grid area) [DIN 25457-7]

Many guides and references focus on the back end of decommissioning (e.g. release of regulatory control) whereas the INSIDER project is more looking at the front end (pre-decommissioning characterization). Statistical techniques used are often based on stand-alone data. They do not systematically combine various data (e.g. historical, non-destructive, destructive data), through multi-variate analyses for example.

1.4 Examples of implementation

The Annex II of [3] describes some problems encountered in the characterization of nuclear reactors:

- Overestimation of the activity of concrete bioshields by computer codes, by factors ranging from five to ten, leading to waste overclassification;
- Discrepancies between calculated and measured activation concentrations;
- Unequal activation of key reactor components (stainless steel components that may derive from separate components with widely differing impurity content): biased results due to inappropriate sampling plan;
- Discrepancies between calculated and measured doses;
- Inadequate planning for characterization (UHTREX reactor, US).

The IAEA [18] reports on lessons learned in decommissioning, with a focus on the following examples:

- Decontamination and decommissioning of a 60” Cyclotron facility at Argonne National Laboratory — East:
  - Inaccurate and insufficient information provided in the characterization report, which resulted in unforeseen radiological problems;
  - Questioning of assumptions and interpretations of radiological characterization reports.
- Connecticut Yankee Power Plant:
  - Mistrust about site characterization, discovery of unexpected radioactive sources and unauthorized transport of materials off site.
- Robust estimation of fissile material quantity in a dissolver (France/COGEMA): a continuous dissolution process was used for the dissolution of the gas cooled reactor spent fuel in the reprocessing plant. The basic decontamination parameters derived from values obtained 15 years before, when a dissolver was replaced and dose rate measurements were taken. A first rinsing operation with nitric acid took place without any prior chemical analysis of the deposits at the bottom of the dissolver and unfortunately the quantity of plutonium removed

was much higher than the forecasted amount, which posed a problem. The troubles encountered in rinsing the continuous dissolvers resulted from a poor knowledge of the initial state, mainly due to a lack of quantitative and qualitative characterization of the deposits contained in the dissolvers.

2 Preliminary statistical analysis

Radioactivity levels may show a wide range of fluctuations. Different causes can explain this variability:

- Natural variations due to local geology, as the quantity of natural nuclides (Uranium, Thorium, Radon…) varies according to the geological type;
- Anthropic variations due to surface matrices (made ground, concrete, roads…), to artificial nuclides now belonging to background (radioactive fallouts from nuclear accidents and atmospheric tests) and finally to direct pollution from nuclear plants (leakage, activation, contamination…);
- Measuring variations due to device configuration (height from ground level, measured surface…), in particular concerning gross counting and dose rate;
- Interpretation variations: calibration applied in laboratories, consideration of radioactive decay (for short-lived nuclides), preexisting fingerprints, detection limits, etc.

For all these reasons, radiological characterization requires the identification of the different sources of heterogeneity and/or bias in order to quantify radioactivity variations properly.

2.1 Physical quantities and statistical indicators to be estimated

2.1.1 Statistical indicators

This section introduces a general point of view about statistical indicators that characterize a series of statistical data and that are used in every statistical development. In dismantling studies, it is not rare that these statistical indicators are the only data provided.

Variables

A characteristic that can take different values is called “a variable”, i.e. a variable is any characteristic that varies from one individual member of the population to another.

Examples of variables for dismantling: density, activity level, dose rate, historical data, radiological spectrum, type (e.g. 200-L drum, 400-L drum, soil...).

The first three of the abovementioned variables yield numerical information (numerical measurements) and are examples of quantitative (or numerical) variables. The last three variables yield non-numerical information (non-numerical measurements) and are examples of qualitative (or categorical) variables. Moreover quantitative variables can be classified as either discrete or continuous.

Sample and Population Distributions

For a variable, frequency distributions apply both to a population and to samples of that population. The first type is known as “the population distribution of the variable”, and the second type, as “a sample distribution” (see Figure 3). In a sense, the sample distribution is a blurry photograph of the population distribution. As the sample size increases, the sample relative frequency in any class interval gets closer to the true population relative frequency. Thus, the photograph gets clearer and the sample distribution looks more like the population distribution.

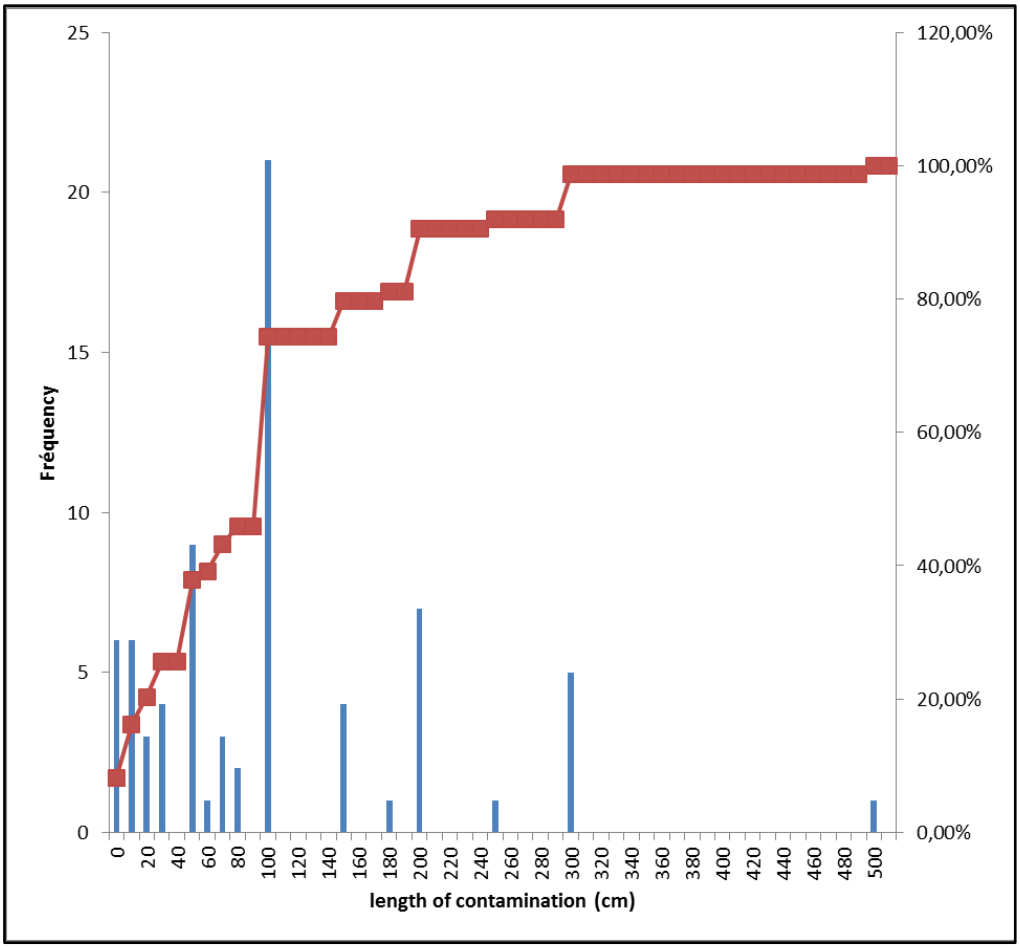


Figure 3: Example of sampling distribution in dismantling studies (ONET Technologies, 2017) with the frequencies (in blue) and the cumulated values (in red).

Mode

The sample mode of a qualitative or discrete quantitative variable is the value of the variable which occurs most frequently in a data set. In the example shown in the Figure 3, the mode value is 100 cm.

Range

The sample range is obtained by computing the difference between the largest and smallest observed value of the variable in a data set. The sample range of the variable is the difference between its maximum and minimum values in a data set:

$$Range = Max - Min.$$

In the example shown in the Figure 3, the range value is: 500-1 = 499 cm.

Mean

The most commonly used measure of center for quantitative variables is the (arithmetic) sample mean. When people speak about taking an average, they most often refer to “mean”. The sample mean of the variable is the sum of observed values divided by the number of observations.

In the example shown in the Figure 3, the mean value is about 117 cm.

Median

The sample median of a quantitative variable is the value of the variable in a data set that divides the set of observed values in half, so that the observed values in one half are less than or equal to the median value and the observed values in the other half are greater or equal to the median value.

To obtain the median of the variable, the observed values in a data set are sorted in increasing order and then, the midpoint value corresponds to the median.

In the example shown in the Figure 3, after sorting the data into the distribution classes, the median value is 100 cm.

Interquartile range

Before we can define the sample interquartile range, we have to define the percentiles, the deciles and the quartiles of the variable in a data set. The percentiles of the variable divide observed values into hundredths, or 100 equal parts. Roughly speaking, the first percentile (P1) is the number that divides the bottom 1% of the observed values from the top 99%. The second percentile (P2) is the number that divides the bottom 2% of the observed values from the top 98%; and so forth. The median is the 50th percentile. The first quartile (Q1) is defined as the middle number between the smallest number and the median of the data set. The second quartile (Q2) is the median of the data. The third quartile (Q3) is the middle value between the median and the highest value of the data set. The interquartile range (IQR) is the difference between the third and the first quartiles.

In the example shown in the Figure 3, the interquartile ranges are symbolized by the cumulated plot (values on the right axis).

Variance

The sample variance is the most frequently used measure of variability. It can be considered as a kind of average of the absolute deviations of observed values from the mean of the variable considered:

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1},$$

where

- $\bar{x}$ is the sample mean of the variable x,
- *n* is the number of samples.

2.1.2 Physical quantities

The decommissioning projects are based on the quantities defined thereafter.

2.1.2.1 Activity level

The activity level (min., average, and max.) is a value used to evaluate radioactive quantity. The quantity *A* is defined [18] as follows, for an amount of a radionuclide in a given energy state at a given time:

$$A(t) = \frac{dN}{dt}$$

where dN is the expectation value of the number of spontaneous nuclear transformations from the given energy state in the time interval dt. The SI unit of activity is the reciprocal of second (s^{-1}), termed “Becquerel” (Bq). Formerly expressed in curie (Ci), activity values may still be given in Ci (with the equivalent in Bq in parentheses) if they are being quoted from a reference which uses that unit.

2.1.2.2 Dose rate

The dose rate level (min., average, and max.) is used in decommissioning studies to predict the dose uptake for dismantling operations. The maximum value obtained for an area is usually retained as reference value.

The absorbed dose D ([18]), which is the fundamental dosimetric quantity, is defined as follows:

$$D = \frac{d\bar{\varepsilon}}{dm}$$

where $d\bar{\varepsilon}$ is the mean energy imparted by ionizing radiation to the matter in a volume element and dm is the mass of matter in the volume element. The energy can be averaged over any defined volume, with the average dose being equal to the total energy imparted in the volume divided by the mass in the volume. The dose absorbed is defined at a certain point. For the average dose in a tissue or an organ, see "organ dose". The unit is J/kg, termed "Gray" (Gy) (formerly, the "rad" was used).

2.1.2.3 Equivalent dose

Equivalent dose rate H*(10), expressed in sievert (Sv) by hour (h) is a practical value used by radiation protection services to quantify the radiation hazards for human body. It is usually referred to as "equivalent dose rate" in radiation protection documents, without any further precision-. The equivalent dose is the product of the absorbed dose in a tissue or an organ at a given point and the quality factor associated with the type of radiation absorbed.

$$H_{T,R} = W_R . D_{T,R}$$

where $D_{T,R}$ is the absorbed dose of R type radiation averaged over a tissue or an organ T and W_R is the radiation weighting factor for R type radiation. When the radiation field is composed of different types of radiation associated with different values of W_R, the equivalent dose is given by the formula below:

$$H_T = \sum_R W_R . D_{T,R}.$$

The current value used includes the impact over a 10-mm distance inside the whole human body "H*(10)".

The equivalent dose unit is J/kg, termed "sievert" (Sv). The rem, equal to 0.01 Sv, is also sometimes used as a unit for the equivalent dose and the effective dose.

For characterization purposes, the use of simple dose (Gy) information is preferable.

2.1.2.4 Radioactive contamination

The contamination level (min., average, and max.) is useful to determine intervention conditions or waste area types and it could also be practical to predict soil contamination.

In [19], radioactive contamination refers to the unintended and unwanted presence of a radioactive substance on structural surfaces and/or within equipment, or to the process giving rise to its presence in such locations.

In the case of nuclear reactors, radiation contamination results from an unpredictable release from the fuel, generally made up of enriched uranium and MOX[1], fissile elements, fission products, and their progeny, in normal or unwanted operating conditions (e.g. rupture of fuel rod cladding tubes). It also derives from the corrosion and erosion taking place within the activated materials cooling systems. These materials actually tend to accumulate in the regions where elbows and adjacent wedges are welded into piping loops.

Contamination only refers to the presence of activity. It gives no indication on the magnitude of the hazard involved. In Transport regulations, contamination is defined as follows: the presence of a

[1] MOX is an acronym for Mixed OXides and it consists of plutonium blended with natural or depleted uranium.

radioactive substance on a surface in quantities in excess of 0.4 Bq/cm^2 for beta and gamma emitters and low toxicity alpha emitters, or 0.04 Bq/cm^2 for all other alpha emitters.

A distinction is usually made between:

- Fixed contamination: contamination that cannot be removed from a surface in normal conditions;
- Non-fixed contamination: contamination that can be removed from a surface in normal conditions.

The air can be contaminated with radioactive nuclides in particulate form, which causes a particular inhalation hazard (airborne contamination).

2.1.2.5 Radionuclide vectors and scaling factors

The scaling factor (SF) is useful to determine the origin of radioactive materials. It is commonly used in calculations related to the total activity of radioactive waste and to radiation protection and safety issues. Many radionuclides whose characterization is important for long-term waste management are difficult to measure (DTM) or impossible-to-measure (ITM) from the outside of a waste package. Within the context of waste characterization, the relationship between some key easy-to-measure (ETM) nuclides, such as certain gamma emitters, and DTM/ITM nuclides is used to derive information on the DTM/ITM nuclides of interest. The scaling factor method is an approach that is widely used to evaluate these DTM/ITM nuclides. It consists in developing a correlation between ETM and DTM/ITM nuclides. A combination of different techniques (e.g. alpha spectrometry, mass spectrometry, gamma spectrometry, modeling, etc.) are used to determine the nuclide vector and the scaling factors. The sample and sample locations are important and have to be representative for the material stream considered. The activities of DTM/ITM nuclides in waste packages are then estimated by measuring the ETM nuclides based on gamma measurements taken from outside the package and applying the SFs to obtain the DTM activities ([20] and [21]). The use of scaling factors is in many cases vital for all kinds of end products (see Figure 1). In order to have a real correlation between the two kinds of radionuclides, DTM/ITM and ETM, the following conditions must be satisfied [22]:

- The activity concentration of the key nuclide must be measurable and far above the minimum detection limit;
- The samples chosen for the measurement must be representative of all materials;
- The number of samples must be such as to cover all the areas of the nuclear installation , which are involved in the process. Simultaneously, each sample must be measured several times in order to accumulate significant statistics to have valid results;
- The different waste streams or plant types must be comparable to the investigated case in order to be sure that the results obtained are valid;
- It is, finally, important to verify that there is a functional relationship between the measured activity of the two radionuclides which are related to eachother.

2.1.2.6 Radioactive waste

For legal and regulatory purposes [18], it refers to waste that contains or is contaminated with radionuclides at concentrations or activities exceeding the clearance levels established by the regulatory body. It should be recognized that this definition is purely for regulatory purposes and that a material with activity concentrations equal to or less than clearance levels is still radioactive from a physical viewpoint, although the associated radiological hazards are considered negligible.

2.1.2.7 Hazardous substances

The potential existence of other hazardous substances also needs to be controlled during the pre-decommissioning characterization of nuclear facilities [23]. Above all, let's mention the particular case of asbestos that used to be the preferred electrical and thermal insulator in the past. Among other chemical toxicants that require monitoring, there are: antimony, arsenic, beryllium, boron, cadmium, chromium, cyanides, mercury, nickel carbonyls, lead, polychlorinated biphenyls, selenium and sodium or sodium-potassium alloys. Finally, materials exhibiting both radiological and toxic properties, commonly known as "mixed wastes", may suppose non-trivial problematics regarding dismantling, matrix encapsulation and disposal routes.

2.2 Usable data

As stated in section 1.1, at the various phases of a plant cycle, coordination and synergy with regard to radiological characterization are essential for an effective initial site characterization.

During the operational stage of an installation, various types of data might be available and could be a very valuable input for the (later) sampling strategy. In general, available data are usually limited to results from dose rate and fixed and non-fixed surface contamination measurements. However, available data could originate from a diversity of actions and might sometimes be broader/more detailed:

- Records of routine occupational exposure measurements made on the installation and on workers during normal operation and maintenance. Records of occupational exposure measurements made during inspections, surveys, repairs and replacement activities.
- Records of contamination spills or of other unusual events/incidents.
- Records of potential previous radiological characterization surveys (including results of various types of non-destructive and destructive analyses.
- Records of potential material production.
- Records of operational waste characterization with an history-based list of possible contaminants. In this respect, nuclide vectors/scaling factors (correlation factors between hard-to-measure radionuclides and key nuclides) could already be available at this stage.

Annex A shows an example of a typical cartography of a nuclear building in France. Such documents usually contain the following data:

- Equivalent dose rate,
- Surface contamination,
- Non-fixed surface contamination.

For those types of rough measurement data, no uncertainty is provided by radiation protection services. Results are strongly dependent upon the type of radionuclides measured. Therefore, they are generally given for information purposes as they need more detailed knowledge of the installation for advanced interpretation. Annex B shows the properties of these usable data.

As listed above, records on operational waste characterization are usually available. Results of studies could be an important input to determine SF within the initial site characterization process.

Computer codes are sometimes used to calculate the activity concentration of neutron activated components. Results of such calculations can then be used in the perspective of the scaling factor method and/or a correlation can be derived from destructive or non-destructive analysis results. The basic process of planning, executing and reporting results for itemized component characterizations (point estimation method) based on neutron source estimations and on elemental compositions, physical parameters and usage of components in the reactor is described in the ISO 16966 standard ([24]).

2.3 Sampling strategy: How many measurements and where?

In general, there is not enough time, energy, money, labour/man power, equipment, and access to suitable sites to measure every single item or site within the parent population or the whole sampling frame. For example, in most cases, it is not feasible to characterize every single waste before dismantling operations.

The main purpose of statistical sampling is to collect representative data sets on which statistical analysis will provide responses with regard to the characterization objectives of surface, objects, etc. Therefore, it is important to obtain a small, informative and above all sufficient subset of the population under study. Given the radiological constraint of measurements and costs, the sampling strategy which consists in defining the number and the location of the measurements needs to be optimized in order to provide the most realistic and reliable information with a reasonable size of measurement data. The preliminary phase of sampling strategy needs due attention.

Sampling techniques consider that:

- Larger sample sizes are often more accurate representations of the whole population;
- The sample size chosen is a balance between a statistically valid representation and the time, energy, money, labour, equipment and access available;
- A sampling strategy made with the minimum of bias is the most statistically valid;
- Most approaches assume that the parent population has a normal distribution where most items or individuals clustered close to the mean, with few extremes;
- A 95% probability or confidence level is usually assumed, for example 95% of items or individuals will be within plus or minus two standard deviations from the mean;
- This also means that up to five per cent may lie outside of this. No matter how good it is, sampling can only ever be claimed to be a very close estimate.

The sampling strategy has to answer several questions: how many measurements required, where to measure, what information can be deduced from the measurements and what is the uncertainty. In order to answer to the first question, the methods described in section 2.3.2 can be applied. When a spatial correlation is assumed (to be validated by the measurement data), the geostatistical techniques presented in section 2.3.3 can be used. However, after having collected the data, it is important to lead a rigorous analysis to infer an estimation of the variable of interest (level of contamination, proportion of non-compliance with threshold…). To do so, there are statistical techniques which depend on the sample strategy being used which give estimations of indicators like mean or proportion and of their related uncertainties (see section 2.3.2). Another way to proceed, which is complementary to the previous one, consists in using numerical models. This approach is described in section 2.3.1.

2.3.1 Sampling strategy directed by numerical models

The isotopic characterization of materials from D&D activities can provide interesting benefits: management and reduction of the uncertainties related to the sampling strategy, forecasting of the evolution of source term.

Specific numerical models provide in silico characterization according to a calculation process detailed in the Figure 4. This general scheme ([25]) concerns use cases where the main neutron source comes from a nuclear reactor. When activation or contamination is imputable to derived sources, for example in matter flow cases, this scheme must be adapted (first step).

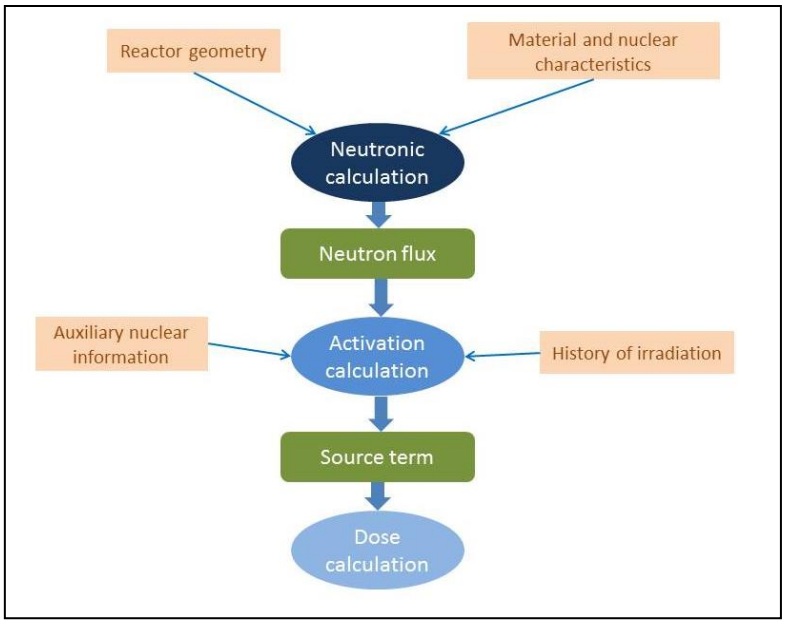


Figure 4 : Different computational steps of the characterization process for D&D.

First, we need to estimate the temporal average of the geometrical distribution of fission sources over the lifetime of the facility to be dismantled. To do so, numerical models for neutron flux assessment at each point of the facility are required. Deterministic [26] or stochastic [27][28][29] codes can be used but deterministic calculations with fixed meshes should be preferred, except when the target is composed of adjacent areas with different sizes, densities or compositions.

The reaction rate provided is then used in the calculation of the fission neutron propagation around the reactor core in the facility (mostly vessel internals and the vessel itself). The results obtained serve as input data in the neutron evolution calculations made by the computer code [30] in order to estimate the radionuclide concentrations in the structures after irradiation and cooling periods, but also the source and spectrum radiations. These elements can then be included in a Monte Carlo code ([27], [31] or [32]) to estimate the equivalent dose rate.

There are other phenomena, like the transport and migration of radionuclides in concrete or other materials that can be more or less modeled by numerical codes. The calculation estimations can then complete data sets (historical background of the facility) to provide a scenario about the production and evolution of radioactive products in the facility. This scenario can serve to drive the sampling strategy and provide an adapted sampling grid with more measurement points on the estimated hot spots for example.

The main difficulty of this approach based on numerical models is to make realistic assumptions about a scenario while avoiding any overestimation or underestimation of the contamination. That is why it should be performed on well-known facilities.

2.3.2 Statistical sampling strategy

How many measurements?

The sample size can be determined according to the estimator (mean, proportion, quantile…) used and the confidence interval required. For example, if the mean is used as estimator, the sample size n is deduced from the following formula:

$$n = \frac{16\sigma^2}{k^2} \quad \textbf{\textit{(Eq. 2.3.2-1)}}$$

where σ is the standard deviation and k, the number of units for the confidence interval.

In general, the sample size can be obtained from the formulation of the error margin resulting from the maximum difference between the observed sample mean and the true value mean of the

population. To do so, we can use the formulation of the confidence interval of the estimation $\hat{\theta}$ of the statistical estimate θ, with the probability $1-\alpha$:

$$P\left(\theta \in \left[\hat{\theta}-m;\hat{\theta}+m;\right]\right)=1-\alpha.$$

This equality can be rewritten as follows for the mean:

$$P\left(\theta \in \left[\hat{\theta}-Z_{1-\alpha/2}\frac{\hat{\sigma}}{n};\hat{\theta}+Z_{1-\alpha/2}\frac{\hat{\sigma}}{n};\right]\right)=1-\alpha$$

$$m \geq Z_{1-\alpha/2}.\frac{\hat{\sigma}}{\sqrt{n}}$$

$$n=\left[Z_{1-\alpha/2}\times\frac{\hat{\sigma}}{m}\right]^2$$

- n : Minimal sample size to obtain significant results for the given hypothesis;
- $Z_{1-\alpha/2}$: Standard law quantile with a confidence level $of\ 1-\alpha/2$ (the level values that are generally used are $\alpha=0.05$ and $Z_{1-\alpha/2}=1.96$);
- σ: Standard deviation;
- $\hat{\sigma}$: Estimation of σ;
- m : Error margin.

Nevertheless, when the objective is to estimate the proportion p of a population that has a given characteristic, the standard deviation formula is as follows:

$$\hat{\sigma}_p=\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\boldsymbol{n=Z^2_{\alpha/2}\times\frac{\hat{p}.(1-\hat{p})}{m^2}} \qquad \textbf{\textit{(Eq. 2.3.2-2)}}$$

- σ_p: Estimation of the standard deviation of the proportion;
- $\hat{p}$: Estimation of the proportion p;
- m : Error margin.

When p is unknown, its value is conservatively set to 0.5.

If the objective is the compliance verification of an expected status for an object under dismantling, a formula based on a hypergeometric probabilistic distribution approach ([33], [34]) can be used to determine the sample size to be checked:

$$n=\left(N-\frac{d}{2}\right)\left(1-\beta^{\frac{1}{(d+1)}}\right) \qquad \textbf{\textit{(Eq. 2.3.2-3)}}$$

where

N : The size of the population;
β : The risk of failing the detection of non-compliance;
d : The maximal number of permitted non-compliances within the population ($\alpha \times N$);
α : The maximal proportion (or risk) of permitted non-compliances.

There is a method that allows the determination of the sample size required to estimate a quantile of order α with a confidence β. It is the Wilks method which is presented in section 5.2 for small data sets.

For example, if the size of the whole population to measure is $N = 1000$, the maximal number of expected noncompliances is $d = 50$ ($\alpha = \frac{d}{N} = \frac{50}{1000} = 0.05 = 5\%$) and the risk of failing the detections of non-compliance is $\beta = 5\%$, the sample size deduced from *(Eq. 2.3.2-2)* is expressed as follows:

$$n = \left(1000 - \frac{50}{2}\right)\left(1 - 0.05^{\frac{1}{(50+1)}}\right) \approx \mathbf{56}.$$

The required size of the (statistic) sample, i.e. the number of measurements, can also be determined from the following approach: the number of samples/measurements can be determined from the maximum permissible fraction of samples/measurements exceeding the clearance level and the probability with which this maximum number must be complied, using the binomial distribution:

$$w(r) = \binom{g}{r} p^r (1\text{-}p)^{g\text{-}r} \qquad \textbf{\textit{(Eq. 2.3.2-4)}}$$

where

p : relative fraction of samples/measurements exceeding the clearance level;
g : number of samples;
r : number of samples/measurements exceeding the clearance level within the statistic sample.

The upper bound of the confidence range V_0 is given by:

$$V_0 = \frac{i.F(n,m,l\text{-}\alpha)}{k+i.F(n,m,l\text{-}\alpha)} \qquad \textbf{\textit{(Eq. 2.3.2-5)}}$$

where

$n = 2\cdot(r+1)$;
$i = n/2$;
$m=2\cdot(g-r)$;
$k=m/2$;
$F(n,m,1\text{-}\alpha)$: quantile of F distribution;
α: probability of error;
n, m: degrees of freedom of F distribution.

If the upper bound of the confidence range does not exceed the predetermined value, the clearance levels are respected.

This statistical procedure can be applied as follows: the size of the statistical sample shall be chosen in such a way that on the 95% confidence level, it can be excluded that more than 5% of exceeded clearance levels remain undetected. The second column of Table 2 shows the probability to find no exceeded clearance level (r = 0) in the statistical sample if in fact p = 5%. The third column shows the probability for the reverse question. The fourth column shows the upper end of the confidence range for the values of the second column. It can be seen that at least 60 samples (last row) have to be taken before the upper end of the confidence range falls below the required probability value of 5%. This means that the statistical sampling size must be 60 or more if no exceeded level is detected. Based on similar considerations, it can be concluded that 95 (125) samples/measurements need to be taken if 1 (2) exceeded level(s) is/are detected.

Table 1: Some examples of statistical sample size determined by a binomial approach.

Number of samples g (1 is added to account for rounding errors)	Probability to find no exceeded clearance level (r = 0) in the statistical sample if in fact p = 5%	Probability for locating at least one exceeded level (r > 0) if p = 5%	Upper end of the confidence range (95% security) if no exceeded level is found in the statistical sample
30	21.5%	78.5%	9.5%
40	12.9%	87.1%	7.2%
50	7.7%	92.3%	5.8%
60	4.6%	95.4%	4.9%

When the size of the data has been determined or fixed by the operational constraints, it is time to pay attention to the sampling strategy. This can be divided in two categories: the random sampling and the systematic sampling.

Which individual or point location to measure?

Random sampling consists in considering few wastes as representative without taking into account any historical or radioactive level. It is the least biased of all the sampling techniques since there is no subjectivity involved: each member of the total population has an equal chance of being selected (for instance, the sizing of a waste recovery operation). This can be done by numbering the fuel wastes and using a mathematical random function to select those which will be part of the sample.

Advantages: It can be used with large sample populations with no information available. It avoids bias.

Limits: It can lead to poor representation of the overall parent population or area if large areas are not covered by the random numbers generated.

Systematic sampling consists of selecting the measurement points in a systematic or regular way. For example in case of radioactive investigations conducted on ventilation ducts before dismantling operations, it consists on sampling every 2 meters of duct in the building, whether or not there are filters, airflows, etc. in these portions.

Other rules can be adapted to the context:

- The measurement points are evenly/regularly distributed in a spatial context, for example every two meters along a duct.
- They can be located at equal/regular intervals in a temporal context, for example every half-hour or at set times of the day for the sampling of an operational effluent discharge pipe.
- They can be regularly numbered, for example every 10th radioactive fuel of a list.

Advantages: It is a more straightforward method than random sampling. A grid does not necessarily have to be used and sampling just has to be performed at uniform intervals. It is easier to achieve a good coverage of the study area than with random sampling.

Limits: It is a more biased method since all the members or points do not have an equal chance of being selected. It may therefore lead to an over- or under-representation of a particular pattern.

As described in [35], if both random sampling and systematic sampling are selected, there are different types of sampling strategy that can be used depending upon the available information (areas of interest, history of contamination incidents, and list of contaminants…). They can be classified into three categories detailed in the following subsections:

- Sampling strategies guided by expert judgement and numerical models (material activation, migration, atmospheric dispersion…);

- Statistical sampling strategies;
- Geostatistical sampling strategies which take into consideration the spatial structure of the studied phenomenon.

The documents [36] and [37] provide descriptions and mathematical developments on the main statistical sampling methods. Some of them directly ensure the representativeness of the sample. Others use auxiliary data with the objective to reach a good accuracy for the estimations calculated from the sample. These different approaches are of different types. Some have been developed for survey theory and others are more empirical methods based on feedback. The first type concerns several statistical methods. The most popular of them are listed below:

- Simple random sampling without replacement scheme for a given sample size *n*.
- Probability-proportional-to-size sampling when auxiliary information is available: separation sampling design; stratified sampling design ([38]) gives an illustration of this approach for the radiological characterization of historical metallic wastes produced at the European Organization for Nuclear Research (CERN).
- Cluster sampling. This approach is well-adapted for area sampling or geographical cluster sampling.

In practice and in a context where very little information is available, it should be recommended to perform an iterative approach or a two-step methodology. An iterative method consists in starting with a regular grid and refining it with new measurements according to the first estimations obtained. Another way is to start with a random design (or not, if there are important difficulties of access) and to validate the sample data representativeness with statistical tests based on some characteristic parameters of the data. In other words, if the population under study is characterized by two parameters, the sample has to be compared to the population for these two parameters using statistical tests like Kolmogorov-Smirnov, Ansari Bradly, … If not, additional measurements are taken to complete the sample and the process is performed again. In [39], this method has been used to characterize the flow rate of radioactive gas contained in waste drums.

Another approach which can be called “two-step methodology” consists in performing first an exhaustive scanning with coarse measurements in order to identify potential hot spots and discontinuities of the phenomenon. Then, new measurements are taken according to a specific sampling design which includes an adapted grid more precise on the areas of interest.

For the random sampling methods presented in this section, there are specific estimators of the quantity of interest associated with an estimation error. For example for stratified proportional sampling, we consider a 200-size sample of ^{137}Cs measurement data taken from a population of 2000 waste drums for which an auxiliary variable related to the production origin exists, with 3 conditions (A, B and C) which give the 3 sampling strata. The objective is to estimate the proportion of drums which exceeds a given threshold. Table 2 gives the description elements of the case study.

Table 2: Case study description.

Production origin	Stratum size N_i	Number n_i of sample individuals in the stratum i	Proportion p_i of non-compliance in the sample
A	100	10	0.1
B	500	50	0.3
C	1400	140	0.4

The theory of proportional stratified sampling gives the following definitions for the estimation of the proportion mean $\hat{p}$, the variance $\hat{\sigma}$ and the confidence interval of $\hat{p}$:

$$\hat{p} = \sum_{i=1}^{3} \frac{N_i}{N} p_i = \frac{100}{2000} \times 0.1 + \frac{500}{2000} \times 0.3 + \frac{1400}{2000} \times 0.4 = 0.36$$

with *N*, the size of the population of drums; N_i, the number of individuals in the stratum i and p_i, the proportion of non-compliances measured in the sample corresponding to the stratum i.

The variance of the proportion estimation is given by:

$$Var(\hat{p}\ \ ) = \sum_{i=1}^{3} \frac{N_i^2}{N^2}(1 - f_i) \times \frac{p_i(1 - p_i)}{n_i}$$

$$Var(\hat{p}\ \ ) = \frac{100^2}{2000^2}\left(1 - \frac{10}{100}\right)\frac{0.1(1 - 0.1)}{200} + \frac{500^2}{2000^2}\left(1 - \frac{50}{500}\right)\frac{0.3(1 - 0.3)}{200} + \frac{1400^2}{2000^2}\left(1 - \frac{140}{1400}\right)\frac{0.4(1 - 0.4)}{200} = 6.10^{-4}$$

with n_i, the number of sample individuals in the stratum i and f_i, the survey rate equal to $\frac{n_i}{N_i}$.

The 95% confidence interval of $\hat{p}$ is then given by:

$$Ic^{95\%}(\hat{p}) = \left[\hat{p} - 1.96 \times \sqrt{Var(\hat{p})}; \hat{p} + 1.96 \times \sqrt{Var(\hat{p})}\right]$$

$$Ic^{95\%}(\hat{p}) = \left[0.36 - 1.96 \times \sqrt{6.10^{-4}}; 0.36 + 1.96 \times \sqrt{6.10^{-4}}\right]$$

$$Ic^{95\%}(\hat{p}) = [0.31; 0.41].$$

Case of segmented matter or material

When a sample originates from a much segmented matter or material, it is possible to forecast the relative precision of a sample to represent the content of a given batch.

Pierre Gy ([40]) has developed a formula relating to the fragment size, the mass of the sample batch and to mineralogical and shape parameters:

$$s_r^2 = \frac{s^2}{a_L^2} = Kl\frac{d^3}{M_e}\left(1 - \frac{M_e}{M_L}\right) \approx Kl\frac{d^3}{M_e} \qquad \textbf{\textit{(Eq. 2.3.2-6)}}$$

A description and discussion about all the parameters can be found in [41].

This approach or formula is correct only for a « random sample », i.e. every fragment or piece in the sample batch has the same probability to be chosen (otherwise there is a bias and the relative variance estimated increases).

This method has been used for determining the radionuclide content in graphite wastes. Given that radionuclide measurements in irradiated graphite exhibit very large discrepancies, only Gy's method explains the results observed ([42]).

2.3.3 Sampling strategy for geostatistical approach

This subsection briefly explains sampling. In the context of a radiological survey, sampling may be achieved using numerous techniques which are broadly classified as direct measurements, scanning measurements or laboratory measurements.

The spatial distribution of sample locations is termed "sampling scheme". A sampling scheme may be systematic (e.g., based on a grid), randomized, judgmental, or a combination of the latter (see § 2.3.2 and Figure 5 [7]). Sampling schemes may be applied to structured or unstructured phenomena, provided some key constraints are considered ([43]):

- Sample spacing should not be greater than the spatial scale of the regionalized variable. When this constraint is not met, the added value of geostatistics compared to statistics is reduced because the spatial structure will be difficult to infer and the estimation uncertainty will become close to statistical uncertainty.
- Sample spacing should not be significantly less than the spatial scale of the regionalized variable. When this constraint is not met, the added value of geostatistics compared to other interpolation methods (e.g., a spline) is reduced.
- Samples should be located throughout the sample domain. A poor coverage at borders of the domain can lead to an increased prediction uncertainty due to extrapolation.

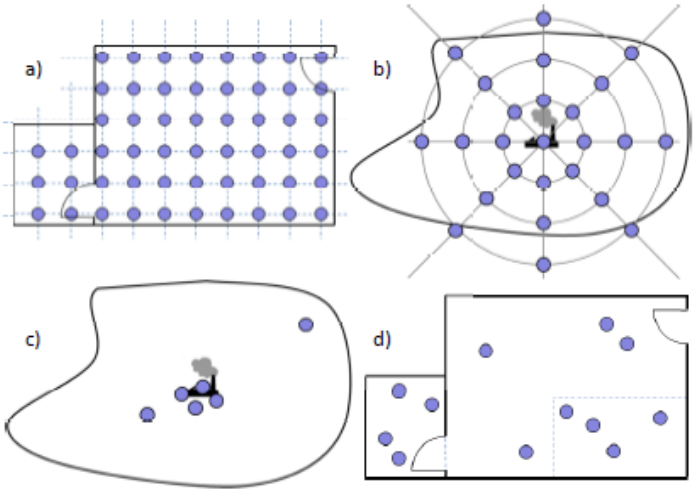


Figure 5: Examples of Sampling schemes: a) Systematic, b) Circular, c) Judgmental and d) Randomized.

For a regionalized variable in a 3D space, such as concrete contamination, the sampling scheme must also span the 3D space. Radiological concentration measurements in a volume (i.e., in the vertical direction) can be achieved with laboratory samples collected by core boring or at well locations, or with borehole gamma logs. In 3D, particular attention should be given to possible heterogeneity and anisotropy (spatial correlation with different properties depending on direction). The mechanisms that lead to variations along the surface may vary significantly compared to those that lead to variations in the vertical direction, for instance.

2.4 Descriptive analysis

The goal of descriptive analysis ([44]) is to describe the distribution of sets of variables and explore the relations between two sets of measured data (with more than two sets, multivariate analysis is used). This first step in data mining includes: the typing of different data (quantitative or qualitative), their distributions, the research of outliers, the homoscedasticity and heteroscedasticity, the discriminating variables, their transformations and the creation of new sets of variables. The conclusions of this first data analysis are often decisive for the future interpretations or conclusions inferred.

Counting and frequency data analysis are often presented in tables. Clustering data is also more convenient to account for the variability. The construction and the determination of the numbers of clusters are tricky operations. The graphical distributions of these clusters are interpreted using histograms. However, too many or too few clusters may result in wrong and inconsistent conclusions. The theory of density estimation can be an appropriate solution or approach for these problems.

For univariate statistics, graphical representations like box plots are useful to show the variation and the range of continuous variables. Therefore, different indicators can be used but it is important to keep in mind that it is always inadequate to sum up a pool of data using only one indicator (or

statistics): size, mean, variance, standard-deviation, median, min, max, range, quartiles (1st, 2nd, 3rd), skewness (asymmetry), kurtosis (flatness), etc.…

For bivariate analysis, the linear coefficient or Pearson coefficient is generally considered. The values of outliers have to be analysed with attention (see § 2.5).

2.4.1 One-point statistics

2.4.1.1 Base map and declustering

The geographical distribution of data points may have a significant impact on the determination of statistical characteristics. A very basic tool to visualize the spatial location of values is a base map. In 2D, it can be considered as the scatter plot of X- and Y-coordinates.

The contribution of the different data points is usually considered as equal, which may distort the statistics in the presence of a spatially correlated phenomenon (spatial redundancy). As radiological levels are not generally spatially independent (see Fukushima example in the Figure 6), it is necessary to handle spatial density or coverage of data points correctly. In reality, it is rare to have data points distributed according to a probabilistic scheme (random strategy or pseudo-regular grid) due to the presence of data clusters and to the lack of information concerning some parts of the field. This may be linked to the combination of different sampling campaigns with distinct characterization objectives and to technical reasons (accessibility, representativeness…).

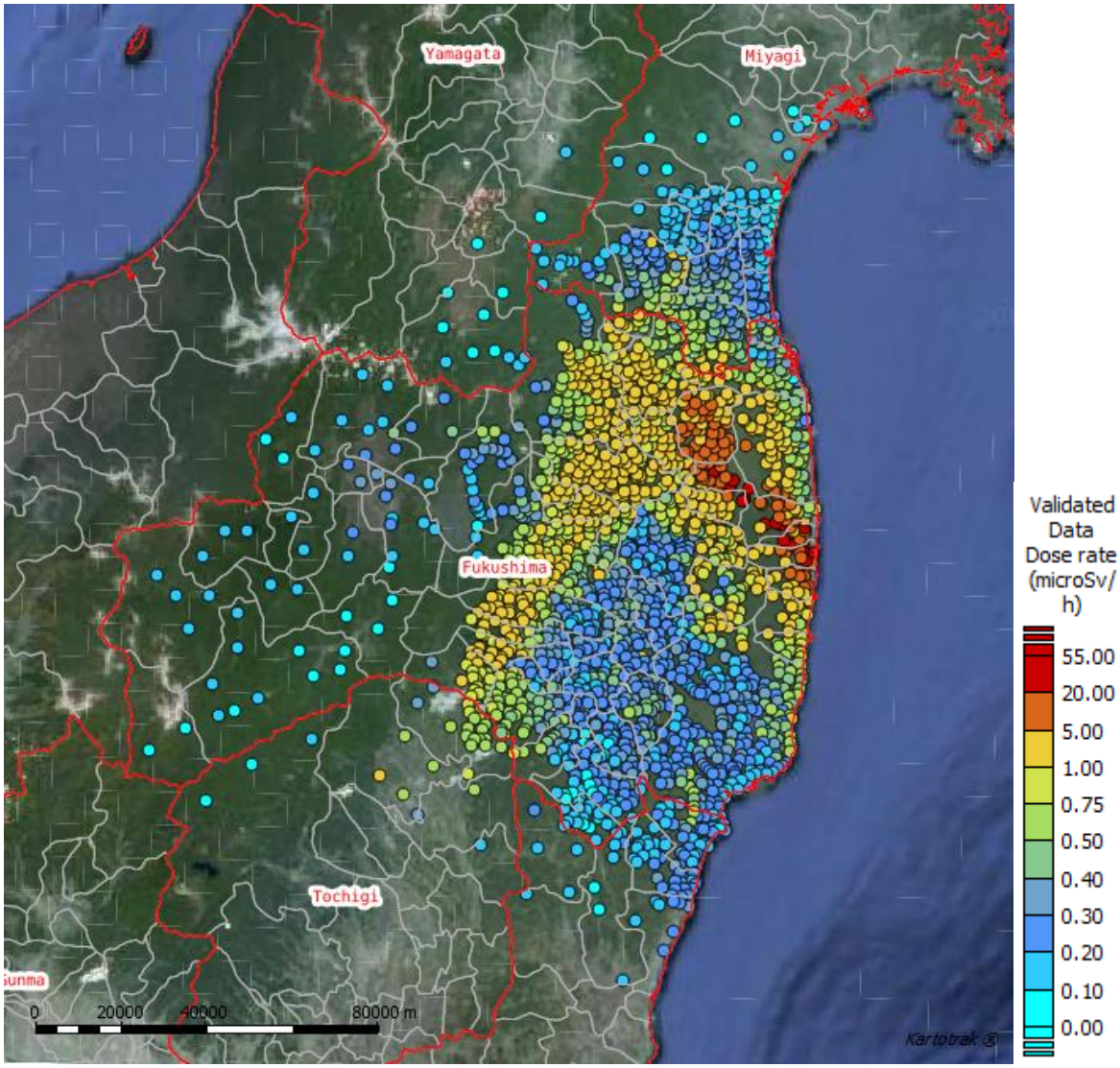


Figure 6: Base map of dose rate measurements on Fukushima site in summer 2011 (MEXT source).

To correct this possible sampling bias, it is necessary to weight each sample value appropriately. For spatial data, this implies using the amount of area that each sample affects. A simple weighting system is obtained by segmenting space according to the distance to the closest point (Voronoi diagram, see Figure 7) using Euclidean metrics.

Another technique consists in defining weights as inversely proportional to the number of data points in each cell of a grid. The size of the grid is generally determined to minimize the global mean because high values are usually more densely sampled (in case of contaminated sites). More advanced weighting systems such as kriging weights of the mean may also be used.

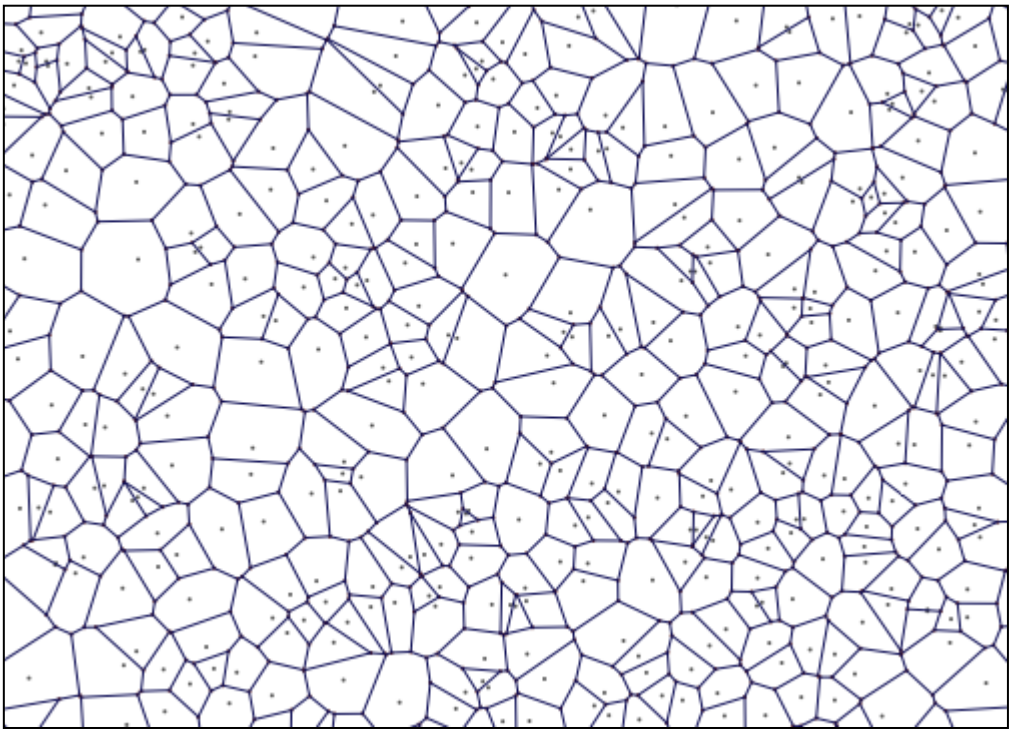


Figure 7: Use of polygons to decluster data by area weighting.

2.4.1.2 Statistical distribution

For visual representations of statistical distributions, box plots and histograms are fundamental analysis tools. A box plot summarizes previous statistics and allows the quick identification of outliers. An histogram is the frequency distribution graph. It rapidly provides a good graphic representation of the data distribution shape.

The most familiar distribution is the normal, or Gaussian, distribution. This distribution is a bell-shaped curve used in many statistical applications. A normal distribution is generally encountered when the value of a phenomenon is the resultant sum of several independent causes. It can be concisely defined and it provides symmetry and other convenient properties for theoretical estimations. For instance the background activity of an homogeneous matrix is expected to follow a Gaussian distribution.

An histogram is useful to check the symmetric distribution of activity levels qualitatively. It is particularly interesting to be able to distinguish two (or more) populations within the same data set. In Figure 8, the histogram shows a bimodal distribution. As values are spatially organized, it is then possible to divide the whole heterogeneous data set into two more homogeneous ones.

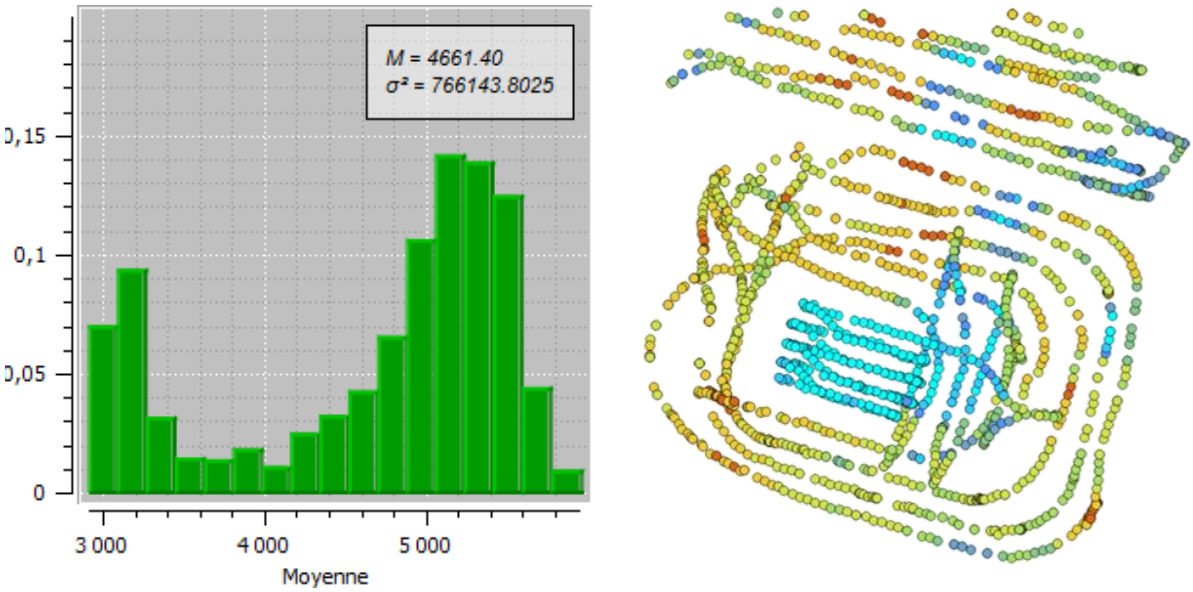


Figure 8: Example of a bimodal histogram and corresponding base map.

This domaining procedure can also be used a priori on the basis of matrix differences or of any other parameter influencing background levels. A difficulty often arises when two populations with almost similar statistics are mixed up. It can then be impossible to distinguish the bimodal histogram. However, the analysis should integrate other pieces of information which deal with homogeneity.

The concept of sampling domains is of critical importance for the correct design and implementation of a successful sampling program and this holds true for both spatial and non-spatial statistics. Mixing domains is an open invitation to both errors and difficulties and it should be avoided to the maximum extent possible and practicable. On the contrary, grouping populations with the same characteristics may simplify the analysis.

2.4.1.3 Analysis of correlations

A correlation is often present between two variables and it can easily be detected visually. A scatter plot presents the points along a line and underlines the relationship between the two variables (Figure 9). If there is a significant correlation, the higher one variable, the higher/lower the second. Such a graph may be used as a primitive prediction model. For nuclides, the correlation is generally positive, i.e. the regression line slope is positive. From a mathematical point of view, linear regression is a least-squares regression or a minimum variance regression. The correlation coefficient is aimed at quantifying how closely values are distributed around a straight line. However, this quantity is commonly used erroneously to measure two entities: the magnitude of the regression line slope and the appropriateness of the straight-line model. That is why the correlation diagram should run parallel to the regression.

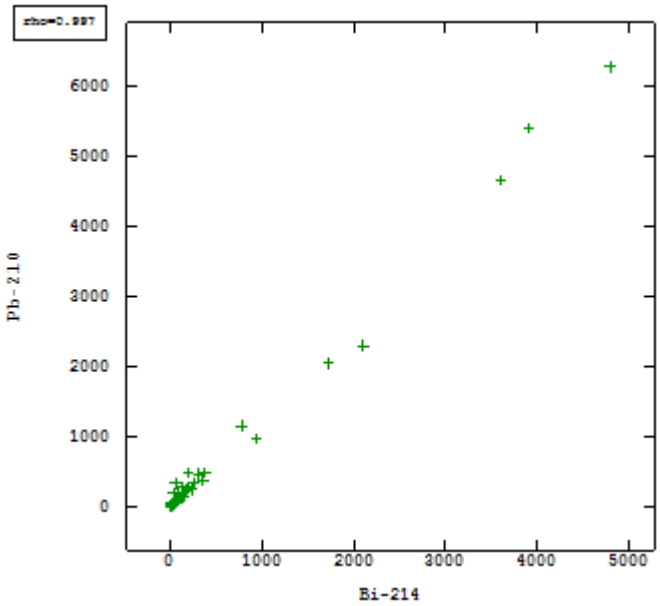


Figure 9: Two nuclides linked by a positive linear correlation.

Linear correlation can easily be found for nuclides within the same decay chain or with similar matrices. Ratios or scaling factors can then be determined to simplify analyses (in particular for hard-to-detect nuclides in correlation with easy-to-measure ones). Both the concentration ranges and the scaling factor ratios exhibit log-normal distributions ([20], [45]). Therefore, a scatter diagram on the logarithmic scale makes it possible to clearly grasp the correlation between radionuclides. Figure 10 shows improving scaling factors using a wide activity concentration range and composite sampling.

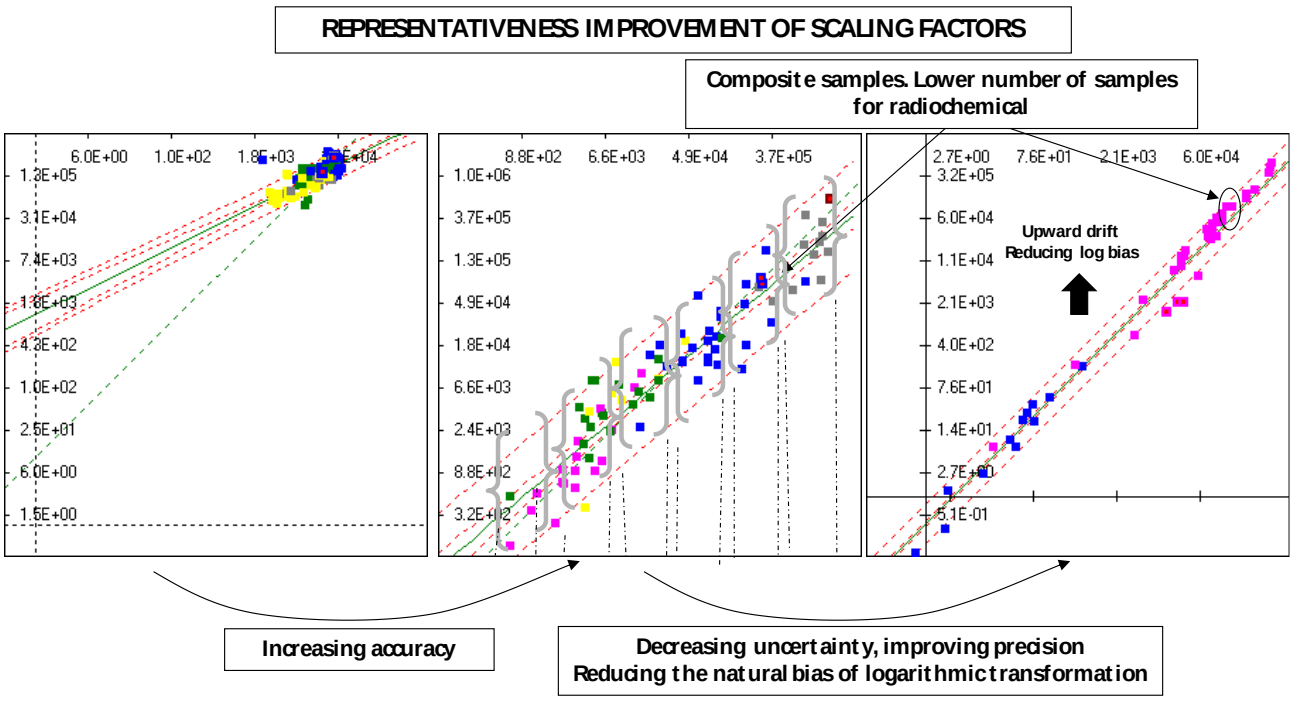


Figure 10: Logarithmic scale scatter plots to improve scaling factors

Distinct fingerprints may be identified on scatter plots (several populations with different linear correlations, as shown on the Figure 11).

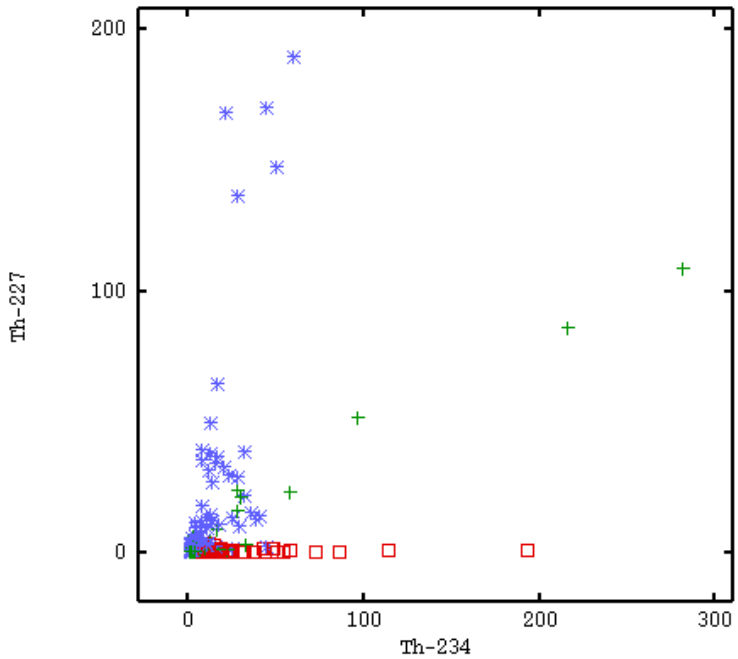


Figure 11: Identification of different groups of correlation.

Similarly, the identification of independent nuclides (background vs. contamination) is an interesting output from scatter plots.

2.4.1.4 Impact of sample geometry on variability

The physical size (volume or mass) of a sampling unit is very important when it comes to characterizing an environmental area. This parameter may have a strong impact on the variable statistical distribution. While mean value remains unaffected, the larger the support, the smaller the variance and standard deviation. This phenomenon is generally known as “support effect” (Figure 12).

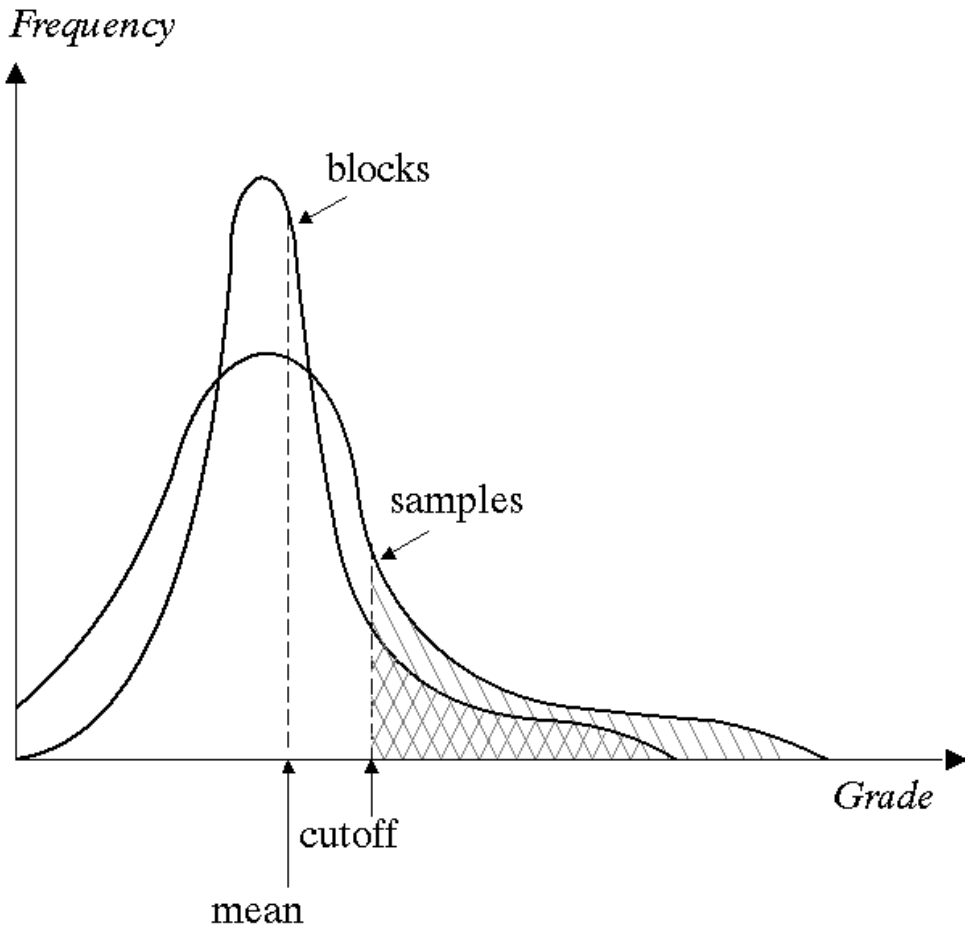


Figure 12: Impact of sample support on statistical distribution.

Natural heterogeneity of materials and radionuclide distribution within the materials produce variations in activity levels. The support effect more or less averages these variations. practically when field samples (already representative of a large volume of materials) reach the laboratory, just a small amount of materials may actually be used to represent the entire sample. Successively smaller volumes of materials are taken as the sample approaches the laboratory analytical device. Thus, the variability of the initial sample data will increase unless appropriate subsampling methods are applied.

Another practical consequence is the prior regularization or correction for the support effect when analyzing a data set combining results from samples with different supports (from different sampling campaigns, from different sampling techniques…).

2.4.2 Two-point statistics

2.4.2.1 Spatial structure analysis (variography)

The spatial variation of a contaminant within a domain can be quantified by the variogram (Figure 13). The variogram is a function that shows how the variation between observations of a variable in two locations depends on the spatial distance between these locations. The variogram is half the mean squared difference between two observations plotted against the distance between them, for all the results in a data set ([46]).

$$\gamma(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2$$

The variogram typically increases with distance until the plot reaches a plateau at a value called the sill variance and at a distance called the range. If the range is very short, this shows that the spatial variation is very intricate. If the range is longer, then it makes sense to create a map of spatial variations from sample observations on a grid.

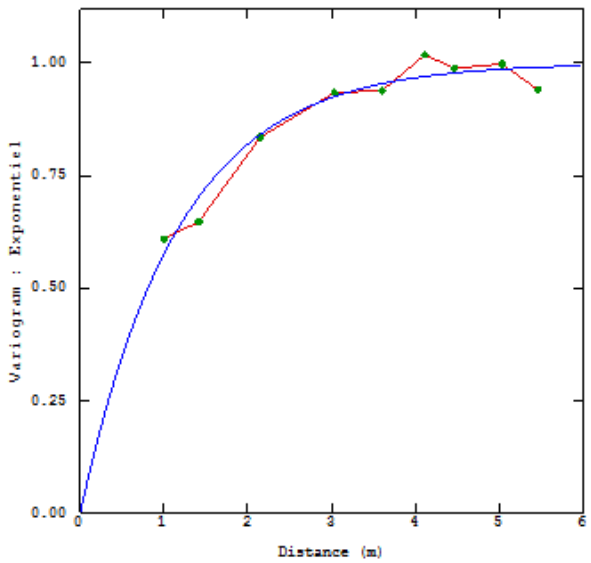


Figure 13: Experimental variogram (green dots) and its fitting (blue curve).

Variography is a very powerful and visual tool for the identification of outliers and boundaries between different spatial populations.

Figure 14 illustrates three phenomena with the same statistical characteristics (shown on histograms). However, they have very different spatial organization (variograms):

- On the left, a spatial random phenomenon with a pure nugget model as variogram, in which the variability equals the experimental variance whatever the distance. Even for two very close points, the variability is very high. It can be the case for contaminated materials from different origins in trenches. In that case, geostatistical techniques will only give the same results as classical statistics.
- In the center, a largely continuous phenomenon with a linear increase in variability at small scale, followed by a sill at 15m range. This case is quite common with radiological contaminations (soils, concrete…).
- On the right, a continuous phenomenon with a progressive increase in variability at small scale, followed by a sill at 15m range. This case can be encountered with phenomena that can be partly modelled deterministically (activation, dispersion plume, contaminated groundwater…).

The variogram, which is based on data, allows the interpretation and modelling of the spatial continuity of the phenomenon. This spatial structure is crucial for the overall geostatistical approach. Contrary to data sets presented in Figure 14, values are generally sparsely available (punctual points, regular grid, random locations…) and the experimental variogram is fitted to interpret and model the spatial structure.

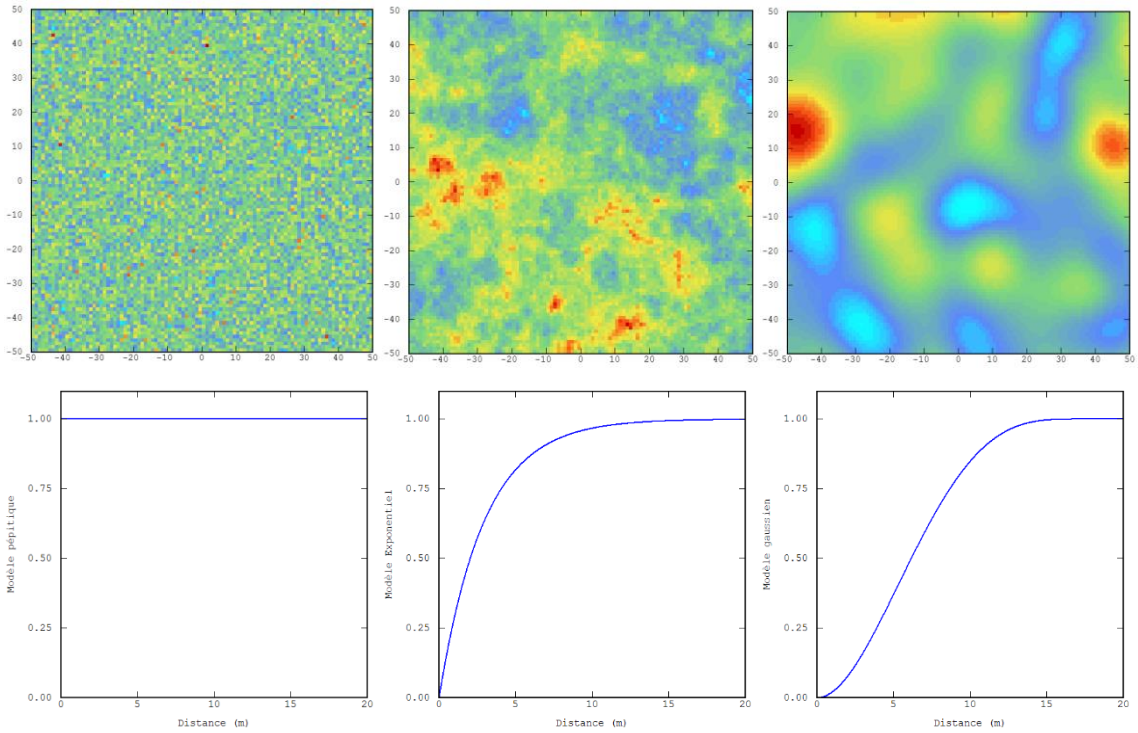


Figure 14: Three phenomena with the same statistical distribution (at the top) but with significantly different spatial structures (corresponding variograms on the bottom part).

Phenomenon anisotropy is tackled using directional variograms.

2.4.2.2 Impact of sample geometry on spatial structure

In addition to the support effect discussion in previous section, the sample geometry not only has consequences on statistics but also on geostatistical interpretation [47]. The two impacts of sample support on spatial structure (Figure 14) are:

- Reduction of the sill value (already mentioned for one-point statistics: variance decrease);
- Modification of the spatial behaviour at small scale (averaging effect): the larger the sample size, the lower the variability at short distances (the phenomenon seems more continuous, the variogram shows a parabolic shape around the origin).

Figure 14 illustrates a unique phenomenon measured on a 0.1m regular grid with 3 different sample sizes: 0.1, 0.5 and 0.9 (from left to right). The impact of sample support on variograms, respectively red, blue and green, underlines the variance decrease and the modification of spatial continuity at small scale.

2.4.3 Spatial trends

The presence of spatial trends can have a significant impact on the sampling strategy to follow. Checking if trends are present is therefore an important part of the preliminary statistical analysis. In many cases, the presence of a trend is obvious from the historical background (*e.g.* activated materials show a trend with the distance from the source of radiation) or by looking at a map showing measurement results. However when there is a doubt, some specific methods can be applied to check that point more thoroughly.

The descriptive two-point statistical analysis outlined above, essentially the experimental variogram, can also be used for the detection of spatial trends. Journel and Huijbregts ([48]) suggest that an experimental variogram increasing as rapidly as $|h|^2$ for large distances h most often indicates the presence of a trend. For a stationary random field, it is actually expected that the semi-variance stabilizes at a certain distance and equals the variance of the data.

Another approach consists in using Kendall's τ test ([49]) to measure the probability of concordance minus the probability of disconcordance. The observation pairs Z(x) and Z(x + h) are concordant if,

for the one-dimension case, an increase in the spatial coordinate corresponds to an increase in the observed value. If it corresponds to a decrease in the observed value, they are disconcordant. Kendall's τ can be calculated for any spatial direction (*e.g.* distance from the source of radiation). If Kendall's τ test is combined with a significance test, it can indicate in which directions significant spatial trends exist.

If spatial trends are present, they can be analyzed in different ways using the geostatistical techniques described in section 3, or regression approaches partly outlined in section 4.3, and considering spatial coordinates, which potentially come in addition to other auxiliary variables, as regressors.

2.5 Factorial analysis for different types of data

The goal of factorial analysis is to describe the relations of the whole pool of data. The pool of data is defined by many types of data depending on the protocol and experimentation used for measurements. Data are collected many times on a number of units and on each unit, they are characterized by many variables which are measured. Because a mere analysis based on simple statistics (or "well known statistics": averages….) may not be sufficient, factorial analysis is chosen when there are many variables and when relations exist between these variables: correlation, trend present in the pool of data. It could be convenient to summarize the pool of data and the related variability to allow a simple representation which would facilitate the 2D representation of data, i.e. a simple projection on a plan. This is the most challenging phase of factorial analysis: giving an easily and comprehensible information from a complex (because multivariate) pool of data. This chapter only describes the most common method, the Principal Component Analysis, But there are some other factorial methods like Correspondence Analysis (FCA), Multiple Correspondence Analysis (FMCA) or Discriminant Analysis (FDA). These latter methods are used when some relations or patterns have been identified between variables ([50]).

Principal Component Analysis (PCA)

PCA is the oldest multivariate methods (1901, Pearson; 1933-36 Hotelling; 1964 Rao). The PCA is a linear method because principal components are linear combinations and their interpretation is based on a linear correlation coefficient. But when non-linear relations exist between variables, this latter coefficient is not very meaningful. To achieve the linearity, transformations of different variables are made to obtain or approach criteria of normality (and use the different subsequent methods).

The first step in the analysis of a principal component is the analysis of the variance-covariance matrix and of the correlation matrix Σ.The first principal component can be defined from Σ ([51]).

The first principal component describes the maximum of variability of the pool data set. Then the second principal component describes the maximum variability of the pool data set with the condition that this second principal component is orthogonal to the first one. And so on for the other principal components (third etc.…), with each principal component explaining a percentage of total variation of the pool data set. Therefore the first two linear combinations (first and second principal component analysis) explain the maximum of variability and can be represented on a 2D diagram. Each principal component is described by different correlations and their contribution to the explanation of the principal component. With a 2D representation, we can thus access to the variation analysis of the pool data set.

When some variables or individuals are suspected to be mismeasured, the PCA may be helpful. By declaring these variables or these individuals as supplementary, they are not contributing to the definition of the principal factors, it is possible to achieve the different relations without any interference on the pool data set.

Because the interpretation of PCA is very easy and useful, the technique has been extended to analyse the outliers or missing values identified in a preliminary descriptive analysis. These data impact the rough estimate and must be analysed carefully ([52]).

- If there are some missing values and if the distribution is unknown, variables have to be left out. If the distribution is known, some imputation methods exist.

- For outliers, a dilemma also exists: leave out values or integrate their information. The process of integration is dependent upon the power of discriminating variable:
  - if the power is low, the value should be a missing value;
  - if the power is high, a model can be created.

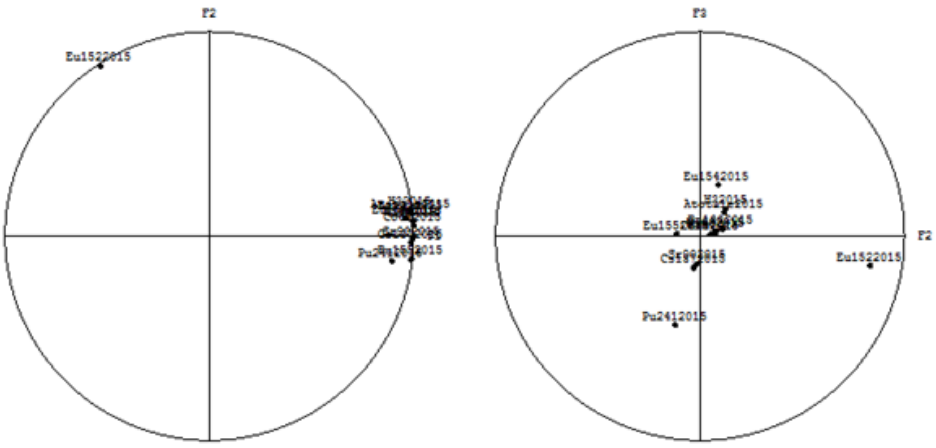


Figure 15: Correlation circles.

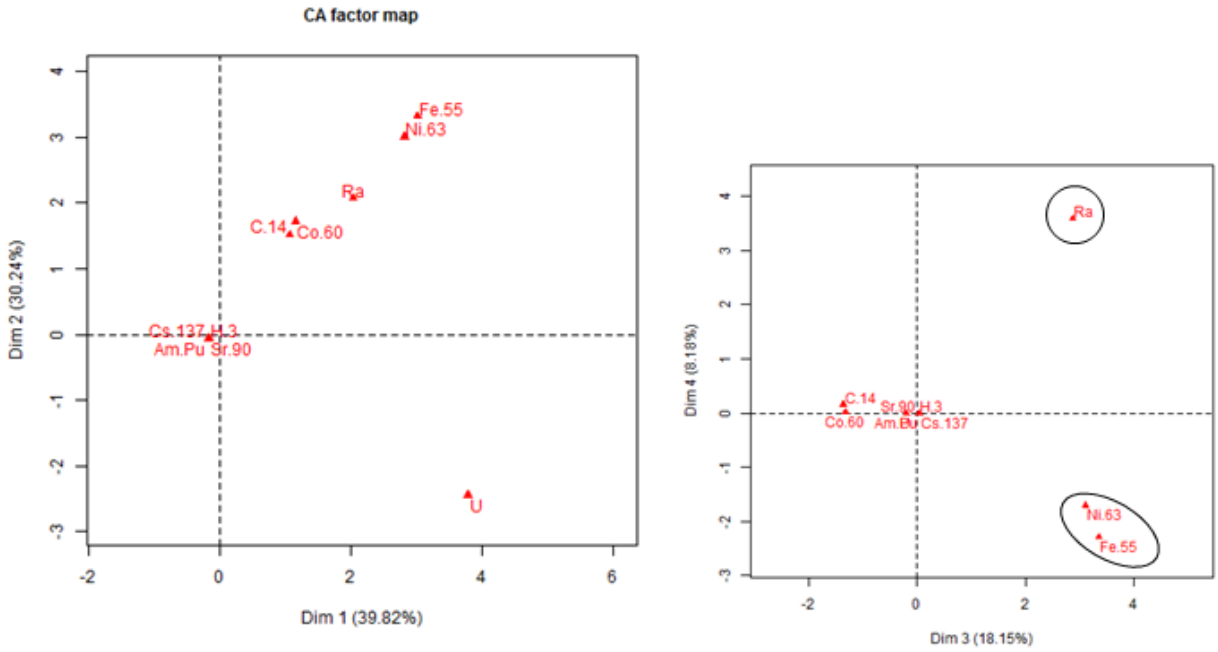


Figure 16: PCA Plan1-2 (left)/Plan 3-4 (right).

Combined with a complete analysis of the projection quality of different data (not given here), the correlation circles (Figure 15) used as first approximations give an interpretation of the different axes 1, 2 and 3 (respectively explaining 70% of the total variability in the plan 1-2 and only 26% of the total variability of the projected data in the plan 3-4 (Figure 16)). Therefore, an interpretation of these illustrations is that Fe-55 and Ni-63 are strongly correlated together but there is no correlation with Ra. For the other radionuclides, the PCA is not informative (all the projections are close to the centre).

Factorial Correspondence Analysis (FCA)

When a link between two qualitative variables is characterized, the pool of data is classically represented on a 2D plan (like PCA) and the distance between individuals is interpreted as a χ^2 distance. Clusters can be made and tests can be performed.

Factorial Multiple Correspondence Analysis (FMCA)

This method can be interpreted as a PCA with individuals described by qualitative variables with several levels. Individuals and variables can be represented in a 2D plan to reveal associations and clusters.

Factorial Discriminant Analysis (FDA)

This method is used when all the variables are assumed to follow a normal (Gaussian) distribution. The distance between individuals can be tested (Mahalanobis distance using an hypergeometric law). A clustering of individuals can be made (total variance = intragroup variance + intergroup variance, due to the Gaussian method used).

3 Statistical approach for data with spatial structure and "enough" data

Geostatistical processing of data is based on all the available pieces of information: historical data, non-destructive measurements and laboratory analyses of samples. The spatial structure modelling is then used to create maps and to estimate the extent of radioactive contamination (surface and depth) ([53], [54]). Quantifications of local and global uncertainties are powerful decision-making tools for an improved management of remediation projects on contaminated sites and for nuclear facility decontamination and dismantling projects. They can be used to identify hot spots, estimate surface and volume contaminations, classify radioactive wastes according to thresholds, estimate source terms, and so on. Theoretical discussions and a more detailed presentation of geostatistics can be found in [46]

3.1 Kriging estimation (+ variance estimation)

With input data and after identification of the spatial structure on the variogram, geostatistical techniques can estimate the variable under study using a method similar to regression analysis called "kriging" (best linear unbiased estimator). This always implies a quantification of the associated uncertainty. As for classic interpolations, the kriging estimation of the surface activity $\hat{z}(x_0)$ at a point 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 classic interpolators);
- The distances between the data (clusters…);
- The spatial structure of the studied phenomenon. Indeed, for equal geometric data configurations, the cartography should integrate by construction the phenomenon spatial continuity (for example, very smooth or heterogeneous).

Knowing the location of the target point, the data configuration and the variogram model, kriging consists in computing the kriging weights by making a matrix inversion. These weights guarantee:

- an unbiased estimation $E(Z - \hat{Z}) = 0$,
- a minimization of the variance of the estimation error $Var(Z - \hat{Z})$, which corresponds to a reduction of the risk of error.

As a result, kriging provides the best linear unbiased estimation and the related uncertainty in the form of a variance of the estimation error.

3.2 Measurement uncertainty considerations

Statistical measurement uncertainty can be considered through the use of a kriging variant that covers this contribution. However, only the stochastic uncertainty should be taken into account, not the deterministic part of measurement uncertainty (calibration, density, homogeneity…) ([55]).

With a systematic error (calibration factor for instance), the corresponding coefficient only impacts variances and sills in the geostatistical analysis (but the spatial structure remains the same). Of course, estimation results are impacted by the same bias.

Kriging with measurement errors ε_i assumes:

- Errors are non-systematic, $E[\varepsilon_i] = 0$
- Errors are uncorrelated with the random function, $E[\varepsilon_i Z(x)] = 0$
- Errors may be correlated between themselves, $E[\varepsilon_i \varepsilon_j] = S_{ij}$.

The variographic analysis is then slightly modified to consider this variance contribution. It can also be seen as a particular case of multivariate approach. In case of uncorrelated errors, kriging estimation will filter errors as the standard kriging system is the same as an enlarged variogram model, except at data points.

The difficulty with nuclear measurements comes from the fact that laboratory results are presented with an associated uncertainty that mixes up systematic and random errors. In addition, these latter are generally negligible compared to sampling and systematic errors.

3.3 Multi-variable kriging

Multivariate geostatistical techniques allow different kinds of information to be combined in order to improve estimates using the spatial correlations between variables. Physical and historical data and non-destructive measurement results (for example dose rate or in situ gamma spectrometry) are thus combined to improve understanding and the prediction of the main variable (results of laboratory analysis, for example) while reducing the estimation uncertainty.

The geostatistical approach is simply extended to a multivariate version:

$$Z^*(x_0) = \sum_{\alpha} \lambda_\alpha Z(x_\alpha) + \sum_{\beta} \lambda'_\beta Aux(x_\beta)$$

The variographic analysis becomes larger as cross-variograms come in addition to simple variograms. A linear model of coregionalisation allows the joint modelling of all the spatial structures.

Similarly, outputs from numerical models (activation, flow, transport, dispersion…) can be adequately integrated as a dense auxiliary variable thanks to a geostatistical multivariate approach and its variants (bundled collocated cokriging or cosimulations, external drift, universal kriging…).

A systematic variation of radiological contamination may occur in practice. For instance, for radionuclides/media with low diffusivity (e.g., ^{137}Cs migration in soil or concrete), concentration may systematically vary as a function of penetration depth. Another example is given by activated materials due to a neutron flux. In that case, a non-stationary model can be adjusted deterministically (with the migration profile as a linear or exponential decrease) and/or geostatistically (universal kriging or external drift kriging). Outputs from numerical models can be easily integrated in the geostatistical process as a drift and/or an auxiliary variable that explains and models the long-range spatial continuity.

3.4 Conditional simulation and risk analysis

More advanced and sophisticated geostatistical methods can be used to quantify different uncertainties, such as the risk of exceeding the threshold (see Figure 19), which is a non-linear

quantity for instance. These estimates are powerful decision-making aids when it comes to classifying surfaces and volumes before decontamination starts (based on different thresholds and considering the remediation support impact).

A conditional expectation requires a normal score transformation of the variable thanks to a Gaussian anamorphosis ([56]). Geostatistical simulations can be seen as a Monte-Carlo method that respects the spatial structure. Corresponding algorithms are abundant (turning bands, sequential Gaussian …).

Change-of-support issue is a useful extension of geostatistical outputs when the support of interest is intermediate in scale between the measurement support and the entire domain. Block kriging and block simulation are then relevant to take decisions about larger areas (industrial constraint for remediation/decontamination). An illustration is presented on Figure 17 and Figure 18for the local probability levels and the corresponding global risk curves.

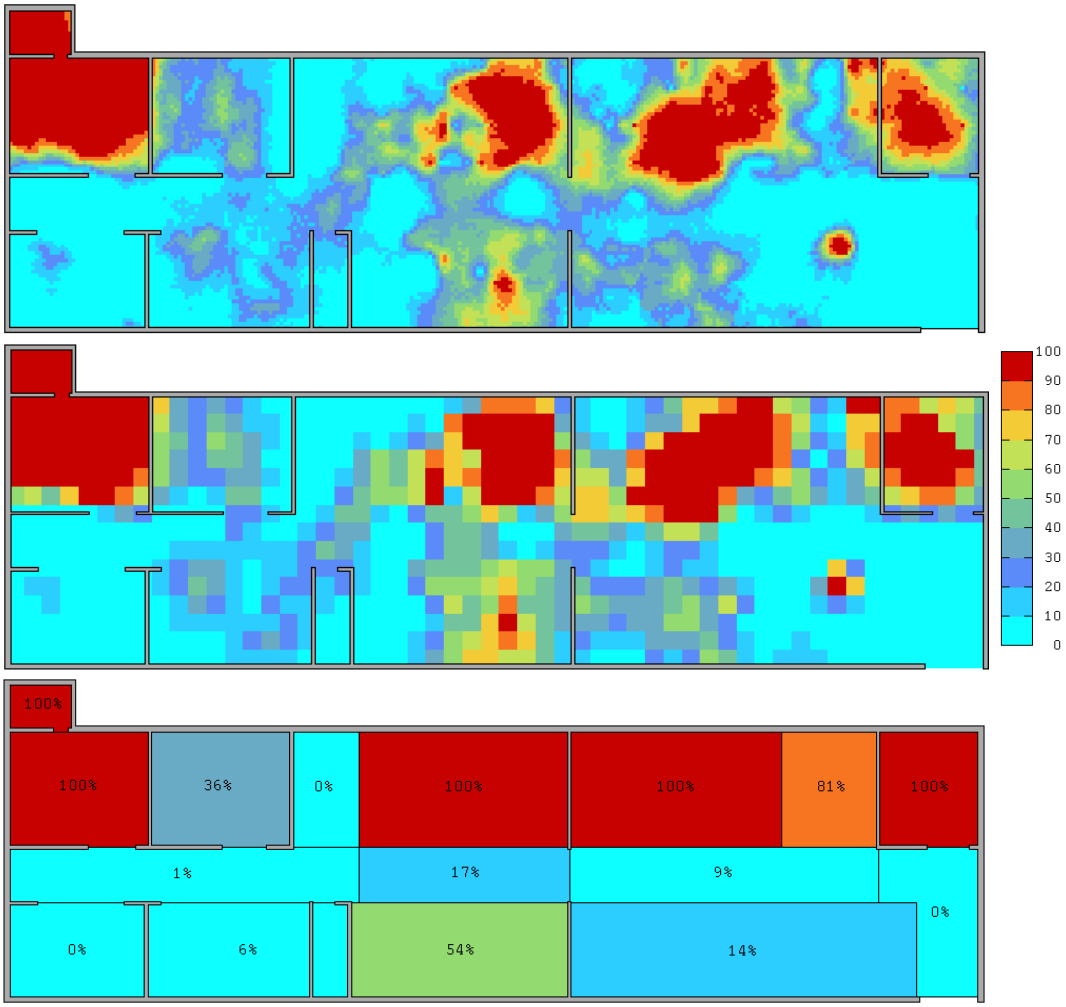


Figure 17: Probability of exceeding a threshold for a punctual support (top), a 1 m² decision unit (middle) and a workstation area (bottom).

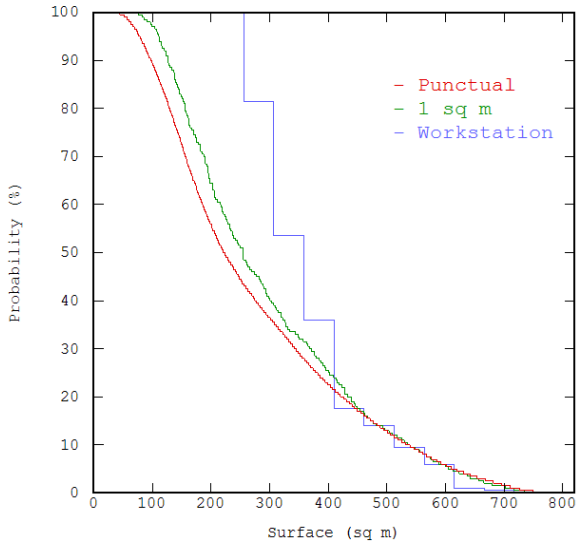


Figure 18: Corresponding global risk curves.

3.5 Specific sampling strategy

The spatial structure of radioactive contamination makes the optimization of sampling (number and position of data points) particularly relevant. Geostatistical methodology can help determine the initial mesh size and reduce estimation uncertainties. Three sequential phases are optimally used to characterize contamination:

- Historical and functional analysis, based on operation records and operators' statements.
- Radiation maps from non-destructive measurements, where possible. These are usually semi-quantitative values which provide a spatial distribution of contamination.
- Characterization of activity levels and depth profiles by collecting samples by coring, hammering and so on, for laboratory analysis purposes.

Each step can be repeated, to reach acceptable levels of confidence or precision. With radiation mapping, prior information about spatial structures of radioactive contamination is used to determine the initial sampling mesh. By default, for contaminated soils around nuclear facilities, 200 data points per hectare (corresponding to a 7 m mesh size) allows a first analysis of the phenomenon spatial structure. This initial mesh may range from 5 m to 10 m according to the available information (historical and functional analyses, prior measurements…). More generally, for larger sizes, the tenth or the twentieth of the site size is a good candidate for the initial mesh size of the regular grid. For building structures (mainly concrete), a 1 m mesh size may be relevant to conduct a first geostatistical analysis. Similarly, this size has to be adapted to the size of the building, the stakes and the expected spatial structure.

Then adding extra data points is a good way to reduce estimation uncertainty. This can be quickly obtained by analyzing early mapping results. For geometric uncertainties, the kriging error variance easily identifies areas with a lower sampling density. For high variability areas, the confidence interval around the estimated value is used to detect the boundaries of contaminated areas ([57]), for instance

For probability results, the risk of exceeding a given threshold allows surfaces or volumes to be categorized in order to optimize radioactive waste management. Figure 19 illustrates the use of a probability of exceeding a threshold to identify an area misclassification risk ([3]). This nuclear facility surface area is 800 m^2. Surface contamination measurements were made using a regular grid with a 66 cm mesh. Fifty sampling points were used to collect samples for laboratory analysis purposes. Here, the risk under study is a false negative, since an area is declared to be below the threshold based on estimation results, but in reality it exceeds the threshold. Additional sample points can be

intelligently allocated. Depending on the threshold, acceptable risks may vary and identified areas may change.

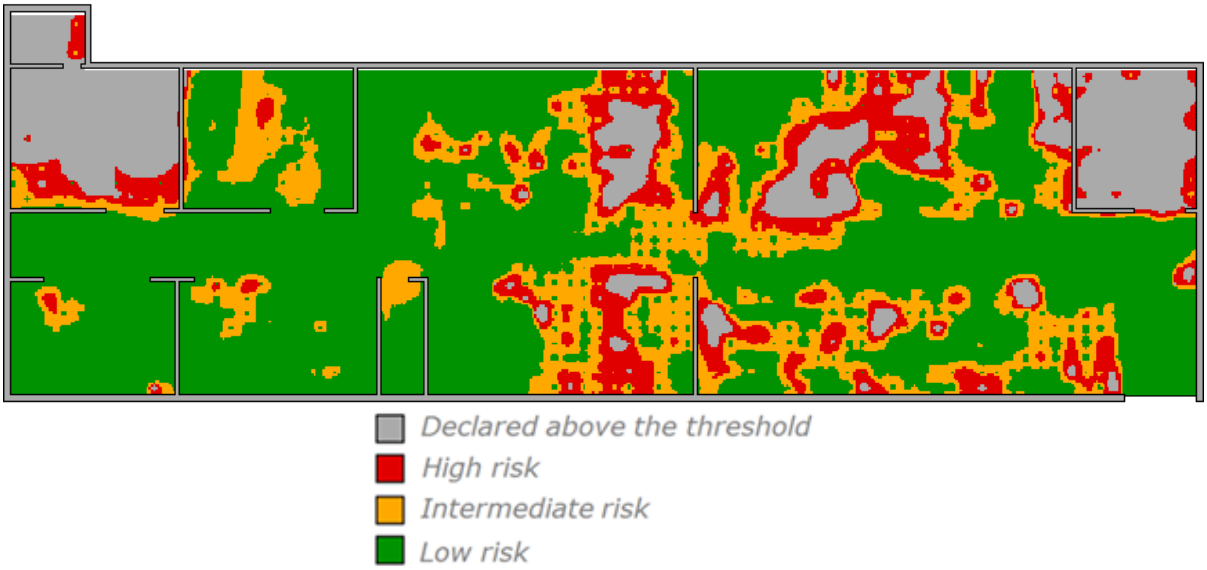


Figure 19: Map of the false negative risk (contaminated area declared as clean).

The number and positions of the destructive samples can be optimized by using the 2D radiation map to identify relevant areas. In general, hot spots are strongly investigated even though they are not representative of the rest of the facility, in particular for low-level thresholds. By taking advantage of the multivariate geostatistical approach, a sequential sampling strategy can be conducted:

- First to correlate non-destructive measurements (2D radiation map) with surface samples. It is then important to calibrate this statistical link for the full range of values.
- Second to delineate areas exceeding specific threshold for waste optimization.
- In parallel to determine migration distribution or profile in depth.

Quick gamma-scans or spectrometric measures on collected samples allow this iterative process and a sufficient reactivity as the complete set of laboratory results (α, β…) could represent a significant period of time.

4 Statistical approach for data with no spatial structure

In many cases, the measurement data to be analysed or collected are not characterized by a spatial or a temporal structure. It is notably the case for radioactive drums or individual objects like metallic pieces. The difficulties in collecting measurements lead to data sets obtained without any sampling strategy. For this type of configurations, geostatistical techniques cannot be used and they can provide results with large uncertainties whereas selected statistical techniques can give robust and rich global information. The first step consists in studying the possibility of fitting a probability distribution onto the sample data. Sometimes, in particular when the data have not been collected according to a sampling strategy, it may be useful to check the representativeness of the data set. In some cases, the information associated with the data measurements related to the physical quantity (parameter values) is available, which allows a regression analysis.

In MARSSIM standard, methods based on rank tests are recommended to demonstrate compliance during the final status survey. In section 4.4, it is shown that this methodology is well adapted to a random distribution of the contamination but is not suitable in the context of initial characterization.

4.1 Probability distribution fitting

When a large sample size (greater than 30 cf. section 5) needs to be studied, different distribution fitting techniques exist: the parametric methods, by which the distribution parameters are estimated

from the data; the non-parametric methods like histogram kernel fitting method or the regression method using a transformation of the cumulative distribution function to find a linear relation between the cumulative probability and the data (see section 4.3).

First of all, creating a histogram of the data is very informative because its shape should orientate the selection of the probability distribution (uniform, Gaussian, lognormal, Gumbel, exponential, Weibull…). If a selection is possible, the distribution parameters have to be determined from the data. To do so, the main parametric methods are the moment method, the maximum likelihood method and the L-moment method. The next step, which is very important, consists in validating the distribution fitting onto the data: graphical validation with probability-probability plot and/or quantile-quantile plot; statistical tests like Kolmogorov-Smirnov Test (more sensitive around the median and for outliers); Cramer Von Mises Test (better to account for the whole data set) and Anderson-Darling Test (better sensitivity for extreme values).

To illustrate how to perform these successive steps, a sample of 40 real data values has been used to build the Figure 20 histogram. The shape looks like a lognormal distribution, thus the moment method and the maximum likelihood method are used to determine the parameters $\theta = (\mu, \sigma)$ of the distribution law.

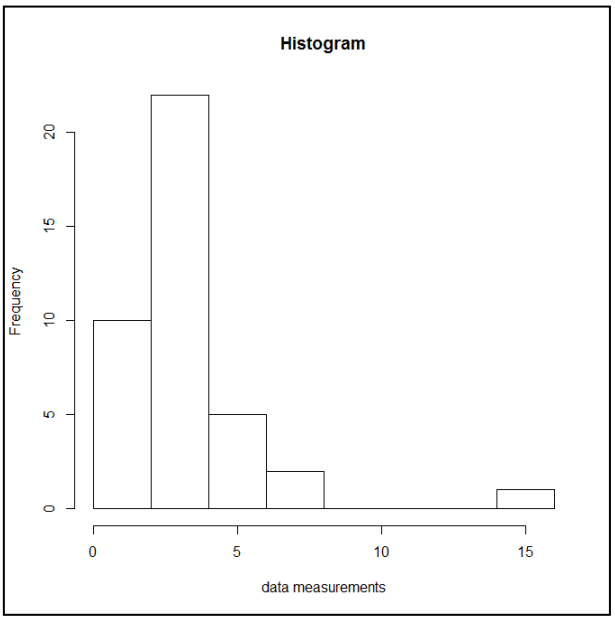


Figure 20 : Histogram of the data measurements.

Moment Method

The first and second moments of the distribution are used:

$$m_1 = E(X|\theta) = e^{\mu+\sigma^2/2} \approx \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{40}\sum_{i=1}^{40} data_i = 3.32$$

$$m_2 = Var(X|\theta) = \left(e^{\sigma^2} - 1\right)e^{2\mu+\sigma^2} \approx \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{1}{40-1}\sum_{i=1}^{40}(data_i - \mu)^2 = 5.44$$

Then the solutions[2] can be deduced: $\theta = (\mu, \sigma) = (1.004, 0.)$

Maximum Likelihood method

The likelihood formula to maximise is expressed as follows:

[2] To solve this type of problems, the R package « fitdistrplus » can be used.

$$L[(x_1, x_2, \cdots, x_{40}), (\mu, \sigma)] = \prod_{i=1}^{40} f(x_i, \mu, \sigma)$$

where $(x_1, x_2, \cdots, x_{40})$ are the data, (μ, σ) the parameters of the distribution and f the density function for the lognormal distribution. The previous formula is simplified by log-transformation and becomes:

$$log\big[L[(x_1, x_2, \cdots, x_{40}), (\mu, \sigma)]\big] = \sum_{i=1}^{40} log[f(x_i, \mu, \sigma)].$$

The parameters (μ, σ) of the fitted lognormal distribution are the solution of the following optimization problem:

$$(\hat{\mu}, \hat{\sigma}) = Argmax_{(\mu,\sigma)} \Big[log\big[L[(x_1, x_2, \cdots, x_{40}), (\mu, \sigma)]\big]\Big].$$

The solutions[2] can thus be deduced as follows: $\theta = (\mu, \sigma) = (1.04, 0.54)$.

The next step consists in validating the estimation of the distribution parameters using a quantile-quantile graph and statistical tests.

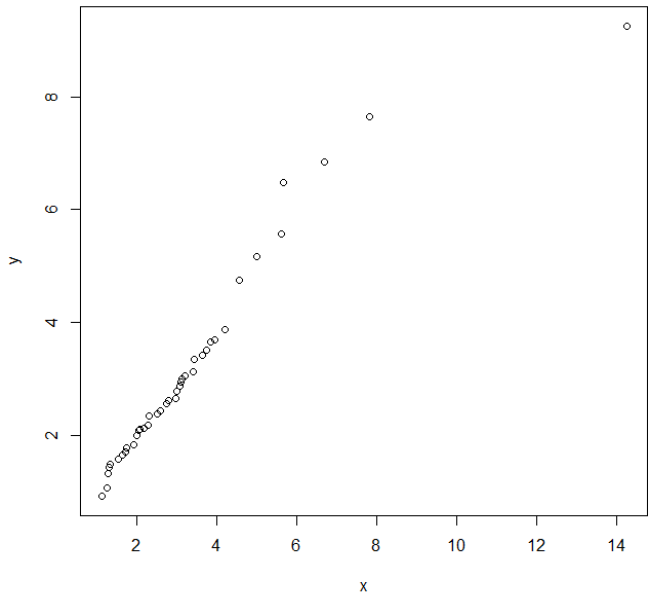


Figure 21: Quantile-quantile graph to compare data to theoretical lognormal distribution.

Table 3: Statistical test results for the lognormal distribution hypothesis.

Test	***p-value***	***Lognormal distribution hypothesis***
Kolmogorov-Smirnov	0.49	Not rejected
Anderson-Darling	0.77	Not rejected
Cramer Von Mises	0.86	Not rejected

In Figure 21, the quantile-quantile graph shows a good fitting of the data to the lognormal distribution with the estimated parameters, except for the extreme quantiles. To complete the analysis, Table 3 presents the results of the statistical tests and none of these results suggest that the lognormality hypothesis has to be rejected. It can thus be concluded that the data represent a random variable which follows a lognormal distribution with the parameters $(\mu, \sigma) = (1.04, 0.54)$.

For the cases where it is not possible to fit a probabilistic distribution onto the data, it should be useful to model the data dispersion represented on the histogram. To do so, non-parametric methods like the kernel method ([58]) can be used to estimate the density of the random variable behind the data:

$$\hat{f}_h(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right)$$

with N the sample size, h the window width (fitting parameter) and K the kernel selected (Gaussian, uniform…: e.g. $\mathrm{K(u)} = \frac{1}{\sqrt{2\pi h}} e^{-\frac{1}{2}u^2}$). This method requires a large number of data. Figure 22 shows example results for this method.

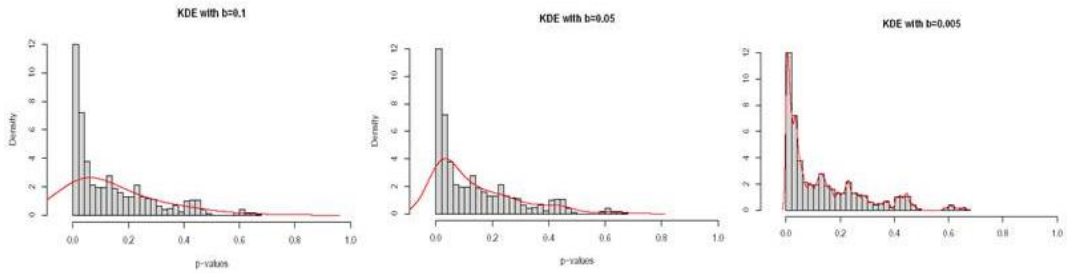


Figure 22: Kernel smoothed results for density function estimation [57].

4.2 Representativeness

Sometimes because of access constraints, measurement costs are such that it is unconceivable to achieve a lot of measurements, but it still remains important to assess the sample representativeness before any statistical analysis. The sample representativeness can be studied through the evolution of bootstrap statistical indicators like the mean or the standard deviation with replicate size varying from a minimum to the size of the reference sample ([34], [59]).

Reference statistical books usually give techniques to study the statistical properties of a sample but these are often valid only for Gaussian distribution data samples/sets or for large-size samples (typically greater than 30) because of the central limit theorem. In case of non-Gaussian small- or medium-size samples, the representativeness study is essential to assess whether or not the sample size is enough to estimate valid statistical indicators. In this context, bootstrap techniques (see section 5.4) are of interest to provide estimations and robust confidence intervals (key information for small-size samples) of statistic indicators. The next point of interest is about confidence interval stability: would the addition of few individuals contribute to improve significantly the confidence interval accuracy?

In [34], this problem is dealt with through convergence graph visualization. In order to determine the sample size impact, a smaller random subsample is generated to calculate the related statistical moments and their bootstrap confidence intervals. This process is repeated with greater subsample sizes, until the reference sample size is reached.

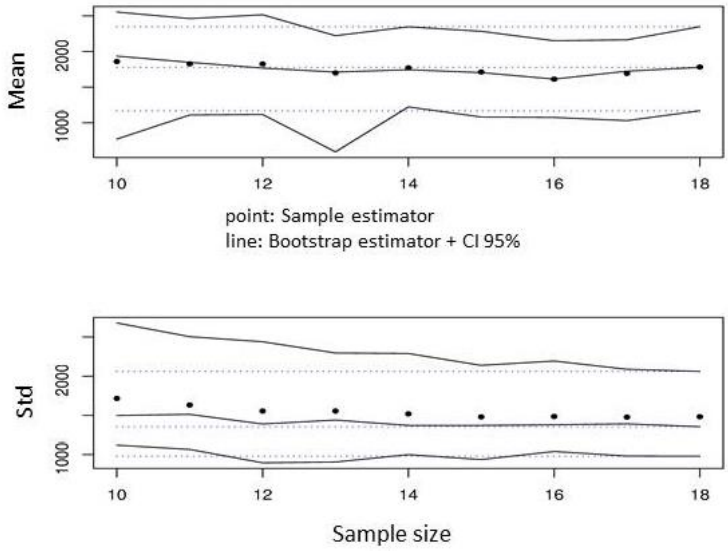


Figure 23 : Bootstrap convergence graph for ^{137}Cs activity measurement data.

In the present example, 18 measurement data of ^{137}Cs concentration in mud are collected on a dismantling site. In the Figure 23, the bootstrap estimations of the mean and its confidence interval are nearly stable. However, the bootstrap estimations of the standard deviation and its confidence interval slowly decrease. We can conclude that some additional data should ensure the representativeness of this sample.

4.3 Linear regression

When one or several parameters are linearly related to measured data, it can be interesting to build a predictive model with a simple linear regression or a multilinear regression. The quality of the regression model may be approached by the determination coefficient R^2 (the square of the sample correlation coefficient between the observed outcomes and the observed predictor values). The more R^2 is close to 1, the better the linear regression is in quality. Nevertheless, R^2 is sensitive to the number of variables and it is sometimes useful to calculate an adjusted R^2 (ratio: $(1-R^2)(n-1)/(n-p-1)$ where *n* is the number of data and *p* is the number of model parameters). Other criteria are used to find the best predictive model related to the maximum likelihood function (see §5.1 in [60]):

- Akaïke: the best model is the model which minimizes the AIC[3] criterion.
- Swartz: this criterion is adapted for large samples. The best model is the model which minimizes the BIC[4] criterion.

The sum-of-squares method is very sensitive to outliers. It is therefore important to analyze these outliers and to find a robust model with another estimator. To do so, it is possible:

- To transform the variable: using log or square to improve the quality of linearity;
- To use kernel estimator (Nadaraya-Watson for example…).

If many explanatory variables exist, it is practical to adopt the General Linear Model (GLM) or multiple regressions. The best models depend on the selection of variables (see adopted techniques, such as stepwise methods).

[3] The **Akaike information criterion** (**AIC**) is an estimator of the relative quality of statistical models for a given set of data. The AIC value of a given model is the following: $AIC = 2k - 2ln(\hat{L})$ where k is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function for the model.

[4] The **Bayesian information criterion** (**BIC**) or **Schwarz criterion** is given by: $BIC = ln(n)k - 2ln(\hat{L})$ where n is the number of data, k is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function for the model.

A generalized linear model may be useful to show different phenomenological behaviors. If it is combined with PCA, clusters can be determined.

A PLS regression (or Partial Least Squares regression) gives very good predictions with small or large estimations.

4.4 MARSSIM (order statistics)

The "Multi-Agency Radiation Survey and Site Investigation Manual" ([6]) provides detailed guidance on how to demonstrate that a site complies with a radiation dose- or risk-based regulation level. The MARSSIM focuses on the demonstration of compliance which is generally done during the final status survey following scoping, characterization and any necessary remediation actions.

The MARSSIM approach uses two statistical tests of hypothesis rejection ("Sign test" and "Wilcoxon Rank Sum test") to validate the conformity of a remediation with respect to a reference value (DCGL in the American terminology). The difference between these two tests is the presence or absence of radionuclides of interest in the environment. They are nonparametric tests, which makes it possible to treat variables whose statistical distribution law is not explicit (the real values rarely follow a Gaussian distribution). They are also based on order statistics, which allows the integration of values at the limit of detection if they do not represent most of the data.

The underlying assumptions are a spatial independence of the values and a not too dissymmetrical statistical distribution (histogram). These tests are particularly interesting when the average value is below the threshold of interest, with a few point values above.

The null hypothesis is made to assume that the median value of the activity levels is greater than the reference value. The test aims to reject this hypothesis for a given confidence level and thus concludes that the median is not statistically different from the reference threshold or below that threshold. If the data are not too dissymmetrical, it is possible to draw the same conclusion with the mean value.

The statistical quantities of the tests (rank sum of numbers of positive differences) are then compared to critical values to validate or reject the null hypothesis:

$$k = \frac{m(m+n+1)}{2} + q_{1-\alpha}\sqrt{\frac{mn(m+n+1)}{12}} \quad \text{or} \quad k = \frac{N}{2} + \frac{q_{1-\alpha}}{2}\sqrt{N}$$

As for sampling scheme, different quantities or parameters are used to determine the minimum number of samples necessary for carrying out both tests:

- The decision errors α and β, to deal with false negative and positive risks,
- The width of the uncertainty zone (especially the lower bound ("LBGR" in the guide) since the upper bound corresponds to the reference threshold),
- The a priori standard deviation.

$$N = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{(3(P_r - 0,5)^2)} \quad \text{or} \quad N = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{4(Signp - 0,5)^2}$$

The relative amplitude is the ratio between the width of the uncertainty zone and the a priori standard deviation. It should ideally be comprised between 1 and 3. In the case of a standard deviation that is very small compared to the reference threshold, the relative amplitude is taken equal to 3 and the number of points to be collected is 20. By default, a 30% coefficient of variation is considered.

5 Statistical approach for small data sets

Classic statistical methods may not be suitable to determine risk prediction bounds related to the level of contamination because they are based on strong assumptions (e.g. that the underlying distribution is Gaussian) which cannot be checked. A radiological characterization based on a small data set is a complex task which belongs to a quite general class of problems: the statistical analysis

of small-size data samples ([61], [62]). For example, if the estimation of the mean value is of interest, according to the central limit theorem, normal-distribution based bound can only be used in the asymptotic limit of a very large sample. Moreover, in that case, the convergence can be very slow in the presence of a noticeably skewed actual data-generating distribution. Therefore the Gaussian distribution hypothesis can be either non valid, or non-justifiable.

As mentioned in section 2.3.2, classic inferential statistics provide results for which the accuracy is strongly dependent upon sample data size. The analysis of small data sets (typically < 30) can lead to estimations with high variability and high sensitivity to outliers.

To face this class of problems, several alternative statistical tools exist. The most interesting are presented thereafter, in terms of weak hypothesis and ease of implementation.

5.1 Bayesian inference

The inferential frequentist statistical approach is aimed at detailing a θ parameter from a sample of observed data [63]. As in section 4.1, the likelihood $L(x_1, \cdots, x_n, \theta) = \prod_{i=1}^{n} f_X(x_i, \theta)$ enables to quantify the value of the parameter which maximises the "probability" to observe the given data sample. Therefore $L(x_1, \cdots, x_n, \theta)$ summarises all the information provided by the data on the θ parameter. However, there is a problem when the sample size is small because the modelling of $f_X(x, \theta)$ may not be consistent. The Bayesian approach consists in using both information: data through the likelihood function $L(x_1, \cdots, x_n, \theta)$ and complementary information (expert judgement for example), namely the prior density of θ. The uncertainty on θ is expressed by the probability $\pi(\theta)$ which represents an a priori opinion on different plausible values for the θ parameter (expert judgements, experimental information…). The unknown parameter is then considered as a random variable like the observed data. The distribution law of the data is interpreted as f_θ, the conditional law of the data given θ:

$$f \quad (x|\theta) = f_\theta(x).$$

The Bayesian inference consists in using the prior law on θ, $\pi(\theta)$ and the observed data with the law $f \quad (x|\theta)$ to revise the law of θ in $\pi(\theta|x)$ in accordance with Bayes theorem ([64]):

$$\pi(\theta) f(x|\theta) = \pi(\theta)\mathcal{L}(x_1, \cdots, x_n, \theta) \propto \pi\big(\theta|(x_1, \cdots, x_n)\big).$$

This prior density $\pi(\theta)$ summarizes both the a priori information on the θ parameter and the uncertainty on this information. This law is often built using the mean as an estimation of the parameter and the variance as the uncertainty on the estimation of the mean. For classic distributions laws, Table 4 gives the corresponding posterior distribution law deduced from Bayes theorem.

Table 4 : Posterior distribution laws for classic prior distribution laws.

Conditional law/likelihood $f(x\|\theta)$	Prior distribution law $\pi(\theta)$	Posterior distribution law $\pi(\theta\|x)$
Normal $\mathcal{N}(\theta,\sigma^2)$	Normal $\mathcal{N}(\mu,\tau^2)$	Normal $\mathcal{N}\left(\frac{\sigma^2\mu+\tau^2x}{\sigma^2+\tau^2},\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$
Binomial $\mathcal{B}(n,\theta)$	Beta $\mathcal{B}e(\alpha,\beta)$	Beta $\mathcal{B}e(\alpha+x,\beta+n-x)$
Poisson $\mathcal{P}(\theta)$	Gamma $\mathcal{G}(\alpha,\beta)$	Gamma $\mathcal{G}(\alpha+x,\beta+1)$
Normal $\mathcal{N}\left(\mu,\frac{1}{\theta}\right)$	Gamma $\mathcal{G}(\alpha,\beta)$	Gamma $\mathcal{G}\left(\alpha+0.5,\beta+\frac{(\mu-x)^2}{2}\right)$
Gamma $\mathcal{G}(\nu,\theta)$	Gamma $\mathcal{G}(\alpha,\beta)$	Gamma $\mathcal{G}(\alpha+\nu,\beta+x)$
Negative Binomial $\mathcal{N}eg(m,\theta)$	Beta $\mathcal{B}e(\alpha,\beta)$	Beta $\mathcal{B}e(\alpha+m,\beta+x)$

To illustrate this method, we consider the problem consisting in estimating the proportion p of non-compliance when an exhaustive survey cannot be performed. Only a set of n measurements are available and X represents the number of non-compliances in this set. The question is: for a given X, what can we say about p?

Without any hypothesis on p, we can use the uniform law to infer a prior law for p:

$$p\sim\mathcal{U}([0,1])\sim\mathcal{B}e(1,1)$$
$$\pi(p)=\mathbb{I}_{[0,1]}(p).$$

For the observation X, we can use the binomial law, $X\sim B(n,p)$:

$$P(X=x|p)=\binom{n}{x}p^x(1-p)^{n-x}.$$

We can deduce the posterior law for p, $p|x\sim Be(x+1,n\text{-}x+1)$, from Table 4:

$$\pi(p|X=x)\propto P(X=x|p)\pi(p)=p^x(1-p)^{n-x}\mathbb{I}_{[0,1]}(p).$$

If the mean of p is $\frac{1}{2}$ (the prior law of p is $U([0,1])$), the observed number of non-compliances is $x=240$ and the size of the survey set is $n=500$, then the mean of p for the given x is:

$$\mathbb{E}(p|x)=\frac{x+1}{n+2}=\frac{241}{502}=0.48.$$

The information can be updated by any new information available, which means the calculations can be iterative. For each step, the posterior law of p calculated at the previous step becomes its prior law and new observations are used to infer the new posterior law of p. For further information, refer to [65] and [66].

5.2 Wilks method

The Wilks method was first introduced to the nuclear engineering community by the German nuclear safety institute (GRS) at the beginning of the 1990s ([67]). Then it has been used in various safety

assessment problems. This method ([68], [69]) based on order statistics allows the user to determine the required sample size precisely in order to estimate, for a random variable, a quantile of order α with a confidence level β. The great interest of this method is that it is robust and it requires no hypothesis.

The explanation below is restricted to the one-sided case. Let's suppose an i.i.d. (independent identically distributed) n-sample $X_1, X_2, \cdots, X_n$ drawn from a random variable X. M is written as follows: $M = \max_i(X_i)$. For M to be an upper bound for at least $\gamma.100\%$ of possible values of X with a given confidence level β, the following equation needs to apply:

$$P[P(X \leq M) \geq \gamma] \geq \beta \qquad \textbf{\textit{(Eq. 5.2.1).}}$$

According to the Wilks formula, the sample size n must therefore satisfy the following inequality:

$$1 - \gamma^n \geq \beta \qquad \textbf{\textit{(Eq. 5.2.2)}}$$

Table 5 : Examples of values given in the first-order case by Wilks formula.

γ	0.9	0.9	0.9	0.95	0.95	0.95	0.95	0.99	0.99
β	0.5	0.9	0.95	0.4	0.5	0.78	0.95	0.95	0.99
n	7	22	29	10	14	30	59	299	459

The Table 5 shows several consistent combinations of the sample size n, the quantile order γ and the confidence level β. For example, to have an estimation of the median (0.5-quantile) with a level of confidence of 90%, the Wilks formula requires a 5-size sample and the corresponding Wilks first-order 0.5-quantile should be the maximum of the sample.

The equation 5.2.1 is a first order equation because the upper bound is set equal to the maximum value of the sample. To extend Wilks formula to higher orders, the n-sample of the random variable X is sorted into the increasing order: $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(r)} \leq \cdots \leq X_{(n)}$. For all $1 \leq r \leq n$, we set:

$$G(\gamma) = \mathbb{P}\left[\mathbb{P}\left(X \leq X_{(r)}\right) \geq \gamma\right].$$

According to the Wilks formula, the previous equation can be redefined as follows:

$$G(\gamma) = \sum_{i=0}^{r-1} C_n^1 \gamma^2 (1-\gamma)^{n-i}.$$

The value $X_{(r)}$ is an upper bound of the γ-quantile with a confidence level β if $1 - G(\gamma) \geq \beta$

Increasing the order of Wilks formula helps reduce the variance in the quantile estimator, but it requires a larger n (according to the formula: $1 - G(\gamma) \geq \beta$). This Wilks formula can be used in two different ways: when the goal is to determine the sample size required to estimate a γ-quantile with a given confidence level β or when a sample size is already available, to determine the couple (γ, β) and the order for the estimation of the Wilks quantile.

To illustrate what kind of statistical information can be inferred from given data using this method, this latter is applied to a use case aiming at characterizing the ^{137}Cs radiological activity of a large-size population of waste objects belonging to a suitable waste category (e.g. low-activity or high-activity waste). The reliability of this classification is all the more important since it directly affects the total waste management costs.

Only 21 measurements have been made. In terms of statistics, this small-size sample has been randomly chosen. The summary statistical parameters which are estimated from these data are: $\text{mean} = 31.45\text{Bq/cm}^2$, $\text{median} = 15.4\text{Bq/cm}^2$, $\text{standarddeviation} = 36.11\text{Bq/cm}^2$, $\text{Min} = 0.83\text{Bq/cm}^2$, and $\text{Max} = 156.67\text{Bq/cm}^2$.

For this sample of size $n = 21$, two types of quantile can be estimated:

- An unilateral first-order γ-quantile with a level of confidence β. Then $\alpha = 1\text{-}\gamma$ and β have to be deduced from the Wilks formula. It results in the following solutions:

1) $P[P(X \leq 156.67) \geq 0.896] \geq 0.9; (\alpha, \beta) = (10.4\%, 90\%)$

which means that 89.6% of the population do not exceed the threshold $S = 156.67 Bq/cm^2$ (the maximum of the sample) with the level of confidence 90%.

2) $P[P(X \leq 156.67) \geq 0.867] \geq 0.95; (\alpha, \beta) = (13.3\%, 95\%)$

which means that 86.7% of the population do not exceed the threshold $S = 156.67 Bq/cm^2$ (the maximum of the sample) with the level of confidence 95%.

- An unilateral second order γ-quantile with a level of confidence β.,Then $\alpha = 1\text{-}\gamma$ and β have to be deduced from the Wilks formula. It results in the following solutions:

1) $P[P(X \leq 79.67) \geq 0.827] \geq 0.9; (\alpha, \beta) = (17.3\%, 90\%)$

which means that 82.7% of the population do not exceed the threshold $S = 79.67 Bq/cm^2$ (the second maximum of the sample) with the level of confidence 90%.

2) $P[P(X \leq 79.67) \geq 0.793] \geq 0.95; (\alpha, \beta) = (20.7\%, 95\%)$

which means that 79.3% of the population do not exceed the threshold $S = 79.67 Bq/cm^2$ (the second maximum of the sample) with the level of confidence 95%**Erreur ! Cela devrait être un chiffre.**.

5.3 Robust probabilistic risk bounds

In this section, the aim is to estimate prediction intervals which can be directly interpreted as probabilistic risk bounds ([69], [70]). Once a unilateral prediction interval has been determined, it is possible to define the limit value that a variable cannot exceed (or reach, depending on the context) with a given probability level. In a practical radiological context, it can then be used to estimate the quantity of contaminants which does not exceed a safety threshold value, based on a few contaminant measures. Mathematically, a unilateral prediction interval for a random variable $X \in R$ is expressed as follows:

$$\mathbb{P}(X \geq S) \leq \alpha$$

where $S \in R$ is the threshold value and $\alpha \in [0,1]$ is the risk probability.

In probability theory, the concentration inequalities relate the tail probabilities of a random variable to its statistical moments[5]. Therefore, they provide bounds for the deviation of a random variable about a given value (for example its mean value). The various inequalities come from the information given about the random variable (mean, variance, bounds, positiveness…). There are many results described in [71] but we focus here on three classic inequalities that seem to be the most useful for the radiological characterization problems related to small samples.

The general formulation of these inequalities is as follows:

$$P(X \geq \mu + t) \leq \left(1 + \frac{t^2}{k\sigma^2}\right)^{-1} \qquad \textbf{\textit{(Eq. 5.3.1)}}$$

where μ is the mean, σ is the standard deviation, $t \geq 0$ and k is a positive constant.

1. Bienaymé-Chebytchev inequality

The Bienaymé-Chebytchev (BC) inequality is expressed in the following terms:

$$\forall t \geq 0, \mathbb{P}(X \geq \mu + t) \leq \left(1 + \frac{t^2}{\sigma^2}\right)^{-1} \qquad \textbf{\textit{(Eq. 5.3.2)}}$$

which corresponds to (Eq. 5.3.1) with $k = 1$. As μ and σ are unknown in practical applications, they are replaced by their empirical counterparts. **This inequality does not require any hypothesis on the probability distribution of X.**

[5] The first moment of a random variable X is the mean $\mu = E(X)$. The second moment is the variance $\sigma^2 = E((X-\mu)^2)$ The third moment is the skewness coefficient $\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$.; The fourth moment is the kurtosis $\gamma_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$.

2. Camp-Meidell inequality

The Camp-Meidell (CM) inequality is written as follows:

$$\forall t \geq 0, \mathbb{P}(X \geq \mu + t) \leq \left(1 + \frac{9t^2}{4\sigma^2}\right)^{-1} \quad \textbf{\textit{(Eq. 5.3.3)}}$$

which corresponds to (Eq. 5.3.1) with $k = 4/9$. As μ and σ are unknown in practical applications, they are replaced by their empirical counterparts.

It is interesting to note that in its two-sided version, this inequality justifies the so-called "three-sigma rule". This rule is generally used in manufacturing processes as it states that 95% of a scalar-valued product result X are located within the interval $[\mu\text{-}3\sigma, \mu + 3\sigma]$. In fact, the document [72] shows that this rule and **this inequality can be applied to all the unimodal continuous probability laws** used in practice (e.g. uniform, Gaussian, triangular, lognormal or Weibull laws).

3. Van Danzig inequality

The Van Danzig (VD) inequality is expressed in the following terms:

$$\forall t \geq 0, \mathbb{P}(X \geq \mu + t) \leq \left(1 + \frac{8t^2}{3\sigma^2}\right)^{-1} \quad \textbf{\textit{(Eq. 5.3.4)}}$$

which corresponds to (Eq. 5.3.1) with $k = 3/8$. As μ and σ are unknown in practical applications, they are replaced by their empirical counterparts. **This inequality can be applied to the convex part of all the unimodal continuous probability laws.** Even though the tail of most classic distribution laws is convex (e.g. exponential, triangular, Gaussian, Weibull), it is not valid for uniform variables.

These three inequalities are applied to the example studied in section 5.2, which consists of 21 measurements of ^{137}Cs radiological activity with a threshold value $S = 100\mathrm{Bq/cm^2}$. Given that the distribution law of the data sample is continuous, unimodal and convex, the following results can be inferred:

- Using **BC inequality**, we obtain:

$$\mathbb{P}(X \geq \mu + t) \leq \left(1 + \frac{t^2}{\sigma^2}\right)^{-1}$$

with $= S\text{-}\mu$:

$$\mathbb{P}(X \geq S) \leq \left(1 + \frac{(S-\mu)^2}{\sigma^2}\right)^{-1} = 0.2172.$$

Then we coarsely estimate that less than 21.7% of the population has an activity higher than $\mathrm{dBq/cm^2}$ unit.

- Using **CM inequality**, we obtain:

$$\mathbb{P}(X \geq \mu + t) \leq \left(1 + \frac{9t^2}{4\sigma^2}\right)^{-1}$$

with $= S\text{-}\mu$:

$$\mathbb{P}(X \geq S) \leq \left(1 + \frac{(S-\mu)^2}{\frac{4}{9}\sigma^2}\right)^{-1} = 0.1098.$$

Then we coarsely estimate that less than 11% of the population has an activity higher than $\mathrm{dBq/cm^2}$ unit.

- Using **VD inequality**, we obtain:

$$\mathbb{P}(X \geq \mu + t) \leq \left(1 + \frac{8t^2}{3\sigma^2}\right)^{-1}$$

with $= S\text{-}\mu$:

$$\mathbb{P}(X \geq S) \leq \left(1 + \frac{(S-\mu)^2}{\frac{3}{8}\sigma^2}\right)^{-1} = 0.0942.$$

Then we coarsely estimate that less than 9.4% of the population has an activity larger than $\mathrm{dBq/cm^2}$ unit.

This simple application illustrates the gain that can result from the use of CM or VD inequalities instead of BC inequalities.

5.4 Bootstrap methods

The bootstrap method is a variation on the traditional Monte Carlo method. Bootstrap (introduced and further developed by B. Efron, [59]) refers to a class of methods that resamples from the original data set, i.e. creating many samples from a single sample. Bootstrap methods are robust nonparametric statistical methods that can be used to build approximate confidence limits for the population mean, to estimate the bias and variance of an estimator or to calibrate hypothesis tests. These methods only assume that the sample data are representative of the underlying population. Bootstrap requires no assumption regarding the statistical distribution (e.g., normal, lognormal, gamma) of the underlying population and it can be applied to a variety of situations ([73]).

Bootstrap is formally defined as follows: the sample data points $\mathrm{X} = \{\mathrm{X_1}, \cdots, \mathrm{X_n}\}$ are assumed to be a representative random sample of size n from some unknown probability distribution F. The parameter of interest, θ, is a characteristic of the distribution $\mathrm{F}, \theta = \mathrm{f(F)}$, such as the mean, variance, shape or scale parameter, or any quantile of the distribution F. An estimate of θ is the statistic $\hat{\theta}$, which is determined from the data set $\hat{\theta} = \mathrm{f(X)}$. Using the data set, **X**, the distribution $\hat{\mathrm{F}}$ is defined as an estimate of the unknown distribution *F*. The distribution $\hat{\mathrm{F}}$ may be defined as either an empirical distribution law or a parametric distribution (i.e. parametric bootstrap). The bootstrap method addresses the uncertainty due to sampling error by assuming first that the original data set **X** is a randomly drawn, representative sample from the distribution $\hat{\mathrm{F}}$ and then by generating bootstrap samples repeatedly. A bootstrap sample, $\mathrm{X^*}$, is defined as a numerically simulated random sample of size n taken from the distribution $\hat{\mathrm{F}}$. A random simulation method, such as Monte Carlo method, must be used for that purpose. A large number, *B*, of independent bootstrap samples $(\mathrm{X^{*1}}, \mathrm{X^{*2}}, \ldots, \mathrm{X^{*B}})$ are selected from the distribution $\hat{\mathrm{F}}$. From each of the *B* bootstrap samples, a new statistical parameter, θ^*, is computed so that: $\hat{\theta}^{*\mathrm{i}} = \mathrm{f}(\mathrm{X}^{*\mathrm{i}})$ for $\mathrm{i} = 1,2, \ldots, \mathrm{B}$, where $\hat{\theta}^{*\mathrm{i}}$ is a bootstrap estimate of $\hat{\theta}$ for the replication i.

Several bootstrap method implementation schemes are used to characterize a site radiologically, depending on the specific initial data and on the required tasks: regular bootstrap, double bootstrap, parametric bootstrap, Bayesian bootstrap, percentile method, Monte Carlo Bootstrap, among others (see e.g. [74] and [75]).

The advantages of bootstrap methods can be summarized as follows ([59]):

- They save the costs of taking new samples (estimation of a sampling distribution when a sample with limited number of measuring points is available);
- They can be used when parametric assumptions cannot be made or are very complex;
- They enable the generation of informative resampling distributions for graph plotting;
- They allow the verification of parametric assumptions;
- They allow the estimation of variance in quantiles;
- They provide an adequate, if not better, estimation of the standard error of nonparametric model predictions, in comparison with alternative asymptotic techniques.

One of the main shortcomings of resampling a data set is that the minimum and maximum values obtained are limited by the minimum and maximum values within the data set. When only small data sets are available, this can lead to biased representations of a given model input (e.g. failure to consider potential large values that are not present in the limited data set).

Parametric distributions can be used to allow for the possibility that smaller or larger values than those contained in the data set may occur in the real system being modeled. In [76], the results of the parametric bootstrap show that the uncertainty in the tails of variability distributions can be large (especially in case of small data sets), which appropriately reflects the limitations of extrapolation from a small data set to the tails of a parametric distribution.

If the number of bootstrap samples is big, a computer intensive analysis is required. According to [77], the analyst must continue increasing the number of bootstrap samples until a convergence with the data population is reached. In the abovementioned document, this type of analysis is recommended for an initial value of 100 samples to a maximum of 100,000 samples.

Applications of bootstrap methods include: error rate estimation in discriminant analysis, subset selection in regression, logistic regression and classification problems, cluster analysis, kriging, nonlinear regression, time series analysis, complex surveys, p-value adjustment in multiple testing problems, and survival and reliability analysis.

An illustration of bootstrap applications is shown in [78]. In order to reconstruct the average dose magnitudes of a contaminated field using a small data set, the bootstrap method is used to develop a model of the dose distribution field. The next figure shows the model testing results obtained from experimental data related to the Chernobyl area.

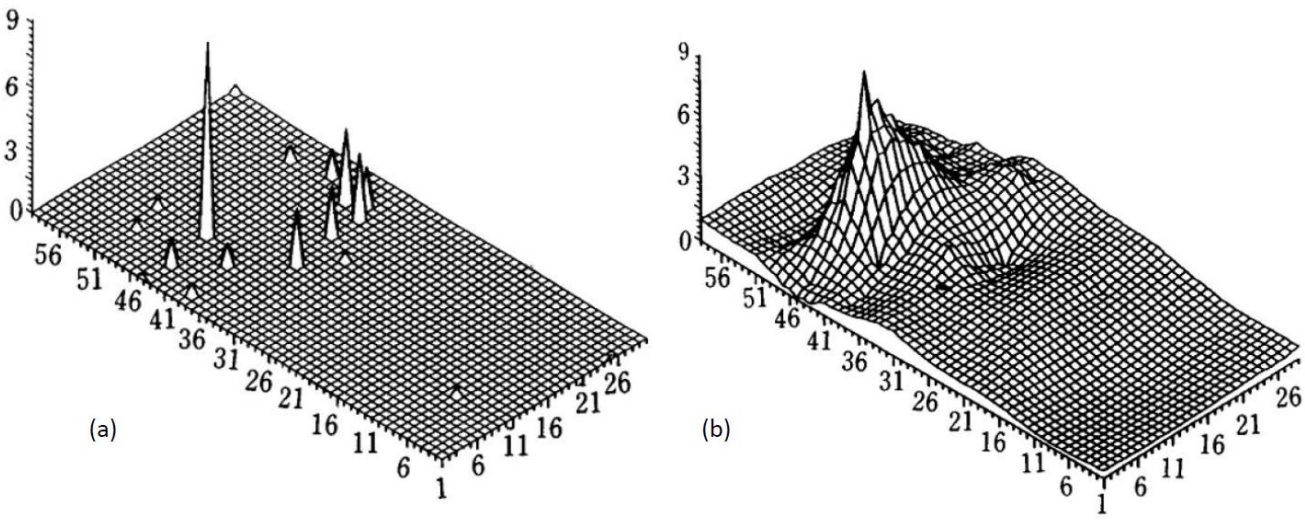


Figure 24: Dose rate distribution of a contaminated area near Chernobyl: (a) measured points; (b) reconstructed dose distribution after smoothing.

Even with a small number of measurement points, the method allows to draw a clear picture of the contaminated field. If the experimental results for the different measurement points badly correlate with each other, this method does not give accurate results. However, even in this case, the distribution obtained by the bootstrap method will provide useful information about the potential dose levels. The following recommendations are finally formulated: to ensure the reliability of the results, the number of measuring points should be about 20 or more; the monitoring or measuring points should be uniformly distributed over the field of interest. If this is not possible, measurements should be made in or near uninhabited locations. It is also emphasized that the main advantage of Bootstrap applications in the context of sampling is a possibility to reconstruct the dose distribution using restricted data. The results of the study [79] show a tendency of the bootstrap distributions obtained towards normality. Therefore, the method can be used in some applications for statistical analyses in the context of radioecology. This method can also be applied to the reconstruction of the spatial and angular distributions of any other physical field (for example, to the analysis of particle beam profiles).

6 Conclusion

The main aim of WP3 within the INSIDER project is to draft a sampling guide for initial nuclear site characterization in constraint environments before decommissioning. The process followed to reach this goal consisted of four stages: status, development, implementation and guidance. This document corresponds to the first stage of the process. It provides an overview on sampling design methods and state-of-the-art statistical techniques for preliminary analyses and data processing.

Many guides and references in the specific field of decommissioning focus on the back end of decommissioning (e.g. release of regulatory control) whereas the INSIDER project is more looking at the front end (pre-decommissioning characterization), with the aim of applying a waste-led approach. Moreover, many of the generic sampling design techniques and state-of-the-art statistical techniques used in preliminary analyses and data processing are often considered as stand-alone methods. In many cases, an integrated and overall approach of pre-decommissioning characterization which notably consists in evaluating historical data, making on-site measurement campaigns, sampling and analysing, developing scaling factors and applying numerical codes, is missing.

This document is not aimed at providing a comprehensive overview of all the existing statistical methods or guidelines to the end user. The sampling design methods and state-of-the-art statistical techniques described in the present document are preselected methods that will be used as main inputs for the next step of the process, strategy development. In this next step, WP3 participants will develop a decision tree to guide the user throughout the radiological process and the choice regarding sampling scheme as well as the selection of the statistical tools for the preliminary analyses and data processing.

Bibliography

[1] Radiological characterisation from a waste and materials end-state perspective: Practices and Experience. Task Group on Radiological Characterisation and Decommissioning (TGRCD)), NEA No. 7373. Nuclear Energy Agency/Organisation for Economic Co-operation and Development, 2017.

[2] Classification of radioactive waste. General Safety Guide. IAEA Safety Standards Series No. GSG-1. (2009). International Atomic Energy Agency. http://www-pub.iaea.org/MTCD/publications/PDF/Pub1419_web.pdf.

[3] AIEA (1998) Radiological Characterization of Shut Down Nuclear Reactors for Decommissioning Purposes (IAEA-TRS-389), 1998.

[4] INSIDER Project, Work Improved Nuclear SIte characterization for waste minimization in DD operations under constrained EnviRonment Research and Innovation action NFRP-2016-2017-1, Criteria for characterization, radionuclides & materials-cartography Deliverable D2.2.

[5] Radiological Characterisation for Decommissioning of Nuclear Installations. Final Report of the Task Group on Radiological Characterisation and Decommissioning (RCD) of the Working Party on Decommissioning and Dismantling (WPDD). NEA/RWM/WPDD(2013)2. Nuclear Energy Agency/Organisation for Economic Co-operation and Development. https://www.oecd-nea.org/rwm/docs/2013/rwm-wpdd2013-2.pdf.

[6] Multi-Agency Radiation Survey and Site Investigation Manual. NUREG-1575, Rev.1. EPA-402-R-97-016, Rev. 1. DOE/EH-0624, Rev. 1 (2001). https://www.epa.gov/radiation/multi-agency-radiation-survey-and-site-investigation-manual-marssim.

[7] Guidance for Data Quality Assessment – Practical Methods for Data Analysis. EPA QA/G-9. EPA/600/R-96/084. (July 2000). U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-06/documents/g9-final.pdf.

[8] Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan. EPA QA/G-5S. EPA/240/R-02/005. (December 2002). U.S. Environmental Protection Agency. https://www.epa.gov/quality/guidance-quality-assurance-project-plans-epa-qag-5.

[9] Data Quality Assessment: Statistical Methods for Practitioners. EPA QA/G-9S. EPA/240/B-06/003. (February 2006). U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-08/documents/g9s-final.pdf.

[10] Data Quality Assessment: A Reviewer's Guide. EPA QA/G-9R. EPA/240/B-06/002. (February 2006). U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-08/documents/g9r-final.pdf.

[11] ProUCL Version 5.0.00 - Technical Guide. Statistical Software for Environmental Applications for Data Sets with and without Non-detect Observations. EPA/600/R-07/041. September 2013. U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-03/documents/proucl_v5.0_tech.pdf.

[12] Characterisation principals for soils, buildings and infrastructures contaminated by radionuclides for remediation purposes, ISO/DIS 18557:2017.

[13] Deutsches Institut für Normung e.V., DIN 25457-4, Activity measurement methods for the clearance of radioactive substances and nuclear facility components - Part 4: Contaminated and activated metal scrap, 2013.

[14] Deutsches Institut für Normung e.V., DIN 25457-6, Activity measurement methods for the clearance of radioactive substances and nuclear facility components - Part 6: Rubble and buildings, 2017.

[15] Deutsches Institut für Normung e.V., DIN 25457-7, Activity measurement methods for the clearance of radioactive substances and nuclear facility components - Part 7: Ground surfaces and excavated soil, 2017.

[16] Guide de l'ASN n°14 : Assainissement des structures dans les installations nucléaires de base, Autorité de Sûreté Nucléaire, 2016.

[17] Council Directive 96/29/Euratom of 13 May 1996 laying down basic safety standards for the protection of the health of workers and the general public against the dangers arising from ionizing radiation; Official Journal of the European Communities, ISSN 0378-6978, L 159, Vol. 39, 29.06.96.

[18] Planning, managing and organizing the decommissioning of nuclear facilities: lessons learned. (May 2004). IAEA-TECDOC-1394. ISBN 92–0–104404–6. International Atomic Energy Agency.

[19] **https://www.iaea.org/ns/tutorials/regcontrol/intro/glossary**

[20] Determination and Use of Scaling Factors for Waste Characterization in Nuclear Power Plants. IAEA Nuclear Energy Series No. NW-T-1.18. 2009. http://www-pub.iaea.org/MTCD/Publications/PDF/Pub1363_web.pdf.

[21] IAEA (2006). Management of problematic waste and material generated during the decommissioning of nuclear facilities. International Atomic Energy Agency, Technical Report Series No. 441.

[22] S. M. Jerome et al. (2018). Evaluation and requirements for scale factors from EMRP project MetroDECOM (ENV 54), Journal of Radioanalytical and Nuclear Chemistry (2018) 318:381–385, https://doi.org/10.1007/s10967-018-6050-7

[23] IAEA (2009). Determination and use of scaling factors for waste characterization in nuclear power plants. International Atomic Energy Agency, Nuclear Energy Series No. NW-T-1.18.

[24] Theoretical activation calculation method to evaluate the radioactivity of activated waste generated at nuclear reactors. ISO/DIS 16966.

[25] Le Loirec C., (2017) CEA internal communication.

[26] Sanchez R. et al., APOLLO2 year 2010, Nuclear Engineering and Technology 42: 474-499 (2010).

[27] Brun E. et al. (2015), TRIPOLI-4® CEA, EDF and AREVA reference Monte Carlo code, Annals of Nuclear Energy 82, 151-160.

[28] "MCNP6.1 home Page". LANL. 5 August 2013. Archived from the original on March 30, 2012. Retrieved 7 August 2013.

[29] GEANT and Application software group of CERN (1993), GEANT, Application software group of CERN, Detector description and simulation Tool, CERN; Program library long write-up, W5013, Genova.

[30] Tsilanizara A. and al. (2000), DARWIN: an evolution code system for a large range of applications, Journal of Nuclear Science and Technology 37:845-849.

[31] "MCNP6.1 home Page". LANL. 5 August 2013. Archived from the original on March 30, 2012. Retrieved 7 August 2013.

[32] GEANT and Application software group of CERN (1993), GEANT, Application software group of CERN, Detector description and simulation Tool, CERN; Program library long write-up, W5013, Genova.

[33] Guidance on the selection and usage of acceptance sampling systems for inspection of discrete items in lots, ISO/TR 8550-1:2007-06.

[34] N. Pérot, B. Iooss (2008), Quelques problématiques d'échantillonnage statistique pour le démantèlement d'installations nucléaires, Conférence λμ16, Avignon, France.

[35] Characterisation principals for soils, buildings and infrastructures contaminated by radionuclides for remediation purposes, ISO/DIS 18557:2017.

[36] Tillé, Y. (2001), Théorie des sondages, Echantillonnage et estimation en populations finies, Dunod.

[37] Hansen, Morris H., William N. Hurwitz, and William G. Madow (1953), "Sample Survey Methods and Theory.»

[38] B. Zaffora and M. Magistris and G. Saporta and F. La Torre (2016), Statistical sampling applied to the radiological characterization of historical waste, Journal *EPJ Nuclear Sci. Technol.* , vol. 2, pp. 11. https://www.epj-n.org/articles/epjn/pdf/2016/01/epjn150085.pdf.

[39] Le Cocguen A., Dauvois V., Duhart-Barone A., Pérot N. (2014), Validation du caractère enveloppe du MOP par mesure de production de dihydrogène de radiolyse de fûts de déchets primaires alpha Pu PEGASE, internal CEA report CEA/DEN/CAD/DSN/SEEC/LECD DO 67 23/07/2014.

[40] Francis F. Pitard, Pitard F. Pitard. 1993, "Pierre Gy's Sampling Theory and Sampling Practice, Second Edition: Heterogeneity, Sampling Correctness, and Statistical Process Control". 2nd edition, 488 pages (Hardback).

[41] Bastiaan Geelhoed, 2000. "Approaches in Material Sampling" – Page 107-111.

[42] Bernard Poncet, Laurent Petit. 2013, "Method to assess the radionuclide inventory of irradiated graphite waste from gas-cooled reactors. Journal of Radioanalytical and Nuclear Chemistry", November 2013, Volume 298, Issue 2, pp 941–953.

[43] Guidance for Using Geostatistics in Developing a Site Final Status Survey Program for Plant Decommissioning, EPRI report 3002007554 (2016) (www.epri.com/#/pages/product/3002007554/).

[44] *S. Tufféry (2007) Data Mining et Statistique décisionnelle, Editions Technip.*

[45] Scaling factor method to determine the radioactivity of low- and intermediate-level radioactive waste packages generated at nuclear power plants, ISO 21238:2007

[46] Chilès JP., P. Delfiner (1999) *Geostatistics – Modeling Spatial Uncertainty*, Wiley series in Probability and Statistics, New-York.

[47] Desnoyers Y. (2015) Sampling considerations for characterization of radioactive contamination using geostatistics. In Proc. of World Conference on Sampling and Blending, Bordeaux, France.

[48] Journel, A.G. and Huijbregts, C.J. (1978) Mining Geostatistics. Academic Press, London, 600 p.

[49] Kendall, M. (1938). "A New Measure of Rank Correlation". *Biometrika*. **30** (1–2): 81–89. doi:10.1093/biomet/30.1-2.81.

[50] *Khattree et al, (2000), Multivariate data reduction and discrimination, 558 page.*

[51] *Borra, S., Rocci, R., Vichi, M., Schader, M. (2001), "Advances in Classification and Data Analysis".*

[52] *Ravindra, Khattree et al (2003), Applied Multivariate Statistics, 363 pages.*

[53] Desnoyers Y., Dubot D. (2014) Characterization of radioactive contamination using geostatistics. Nuclear Engineering International. Volume 59, No. 716, March 2014, Pages 16-18. Sidcup, United Kingdom.

[54] Guidance for Using Geostatistics in Developing a Site Final Status Survey Program for Plant Decommissioning, EPRI report 3002007554 (2016) (www.epri.com/#/pages/product/3002007554/).

[55] J. C. Myers, 1997. Geostatistical error management: Quantifying uncertainty for environmental sampling and mapping. Van Nostrand Reinold (New York).

[56] Desnoyers Y., Chilès J.-P., Dubot D., Jeannée N., and Idasiak J.-M. (2010), Geostatistics for radiological evaluation: study of structuring of extreme values. In Stochastic Environmental Research and Risk Assessment: Volume 25, Issue 8 (2011), Page 1031-1037 and in Proc. of GeoENV 2010 congress, Gent, Belgium.

[57] Bechler A., Romary T., Jeannée N., Desnoyers Y. (2013) Geostatistical sampling optimization of contaminated facilities. In Stochastic Environmental Research and Risk Assessment: Volume 27, Issue 8 (2013), Pages 1967-1974.

[58] Parzen E. (1962). On estimation of a probability density function and mode, Ann. Math. Stat. **33**, pp. 1065-1076.

[59] B. Efron and R.J. Tibshirani (1993), An introduction to the bootstrap, Chapman & Hall.

[60] *Borra, S., Rocci, R., Vichi, M., Schader, M. (2001), "Advances in Classification and Data Analysis".*

[61] Brazzale A.R., Davison A.C., and Reid N. (2007), Applied asymptotics-Case studies in small sample statistics, Cambridge University Press.

[62] Pupion P.C., Pupion G. (2010), Méthodes statistiques applicables aux petits échantillons, Hermann.

[63] Bousquet N. (2010), Eliciting vague but proper maximal entropy priors in Bayesian experiments, Statistical Papers, Springer, 51:613-628.

[64] Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (2014), Bayesian Data Analysis, Chapman and Hall Texts in Statistical Science Series.

[65] Andrew Gelman, John B. Carlin, Hal S. Stern and Donald B. Rubin (2013), Bayesian Data Analysis, Chapman CRC, 2013, 3e éd., 675 p.

[66] Christian Robert (2006), *Le Choix bayésien*, Springer Verlag, coll. « Statistiques et probabilités appliquées », 658 p.

[67] E. Hofer (1993), Probabilitische unsicherheitsanalyse von ergenissen umfangenreicher rechenmodelle, GRS-A-2002.

[68] Wilks S.S. (1941), Determination of sample sizes from setting tolerance limits, Anals of Mathematical Statistics.

[69] Blatman G., Delage T., Iooss B., Pérot N (2017), Probabilistic risk bounds for the characterisation of radiological contamination, EPJ Nuclear Sci. Technol. 3, 23. https://www.epj-n.org/articles/epjn/pdf/2017/01/epjn160031.pdf.

[70] G. Blatman and B. Iooss (2012), Confidence bounds on risk assessments - application to radiological contamination. In Proceedings of the PSAM11 ESREL 2012 Conference, pages 1223–1232, Helsinki, Finland.

[71] I.R. Savage (1961), Probability inequalities of the Tchebycheff type. Journal of Research of the National Bureau of Standards-B. Mathematics and Mathematical Physics, 65B:211–226.

[72] D.F. Vysochanskii and Y.I. Petunin (1980). Justification of the 3σ rule for unimodal distribution. Theory of Probability and Mathematical Statistics, 21:25–36.

[73] E. J. Dudewicz, (1992) “The Generalized Bootstrap,” In: K.-H. Jöckel, G. Rothe and W. Sendler, Eds., Bootstrapping and Related Techniques, Springer-Verlag, Berlin, , pp. 31-37.

[74] HARDING, R. (1998) “Environmental Decision Making the roles of scientists, engineers and the public”, The Federation Press, Sydney.

[75] A. C. Davison and D. V. Hinkley (1997), “Bootstrap Methods and Their Application”, Cambridge University Press, Cambridge.

[76] H. C. Frey and D.S. Rhodes (1999) “Quantitative Analysis of Variability and Uncertainty in Environmental Data and Models. Volume 1. Theory and Methodology Based Upon

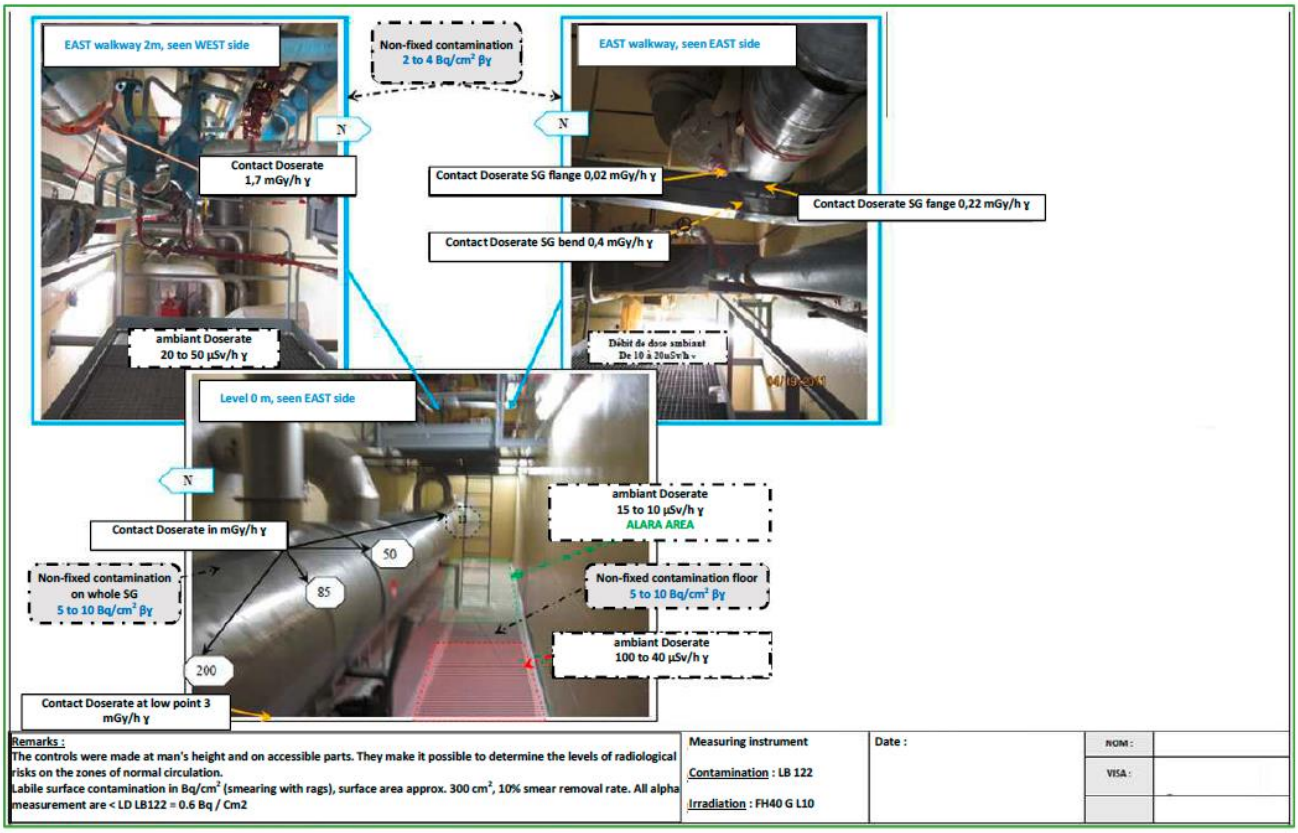
Bootstrap Simulation”. Report No.: DOE/ER/30250--Vol. 1

[77] Michael R. Chernick (2007). Bootstrap Methods: A Guide for Practitioners and Researchers, 2nd Edition.

[78] S.Mikheenko, S.Erofeeva, T. Kosako (1994). Reconstruction of Dose Distribution by the Bootstrap Method Using Limited Measured Data. Radioisotopes, 43, 595-604.

[79] A.N. Casado da Silva, Romilton dos Santos Amaral, Jose Araujo dos Santos Junior, Jose Wilson Vieira, Romulo Simoes Cezar Menezes (2015). Statistical analysis of discrepant radioecological data using Monte Carlo Bootstrap Method. Article in Journal of Radioanalytical and Nuclear Chemistry.

Annex A: Typical cartography of a nuclear building in France



This project has received funding from the Euratom research and training programme 2014-2018 under the grant agreement n°755554

Annex B: Properties of usable data from a typical cartography of a nuclear building in France

Available data	Physical quantities	Uncertainties
Equivalent dose rate	Equivalent dose rate H*(10), expressed in sievert (Sv) by hour (h), is a practical value which is used by radiation protection services to quantity the radiation hazards for human body. It is usually referred to as "equivalent dose rate" in radiation protection documents, – without any further precision-. It is a very complex notion which includes a lot of factors dependent upon energy of radiation, type of radiation, and impact on human body. The current value used considers the impact over a 10-mm distance inside the whole human body "H*(10)" and it can be approximated as follows, for basic considerations: H*(10) = $\sum_R w_R * D_{TR}$, expressed in Sv.h^{-1}, where • w_R = radiation impact on the body as given in tables • D_{TR} = mean absorbed dose rate. The dose rate is commonly measured by radiometers. Physically, but still basically, the measurement depends on the type and energy of the radiometers as follows: • $D_{TR} = \int \frac{n}{\varepsilon_i}$, expressed in Gy.h^{-1} • n = total counting rate measured in seconds in powerless conditions for a given energy, • ε_i = performance of the instrument for a given energy. Count rates are provided with typical uncertainty of $\pm$ 10 %.	Dose rate data include an internal variance due to the statistical variability associated with the term of the equation, as well as an external variance. They are particularly sensitive to the measured energy.
Surface contamination	The surface (alpha or beta) activity, of fixed and non-fixed contamination on the controlled surface, is measured by a detection calibrated system called "contamination controller", and it is given according to the measured counting rate as follows: $As = \frac{n-n_B}{\varepsilon_i * W * \varepsilon_s}$, $(Bq.cm^{-2})$, where • n = total count rate measured (counts per second) • n_B = background count rate (counts per second) • ε_i = efficiency of the instrument (depending on the radionuclide measured) • W = surface window of the measuring instrument (cm²) • ε_s = yield of the source of contamination. Count rates are provided with typical uncertainties of $\pm$ 10 %.	Contamination data usually include internal variances due to the statistical variabilities associated with the terms of the equations as well as external variances. They are local values specific to the area under analysis.
Non fixed surface contamination	It refers to the contamination collected by taking a smear sample, where measurements are then made using a detection calibrated system called "contamination controller". The calibration depends on the smear surface area and on the radiological spectrum as follows: $As = \frac{n-n_B}{\varepsilon_i * S * F * \varepsilon_s}$, $(Bq.cm^{-2})$, where • n = total counting rate measured (counts per second) • n_B = background count rate (counts per second) • ε_i = performance of the instrument • F = sampling factor • S = rubbed surface area, in square centimeters • ε_s = yield of the source of contamination. Count rates are provided with typical uncertainties of $\pm$ 10 %.	Contamination data include internal variances due to the statistical variability associated with the term of the equation, as well as an external variance. They are local values specific to the area under analysis.