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

Research and Innovation action
NFRP-2016-2017-1

Statistical approach guide

Deliverable D3.7

Version n° 1

Authors: Rogiers Bart (SCK CEN), Yvon Desnoyers (Geovariances), Nadia Pérot (CEA), Gunhild von Oertzen (Brenk Systemplanung), Oleksandr Sevbo (Energorisk), Séverine Demeyer (LNE), Sven Boden (SCK CEN).

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 D&D operations under constrained EnviRonment
Project Acronym:	INSIDER
Project Start Date:	01 June 2017
Related work package:	WP 3: Sampling strategy
Related task(s):	Task 3.4: Statistical approach guide
Lead Organisation:	Brenk Systemplanung GmbH
Submission date:	2021-10-31
Dissemination Level:	Public

History

Date	Submitted by	Reviewed by	Version (Notes)
2021-09-28	Sven Boden	End User Group • José Luis Leganés Nieto (ENRESA) • Mathieu Le Coz (IRSN)	Draft 1
2021-10-18	Sven Boden	Danielle Roudil	1

Table of contents

List of Figures

List of abbreviations

ALARA	As Low As Reasonably Achievable
D&D	Decommissioning and Dismantling
EURATOM	European Atomic Energy Community
INSIDER	Improved Nuclear SIte characterization for waste minimization in Decommissioning under constrained EnviRonment
IQR	Interquartile range
WP	Work Package

Summary

The INSIDER project (2017-2021, (LGI, 2017)) developed and validated a new and improved integrated characterisation methodology and strategy during nuclear decommissioning and dismantling operations (D&D) of nuclear power plants, post-accidental land remediation or nuclear facilities under constrained environments. One of the important outcomes of this strategy development is a statistical approach guideline that is transformed into a web tool, which serves as a user-friendly interactive interface.

The web tool, named STRATEGIST (Sampling Toolbox for Radiological Assessment To Enable Geostatistical and statistical Implementation with a Smart Tactic), intends to guide the expert in handling the problem definition and applying a strategy based on proper data analysis and sampling design.

This document is a blueprint for the STRATEGIST web tool: https://strategist.sckcen.be/.

1 Introduction

STRATEGIST

Sampling Toolbox for Radiological Assessment To Enable Geostatistical and statistical Implementation with a Smart Tactic

- **Guides the expert** in handling the problem definition and applying a strategy based on proper data analysis and sampling design.
- Promotes the application of **an integrated characterisation methodology and strategy** during nuclear decommissioning and dismantling operations (D&D) of nuclear power plants, post-accidental land remediation or nuclear facilities under constrained environments.
- Relies on **state-of-the art statistical techniques** for preliminary analysis and data processing.

The STRATEGIST web tool provides concise descriptions, lists relevant theoretical references, case studies and software implementations for **helping the end user to get started,** using the following diagrams (Figure 1):

- The overall strategy or generic workflow (Section 2), outlining the general steps between a request for initial characterization and reporting of the final results, after all objectives have been reached.
- The data analysis and sampling design workflow (Section 3).
- Venn diagram of the main methods for data analysis (Section 4)
- Venn diagram of the main methods for sampling design (Section 5).

How should this strategy be applied?

- For the application of this strategy, we recommend all involved parties to **familiariz**e themselves with it, or **at least the general workflow** (Section 2). This will ease the discussion on the different objectives and constraints, and create more realistic expectations on the work that has to be done.
- People involved in the data analysis, and especially the selection of the appropriate methods, should at least have some **general notions on all different types of methods** discussed here, to enable proper judgement of the different options, and selection of the most appropriate one.
- Please take into account the following important remarks:
    - This strategy is NOT intended to provide the nonspecialist with a comprehensive mode of operation for the complete process of initial nuclear state characterisation in view of decommissioning.
    - This is **only a guideline**, and should not be blindly followed. Special circumstances often ask for special solutions, which cannot all be covered by a generic strategy.
    - This strategy can be used to inform people with no or very little experience in statistics about the complexity of the issue, and provide them some relevant background, but it cannot justify not involving people experienced in the matter.

For the vocabulary used in thei STRATEGIST tool we refer to the VIM (BIPM, IEC, IFCC, ILAC, ISO, IUPAC , IUPAP et OIML, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), 3rd edition éd., Joint Committee for Guides in Metrology, JCGM 200:2012, 2012).

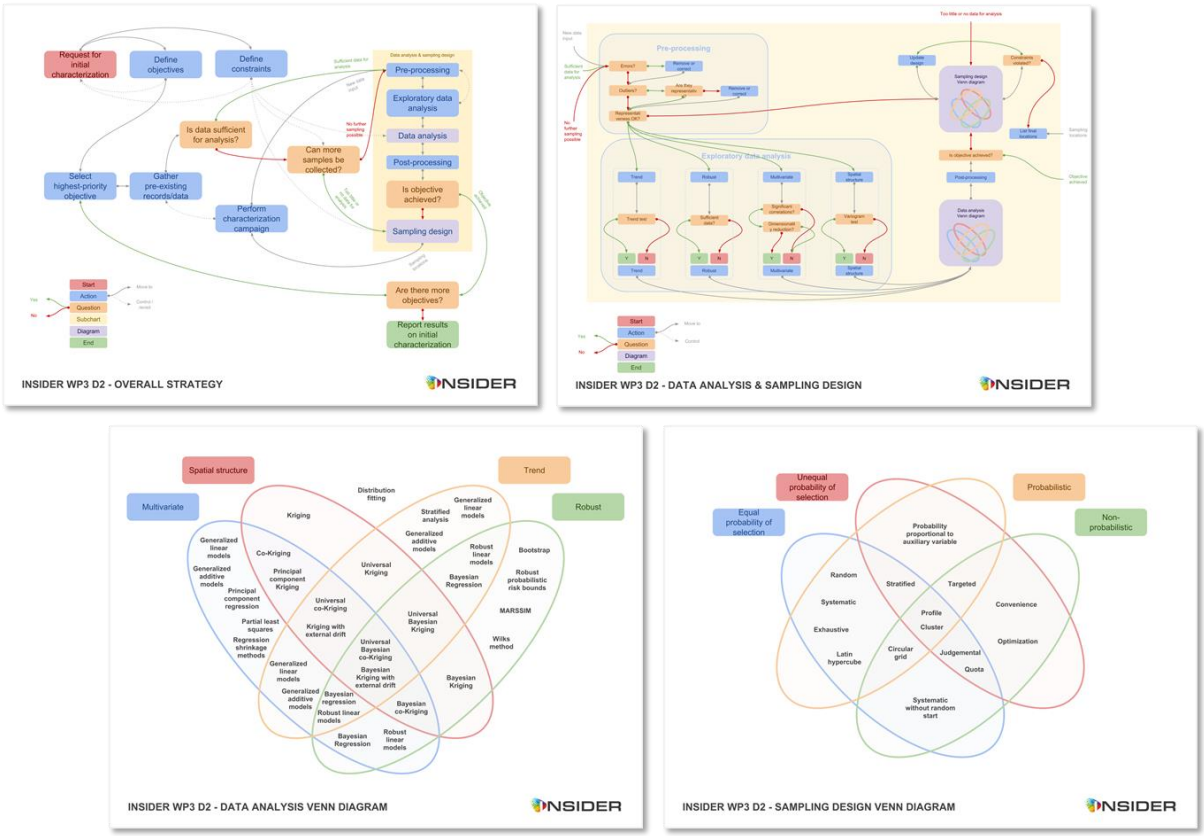


Figure 1. Overview of the diagrams illustrating the different parts of the developed strategy for data analysis and sampling design in the framework of initial nuclear site characterization in view of decommissioning.

In section 6 you can find details on the **implementation of the strategy in four different use cases** as well as a summary of the **lessons learnt**.

2 Overall strategy

While the data analysis and sampling design methods that can be applied depend strongly on the situation and specific goals of initial nuclear site characterization, the overall strategy often takes the form of the generic workflow illustrated in Figure 2.

The starting point we consider here is the **request for initial nuclear site characterization** to a radiological characterization team. Such a request can come from different kinds of actors, and can come with different amounts of detail. Following this request, a clear list of all **objectives** and identification of the **constraints** is absolutely required, and might require some iterations with the applicant to agree on the goals and priorities:

- Very clear and quantifiable objectives allow the development of an effective sampling plan, including the selection of appropriate measurement techniques and the up-front definition of criteria for the measurements (e.g. detection limits, uncertainties).
- On the contrary, the absence of explicit objectives complicates all aspects of designing and implementing the characterization strategy and the planning down to the selection of appropriate measurement techniques, determination of minimal detection limits, etc.

Decommissioning is a multi-disciplinary operation and the involvement of specialized staff performing the next stages of the decommissioning project can be highly beneficial. Technical feasibilities/constraints in the next decommissioning stages might strongly influence the initial characterization program. Effective communication and a common basis of understanding are essential. Extensive compartmentalizing might result in misinterpretation, non-optimal solutions and wrong decisions.

The **highest-priority objective** should be tackled first in most cases, and the cycle along the different objectives is started.

If feasible, data sets containing large amounts of data below **detection limit** should be avoided for a proper statistical analysis and whenever possible tackled during the strategy development. However, this is not always possible; for example due to low threshold values, the limitations of measurement techniques, but also due to unclear initial objectives or potentially changing objectives or thresholds during the characterisation process. Depending on the case, it might be very wise to use advanced statistical methods for dealing with samples below detection limit ((Kim, Hornibrook, & Yim, 2020), (France Patent No. 20 10176, 2020)).

All **prior information** that is available and relevant for the investigated case should be gathered as a first step (historical records, mappings, incidents, etc.). If some radiological data would already be available, a first analysis to **check if the objective is achieved** is probably very useful, even if the results come with lots of uncertainty. In D&D, such prior information is nearly always available. We are working on historical installations and/or sites that have been shut down, or are going to be. Therefore, there is always a history of the exploitation phase, with available data, so this initial data-gathering step is of vital importance.

The data analysis succeeding the data collection consists, in general, of the following steps: **Pre-processing**, **exploratory data analysis**, the actual **data analysis**, and potentially a **post-processing** step. If the objective is not achieved, a **sampling design** should be proposed using the most appropriate method(s) given all prior information and the data analysis result. Following the design, the corresponding **characterization campaign** should be performed. Additional characterization can reveal unexpected issues, and often revisiting the gathering of prior information is then useful. After the additional characterization, the updated dataset is again analysed, and this iterative procedure is continued until the objective is finally reached. The entire process can then be repeated to tackle the **remaining objectives**. Once all objectives have been achieved, the initial characterization study should be **reported** in a transparent way, making clear what has been measured, which results were obtained from the data analysis, and how large the corresponding uncertainty is. The different steps are more extensively discussed one by one below.

Performing a radiological characterization program in two or more **stages/phases** can be efficient and effective to tackle areas with higher uncertainties. Unfortunately, this might not always possible due to planning constraints.

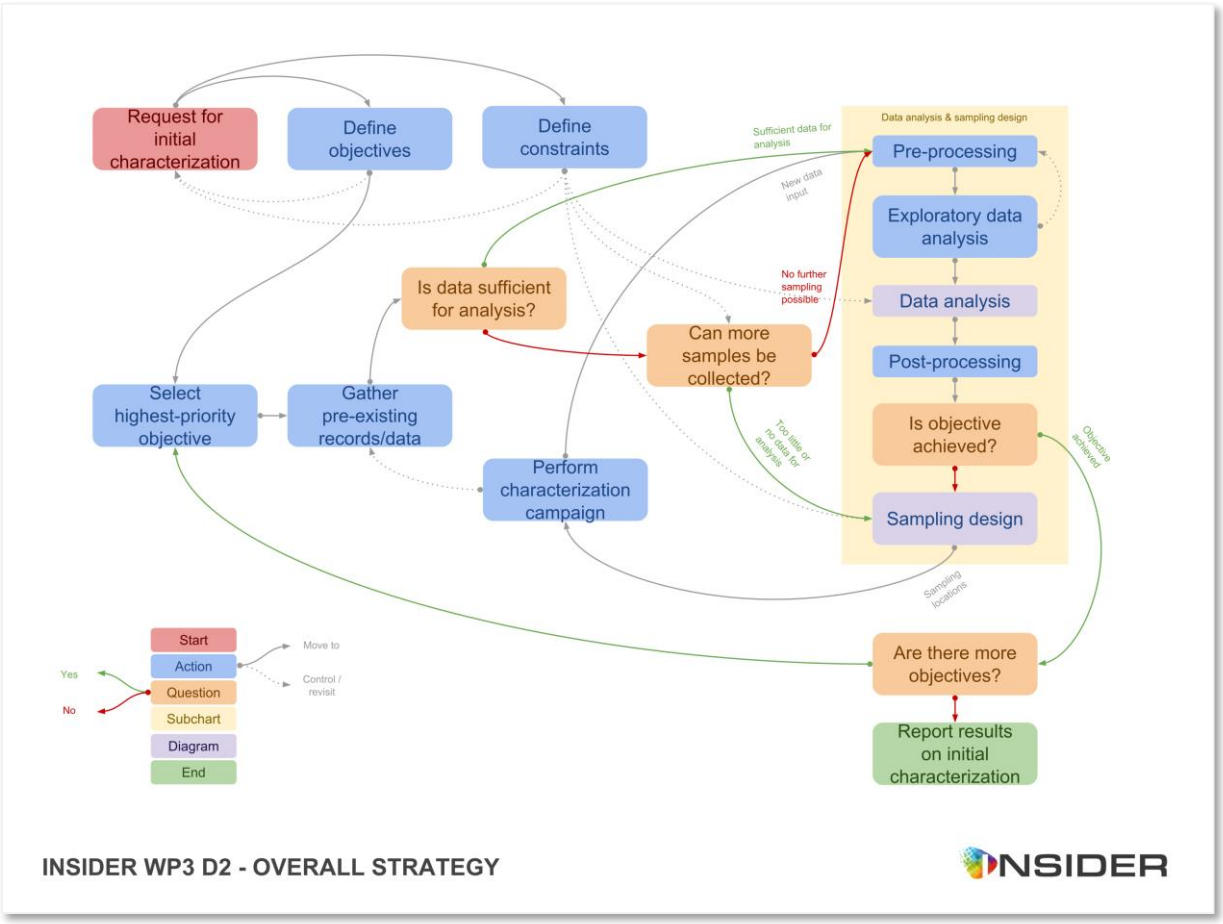


Figure 2. Overall strategy flow chart. See Annex A for a larger version.

2.1 Request for initial characterization

The **starting point** we consider here is the request for initial nuclear site characterization to a radiological characterization team. Such a request can come from different kinds of actors, and can come in different amounts of detail. In many cases however, it probably comes with a description of the **general purpose that the initial characterization should serve**, but this does not justify getting started right away with collecting data, performing data analysis and sampling design. Instead, the request and the general purpose should be **discussed thoroughly**, so **all involved parties** are on the same page, and the request and general purpose can be translated to specific objectives and constraints in a next phase.

Once the **specific objectives and constraints** have been defined, **feedback** from the applicant is required, potentially within an iterative process. The request can then be reformulated, or more resources can be made available, if the objectives would not meet the **expectations** of the applicant, or the constraints would interfere with the **feasibility** of the required work.

2.2 Define objectives

As introduced above, the first important task should be the **translation of the request for initial characterization** into a list of **specific objectives** that have to be fulfilled. Objectives should be

well defined in order to **enable well-informed decision making**, once they are reached. The lack of clear objectives can result in useless data, over-characterization and a waste of budget, time and efforts. In principle, every objective can be expressed by defining **two aspects**:

1. Which **variable(s) of interest** is (are) to be investigated?
2. Which **statistical indicator(s)** should be quantified?

In practice however, these aspects can take very different forms. We provide a non-exhaustive list below, to provide an idea on the range of possibilities:

Variables of interest

- Radiological status qualification
  - Location of fixed and non-fixed contamination / activation
  - Dose rates / equivalent doses
- Material inventory quantification
  - Subject under study
    - Batch of discrete objects
    - 2D surface
    - 3D volume
  - Classification (waste (type), (un)conditional release)
  - Mass-, volume- or area-specific activities
  - Isotopic composition (which radionuclides, non-radioactive elements, scaling factors)
  - Assessments through DA and / or NDA
- Physico-chemical conditions
  - Hazardous substances

Statistical indicators

- Measures of central tendency
  - Mean
  - Median
  - Mode
- Measures of dispersion
  - Standard deviation / variance
  - Extremes (minimum, maximum)
  - Quantiles (IQR, percentiles)
- Measures of association
  - Covariance / correlation
- Measures of shape
  - Skewness
  - Kurtosis
- Measures of risk
  - Probability of exceeding a threshold

If multiple objectives are defined, a **prioritization** should be agreed on as well, as in most cases not all the work can be executed in parallel. This is especially true for constrained environments. On the other hand, reaching certain major objectives might already provide the necessary information for accomplishing other objectives as well.

2.3 Define constraints

Next to the definition of objectives, a list of constraints that have to be honoured has to be created as well. Some constraints can be part of the initial request. These are often related to the aspects managed by the applicant, like the required **timing** and available **budget**. Other constraints are dictated by the **legislation** that is applicable, and/or the control bodies that oversee the executed work. Furthermore, the constrained environment itself might pose additional **operational constraints** as well. We provide a non-exhaustive list below, to provide an idea on the range of possibilities:

- Budget
- Timing
- Operational constraints
  - Non radiological & radiological hazards & safety aspects
  - Radiation levels and removable contamination (ALARA)
  - Worker protection
  - Available equipment
  - Accessibility

    - Cross contamination during sampling
    - Measurement / lab procedures
        - Minimum number of replicates
        - Required volume of material
    - QA, QC, QM
- Legislation
    - Decontamination requirements
    - Decommissioning requirements (temporary storage, treatment)
    - Sampling requirements
    - Tolerable risk
    - Tolerable uncertainty
    - Recycling options

2.4 Select highest-priority objective

As it is often difficult in constrained environments to perform different tasks in parallel, the highest-priority objective should be addressed first. The priorities should be defined together with the objectives, in order to let the iterative process run smoothly. It is very likely that by addressing major objectives, the necessary data will also be gathered for fulfilling other, minor objectives. In that sense, it is relevant to stress here that **working on a single objective** does not at all mean that the other objectives can be temporarily disregarded. Instead, **potential synergies between different objectives** should be checked at every step of the process, in order to reduce the amount of data that has to be gathered, and/or the number of iterations required to reach the predefined goals.

2.5 Gather pre-existing records/data

A first step in tackling a specific objective is the **collection of pre-existing records and/or data**. These types of information can be extremely useful for framing the problem at hand, by creating **reasonable expectations** on the site and materials to be investigated (in terms of possible radiological hazards and potential contamination/activation characteristics), and can often provide a **basis for more efficient sampling** than a purely random approach. If sufficient quantitative data is already available, and the quality is adequate, also a **first data analysis** step can be performed, which again allows for a more appropriate sampling design, if additional data would be required. The types of possibly useful information are manifold, however, so we provide only a non-exhaustive list below, to provide an idea on the range of possibilities:

- Site context
    - Current activities
    - Plans (initial but preferably “as-built”)
    - Maps
- Historical archives and/or data bases
    - Operation history
    - Functional analysis
    - Lists of incidents
    - Routine occupational exposure measurements
    - Production records
    - Operational waste characterization
- Collective knowledge of the employees
    - Interviews
    - Testimonies
    - Site visits
- Previous studies
    - Characterization campaigns
    - Calculations/modelling exercises

2.6 Is data sufficient for analysis?

The pre-existing data and/or records might not provide **sufficient amounts of quantitative data**, or the **data quality** might not be adequate, justifying a first data analysis step. They might however provide very relevant **qualitative information that can be leveraged directly for sampling design**. In such cases, the first data analysis step can effectively be skipped, and a first sampling design step and characterization campaign should be performed, after which the normal iterative procedure can be followed.

2.7 Can more samples be collected?

It is possible that the **collection of additional samples is not an option**. There might be various reasons for this (*e.g.* budgetary, safety, *etc.*), but these should always be part of the **identified constraints**. In such cases, we are obliged to work with the samples and/or data that is available, and a data analysis has to be performed anyway. The **uncertainties** related to the likely **very conservative analysis** should then reflect the lack of samples required for a more adequate, properly data-driven analysis.

2.8 Data analysis & sampling design

The data analysis and sampling design steps make up the heart of this strategy. Due to their importance and complexity, they are more thoroughly discussed in Section 3. The basic idea here is that all the **available information** serves as an **input** to a **pre-processing** and **exploratory data analysis** phase, after which the actual **data analysis** is performed. Here, an appropriate statistical method should be used. Afterwards, **post-processing** of the results may be required in order to judge if the **objective is achieved or not**. If it is not, the results can be used to inform the **sampling design** for a new characterization campaign, after which the process can be repeated.

2.9 Perform characterization campaign

When a sampling design is created, the corresponding characterization campaign should of course be performed. The two are clearly split into two separate steps here, as the **sampling design** should basically represent how we want the samples to be taken, **in theory**, according to all the practical limitations and constraints that are known to us. In practice, however, all kinds of **unexpected issues** might be encountered which make that the **characterization campaign** will not be able to honour the sampling design perfectly. The **deviations** should be discussed with all involved parties, if possible, and **properly documented**, in order to make the most out of the data gathered.

2.10 Are there more objectives?

After realizing an objective, the **remaining objectives should be revisited** first, accounting for the newly gathered data (if any). Some other objectives might in fact be fulfilled by the **new data or available knowledge**. The remaining highest-priority objective should then be selected, and the whole procedure starts again.

2.11 Report results on initial characterization

After tackling all objectives, the results should be reported in a way that makes clear **how the problem was approached**, **which assumptions were made** (and hence what the limitations are of the reported study), and **what data was used**. In this way, the applicant can have a clear idea

on the uncertainty sources that are covered by **the obtained results.** This provides a **solid basis for** determining **the next steps** in planning the decontamination or decommissioning process.

3 Data analysis & sampling design

While the overall strategy often takes the form of the generic workflow, as illustrated in Figure 2, the data analysis and sampling design **depend strongly on the situation, the available data and the specific goals and constraints**. We however attempt to bring some structure to the typical workflow here, by dividing the process into a pre-processing step, the exploratory data analysis, the actual data analysis, potentially a post-processing step and the sampling design, as illustrated in Figure 3. Checking if the objective is achieved, and whether the constraints are violated, is of course also necessary at some point.

The **pre-processing** step is still relatively straightforward, and encompasses checking for errors and outliers, possibly making corrections and removing parts of the data irrelevant to the problem at hand. The **exploratory data analysis** can also still be structured in a way that it is applicable to most problems in D&D for constrained environments, by looking into four aspects of the data:

1. Is this a **univariate or multivariate** problem?
2. Is this a problem involving **spatial trends**?
3. Is this a problem involving **spatial structure**?
4. Is this data requiring **robust methods**?

We use the outcome of the exploratory data analysis, *i.e.* the answers to the above questions, to bring some structure in the range of methods for the actual **data analysis**, possibly applicable to the problem at hand. As the range of situations and methods is vast, we cannot discuss every possible road to selecting a specific approach. This only provides some guidance on the type of methods to use, and the expertise of the user of this strategy comes in at this point to make a final decision, based on the individual method descriptions, suggestions and remarks contained in this document.

A **post-processing** step can be required to translate the results obtained into the required information for checking if the **objective** is achieved. If it is not, a **sampling design** approach has to be selected. Similar to the data analysis, we can also only provide general recommendations here, and use a classification in terms of:

1. Is this sampling design **probabilistic** or not?
2. Are the **selection probabilities equal** or not?

We do try to relate the different approaches to the four aspects of the data defined above as well, in the list of approaches in Section 5. When a sampling design is proposed, a final check on violation of the defined **constraints** should be made, before moving to the corresponding characterization campaign. The different steps outlined here are further discussed in detail one by one below.

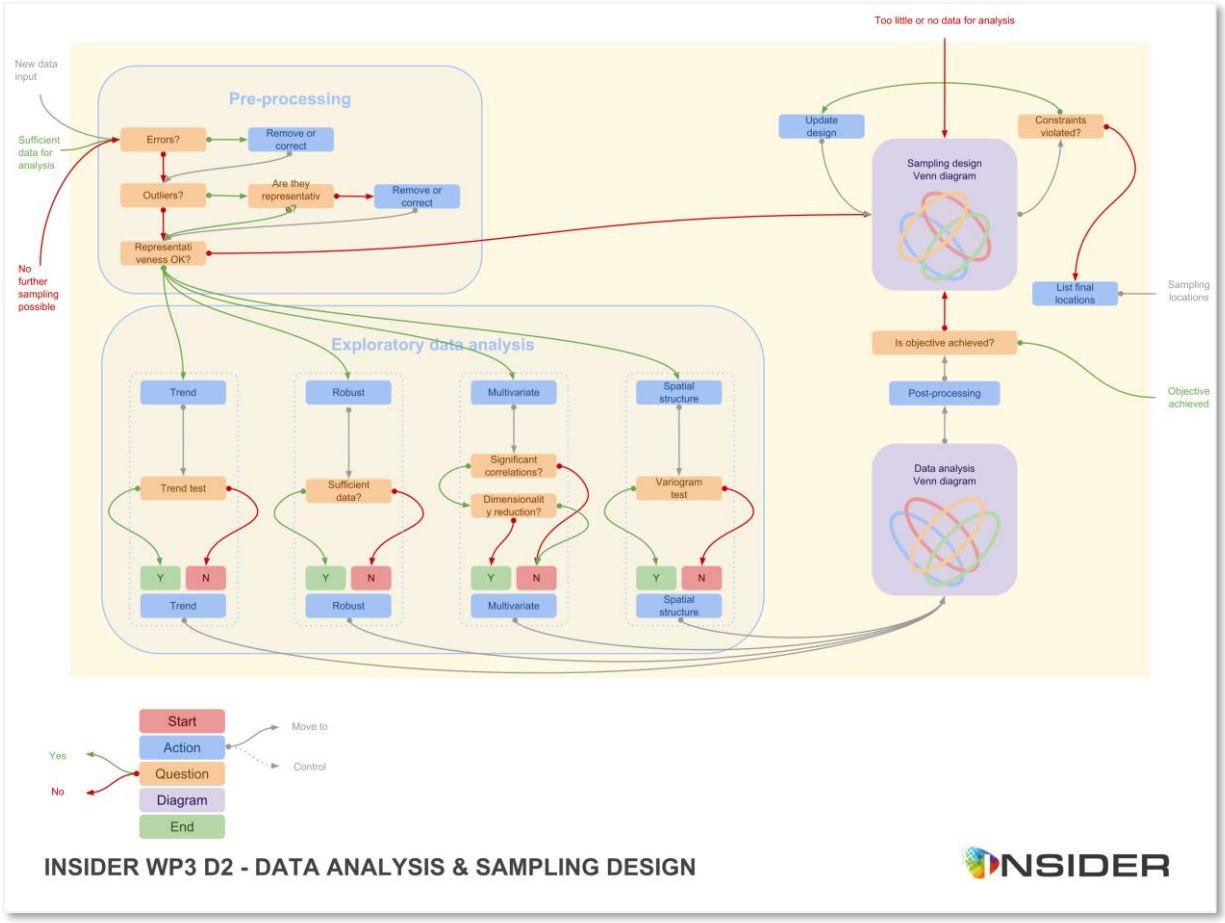


Figure 3. Data analysis & sampling design flow chart. See Annex B for a larger version.

3.1 Pre-processing

3.1.1 Check the data for errors

Gross errors or anomalous measurements of the data set may arise *e.g.* due to:

- Transcription mistakes,
- Data coding errors,
- Undetected equipment failures or malfunction,
- Calibration mistakes,
- Data evaluation errors, and
- Undetected interfering signals.

The selection of a good measurement and data collection strategy should be able to minimize the number of data errors, but cannot fully avoid them. Therefore, a proper **data quality assurance procedure** is absolutely required. Such a procedure can take various forms, depending on, *e.g.*:

- The objective of the measurements,
- The number of measurements within one evaluation,
- The measurement method itself,
- The way of data transfer, and
- The possible automation of data quality checks.

Depending on the objective of the measurements, the data quality assurance procedure should be documented appropriately. Common **quality assurance measures** are:

- Verification of all necessary data transfer and data evaluation software,
- Periodic (yearly or monthly base) detailed technical inspection of measurement devices and data transfer systems,
- High frequency cross-check of measurement device functionality,
- Verification of calibration procedures,
- Periodic (daily base) calibration,
- Random countercheck of data transcription processes (usually with increasing frequency in the case errors are found),
- Certification procedures for external laboratories,
- Blind tests and interlaboratory comparisons, and
- Automated data checks (*e.g.* checking sign, order of magnitude or compare with realistic range of results).

The objective of the data quality assurance is to minimize the occurrence of errors and ensure proper statistical inference and adequate decisions can be made based on the data.

3.1.2 Remove or correct errors

Data errors are, unlike outliers, not acceptable at all in a data set. This means they have to be **removed from the data set**, **flagged as missing values**, or if this would result in a significant loss of information, the erroneous **values can be imputed**. In some cases, however, **correction of the errors** might still be possible, when the uncorrupted raw data would still be available. One example is a wrong calibration of certain equipment. If the raw data is still available, it can be re-processed after the calibration is corrected.

For errors in a multivariate context, imputation techniques can be used to complete the dataset if required. If such corrections are not possible, or if they tend to affect the relation between variables, the original dataset may be used in a restricted manner, with missing values, within an available case analysis approach.

3.1.3 Check the data for outliers

After removing, flagging and/or correcting the errors in the data, the remaining outliers (a.k.a. natural outliers) are observations with a unique combination of characteristics identifiable as **distinctly different** from the other observations. However, the distinction between outliers and extreme values in a probability distribution for a random variable, which occurs quite naturally but not frequently, is not always straightforward. The most straightforward way of detecting outliers is by visualizing your data. If this is not straightforward, dimensionality reduction techniques might be of help (see 3.2.1.2). Many approaches that are more specialized exist however, but discussing these is out of scope here.

3.1.4 Check representativeness of outliers

Outliers cannot be categorically characterized as either beneficial or problematic, but instead must be viewed within the context of the analysis and should be evaluated by the types of information they may provide. When beneficial, outliers – although different from the majority of the sample – may be indicative of characteristics of the population that would not be discovered in the normal course of analysis. In contrast, problematic outliers, not representative of the population, are counter to the objectives of the analysis, and can seriously distort statistical tests.

3.1.5 Remove or correct outliers

If the outliers are natural, not artificial in any kind of way, and if they are representative of the problem at hand, they should remain part of the dataset, in an unaltered way. However, if they would not be representative, and if this would bias the analysis later on, they are better dealt with at this point. Similar to the uncorrectable errors, they can be **removed from the dataset**, **flagged as missing values**, or the values can be **imputed** in some way to make most out of the available data.

Censored data do not necessarily belong to the class of outliers, but it is sometimes useful to treat them in the same way. Different methods exist to work with left- and/or right-censored data, but if the censored data only represents a minor and irrelevant part of the sample, or other methods are to be used, they can also be removed, flagged as missing or imputed. A description of the use of censored data is given for example by (Helsel & Cohn, 1988).

3.1.6 Check representativeness

A final check of the representativeness of all remaining data at this point is in order to identify potential parts of the dataset not relevant to the problem at hand, which may overly complicate the subsequent analysis, or bias the results. Only the **representative part should be retained**.

Furthermore, also the **representativeness of small sample sizes**, for making inference on the entire population, can be checked using more formal methods here (see (Pérot, et al., 2019), but this should not break the flow of the analysis, as issues with small sample sizes should be reflected in the results of the data analysis later on.

3.2 Exploratory data analysis

After the pre-processing phase, the data should be ready for analysis. The exploratory analysis aims at identifying certain aspects of the data in order to determine the class of methods that is fit to perform the actual analysis.

3.2.1 Univariate versus multivariate

Univariate and multivariate data analysis can **differ greatly in terms of complexity**, and simplifying a problem to the univariate case, or a set of independent univariate problems is often beneficial. If this is not possible, however, appropriate methods have to be used that can account for correlations between different variables, or more complex relationships.

3.2.1.1 Check for significant correlations

Pairwise scatterplots provide a straightforward way to **visualize** the data and check if significant correlations are present or not. When these are not conclusive however, *e.g.* in case of low correlations or little data, more formal assessments on the correlations can be made. One approach is the use of **parametric tests** that provide a more objective view on the presence of correlations. These do however rely on different assumptions on the underlying probability distributions of the sample that is investigated. The test statistic for Pearson's product moment correlation coefficient follows for example a t distribution with the number of samples minus two degrees of freedom, if the samples follow independent normal distributions. These kind of tests can be easily performed in many different software packages. If it is difficult to justify the assumptions made however, **non-parametric tests** can be performed as well, for instance based on resampling.

3.2.1.2 Can dimensionality reduction simplify the problem?

In case significant correlations seem to be present, they should be accounted for. One special way of achieving this is by reformulating the original problem with the correlated variables in terms of a

set of uncorrelated latent variables or principal components. Methods that are suited for achieving this are different kinds of factor analysis, principal component analysis and its supervised variant partial least squares. We refer to (Pérot, et al., 2019) for a brief description on these methods.

The dimensionality of the problem is typically reduced in order to keep the analysis manageable, and focuses on the **latent variables or principal components explaining most of the variance** within the data.

3.2.2 Spatial trends

The presence of spatial trends can have a significant impact on the sampling strategy to follow, and therefore we consider it here as a separate aspect of the available data or problem at hand. When a large-scale trend goes together with spatial structure on top of it, methods exist to handle both at the same time. However, in such cases, it is still useful to check explicitly for the presence of a trend, so the appropriate methods can be chosen, and the sampling design is done accounting for the trend (if it would relate to the objective). There are several approaches for trend detection:

- In many cases, the presence of a trend is obvious from the **historical background** and/or the process that has generated the contamination or activation that is being studied. Even without any data, such expected trends are very relevant, and influence the sampling design. Typical examples are trends with:
  - The distance from a source of radiation.
  - The distance from a spill location.
  - The distance along the travel path of the contaminants.

- If data is available, looking at a map visualizing measurement results is often sufficient to confirm the presence of expected trends. If sufficient data points were gathered, even unexpected trends might be discovered by plain **visual inspection**. In addition, visualization of the data along a one-dimensional path (a.k.a. swath plots) through the investigated area or volume might help in recognizing more complex trends. When there is doubt however, some specific quantitative methods can be employed.
- The **experimental variogram** can be used for the detection of spatial trends. (Journel & Huijbrechts, 1978) 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 spatial variance stabilizes at a certain distance and equals the variance of the data. If this is not the case, a trend is probably present.
- Another approach consists in using **Kendall's τ** test (Kendall, 1938) 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-dimensional 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.
- Linear **regression models**, and in certain cases extensions like generalized additive models (for non-linear situations) or robust linear models (in the presence of outliers), can provide indications on the presence of trends as well. This in fact already relates a lot to the data analysis step in case of a trend, but preliminary testing of these models focusing on the model coefficient statistics, can be useful as well.

3.2.3 Spatial structure

A first qualitative understanding of spatial continuity is performed on the **base map**: location of hot spots, size of the impacted areas (activated, contaminated), spatial distribution of high and low values, *etc*. Then, in order to quantify more precisely the spatial structure of the phenomenon **two-**

point statistics are typically used. They allow the characterisation of the spatial continuity, as a gradient for instance, or more adequately for geostatistics, by the **variogram** that describes the evolution of variability between pairs of points as a function of distance. This change of variability with distance can have different forms:

- It can be constant, which is an indication that there is no spatial structure, or at least it is not present at the considered lag distances.
- In case of spatial structure, it typically starts at a zero or a low value for zero distance, and then increases as the distance increases. When no trend is present, it is then normally expected to stabilize approaching a certain distance (the effective range or correlation length), and remain constant after that.
- It might not stabilize, or show a decrease again at some point, which are indications of trends or periodicity.

The existence of spatial structure, or auto-correlation (*i.e.* the variability is lower at shorter distances) makes **geostatistics** an approach relevant to correctly analyse and model the spatial behaviour of the variable of interest.

Large-scale phenomena, or trends, can be modelled and removed (subtracted) from the data before the detailed analysis of spatial structure. That way only local variations are analysed through the variography.

3.2.4 Robust methods

Different aspects of the data can give rise to the need for working with robust methods. One example is the case of small data sets. Small data sets allow little inference about the studied population, and the analyses should therefore come with large uncertainties. Typical statistical methods are however often not designed to work with small data, and the use of special **robust tools** might be required. In case there is considerable prior knowledge on the problem at hand, **Bayesian methods** can be invoked to make sure this is accounted for as well. Furthermore, when the underlying probability distributions of the data are not known and difficult to infer, **non-parametric methods** should be used, and when outliers are present, but affect the analysis of the mean behaviour too much, methods **robust to outliers** are available as well.

Fixing hard rules for determining if such methods are required or not is however not straightforward, and out of scope here. Instead, the user of this strategy should look at all facets of the problem at hand and its context, and his/her expertise should come into play.

An example are the robust inequalities, applicable without knowing the probability distribution of the studied variable, discussed by (Pérot, et al., 2019). In a practical radiological context, these can be used to estimate the quantity of contaminants, which does not exceed a specific threshold value, based on a few contaminant measures.

3.3 Data analysis

For organizing the different data analysis techniques, we make use of the **Venn diagram** presented in Figure 4. The different categories we use are based on the **four aspects of the data**, studied in the exploratory data analysis step: The requirement for multivariate methods, the presence of spatial structure, the presence of trends, and the requirement for robust methods. The methods that are able to handle two, three or all aspects, are listed in the corresponding intersections. It is also possible none of these aspects apply, in which case we list the method outside of the diagram. The individual methods are briefly introduced one by one in Section 4. **In case it would not be clear which method to use**, or if multiple methods are mentioned in the relevant section of the diagram, it is recommended to go through these brief introductions first. If it would still be unclear what method to use, **further reading** is recommended, or the **consultation of a more experienced person**.

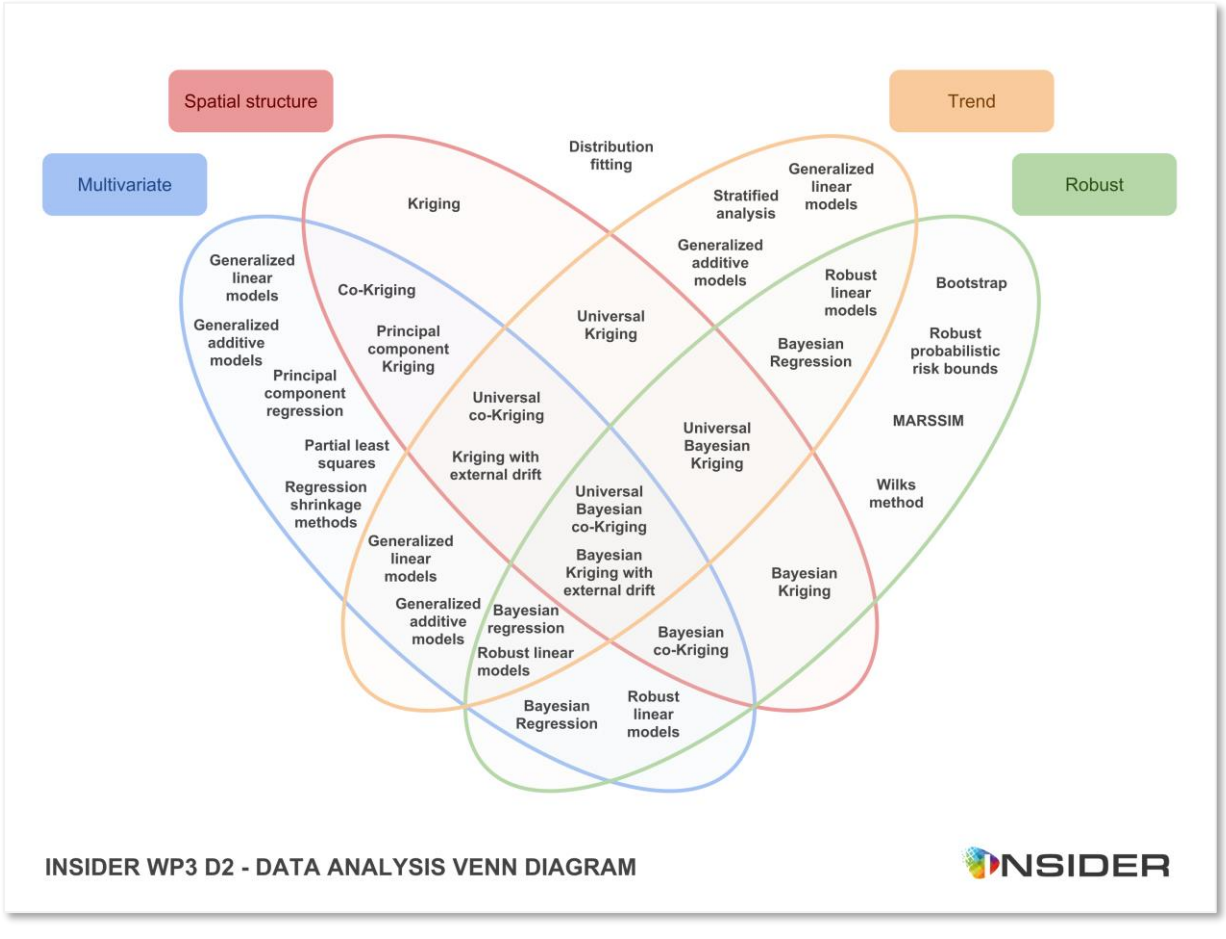


Figure 4. Data analysis Venn diagram. See Annex C for a larger version.

3.4 Postprocessing

The data analysis step does not always result in a direct answer to the question at hand, or direct information on the achievement of the target objective. Some postprocessing might be required, for instance, to

- Average the results over certain areas, volumes or masses,
- Change the spatial support of the outcome to one relevant for the objective,
- Compare estimates of variables of interest with *e.g.* regulatory thresholds,
- Merge the information related to different isotopes, or types of radiation,
- Interpret the outcome in terms of safety or cost,
- etc.

3.5 Is the objective achieved?

If the objective and constraints were clearly defined at the start, including the level of confidence that should be used in a probabilistic context, the answer to the question “Is the objective achieved?” should be as simple as “Yes” or “No” after the post-processing step. If achieved, one can move on to the next objective, if not, more data should be collected.

4 List of methods for data analysis

The different types of methods are discussed here one by one, in no particular order. All methods mentioned on the Venn diagram in Figure 4 are discussed separately or are part of a broader type of methods, in which case they are mentioned in the text. The expert can select one or more suitable methods from the Venn diagram. The same problem can typically be tackled using various methods.

Validation techniques for assessing the results obtained

Cross-validation can be used to assessing how the results of a statistical analysis will generalize to an independent data set. The model is initially fit on a training dataset, which is a set of values used to fit the parameters of the model. Successively, the fitted model is used to predict the responses for the observations in a second dataset called the validation dataset. This hold back sample of the full data should be used to give an unbiased estimate of the skill of the final tuned model by comparing or selecting between final models. A dataset can be repeatedly split into a training dataset and a validation dataset. Cross-validation combines (averages) measures of fitness in prediction to derive a more accurate estimate of model prediction performance. We distinguish two types of cross-validation: exhaustive and non-exhaustive cross-validation. Exhaustive cross-validation methods are cross-validation methods which learn and test on all possible ways to divide the original sample into a training and a validation set. Non-exhaustive cross validation methods do not compute all ways of splitting the original sample.

Attention should be paid to the re-estimation of extreme values as they can significantly bias the cross validation in the case of skewed distribution. As a consequence, a model can be preferred globally from cross-validation by better honouring the extreme values whereas the characterisation objective focuses on a low threshold where the model is not satisfactory.

Uncertainty & sensitivity analysis

The final outcome is anyhow affected by a certain degree of uncertainty, which often strongly impacts the decision-making. Quantifying the uncertainty on the variables of interest is usually a multi-dimensional and therefore complex task. Ideally, a **Bayesian inference** approach would be preferential compared to **Monte Carlo error propagation** or an approach based on first-order Taylor expansion. The latter might suffer from important drawbacks related to non-Gaussian distributions and non-linear expressions, or the omission of systematic uncertainties. Although applying a Bayesian inference approach or even a Monte Carlo error propagation would be the best technical choice, it might not be particularly suitable due to its complexity and the individual character of each model and case. For a one-time case, the following **alternative options** could be considered:

- A qualitative description, examining the most important and uncertain assumptions without performing a quantitative analysis.
- A scenario comparison, comparing two or more alternative scenarios (i.e. best case & worst case scenario).
- A once-at-a-time approach, introducing one at a time changes (i.e. minimum and maximum values for certain parameters) and comparing with a given baseline.

In addition, sensitivity analysis is a valuable instrument, which enriches the quantitative analysis of impact with a deeper investigation and identification of the sources of uncertainty. Among the most relevant techniques:

- Screening methods (Morris, 1991, Campolongo et al., 2011).
- Non-parametric or regression-based approaches (Saltelli & Marivoet, 1990; Helton, 1993).
- Variance-based methods (Sobol', 1993; Saltelli et al., 2008; Iman & Hora, 1990; Sacks et al., 1989).
- Spectral methods (Cukier et al., 1973; Saltelli et al., 1999; Sudret, 2008; Shao et al., 2017), and

- Moment-independent importance measures (Park & Ahn, 1994; Plischke et al.,2013).

In particular, variance- based methods related to Sobol's sensitivity indices are the most popular methods among practitioners due to their versatility and easiness of interpretation.

Uncertainty and sensitivity analysis are crucial as they help identifying the factors (assumptions, variables, data, and uncertainties) at play and provide information on their influence in quantitatively driving the impacts of the various decision options. In particular in view of nuclear site characterization:

- The impact of the sampling design is most fundamental (see section on sampling design).
- Using advanced statistical methods for dealing with samples below detection limit can have an important (beneficial) impact on estimated values and should be considered in case the data set contains a considerable amount of results below detection limit (see section on overall strategy). Obviously, the potential impact depends on the objective (estimated quantity).
- The following observed effects could be occasionally generalised as well, but might be specifically related to of specific use cases that have been tested for the validation of this guide (see section 6; Implementation and validation of the strategy in the following four use cases).
    - In the case of small data sets, the presence of outliers could clearly increase the uncertainty. Additional verification (process driven or error) might be necessary when estimations are becoming close to a decision threshold.
    - Sensitivity analysis shows that:
        - Most important uncertainties for the physical parameters estimated (e.g. volume categorisation, total activity) are due to spatial uncertainty (geostatistical simulations) and heteroscedasticity. Heteroscedasticity occurs more often in datasets that have a large range between the largest and smallest observed values, which is typically the case for radiological data.
        - The impact of sample measurement uncertainty on the volume categorisation is quite limited (if this uncertainty does not come from a systematic bias).
        - The effect on the estimated quantity (e.g. total activity or threshold) is not always the same as on the uncertainty.
        - The integration of non-destructive measurements in a multivariate approach significantly reduces uncertainties when sampling density is reduced.

4.1 Robust probabilistic risk bounds

4.1.1 Workflow

The Bienaymé-Chebytchev, Camp-Meidell and Van Danzig inequalities can be used to estimate the proportion of a population that exceeds a certain threshold, in a robust way. The approach is very much suited for small sample sizes and all inequalities make use of the empirical mean and the standard deviation. The Bienaymé-Chebytchev inequality does not require hypotheses on the probability distribution function for the studied population, and hence is the most conservative. The Camp-Meidell inequality assumes that the probability distribution function follows a unimodal continuous probability law (*e.g.* uniform, Gaussian, triangular, lognormal, Weibull). It is less conservative, but more accurate than the Bienaymé-Chebytchev inequality. The Van Dantzig inequality can only be applied to the convex part of all the unimodal continuous probability laws, and hence is less conservative than the Camp-Meidell inequality. Even though the tail of most classic distribution laws is convex (*e.g.* exponential, triangular, Gaussian, Weibull), it can for instance not be applied to uniform distributions.

4.1.2 Example software implementations

We are not aware of any implementations. All equations are however provided by (Pérot, et al., 2019), and should be straightforward to implement in any programming language or spreadsheet software.

4.1.3 Example applications

Reference	Type
Blatman G, Delage T, Iooss B, Pérot N. 2017. Probabilistic risk bounds for the characterization of radiological contamination. EPJ Nuclear Sci. Technol. 3, 23. https://www.epj-n.org/articles/epjn/pdf/2017/01/epjn160031.pdf	Waste categorization & cost estimation. H2 flow rate estimation.
Blatman G, Iooss B. 2012. Confidence bounds on risk assessments - application to radiological contamination. In: Proceedings of the PSAM11 ESREL 2012 Conference, Helsinki, Finland. pp. 1223–1232	Waste categorization from few contamination measurements of radiological activity in Cesium 137 of a large-size population of waste objects.

Table 1. Examples of case studies that make use of this method.

4.1.4 Theoretical references

Reference
Blatman G, Delage T, Iooss B, Pérot N. 2017. Probabilistic risk bounds for the characterization of radiological contamination. EPJ Nuclear Sci. Technol. 3, 23. https://www.epj-n.org/articles/epjn/pdf/2017/01/epjn160031.pdf
Blatman G, Iooss B. 2012. Confidence bounds on risk assessments - application to radiological contamination. In: Proceedings of the PSAM11 ESREL 2012 Conference, Helsinki, Finland. pp. 1223–1232.

Table 2. Theoretical references, with more details concerning this method.

4.2 Bootstrap

4.2.1 Workflow

Bootstrap refers to a class of methods that makes use of resampling from a particular dataset, to obtain *e.g.* robust estimates on certain population parameters. In the context of nuclear site characterization, it is useful when the number of experimental data is limited (*e.g.* due to complexity of radioactive contamination measurements in the nuclear facility site) or when the exact or asymptotic distributions of data are unknown or at least very uncertain. There should be a minimum sample size, however. Several studies recommend a minimum number of measuring points of 18 to ensure a minimum amount of reliability of the results. The main advantage of this method is that it allows defining an empirical distribution function, which is the maximum likelihood estimator of the distribution for the observations when no parametric assumptions are made.

Different variants of bootstrap exist however. Practical application of the technique usually requires the generation of k bootstrap samples or resamples (i.e., samples obtained by independently sampling with replacement from the empirical distribution). For standard error estimation, k is recommended to be at least 100. If k becomes very large (*e.g.* more than 500) there is very little difference between the regular bootstrap estimator and Monte Carlo approximation, and the term Monte Carlo bootstrap is often used.

The double bootstrap was proposed to improve on the bootstrap bias correction for the apparent error rate of a linear discriminant rule. It was the first application of bootstrap iteration, which includes taking resamples from each bootstrap resample. The double bootstrap confidence interval procedure is a further iteration to the normal bootstrap confidence interval procedure, which would further reduce the order of magnitude of coverage error. A two-sided coverage error

of O (n^{-1}) at a nominal level of α, as a result of further iteration, the confidence interval of the normal bootstrap will further reduce the order of magnitude of the coverage error to O (n^{-2}).

Parametric bootstrap can be viewed as a generalization of Fisher's maximum likelihood approach to the nonparametric framework. Parametric bootstrap may be used in cases where the estimator of interest has a distribution that is difficult to derive analytically or has an asymptotic distribution that does not provide a good small sample approximation, particularly for the variance, which is where the bootstrap is often useful. However, since the existing theory on maximum likelihood estimation is adequate, it is not common to see parametric bootstrap used in practice.

Bayesian bootstrap can be used to make the usual Bayesian - type inferences about parameters based on their estimated posterior distribution, whereas, strictly speaking, the regular nonparametric bootstrap has only the usual frequentist's interpretation about the distribution of the statistic. Bayesian bootstrap is appropriate in some problems but the prior is viewed as restrictive and hence it is not recommended as a general inference tool.

The percentile method is the most obvious way to construct a confidence interval for a parameter based on bootstrap estimates. The main difference between random subsampling and bootstrapping is that bootstrapping involves sampling with replacement from the original sample whereas random subsampling selects without replacement from the set of all possible subsamples. As the sample size becomes large, the difference in the distribution of the bootstrap estimates and the subsample estimates becomes small. Therefore, we expect the bootstrap percentile interval to be almost the same as the random subsample interval. The bootstrap percentile method is not exact (*i.e.* the parameter is contained in the generated intervals in exactly the advertised proportion of intervals as the number of generated cases becomes large). For the median, the percentile method provides nearly the same confidence interval as the nonparametric interval based on the binomial distribution, so the percentile method works well in some cases even though it is not exact.

So the percentile intervals inherit the exactness property of the subsample interval asymptotically (i.e., as the sample size becomes infinitely large). In the case of small samples (especially for asymmetric distributions), the percentile method does not work well.

4.2.2 Example software implementations

Name	Type	What does it do?	URL
boot	Open source R package	Classic parametric & non-parametric bootstrap	https://cran.r-project.org/web/packages/boot/index.html
bayesboot	Open source R package	Bayesian bootstrap	https://cran.r-project.org/web/packages/bayesboot/index.html
kernelboot	Open source R package	Smoothed bootstrap	https://cran.r-project.org/web/packages/kernelboot/index.html

Table 3. Example software implementations of this method.

4.2.3 Example applications

Reference	Type
Zaffora B, Magistris M, Saporta G, La Torre F. 2016. Statistical sampling applied to the radiological characterization of historical waste. EPJ Nuclear Sci. Technol., 2, pp. 11.	Bootstrap, Monte-Carlo bootstrap, to estimate mean, median, standard error, confidence intervals and bias
Mikheenko S, Erofeeva S, Kosako T. 1994. Reconstruction of Dose Distribution by the Bootstrap Method Using Limited Measured Data. Radioisotopes, 43, 595-604.	Regular bootstrap, bootstrap with weight coefficients, to estimate average, standard deviation, intervals, confidence level and weighted coefficients
da Silva ANC, dos Santos Amaral R, dos Santos Junior JA, Vieira JW, Cezar Menezes RS. 2015. Statistical analysis of discrepant radioecological data using Monte Carlo Bootstrap Method. Journal of Radioanalytical and Nuclear Chemistry, 306-571.	Monte-Carlo bootstrap, to estimate mean, standard deviation, confidence intervals, quantiles and range
Wan H, Zhang T,·Zhu Y. 2012. Detection and localization of hidden radioactive sources with spatial statistical method. Ann Oper Res, 192, 87-104.	Bootstrap, for the likelihood ratio statistic, maximum likelihood estimates of strength and location of a source, and the p-value of the likelihood ratio statistic

Table 4. Examples of case studies that make use of this method.

4.2.4 Theoretical references

Reference
Davison AC, Hinkley DV. 1997. Bootstrap Methods and Their Application. Cambridge University Press, Cambridge.
Efron B. 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. SIAM, Philadelphia.
Efron B. 1983. Estimating the error rate of a prediction rule: improvements on crossvalidation. J. Am. Statist. Assoc. 78, 316 – 331.
Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors: Confidence intervals and other measures of statistical accuracy. Statistical Science 1, 54 – 77.
Efron B, Tibshirani R. 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.
Hall P. 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.
Chernick MR. 2008. Bootstrap Methods: A Guide for Practitioners and Researchers. Second Edition. United BioSource Corporation. Newtown, PA.

Table 5. Theoretical references, with more details concerning this method.

4.3 Geostatistics

4.3.1 Workflow

A good geostatistics processing of the data always starts with a deep preliminary data analysis. This pre-processing step is crucial to build a consistent database. First, spatial bias due to non-probabilistic sampling is tackled with declustering techniques (spatial weighting). The joint analysis of the statistical distribution (histogram) may identify heterogeneous populations (spatial, temporal, units, sample support, …) that need to be corrected or separated for a proper understanding. Regularization and deconvolution may be used to take support differences into account. Similarly, a skewed distribution requires a non-linear data transformation (indicator or logarithm eventually but more interestingly Gaussian anamorphosis). In the presence of a correlated variable, a multivariate processing can be implemented at all stages, and is very useful if the auxiliary data is denser than that of the primary variable.

The heart of geostatistics is the analysis and the modeling of the spatial continuity using the variogram. A nugget effect (variability at small scale) can be related to the phenomenon itself, to positioning uncertainty or measurement variability. Anisotropy is generally relevant in 3D, and also

in 2D if there is a specific direction for the phenomenon. An unbounded variogram may point to the necessity of trend modelling (using external drift or a universal kriging approach). In a multivariate approach and when primary data is very scarcely sampled, bundled processing can be selected (spatial structure from auxiliary variable or external drift (generalized covariance)).

The first output of geostatistics is the interpolation, or the kriging estimate. It can be a punctual or a block estimate. More sophisticated regular change-of-support models can also be used. The interpolation neighborhood can be global or moving, also considering anisotropy, and the mean of the variable can be known or not (*i.e.* simple vs ordinary kriging). In the multivariate case, one uses co-kriging. Things can also be simplified, however, in which case univariate kriging combined with *e.g.* principal component analysis (*i.e.* principal component kriging) can provide an adequate solution as well. Finally, different degrees of Bayesian inference can be integrated in the geostatistical approach, for instance for accounting for prior information and the uncertainty on the variogram.

In addition to the kriging estimate, the kriging error variance is a second output. Depending on the geostatistical model, non-linear estimates can be calculated. Risk of exceeding a threshold with prior Gaussian anamorphosis (conditional expectation) is specifically relevant for radiological waste classification. Another approach is to compute conditional simulations (which can be seen as spatially consistent Monte-Carlo simulations). They generally require prior Gaussian anamorphosis. Post-processing of these simulations enables regular (grid) and irregular (polygon) change-of-support modelling (averaging on a larger support) as well as any statistics (mean, variance, probability of exceeding a threshold). Global estimates (volumes according to a threshold, source term) are obtained in the same way.

4.3.2 Example software implementations

Name	Type	What does it do?	URL
gstat	Open source R package	Variography, simple, ordinary and universal point or block (co)kriging, sequential Gaussian or indicator (co)simulation	https://cran.r-project.org/package=gstat
geoR	Open source R package	Traditional, likelihood-based and Bayesian methods	https://cran.r-project.org/package=geoR
georob	Open source R package	Robust and Gaussian (Restricted) Maximum Likelihood methods.	https://cran.r-project.org/package=georob
RandomFields	Open source R package	Inference on and the simulation of Gaussian fields.	https://cran.r-project.org/package=RandomFields
mGstat	Matlab library	Structural analysis, kriging (simple, ordinary, with external drift), conditional simulations, 3D and spatio-temporal	http://mgstat.sourceforge.net/
Earth Volumetric Studio	Commercial	Kriging, indicator kriging (for geology), probability of exceeding a threshold, 2D and 3D	https://www.ctech.com/products/earth-volumetric-studio/
Geostatistical Analyst	Commercial	Only 2D, variography, trend removal, simple, ordinary, universal, indicator, and disjunctive kriging and co-kriging	https://www.esri.com/en-us/arcgis/products/geostatistical-analyst/
Isatis	Commercial	Exploratory data analysis, variography, trend, exhaustive set of kriging and simulations techniques, multivariate, computation faults, risk curves	https://www.geovariances.com/en/software/isatis-geostatistics-software/
Kartotrak	Commercial	Exploratory data analysis, variography, punctual and block kriging, turning band simulations, multivariate, risk curves	https://www.geovariances.com/en/software/kartotrak-software-contamination-characterization/
SADA		Variography, ordinary kriging, indicator kriging, and co-kriging	https://www.sadaproject.net/download.html
SGeMS		Variography, simple kriging, ordinary kriging, kriging with external drift, block kriging, and indicator kriging. Extensive set of conditional simulation methods	http://sgems.sourceforge.net/
Visual Sample Plan	Open source	Only 2D, variography, simple, ordinary and block kriging	https://vsp.pnnl.gov/

Table 6. Example software implementations of this method.

4.3.3 Example applications

Reference	Type
Desnoyers Y. 2015. Sampling considerations for characterization of radioactive contamination using geostatistics. In Proc. of World Conference on Sampling and Blending, Bordeaux, France.	Sampling considerations
Desnoyers Y, Dubot D. 2014. Characterization of radioactive contamination using geostatistics. Nuclear Engineering International 59, 716, 16-18. Sidcup, United Kingdom.	General geostatistical approach and sampling optimization
Desnoyers Y, Chilès J-P, Dubot D, Jeannée N, Idasiak J-M. 2010. Geostatistics for radiological evaluation: study of structuring of extreme values. In: Stochastic Environmental Research and Risk Assessment 25, 8 (2011), 1031-1037	Spatial structure of extreme values
Bechler A, Romary T, Jeannée N, Desnoyers Y. 2013. Geostatistical sampling optimization of contaminated facilities. In: Stochastic Environmental Research and Risk Assessment 27, 8 (2013), 1967-1974.	Sampling optimization
Boden S, Rogiers B, Jacques D. 2013. Determination of Cs-137 contamination depth distribution in building structures using geostatistical modelling of ISOCS measurements. Applied Radiation and Isotopes 79, 25-36.	Combination of different measurement supports.

Table 7. Examples of case studies that make use of this method.

4.3.4 Theoretical references

Reference
Chilès JP, Delfiner P. 1999. Geostatistics – Modeling Spatial Uncertainty. Wiley series in Probability and Statistics, New-York.
Journel AG, Huijbregts CJ. 1978. Mining Geostatistics. Academic Press, London, 600p.
Isaaks EH, Srivastava RM. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York, 561 p.
Goovaerts P. 1997. Geostatistics for natural resources evaluation. Oxford University Press.

Table 8. Theoretical references, with more details concerning this method.

4.4 MARSSIM

4.4.1 Workflow

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). 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.

4.4.2 Example software implementations

Name	Type	What does it do?	URL
Visual Sampling Plan	Freeware	Sign and WRS tests Other statistical tests	https://vsp.pnnl.gov/
Kartotrak	Commercial	Sign and WRS tests Wilks formula	https://www.geovariances.com/en/software/kartotrak-software-contamination-characterization/
stats	Open source R package	wilcox.test: WRS test.	https://www.r-project.org/

Table 9. Example software implementations of this method.

4.4.3 Theoretical references

Reference
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).
Mann HB, Whitney DR. 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Annals of Mathematical Statistics. 18, 1, 50–60.
Wilcoxon F. 1945. Individual comparisons by ranking methods. Biometrics Bulletin. 1, 6, 80–83.

Table 10. Theoretical references, with more details concerning this method.

4.5 Wilks method

4.5.1 Workflow

The Wilks formula computes a quantile (or a tolerance interval) with a given confidence level from an i.i.d. (random) sample, or computes the minimal sample size to estimate a quantile (or a tolerance interval) with a given confidence level (cf. § 5.2. (Pérot, et al., 2019)). This method 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.

4.5.2 Example software implementations

Name	Type	What does it do?	URL
MISTRAL	Open source R package	Computing quantile with Wilks formula (quantileWilks)	https://cran.r-project.org/package=mistral

Table 11. Example software implementations of this method.

4.5.3 Example applications

Reference	Type
Blatman G, Delage T, Iooss B, Pérot N. 2017. Probabilistic risk bounds for the characterization of radiological contamination. EPJ Nuclear Sci. Technol. 3, 23. https://www.epj-n.org/articles/epjn/pdf/2017/01/epjn160031.pdf	Waste categorization & cost estimation. H2 flow rate estimation.
Pérot N. 2018. Sampling strategy for dihydrogen flow rate characterization of radioactive waste, DEM 2018, Avignon.	H2 flow rate estimation

Table 12. Examples of case studies that make use of this method.

4.6 Regression

4.6.1 Workflow

Regression encompasses a large range of methods, and many of them can be applicable in the field of initial nuclear site characterization. All methods however share the basic goal of estimating a certain response variable based on one or more predictor variables. In most cases, the response variable corresponds to the variable of interest discussed in Section 2.2. We consider here two types of predictor variable(s), but their treatment is mathematically the same: Those that are related to the spatial location (*i.e.* spatial coordinates in the broadest sense), and other covariates that teach us something on the predictor, irrespective of the spatial location. This corresponds to the trend and multivariate cases, used to structure the methods in this document. Hence, the same methods might be relevant in the presence of trends and other multivariate cases.

In general, linear models are very well suited for the majority of problems. Generalized linear models extend this class of models to response variables with non-Gaussian error distributions. It is worth mentioning logistic and Poisson regression here, both part of the generalized linear model class, as they are relevant for *e.g.* classification and working with count data.

When linear models are not able to capture the studied predictor-response relations, generalized additive models provide another extension to capture non-linearity. If the functional form of the relationship is (partly) known, however, non-linear regression can also be used to estimate the unknown coefficients. If robustness is required, robust linear models can be used to decrease the impact of outliers, and Bayesian regression variants exist for incorporating prior knowledge in case of small datasets.

A final set of methods is designed to handle cases where predictor variables are heavily correlated, and this multicollinearity would affect the outcome and uncertainty estimates of the standard approaches. Dimensionality reduction can be built into the regression approach, as *e.g.* with principal component regression or partial least squares (the latter is preferred when the response variable is correlated with one or more of the minor factors/components). Another approach is to put restrictions on the regression coefficients with so-called shrinkage methods: Ridge regression tries to keep the regression coefficients small, whereas the lasso approach additionally may set some coefficients to zero, effectively selecting a subset of all predictors in the model. The elastic net approach combines the ridge and lasso methods.

The field of machine learning offers many more techniques that might be appropriate if linear regression and its extensions fail at capturing the predictor-response relationships. However, typically, much more data is required to use these methods, which is likely not available from initial nuclear site characterization, and constraint environments.

4.6.2 Example software implementations

Name	Type	What does it do?	URL
MASS	Open source R package	Linear regression and graphical analysis, robust linear models	https://cran.r-project.org/package=MASS
e1071	Open source R package	Regression diagnostics	https://cran.r-project.org/package=e1071
stats	Open source R package	Linear models, Generalized linear models	https://www.r-project.org/
gam	Open source R package	Generalized additive models	https://cran.r-project.org/package=gam
mgcv	Open source R package	Generalized additive models	https://cran.r-project.org/package=mgcv
glmnet	Open source R package	Penalized linear models (shrinkage methods)	https://cran.r-project.org/package=glmnet

Table 13. Example software implementations of this method.

4.6.3 Example applications

Reference	Type
Zaffora B, Magistris M, Saporta G, La Torre FP. 2016. Statistical sampling applied to the radiological characterization of historical waste. EPJ Nuclear Sci. Technol. 2, 34	Linear regression for scaling factors
Zaffora B. 2017. Statistical analysis for the radiological characterization of radioactive waste in particle accelerators. PhD thesis. https://cds.cern.ch/record/2290521/files/CERN-THESIS-2017-194.pdf	Multiple linear regression for scaling factors
Zaffora B, Magistris M, Chevalier J-P, Luccioni C, Saporta G, *et al.* 2017. Appl. Radiat. Isot. 122, 141-147.	Regression for scaling factors

Table 14. Examples of case studies that make use of this method.

4.6.4 Theoretical references

Reference
Rawlings JO, Pantula SG, Dickey DA. 1998. Applied regression analysis - A research tool. Springer-Verlag New York. 660 pp
James G, Witten D, Hastie T, Tibshirani R. 2013. An introduction to statistical learning – with Applications in R. Springer-Verlag New York. 426 pp
Hastie T, Tibshirani R, Friedman J. 2009. The elements of statistical learning – Data mining, inference, and prediction, Second edition. Springer-Verlag New York. 745 pp

Table 15. Theoretical references, with more details concerning this method.

4.7 Distribution fitting

4.7.1 Workflow

The first step of fitting a distribution on a data set is to represent its histogram, which 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) (cf. § 4.1. of (Pérot, et al., 2019).

4.7.2 Example software implementations

Name	Type	What does it do?	URL
MASS	Open source R package	Distribution fitting (fitdistr)	https://cran.r-project.org/package=MASS
Matlab (Mathworks)	Statistics and curve fitting toolboxes	Distribution fitting	https://fr.mathworks.com/

Table 16. Example software implementations of this method.

4.7.3 Example applications

Reference	Type
Blatman G., Delage T., Iooss B., Pérot N (2017), Probabilistic risk bounds for the characterization of radiological contamination, EPJ Nuclear Sci. Technol. 3, 23. https://www.epj-n.org/articles/epjn/pdf/2017/01/epjn160031.pdf.	Waste categorization & cost estimation. H2 flow rate estimation.

Table 17. Examples of case studies that make use of this method.

5 List of approaches for sampling design

If the objective cannot be achieved with the available data, more information is required, and a proper sampling design should be made before collecting new data. There exists a variety of different ways to approach this, and the main drivers here are the available data, the type of problem at hand (revealed by the exploratory data analysis), the outcome of the data analysis, and the reason why the objective cannot be achieved. As many factors therefore influence what would be the best approach, we take a top down approach here, and describe the individual approaches briefly in this section, discussing typical cases where they are commonly applied. Some aspects related to sample locations, size and support are discussed first below however.

In case it would not be clear which method to use, or if multiple methods seem to be equally adequate based on the Venn diagram, it is recommended to go through the brief introductions first. If it would still be unclear what method to use, **further reading** is recommended, or the **consultation of a more experienced person**.

Note that the list of approaches discussed here is non-limitative. Sampling approaches more applicable in alterntive fields were not considered, and a more advanced set of approaches, which are less commonly used, but might be useful in certain cases, is discussed separately under “optimization”.

Sample locations

A Venn diagram providing an overview of different sampling design approaches is provided in Figure 5. The expert can select one or more suitable methods from the Venn diagram. The same problem can be tackled using various methods. Often, a combination of approaches is being implemented. In any case, the sample locations should be selected so that subsequent extrapolation during data analysis is avoided. Many characterisation projects have the tendency to focus their sampling efforts on the highest affected areas, neglecting areas with lower activity concentration levels. Nonetheless, it is necessary to sample the supposedly least impacted zones as well as the most impacted zones to achieve a realistic understanding of the statistical distribution of the activity concentration. Confirming some non-impacted areas is often as important as (or even more important than) confirming historically impacted areas. From the point of view of waste volume management, transition zones are more critical, since it is difficult to categorize them with respect to the reference thresholds. Uncertainty being the most important in these areas for proper delineation (and limiting misclassification errors), the sampling distribution should favour them over other areas that only require confirmation of impacted or non-impacted. We make a distinction here between probabilistic and non-probabilistic approaches, and designs with equal or non-equal probability of selection:

1. **Probabilistic sampling:** We use the term here to indicate sampling strategies where all elements in a population have a certain probability to be selected. This probability should be known, or easily determined, so proper **inference on the total population** can be made.
2. **Non-probabilistic sampling:** We consider an approach to be non-probabilistic when certain elements in a population have a zero probability of being selected, or the probability cannot be determined. Hence, a non-probabilistic sample cannot be used to do inference on the total population, without making assumptions, and is only **targeted at a specific part** of it.
3. **Equal probability of selection:** We use the term here to indicate sampling strategies in which all obtained samples had the same probability of being selected. The part of the population considered for sampling is therefore **explored in a uniform way**.
4. **Unequal probability of selection:** We use the term here to indicate sampling strategies in which the obtained samples had different probabilities of being selected. The part of the population considered for sampling is therefore **explored in a non-uniform way**.

Intersections in the Venn diagram indicate here that the details of the respective sampling design approaches can be chosen such that they can fall both under the probabilistic or non-probabilistic classes, and the equal or unequal probability of selection classes. The effective implementation of a sampling design approach can however not be probabilistic and non-probabilistic, or equal and unequal probability of selection at the same time.

The different approaches listed in Figure 5 are discussed in greater detail below. It should be noted here however, that in practice, sampling design consists most often of a combination of these approaches, as objectives and/or sampling targets are often multifold in real life.

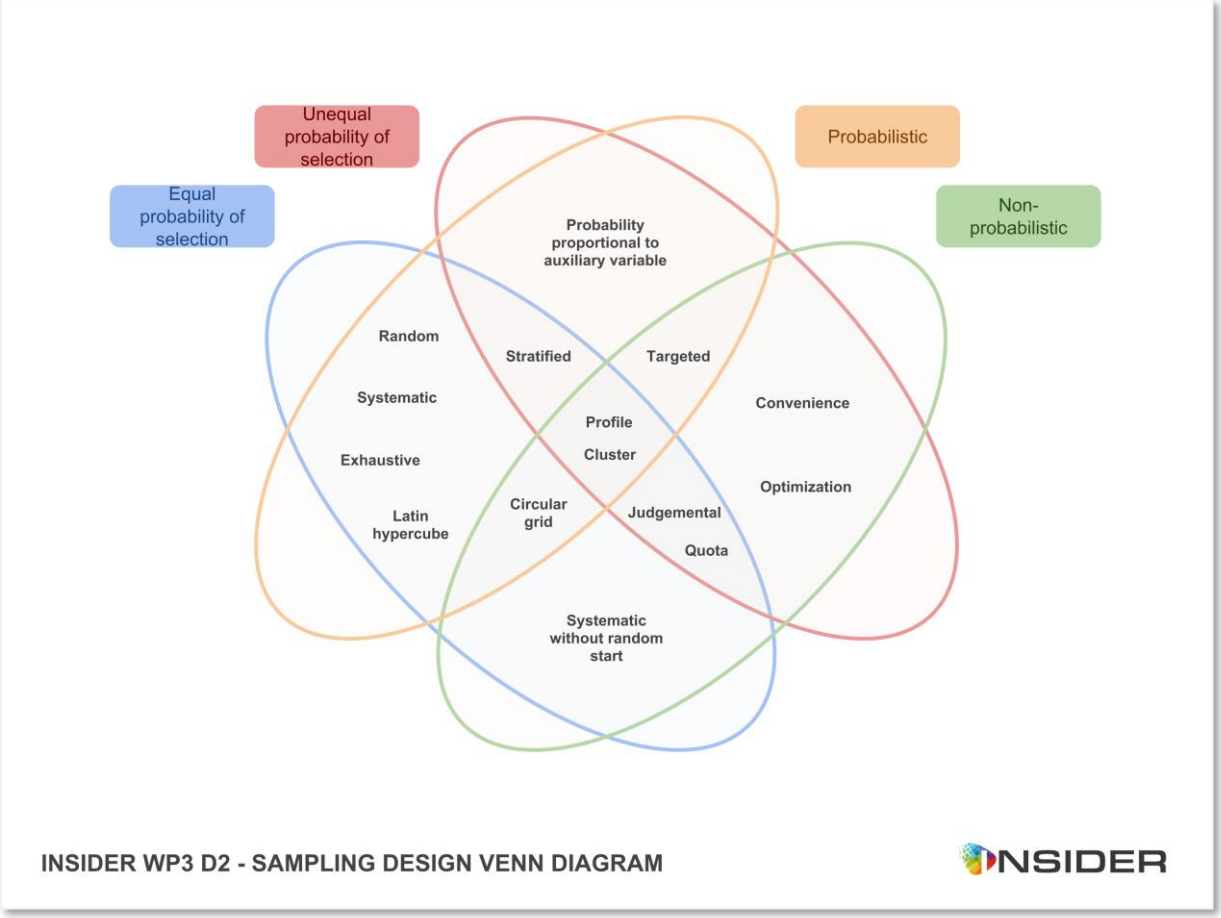


Figure 5. Sampling design Venn diagram. See Annex D for a larger version.

Sample size or density

The sample size can be determined according to the estimator (mean, proportion, quantile, etc.) used and the confidence interval required. 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 (cf. §2.3.2, (Pérot, et al., 2019)).

Sometimes because of access constraints, measurement costs are such that it is unconceivable to achieve many measurements, but it 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 (cf. §4.2, (Pérot, et al., 2019)). If we observe a stabilization of the bootstrap estimator and its confidence interval, we can deduce the data set size is correct. Otherwise, more measurement data are required. Wilks method is

another way to estimate the size of the data set required to estimate quantiles with a given confidence level (cf. § 4.5, (Pérot, et al., 2019)).

Defining sample size might be challenging when several physical parameters need to be assessed (i.e. total activity, activity concentration, thresholds) based on various data sources that might not always be representative. Moreover, a data set might contain considerable amounts of values below detection limit and confidence levels required might not always be unequivocally defined. Reducing the size generally leads to an increasing uncertainty and could result in extreme under/over estimation of the volume exceeding a threshold. Such deviations can be strongly reduced by combining the limited higher quality and costly primary dataset (e.g. in-lab sample measurements) with a large cheap secondary data set (e.g. in-situ measurements).

Sample support

In certain cases, the sample support is not really an issue, as the population consists of discrete objects that are measured in their entirety. In many cases however, samples have to be taken from a continuum of material with a certain spatial support (*i.e.* length, area or volume). The number of possible sample locations is infinite in such a case, and an appropriate sample support should be defined. The selection of a sample support can be influenced by different factors, amongst others:

- The amount of material required in the measurement procedure.
- The selective decontamination spatial support (if decontamination methods are known a priori).
- The relation sample support vs cost.

When the sample support would be far larger than the amount of material required for performing a measurement, homogenization or subsampling techniques can be considered to effectively homogenize the contents of the sample, and reduce the amount of measurements required.

If the most interesting sample support is of a practically infeasible magnitude, composite sampling techniques can be used for homogenization across the targeted support, resulting in a manageable set of samples that carry the targeted information.

Note that when samples with different supports are collected, measured and combined for a certain analysis, prior regularization or correction for the support effect should be performed.

5.1 Random sampling

Random sampling is the most basic probabilistic way of sampling, where all elements in a population have the same probability to be selected. It offers a great starting point when little is known about the population, the presence of trends or a spatial structure is not expected, and there are no clear subpopulations. In principle, it is a uniform way of sampling from the sample space, but systematic and Latin hypercube sampling are more efficient in that respect.

5.2 Systematic sampling

In systematic sampling (a.k.a. regular or grid sampling), the population is ordered in a certain way, and every k^{th} element is selected. The starting point should however be random, so all elements would have a certain probability of being selected (*i.e.* a probabilistic approach). In a spatial context, ordering is often done using the spatial coordinates, and the term grid sampling is often used as well. In many cases, systematic sampling offers a more homogeneous spread of samples over the population than random sampling, but it might be dangerous in case of periodicity, with periods equal or close to the sampling interval. It is a good option for characterizing spatial structure, or when this would be known a priori, it can be used to optimize the sample spacing.

5.3 Stratified sampling

Stratified sampling can be used in case there are clearly different subpopulations. Every subpopulation is sampled independently, to make sure the desired balance between subpopulations is obtained in the final sample. Sampling within a subpopulation is often done randomly, but virtually any approach can be applied. When the sample sizes for all subpopulations are equal, we are dealing with a balanced design. For making inference on the total population, however, sample sizes proportional to the subpopulation sizes make more sense (= special case of probability proportional to auxiliary variable sampling).

5.4 Cluster sampling

Cluster sampling also deals with the case of different subpopulations. In contrast to stratified sampling, however, only a subset of all subpopulations is sampled in a random way. The selected subpopulations can be sampled exhaustively, or with any other sampling approach. If sampling of all subpopulations (*i.e.* a stratified approach) would be infeasible, cluster sampling offers an alternative with a larger set of samples, but larger risk of bias.

5.5 Judgemental sampling

When samples are chosen based on expert judgement, we use the term judgemental sampling (a.k.a. purposive sampling). It is often very useful to explore the behaviour of elements in a population that deviate from the dominant, overall pattern in that population. As it does not tell us much about the total population, however, it is often applied in combination with probabilistic sampling approaches.

5.6 Targeted sampling

Targeted sampling refers to the case where subpopulations are selected using a judgemental approach, after which another sampling approach can be used for selecting elements within that subpopulation. Hence, it cannot be used to make inferences on the total population. However, when the judgemental selection is appropriate, it might still contain the entire relevant part of a population. An example is sampling within a certain distance from a source of contamination or activation. If the new data would reveal a trend suggesting that the neglected part of the population is far below any regulatory thresholds, still relevant inferences can be made with the appropriate assumptions.

5.7 Convenience sampling

Convenience sampling (a.k.a. accidental, grab or opportunity sampling) refers to the case where the selection of elements from a population is done because they are readily available, or the most convenient to sample. Hence, this is a non-probabilistic approach, and might involve quite some bias. It can be useful however for obtaining a quick first set of estimates on a population that is expected to be reasonably homogeneous, or that should exhibit a clear, simple trend.

5.8 Circular grid sampling

Circular grid sampling is a form of systematic sampling, in a spatial context, where a polar coordinate system is used rather than a Cartesian one. This is useful in the context of point sources of contamination, and an isotropic or unknown migration direction.

5.9 Latin hypercube sampling

Latin hypercube sampling is similar to systematic sampling, with the difference that a random sample is taken in every set of k elements, instead of retaining every k^{th} element. In this way, it is more robust with respect to periodicity than systematic sampling. An alternative way of putting this is that for every k^{th} element in a systematic design, we choose the element at a random distance (with a given maximum, typically k/2) from it.

5.10 Profile sampling

Profile sampling (a.k.a. line-intercept or transect sampling) is only applicable in the context of trends or spatial structure, and aims at selecting a set of samples on a line (or close to it), or at least some path throughout the population, whereby the retained set of samples is representative in some kind of way for the entire population. It can be generalized as systematic sampling along a lower-dimensional path through the population (*e.g.* sampling along a plane through a 3D population). It is a non-probabilistic non-uniform way of sampling, but when a trend is indeed revealed inference on the total population can be made under the assumption that the trend is valid everywhere. In terms of spatial structure, it offers an efficient approach for quantifying the spatial structure, which can be used in a later iteration for more appropriate sampling design.

5.11 Quota sampling

Quota sampling is similar to stratified sampling. For sampling within a subpopulation, however, a non-probabilistic approach like judgement sampling is used rather than a probabilistic one like random sampling, which is typically done in a stratified approach.

5.12 Probability proportional to auxiliary variable sampling

Probability proportional to size sampling is the most used variant of sampling based on an auxiliary variable (uncertainty sampling is another). In this case, the auxiliary variable is the size of a given subpopulation, or element, which represents its relative contribution to the population as a whole. In general, however, the auxiliary variable can be any kind of information that could be used for obtaining a more representative or targeted sample. An example is for instance the uncertainty related to a specific element in a population. Sampling elements with larger uncertainty preferentially (*i.e.* using a sampling probability proportional to the uncertainty in some way), will reduce the uncertainty on the population as a whole more efficiently. Another example is the distance to a certain decision threshold. Sampling elements with large uncertainty, but still far away from a decision threshold makes no sense, while sampling elements with small uncertainty, incorporating the decision threshold, can better inform decision-making.

5.13 Exhaustive sampling

In general, a sampling design aims at providing as much as information, with as little as samples possible. In certain cases however, it might be more convenient to just sample all elements within a population. This is especially true when the performed measurements are cheap and quick, and might provide the necessary data to meet the objectives, or at least provide some secondary information (*e.g.* correlated variables, or measurements on a different support) that can contribute to reaching the goal.

5.14 Optimization

Sampling strategy sometimes evolves into an iterative or adaptive approach. Based on a first sampling data set, it can be necessary to collect additional points in order to improve the initial estimation and/or to reduce related uncertainties. This sampling optimization is then strongly impacted by the characterization objective and can follow different rules:

- Statistics: add random points to improve statistics.
- Spatial clustering: add points around initial values that exceed a threshold (or any other criterion) to improve delineation.
- General optimisation: find the best set (number and location) of additional points using computer algorithms (simulated annealing, genetic algorithm…) for a given objective function.

6 About

The STRATEGIST web tool has been developed within work package 3 of the INSIDER EU Horizon 2020 project and received funding from the Euratom Research and Training Programme 2014-2018 under grant agreement No 755554.

The development consisted of:

1. Providing an overview of the available sampling design methods and state-of-the-art techniques.
2. Development of a strategy for data analysis and sampling design, referring to state-of-the-art techniques, and provide guidance to the end user through an application in which the strategy contents can be explored in a user-friendly way.
3. Implementation and validation of the strategy in the following four use cases:
    a. Use case 1 concerns two tanks (VA001 and VA002), each about 50 m^3 in volume, containing low level liquid waste (LLLW) and located in the liquid waste storage facility at the Joint Research Centre site of Ispra, Italy.
    b. Use case 2 involves the biological shield of the Belgian Reactor 3 (BR3), a pilot pressurized water reactor of the SCK CEN (Belgian Nuclear Research Centre).
    c. Use case 3a relates to a nuclear facility that was devoted to radiochemistry on trans-uranium elements. It was under operation until 1992 on a CEA site in France.
    d. Use case 3b covers the graphite moderator and reflector (about 1300 tons) of the G2 UNGG reac-tor localized at the CEA site of Marcoule.

    The lessons learnt are summarized in:
    - The following document.
    - The following publication.
4. Designing the STRATEGIST web tool: blueprint of the tool.

Authors

- bart.rogiers@sckcen.be (SCK CEN, Belgium)
- desnoyers@geovariances.com (Geovariances, France)
- nadia.perot@cea.fr (CEA, France)
- g.vonoertzen@brenk.com (Brenk Systemplanung, Germany)
- sevbo@energorisk.com.ua (Energorisk, Oekraine)
- severine.demeyer@lne.fr (LNE, France)
- sven.boden@sckcen.be (SCK CEN, Belgium)

7 Bibliography

Campolongo, F., Saltelli, A., & Cariboni, J. (2011). From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications*, 182(4), pp. 978-988.

Cukier, R., Fortuin, C., Shuler, K., Petschek, A., & Schaibly, J. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of chemical physics*, 59(8), pp. 3873-3878.

Helsel, D., & Cohn, T. (1988). Estimation of Descriptive Statistics for Multiple Censored Water Quality data. *Water Resources Research Vol. 24, 1997-2004*.

Helton, J. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, 42(2-3), pp. 327-367.

Iman, R., & Hora, S. (1990). A robust measure of uncertainty importance for use in fault tree system analysis. *Risk analysis*, 10(3), pp. 401-406.

Journel, A., & Huijbrechts, C. (1978). *Mining Geostatistics.* Academic Press, London, 600 p.

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

Kim, J. H., Hornibrook, C., & Yim, M.-S. (2020). The impact of below detection limit samples in residual risk assessments for decommissioning nuclear power plant sites. *Journal of Environmental Radioactivity, 222*(ISSN 0265-931X, https://doi.org/10.1016/j.jenvrad.2020.106340), 106340.

Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33, 161-174.

Park, C., & Ahn, K.-I. (1994). A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment. *Reliability Engineering & System Safety*, 46(3), pp. 253-261.

Pérot, N. (2020). *France Patent No. 20 10176.*

Pérot, N., Desnoyers, Y., Augé, G., Aspe, F., Boden, S., Rogiers, B., . . . de Groot, J. (2019). *WP3 –Sampling strategy: state-of-the-art report.* EC H2020 INSIDER project, http://insider-h2020.eu/, Deliverable D3.1.

Plischke, E., Borgonovo, E., & Smith, C. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3), pp.536-550.

Sacks, J., Schiller, S., & Welch, W. (1989). Designs for computer experiments. *Technometrics*, 1(1), pp. 41-47.

Saltelli, A., & Marivoet, J. (1990). Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. *Reliability Engineering & System Safe*, 28(2), pp. 229-253.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., . . . Tarantola, S. (2008). Global sensitivity analysis: the primer. *robability and Statistics. John Wiley & Sons, Chichester NY.*

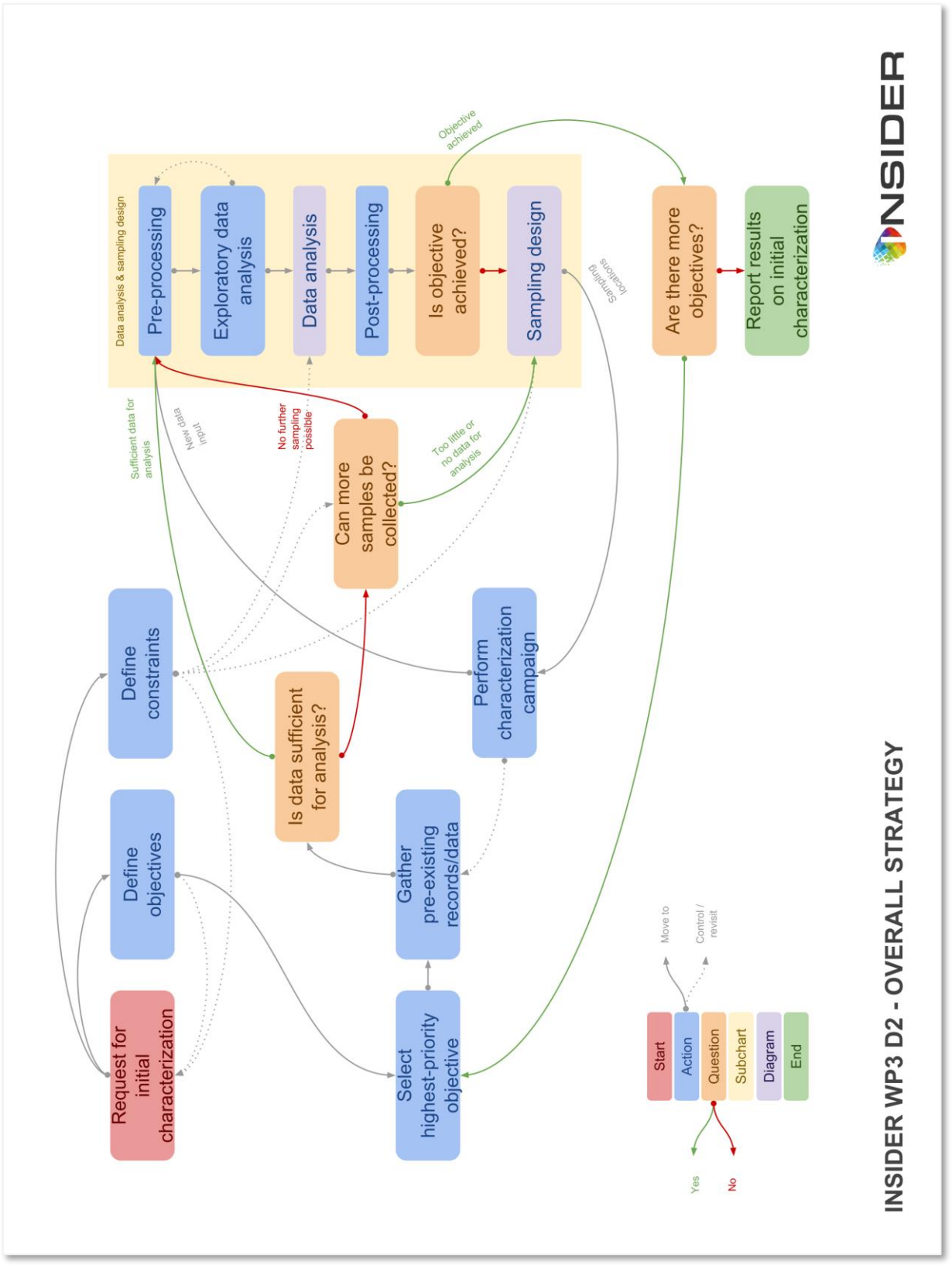
Saltelli, A., Tarantola, S., & Chan, K. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1), pp. 39-56.

Shao, Q., Younes, A., Fahs, M., & Mara, T. (2017). Bayesian sparse polynomial chaos expansion for global sensitivity analysis. *omputer Methods in Applied Mechanics and Engineering*, 318, pp. 474-496.

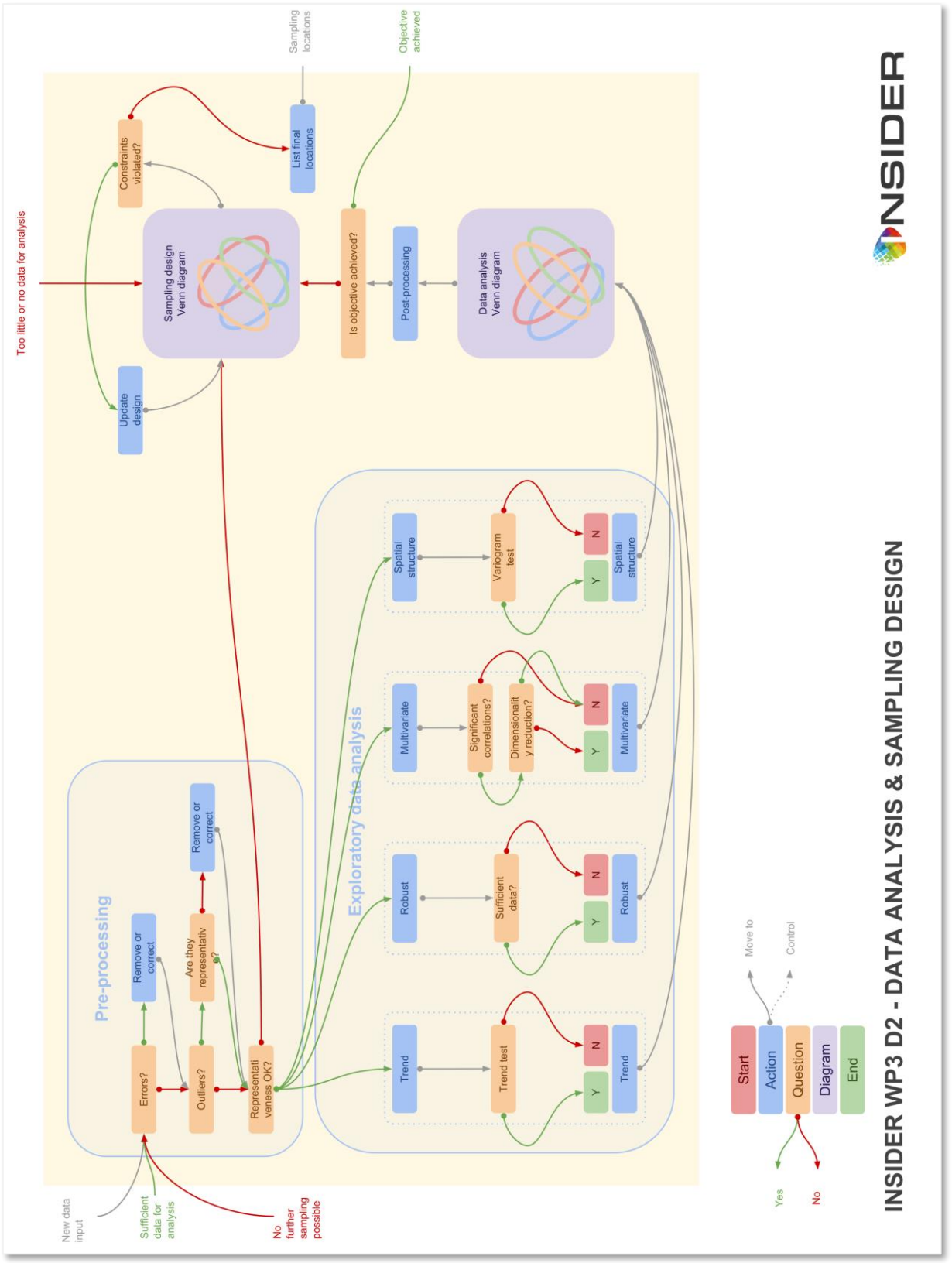
Sobol', I. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1, pp. 407-414.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7), pp. 964-979.

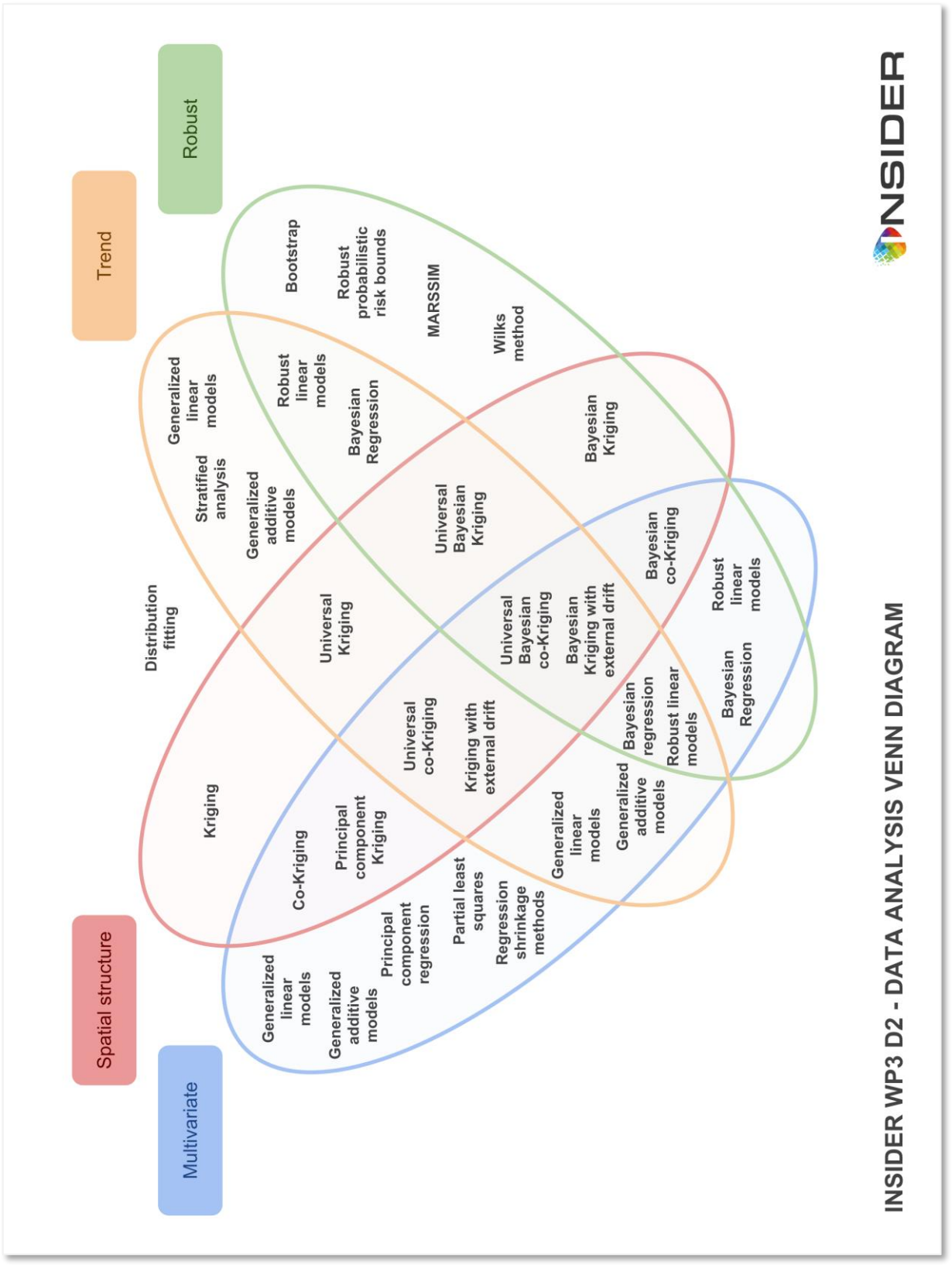
Annex A: Overall strategy flow chart



Annex B: Data analysis & sampling design flow chart



Annex C: Data analysis Venn diagram



Annex D: Sampling design Venn diagram