



**HAL**  
open science

## Statistical characterization of radiological contamination with small data samples

B. Iooss, N. Perot, T. Delage

► **To cite this version:**

B. Iooss, N. Perot, T. Delage. Statistical characterization of radiological contamination with small data samples. DEM 2018, Oct 2018, Avignon, France. hal-02415480

**HAL Id: hal-02415480**

**<https://hal.science/hal-02415480v1>**

Submitted on 17 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Statistical characterization of radiological contamination with small data samples

Bertrand Iooss<sup>1\*</sup>  
Nadia Pérot<sup>2</sup>  
Thibault Delage<sup>1</sup>

<sup>1</sup>EDF R&D, 6 Quai Watier, 78401 Chatou, France  
<sup>2</sup>CEA Cadarache, 1318 Saint-Paul-lez-Durance, France  
Main Author, bertrand.iooss@edf.fr

## Abstract

The radiological characterization of contaminated elements (walls, grounds, objects) from nuclear facilities often suffers from too few measurements. In order to determine risk prediction bounds on the level of contamination, some classic statistical methods may therefore be unsuitable, as they rely upon strong assumptions (e.g., that the underlying distribution is Gaussian) which cannot be verified. Considering that a set of measurements or their average value come from a Gaussian distribution can sometimes lead to erroneous conclusions, possibly not sufficiently conservative. This paper presents several alternative statistical approaches which are based on much weaker hypotheses than the Gaussian one, which result from general probabilistic inequalities. Given a data sample, these inequalities make it possible to derive prediction intervals for a random variable which can be directly interpreted as probabilistic risk bounds. The proposed methods are applied to several data sets obtained from real radiological contamination measurements to provide more realistic risk bound estimation.

**KEYWORDS:** radioactive waste characterisation, small data sample, statistical analysis.

## Introduction

In nuclear engineering, as in most of the industrial domains, one often faces up with difficult decision making processes, especially when safety issues are involved. In order to consider in a rigorous and coherent way the uncertainties associated to the models, input conditions, measurements and sampling in a decision process, the probabilistic framework offers an invaluable help.

In particular, a non-exhaustive sampling of a population of objects induces some uncertainties that have to be known in order to control their effects. In this communication, we are concerned by the radiological characterization of elements (walls, grounds, objects) from nuclear facilities whose contamination is not spatially structured (a geostatistical approach is then useless) and which are based on a small number of measures (from a few to a few tens). Indeed, the characterization of contaminated elements in a nuclear facility may be difficult because of practical and/or strong operation constraints, often limiting the number of possible measurements. Nevertheless, the estimation of radioactivity levels is essential to assess the risk of exposure of nuclear dismantling operators, as well as the risk of environmental contamination (1).

## Small data sample and risk bound estimation

In the small-size sampling cases, the classical statistical tools, which are based on the hypothesis of a Gaussian distribution for the value of a measure, can lead to erroneous conclusions. This paper aims to present and illustrate several alternative statistical tools, called concentration inequalities, which are applicable without making strong hypotheses on the law of probability of the studied variable (2). This kind of tools are particularly relevant in the context of a demonstration before a safety authority. From a sample of data, these concentration inequalities allow to obtain the three different classical statistical intervals (3):

- The confidence intervals for the estimation of the mean (or other probability distribution

parameters) of a random variable. For example, we can determine the size of the set of measurements to make in order to reach a given precision for calculating the average value of various contamination measures. This process allows us to optimize the sampling strategy and offers invaluable economic gains;

- The prediction intervals for a random variable. For example, we can compute the probability that the value of a point contamination is larger than a given critical value. In practice, regulatory threshold values are set for different waste categories. Determination of the probability that the contaminant's value is smaller than a given threshold can be used to predefine the volumes of waste by category.
- The tolerance intervals, which extend the prediction intervals to take into account uncertainty in the parameters of a variable's distribution. A tolerance interval gives the statistical interval within which, with some confidence level, a specified proportion of a sampled population falls.

The main principles of the concentration inequalities will be given. Some numerical tests will show their robustness and their ability to avoid unnecessary conservatism. Indeed, there exists a catalog of inequalities based on a series of hypotheses, from the least restrictive to the most restrictive.

Their practical use will also be demonstrated on several real-cases, as an ensemble of radioactive waste drums which are concerned by a radiolysis reaction. The evaluation of the hydrogen flow rate is then required for their disposal in final waste repositories, while only a small-size sample has been measured.

## Conclusion

In summary, the concentration inequalities allow to determine unilateral intervals, by defining a limit value that a variable cannot exceed (or reach, depending on the context) with a given probability level and a confidence level. This can then be used to estimate, on the basis of a small number of contaminant measures, the quantity of contaminant which does not exceed a safety threshold value. These elements are of first importance for the preliminary categorization assessment of nuclear waste in view of cost-effectively plan and manage a decommissioning project.

## References

- 1) CETAMA, *Soil radiological characterisation methodology*, CEA-R-6386, Commissariat à l'Energie Atomique et aux Energies Alternatives, Direction de l'Energie Nucléaire, Marcoule (2014)
- 2) G. Blatman, T. Delage, B. Iooss and N. Pérot, "Probabilistic risk bounds for the characterization of radiological contamination", *The European Journal of Physics - Nuclear Sciences & Technology (EPJ-N)* 3, 23 (2017)
- 3) G.J. Hahn and W.Q. Meeker, *Statistical intervals. A guide for practitioners*, Wiley-Interscience (1991)