

Statistical sampling applied to the radiological characterization of historical waste

Biagio Zaffora^{1,*}, Matteo Magistris¹, Gilbert Saporta², and Francesco Paolo La Torre¹

¹ CERN, 1211 Geneva 23, Switzerland

² CEDRIC-CNAM, 292 Rue Saint-Martin, 75003 Paris, France

Received: 18 December 2015 / Received in final form: 23 May 2016 / Accepted: 26 July 2016

Abstract. The evaluation of the activity of radionuclides in radioactive waste is required for its disposal in final repositories. Easy-to-measure nuclides, like γ -emitters and high-energy X-rays, can be measured via non-destructive nuclear techniques from outside a waste package. Some radionuclides are difficult-to-measure (DTM) from outside a package because they are α - or β -emitters. The present article discusses the application of linear regression, scaling factors (SF) and the so-called “mean activity method” to estimate the activity of DTM nuclides on metallic waste produced at the European Organization for Nuclear Research (CERN). Various statistical sampling techniques including simple random sampling, systematic sampling, stratified and authoritative sampling are described and applied to 2 waste populations of activated copper cables. The bootstrap is introduced as a tool to estimate average activities and standard errors in waste characterization. The analysis of the DTM Ni-63 is used as an example. Experimental and theoretical values of SFs are calculated and compared. Guidelines for sampling historical waste using probabilistic and non-probabilistic sampling are finally given.

1 Introduction

The evaluation of the activity of the radionuclides in radioactive waste is required for its disposal in final repositories. The characterization of radioactive waste includes establishing the list of radionuclides, together with their specific activity, inside each package.

For historical waste, which is defined as waste collected before the implementation of a traceability system [1], the radiological characterization process is complex. This is due to limited or missing information about the radiological history of the waste. Some of the radionuclides are easy-to-measure (ETM) from outside the waste package by means of nuclear non-destructive assay, such as γ -spectrometry. Other radionuclides, such as pure- β , α and low-energy X-rays, are difficult-to-measure (DTM) or impossible-to-measure (ITM) by non-destructive techniques. When an experimental statistical correlation can be established between an ETM and DTM radionuclides, the scaling factor (SF) method can be applied to quantify the specific activity of DTMs [2]. The scaling factor method consists of evaluating the activity of a radionuclide by applying a multiplicative factor (the so-called “scaling factor”) to the activity of the dominant gamma emitter. ETM radionuclide statistically correlated to a DTM is defined the tracer or the key nuclide (KN).

A statistical correlation can be checked only if the sampling technique adopted is probabilistic. In the present article, we introduce various techniques, including simple random, systematic and stratified sampling, to estimate average specific activity of Ni-63 on copper shreds from power and signal cables activated at CERN.

Section 2 describes the SF method, the sampling techniques tested, the resampling technique called bootstrap, measurement and calculation tools for activity quantification. Section 3 presents the waste populations used to validate and compare statistical methods for sampling. Section 4 presents the implementation of the experiments, the calculations performed and the comparison of the various techniques. Conclusions are finally given in the last section.

2 Methods

2.1 Scaling factors, linear regression and mean activity method

The scaling factor method is described in references [1,2]. Its applicability can be checked by either studying the production mechanisms of the radionuclides and by observing their correlation or by using statistical methods. For historical waste it is often impossible to check the activation conditions of materials and, consequently,

* e-mail: biagio.zaffora@cern.ch

production mechanisms. Only statistical correlations can therefore be tested, based on experimental data obtained from a sample.

When measurements of DTMs and a KN are performed, the scaling factor SF_i for the i th pair DTM/KN is given by:

$$SF_i = \frac{a_{\text{DTM},i}}{a_{\text{KN},i}}, \quad (1)$$

where $a_{\text{DTM},i}$ is the specific activity of the DTM in the i th sample (in Bq/g) and $a_{\text{KN},i}$ is the specific activity of the KN in the i th sample (in Bq/g). If many samples are collected from a waste population the distribution of the SFs can be calculated together with the correlation r of the random variables a_{DTM} and a_{KN} . Only values of activity above the detection limit should be used.

Based on the strength of the correlation r , different methods can be used to evaluate the activity of the DTM nuclides. For the present study we considered linear regression, mean and geometric mean of the scaling factors and the so-called ‘‘mean activity method’’.

The general equation of the linear model between the activities of the pair of radionuclides DTM and KN is:

$$a_{\text{DTM}} = \beta_0 + \beta_1 \times a_{\text{KN}}, \quad (2)$$

where β_0 and β_1 are respectively the intercept and the slope of the regression line. The hypothesis $\beta_0 = 0$ is often considered [2]. In this case β_1 represents the scaling factor that, multiplied by the activity of the KN, allows us to estimate the activity of the DTM nuclide. The validity of the linear model can be checked using the p-value for parameter importance and the F-statistic for appreciation of the overall model.

A second technique to estimate the scaling factor is based on the hypothesis that the underlying distribution of SFs is often log-normal. If scaling factors are log-normally distributed, the geometric mean \overline{SF} is a robust central tendency estimator:

$$\overline{SF} = e^{\left(\frac{\sum_{i=1}^n \ln(SF_i)}{n} \right)}, \quad (3)$$

where SF_i is given by equation (1) and n is the number of units in the sample collected.

The geometric standard deviation around the geometric mean, called dispersion D , can be calculated as follows:

$$D = e^{\left(\sqrt{\frac{\sum_{i=1}^n [\ln(SF_i) - \ln(\overline{SF})]^2}{n-1}} \right)}. \quad (4)$$

The IAEA technical report in reference [2] suggests that, for the geometric mean to be applicable, the coefficient of determination R^2 should be above 0.5. If the distribution of SFs is approximately normal the mean scaling factor should be used.

Finally, if a statistical correlation between DTMs and KN is not found, the so-called ‘‘mean activity method’’ can be applied. This technique consists of calculating the arithmetic mean activity of each DTM nuclide from a

sample, including values which are below the detection limit DL. The mean value so found is applied to the entire population. It must be stressed however that the use of the arithmetic mean can be biased, especially when the activity distribution is skewed. This is particularly true when more robust average content estimators (such as median or geometric mean) are considered. A detailed description of these methods and practical applications will be given in the following sections.

2.2 Sampling techniques

2.2.1 Simple random and systematic sampling

In most practical situations census data, which are data of all the units in a population, are impossible or too expensive to collect. Simple random (SRS) and systematic sampling (SYS) are often used to collect samples in order to estimate the true value of a parameter of a population. A complete mathematical treatment of these sampling techniques can be found in references [3,4].

In SRS each member of the population has an equal probability of being included in the sample. In practice, the units of the population are numbered from 1 to N . A series of random numbers between 1 and N is drawn without replacement. The sampling units associated to the random numbers drawn are selected for sampling.

SRS can be impractical when sampling radioactive waste because not all the units of a population are necessarily accessible during the sampling campaign. SYS is often used instead.

SYS is a statistical process that allows the analyst to choose n samples over a population of N units, with samples spaced by a factor k . If the N units of the population are numbered between 1 and N and n samples must be collected, k is calculated as the ratio N/n . A random sample between 1 and k and then every k th unit thereafter are taken. SYS may be affected by the order of the sampling units in the population file but is very practical in a continuous industrial production of packages of radioactive waste.

2.2.2 Multi-stage stratified sampling

In stratified sampling the population of N units is divided into non-overlapping subpopulations of N_1, N_2, \dots, N_L units, called strata. A sample is then randomly selected from each stratum.

If multiple samples can be collected from each sampling unit (a waste package for instance), we can apply a second sampling stage that allow us to select secondary samples from the units of each stratum. This strategy is called 2-stage stratified sampling and is a special case of the so-called ‘‘multi-stage stratified sampling’’.

A common strategy to chose the number of samples n_h to collect per single stratum h is the Neyman allocation [3]:

$$n_h = \frac{nw_h s_h}{\sum_{h=1}^L w_h s_h}, \quad (5)$$

where n is the total number of samples to collect, w_h is the weight of the stratum h , s_h is the standard deviation of the population parameter to quantify (such as the

specific activity) in the stratum h and L is the number of strata. The standard deviation s_h on a stratum can be estimated from previous studies and conservative hypothesis can also be used.

Equation (5) states that more samples must be collected in strata with a higher weight or a higher dispersion. For waste characterization this implies that more samples should be collected in strata where the activity is higher and the dispersion of data is highly variable.

Once the number of samples to collect per stratum is calculated, we can use SRS to chose samples into each stratum. When using 2-stage stratified sampling and the strata have different sizes, an unbiased estimator of the average specific activity \bar{a} of a radionuclide is [4]:

$$\bar{a} = \frac{\sum_{h=1}^L N_h \bar{a}_h}{\sum_{h=1}^L N_h}, \quad (6)$$

where N_h is the number of primary units in the stratum h and \bar{a}_h is the average specific activity calculated from the samples of the stratum h . A detailed mathematical treatment of stratified sampling can be found in reference [3].

2.2.3 Authoritative sampling

Authoritative sampling is a non-statistical sampling design because it does not assign an equal probability of being sampled to all portions of the population.

Authoritative sampling may be appropriate under the following circumstances:

- preliminary information is needed about the waste or site to facilitate planning or to gain familiarity with the waste matrix for analytical purposes;
- only a small portion of the population is accessible and judgement is applied to assess the usefulness of samples drawn from the small portion;
- extremes values are searched for the calculation of the worst case scenario.

In the present study, we used the so-called judgemental sampling [5], which is a type of authoritative sampling, to estimate preliminary standard deviations needed for the stratified sampling and to estimate extreme values. More information on the application of authoritative sampling is given in Sections 3 and 4.

2.3 The bootstrap

The bootstrap is a resampling method that can be used to estimate the (unknown) distribution of a parameter θ of a population, such as the average specific activity of a radionuclide in a radioactive waste batch. When a sample of n units is withdrawn from a population, a high number of replicates of the sample are generated via sampling with repetition from the original sample. For each replicate, also of size n , we calculate the bootstrap parameter $\hat{\theta}^*$ which is an estimation of the true population parameter θ . The population parameter calculated from the sample is indicated with $\hat{\theta}$ [6,7]. With this technique, instead of

evaluating via a single value the parameter θ , we construct an experimental distribution for the same parameter which is otherwise unknown.

The bootstrap is commonly used to estimate mean, median, standard error, confidence intervals and bias.

We applied this computation technique to evaluate the specific activity of DTM nuclides and to estimate the standard deviation in stratified sampling.

2.4 Measurements techniques

Techniques for γ -ray detection and for activity quantification of γ -emitters are well known and documented in many references, such as in [8–10]. In the present study, two classes of instruments are proposed for the quantification of the activity of ETMs, namely total- γ counters and γ -spectrometry detectors. The first class of counters is mainly used for the quantification of the specific activity of waste packages. The second class of detectors is used for a more precise measurement of the ETMs specific activity. In particular, the activity measurements of γ -emitters for SF estimation are carried out using γ -ray spectrometers.

At CERN, two total- γ counters are currently in use: the first counter consists of 6 detectors in a 4π geometry with internal volume 0.44 m^3 and 50 mm of lead shielding and the second counter consists of an array of 24 detectors in a 4π geometry with internal volume 1.82 m^3 and 70 mm of lead shielding. For both instruments the counting time is very short (generally below 5 min) and the measurable γ -activities can reach $\sim 10^{-4} \text{ Bq/g}$. For the present study a fingerprint 100% Co-60 was used, which means that each photon collected by the counter was considered as emitted by a Co-60 nucleus. Detailed information on the calibration of total- γ counters can be found in [11].

The second class of instruments, based on Germanium technology, is used to perform γ -ray spectrometry either for low background or in-situ measurements. Several γ spectrometers, cooled either electrically or by liquid nitrogen, are presently used at CERN. Their relative efficiency for the Co-60 at 1.33 MeV ranges from 30% up to 60%.

The specific activity of pure β -emitters is evaluated via radiochemical analysis performed on samples. The β -emitters are defined DTM [1] because their quantification requires complex multi-stage techniques involving acid digestion, separation, filtration through resins or columns and measurement. A complete description of the chemical treatment of samples can be found in [12]. The description of the liquid scintillation technique, used for the measurement of the activity of DTMs is given in [10].

Common values of the detection limits for the DTMs considered in the present study are in the range $0.1\text{--}0.5 \text{ Bq/g}$.

2.5 Simulation codes

Actiwiz is a software developed at CERN to build a radiological hazard assessment for an arbitrary material exposed to the radiological environment of the accelerator complex [13,14]. The application was developed to give

quick answers to general questions about radiological hazards without the need for the user to implement complex input files with a Monte Carlo code such as FLUKA [15,16].

The developers have run thousands of FLUKA simulations [15,16] of nuclide inventories on different materials for 42 typical hadronic spectra and for various positions inside the accelerators' tunnels. The results of these simulations are stored as a database in Activiz [13,14] and the user can run calculations on predefined simulated scenarios.

The radiological environments available for calculations represent all the accelerators in CERN's complex and include the Linac4 (160 MeV), the PS Booster (1.4 GeV), the PS (14 GeV/c), the SPS (450 GeV/c) and the LHC (7 TeV).

Amongst the information obtained by running Activiz [13,14], the interest for the present study lies mainly in the establishment of expected radionuclide inventories and calculation of theoretical scaling factors. The radionuclide inventory is defined as the complete list of radionuclides, together with their activity, produced by activation of a given material.

For the present study, extensive Activiz [13,14] calculations were carried out using the chemical composition of copper CuOFE from CERN's material catalogue [17]. This composition was exposed to typical CERN irradiation scenarios. The traces present as weight fraction in copper CuOFE are bismuth (0.1%), cadmium (0.01%), lead (0.1%), mercury (0.01%), oxygen (0.05%), sulfur (0.18%) and zinc (0.01%). The balance is copper. The results of the calculations performed with these tools are presented in Section 4.1.

3 Waste populations

We identified 2 populations of low-level radioactive copper to test the methods introduced in Section 2. These populations consist of copper cables dismantled from CERN's different installations. The cables' core was shredded and separated from the insulating layers with the purpose of diminishing their heterogeneity. In the following sections the 2 waste populations are indicated as campaign 1 and campaign 2.

3.1 Campaign 1

A summary of the main information describing the waste population of campaign 1 is given in Table 1. The shredded copper is collected in drums which represent the primary sampling units. Each drum was measured via total- γ counting and the summary statistics of the specific activity of the key nuclide Co-60 are given. In the following sections we use SE to indicate the standard error of the mean (which is the ratio of the standard deviation and the square root of the sample size) and I.Q. for the interquartile range (difference between the 75th and 25th percentiles).

The waste population of campaign 1 consists of 87 drums. Each secondary sample taken from a drum is considered representative of the entire drum. This hypothesis can be

Table 1. Summary statistics of the specific activity of the key nuclide Co-60 for campaign 1.

Number of drums	87
Mean weight per drum (kg)	98
Total weight (kg)	8522
Mean $a_{\text{Co-60}}$ in Bq/g	0.044
SE of $a_{\text{Co-60}}$ at $k=1$ in Bq/g	0.004
Median $a_{\text{Co-60}}$ in Bq/g	0.039
I.Q. range in Bq/g	0.026
Minimum $a_{\text{Co-60}}$ in Bq/g	0.0046
Maximum $a_{\text{Co-60}}$ in Bq/g	0.31

Table 2. Summary statistics of the specific activity of the key nuclide Co-60 for campaign 2.

Number of drums	229
Mean weight per drum (kg)	97
Total weight (kg)	21,487.7
Mean $a_{\text{Co-60}}$ in Bq/g	0.095
SE of $a_{\text{Co-60}}$ at $k=1$ in Bq/g	0.007
Median $a_{\text{Co-60}}$ in Bq/g	0.059
I.Q. range in Bq/g	0.12
Minimum $a_{\text{Co-60}}$ in Bq/g	0.0002
Maximum $a_{\text{Co-60}}$ in Bq/g	0.58

made because multiple samples were collected from each drum, mixed and composited into a final representative sample.

As further discussed in Section 4, we use the population of campaign 1 to compare the specific activity of the DTM Ni-63 from census data with estimations obtained applying SRS, SYS and the bootstrap. The comparison is performed on both specific and total activity of Ni-63.

3.2 Campaign 2

The preliminary information available for campaign 2 is given in Table 2. As for campaign 1, each drum of campaign 2 was measured via total- γ counting and a statistical summary of the activity of Co-60 is given.

We applied multi-stage stratified sampling to select samples for the estimation of Ni-63 content. As discussed in Section 2.2.2, when this technique is used, we need a preliminary estimation of the standard deviation to calculate the number of samples per stratum, as in equation (5). Within this frame, we used 13 authoritative samples on activated high-dose copper cables and measured the content of Ni-63 via radiochemical analysis. A summary of the results is presented in Table 3.

Campaign 2 consists of 229 drums of shredded copper. Each drum is a sampling unit from which we can withdraw secondary samples. Multi-stage stratified sampling techniques allows us to take into account the

Table 3. Summary statistics of Ni-63 via analysis of 13 authoritative samples.

Mean $a_{\text{Ni-63}}$ in Bq/g	1.28
SE of $a_{\text{Ni-63}}$ at $k=1$ in Bq/g	0.71
Median $a_{\text{Ni-63}}$ in Bq/g	0.1
I.Q. range in Bq/g	0.44
Minimum $a_{\text{Ni-63}}$ in Bq/g	<0.1
Maximum $a_{\text{Ni-63}}$ in Bq/g	8.82

Table 4. List of potential key nuclides for low-level radioactive copper, type CuOFE [17].

Key nuclide	Half-life (y)	Main γ -emitters
Na-22	2.603	1275 keV
Ti-44	58.9	1157 keV (from Sc-44)
Co-60	5.2711	1173 keV, 1332 keV

potential variations of the activity of Ni-63 (within a sampling unit) when no prior information is available on the heterogeneity of a drum.

4 Simulations and experimental results

In this section, we present the results from Actiwiz calculations and from the measurements of Ni-63 performed on the collected samples.

4.1 Activation studies

To consider a comprehensive amount of activation scenarios we simulated the irradiation of copper CuOFE [17] on all the scenarios available in Actiwiz, using 17 irradiation times (from 0.25 up to 30 years) and 16 decay times (from 1 up to 30 years). The total number of scenarios studied is 11,424.

We used these calculations to establish the radionuclide inventory for the 2 waste populations considered, to identify potential key nuclides and to calculate preliminary, theoretical scaling factors.

A non-comprehensive list of radionuclides obtained from Actiwiz simulations includes H-3, C-14, Na-22, Ca-41, Ti-44, Mn-54, Fe-55, Co-57, Co-60, Ni-63 and Zn-65. Amongst these radionuclides, only a limited number respect the criteria for being selected as a key nuclide, following the indications of [18]. Some properties of the potential KNs for the characterization of shredded copper cables are given in Table 4.

Ti-44, whose main γ lines (68 keV and 78 keV) are difficult to use to estimate its activity (mainly due to multiple interferences with naturally occurring radionuclides) is quantified via measurement of its daughter's γ -line, the Sc-44 ($E_\gamma = 1157$ keV).

For the present study, we chose Co-60 as a key nuclide when carrying out the calculations. This choice is justified by the systematic detection of Co-60 in each single drum and samples from both campaigns.

With respect to DTM nuclides, the present study focuses on Ni-63. Measurements of H-3 and Fe-55 were also performed but the value of their activity was often below the detection limit and could not be used to evaluate scaling factors. We illustrate the estimation of Ni-63 as an example. The specific activity of other DTM nuclides can be estimated either by the mean activity method or by calculation.

Figure 1 shows the distributions of the logarithm of Ni-63 and Co-60 activities and the distribution of the logarithm of their ratios (theoretical scaling factors). The histograms summarize the results obtained from the 11,424 irradiation scenarios considered.

As can be seen in Figure 1, the log-transformed activity of both Ni-63 and Co-60 shows a normal distribution. Moreover Ni-63 and Co-60, respectively DTM and KN, have similar production mechanisms when activating copper at hadron accelerators. In particular, nuclear reactions of the type (n, p) or (γ , pn) are responsible for the production of Ni-63 from naturally occurring isotopes of copper, such as Cu-63 and Cu-65. Similar reactions are responsible for the production of Co-60 from copper via the intermediate production of nickel isotopes. Spallation mechanisms can also be involved.

The summary statistics of the theoretical SFs obtained by calculation are given in Table 5. The dispersion (see Eq. (4)) is a multiplicative term and therefore dimensionless.

4.2 Sampling and analysis of campaign 1

4.2.1 Sampling and results

The sampling strategy of the waste of campaign 1 represents the uncommon case of census because a sample per drum was collected. Furthermore each sample is considered as being representative of the entire drum because it was collected by compositing multiple sub-samples, from different layers of a given sampling unit, into the final sample. We measured the specific activity of Ni-63 on 87 units and compared the results from census with the results from SRS, SYS and bootstrap estimation (Boot).

Of the 87 samples collected, 23 have Ni-63 specific activity below the detection limit. The calculations for the average amount of Ni-63 were performed twice, with and without values below the DLs. The relative error ε associated to each technique is calculated as follows:

$$\varepsilon = \frac{\bar{a} - \bar{a}_{\text{census}}}{\bar{a}_{\text{census}}}, \quad (7)$$

where \bar{a} is the average specific activity of the Ni-63 calculated by SRS, SYS or Boot and \bar{a}_{census} is the average specific activity of Ni-63 from census data.

Table 6 summarizes the results obtained using the different statistical sampling techniques. The first column represents the sampling strategy used. For example, the first line after census indicates that SRS was used and that 5% of the units were selected for sampling. This means that 4 samples were selected from the population that includes values below the DL (5%, $n=4$) and 3 samples were selected from the population that does not include values

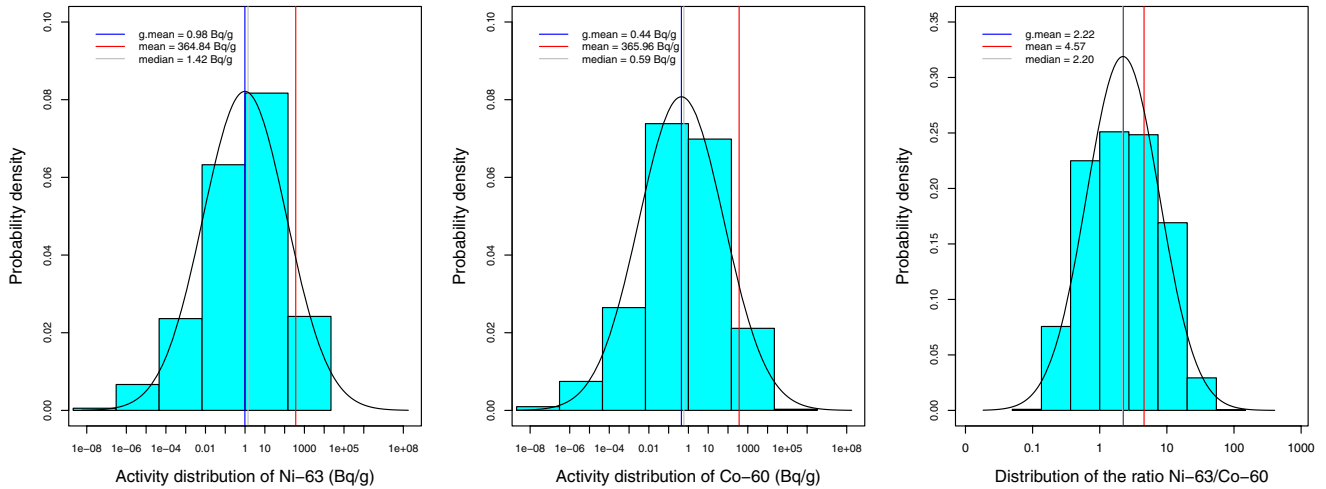


Fig. 1. Distributions of the logarithms of Ni-63 and Co-60 specific activities and distribution of the logarithm of their ratio obtained by simulating the irradiation of copper CuOFE [17] for common CERN scenarios. The average content statistics geometric mean, mean and median are represented together with the normal curve. The x -axes are in log-scale and the distributions are approximately log-normal.

Table 5. Summary statistics of the calculated, theoretical scaling factors of Ni-63 and Co-60 for copper CuOFE.

Mean	4.57
SE ($k = 1$)	0.056
Median	2.2
I.Q. range	5.15
Geometric mean	2.22
Dispersion ($k = 1$)	3.49
Minimum	0.11
Maximum	74.03

below the DL (5%, $n = 3$). For the bootstrap, n and N represent respectively the number of values that are considered for each sampling and the total number of times that a resampling occurs.

For the population waste of campaign 1, correlation between the activities of Ni-63 and Co-60 cannot be established (the correlation coefficient is 0.27). The estimation of the concentration of Ni-63 is performed using the mean activity method (see Sect. 2.1). For SRS and SYS methods, the specific activity of Ni-63 is calculated as the average of samples measurements extracted from census data. The total amount is calculated as the product of the specific activity and the total weight of the batch. For the bootstrap, the average activity of Ni-63 is estimated as the average of the N sampling extractions (with repetition) from census data, using n samples. The test is repeated for $n = 5, 10$ and 20 and $N = 250, 500$ and 1000 .

4.2.2 Analysis and discussion

The content of Ni-63 from census data is compared to the average content estimated using SRS, SYS and bootstrap.

When SRS is considered, more than 25% of the population must be sampled to achieve a relative error below 10%. The maximum relative error found was 47%

(SRS of 10% of the population with values below DL). A large sample is needed to obtain a content of Ni-63 close to the true value.

SYS performed was efficient in predicting Ni-63 content. The relative error for 9 out of 10 scenarios considered is below 14%. In 5 scenarios the relative error is below 5%.

For the present study, we considered 18 bootstrap scenarios. We found that when the number of repetitions is $N = 250$, only the case of $n = 20$ has a relative error below 6.5%. For resampling number $N = 500$, the relative error is systematically below 25%. In 10 out of 18 scenarios considered, the relative error is below 10% and 3 of these scenarios (obtained using $N = 1000$) have a null relative error.

The bootstrap technique can be considered as a complementary way of calculating average content estimators of the activity for DTM nuclides. The bootstrap performs better when a robust sampling technique is used.

From Table 6 it seems that SYS performs better than SRS. The relative errors are nevertheless obtained from a specific random process and repeating the experiment with different random numbers could generate different results. The difference between the 2 sampling techniques diminishes when the number of samples increases. SYS is however easier to implement in practice, especially when similar sampling units must be sampled. This is the case for example for drums with the same weight containing particulate waste with similar chemical and physical properties. For a limited-size batch of waste, with low-heterogeneous characteristics, differences between SRS and SYS sampling should not be expected.

The bootstrap predicts very well the true activity of Ni-63, especially when data without values below DL is used. Increasing the number of repetitions is also useful to lower the relative error. This technique can be used to increase the confidence of calculated average content estimators (such as the mean and the median specific activity) for data samples of medium or low size. We recall here that the standard error of the bootstrap mean is simply the standard deviation of the distribution of the bootstrap mean.

Table 6. Average and total content of Ni-63 calculated for the population of 87 drums using simple random sampling (SRS), systematic sampling (SYS) and bootstrap (Boot). The relative error ε is calculated twice, with and without values below DL.

Sampling techniques	Population with DL			Population without DL			ε	
	$a_{\text{Ni-63}}$ (Bq/g)	SE (Bq/g)	Tot. Ni-63 (MBq)	$a_{\text{Ni-63}}$ (Bq/g)	SE (Bq/g)	Tot. Ni-63 (MBq)	With DL	Without DL
Census	0.53	0.04	~4.52	0.65	0.05	~5.54	–	–
SRS (5%, $n=4$), (5%, $n=3$)	0.705	0.07	~6	0.53	0.13	~5.37	33.02%	–18.46%
SRS (10%, $n=9$), (10%, $n=6$)	0.28	0.05	~2.39	0.58	0.07	~5.28	–47.17%	–10.77%
SRS (25%, $n=22$), (25%, $n=16$)	0.52	0.11	~4.43	0.69	0.12	~5.37	–1.9%	6.15%
SRS (50%, $n=44$), (50%, $n=32$)	0.5	0.04	~4.26	0.68	0.05	~5.88	–5.7%	4.61%
SYS (5%, $n=4$), (5%, $n=3$)	0.65	0.22	~5.54	0.66	0.13	~5.6	22.64%	1.54%
SYS (10%, $n=9$), (10%, $n=6$)	0.48	0.06	~4.09	0.65	0.08	~5.54	–9.43%	0
SYS (25%, $n=22$), (25%, $n=16$)	0.46	0.06	~3.92	0.57	0.04	~4.86	–13.21%	–12.31%
SYS (50%, $n=44$), (50%, $n=32$)	0.56	0.07	~4.77	0.71	0.08	~6.05	5.66%	9.23%
Boot ($n=5$, $N=250$)	0.33	0.08	~2.81	0.51	0.07	~4.36	–37.74%	–21.23%
Boot ($n=10$, $N=250$)	0.75	0.21	~6.39	0.6	0.05	~5.11	41.51%	–7.69%
Boot ($n=20$, $N=250$)	0.54	0.08	~4.6	0.61	0.05	~5.2	1.89%	–6.15%
Boot ($n=5$, $N=500$)	0.415	0.11	~3.54	0.56	0.22	~4.77	–21.7%	–13.84%
Boot ($n=10$, $N=500$)	0.665	0.11	~5.67	0.64	0.04	~5.45	25.47%	–1.54%
Boot ($n=20$, $N=500$)	0.52	0.06	~4.43	0.6	0.07	~5.11	–1.89%	–7.69%
Boot ($n=5$, $N=1000$)	0.63	0.2	~5.39	0.65	0.13	~5.54	19.24%	0
Boot ($n=10$, $N=1000$)	0.4	0.04	~3.41	0.65	0.11	~5.54	–24.53%	0
Boot ($n=20$, $N=1000$)	0.53	0.07	~4.52	0.68	0.09	~5.79	0	4.61%
Min	0.28	–	2.39	0.51	–	4.36	0	0
Max	0.75	–	6.39	0.71	–	6.05	–47.17%	–21.23%

Table 7. Stratification of the waste population for campaign 1 based on the total- γ activity of Co-60. N_h represents the number of drums in the stratum h .

$a_{\text{Co-60}}$ in Bq/g	Stratum	N_h	Weight (kg)	Weight (%)	Tot. activity (kBq)	Tot. activity (%)
$a_{\text{Co-60}} \leq 0.01$	1	49	4743.8	22.08	27.3	1.37
$0.01 < a_{\text{Co-60}} \leq 0.1$	2	99	9289	43.23	425.2	21.38
$0.1 < a_{\text{Co-60}} \leq 0.3$	3	68	6259.6	29.13	1036	52.09
$a_{\text{Co-60}} > 0.3$	4	13	1195.3	5.56	500.4	25.16
Total	–	229	21,487.7	100	1988.9	100

4.3 Sampling and analysis of campaign 2

4.3.1 Sampling and results

For the waste population of campaign 2 we used 2-stage stratified sampling in order to concentrate the sampling effort on the strata of the population having higher total γ -activity. This means that the number of samples to collect in a stratum h is calculated using as weight w_h the total γ -activity in the stratum h (see Sect. 2.2.2). The details of the stratified population are given in Table 7. The total number of samples to collect (40) was fixed by project constraints.

An extra 24 samples were further collected and the robustness of stratified sampling was tested.

As previously discussed in Sections 2.2.2 and 3.2, we estimated the standard deviation s_h of Ni-63 in the stratum h , via the results from 13 authoritative samples. The 13 samples were split into 4 strata, following their γ -activity, and each s_h was calculated from the variance of Ni-63 activities in the stratum h . We also estimated the standard deviation via the bootstrap technique obtaining comparable values.

Using equation (5) we calculated the number of samples n_h to collect in each stratum h . The results are presented in Table 8. N_h is the number of drums in the stratum h .

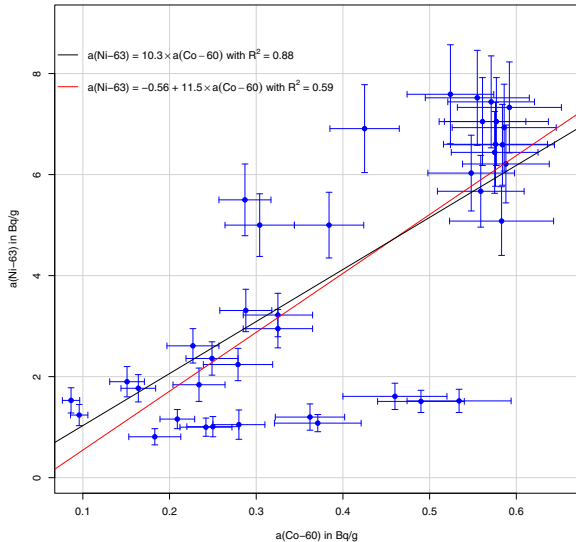
Once n_h was calculated, we randomly identified the drums from which to collect the samples. Due to the very low total- γ activity of stratum 1, no samples were collected

Table 8. Strata and number of samples per stratum chosen for the characterization of the population of campaign 2.

$a_{\text{Co-60}}$ in Bq/g	Stratum	N_h	n_h
$a_{\text{Co-60}} \leq 0.01$	1	49	–
$0.01 < a_{\text{Co-60}} \leq 0.1$	2	99	2
$0.1 < a_{\text{Co-60}} \leq 0.3$	3	68	4
$a_{\text{Co-60}} > 0.3$	4	13	34

Table 9. Summary statistics of the ratio Ni-63/Co-60.

Mean	9.96
SE ($k=1$)	0.7
Median	11
I.Q. range	5.86
Geom. mean	8.76
Dispersion ($k=1$)	1.75
Minimum	2.85
Maximum	19.16

**Fig. 2.** Scatterplot of Ni-63 vs. Co-60. The regression lines with and without intercept are represented in black and red.

in that sub-population (according to Eq. (5)). In strata 2 and 3, we collected 2 and 4 samples respectively. For stratum 4 the number of samples n_h is above the number of sampling units N_h . For this stratum we collected multiple samples from each drum. The samples were chosen randomly according to the rules of 2-stage sampling. In particular, we recall here that the copper waste is in the format of particulate material and that from a single drum we can identify up to 5000 different secondary samples (the mass of a sample for the radiochemical determination of Ni-63 is in the range 20–70 g).

Using equation (6) we calculated the stratified average specific activity $a_{\text{Ni-63}}^{\text{str}} = 0.98$ Bq/g (the standard error for $k=1$ is 0.095 Bq/g) and derived the total activity of Ni-63.

Table 10. Summary of average content of Ni-63 obtained applying different estimation techniques. C.I. stands for confidence interval for a standard error at $k=1$.

Method	Authoritative	Stratified	Geom. mean SF (8.76)	Mean SF (9.96)	Linear model ($\beta_0 = 0$)
Mean $a_{\text{Ni-63}}$ in Bq/g	1.28	0.98	0.83	0.94	0.97
SE ($k=1$) in Bq/g	0.71	0.095	0.063	0.072	0.06
Total Ni-63 in MBq	~27.5	~21.1	~17.4	~19.8	~20.5
C.I. total Ni-63 in MBq	[12.3, 42.8]	[19.0, 23.1]	[16.5, 19.2]	[18.7, 21.8]	[19.3, 21.7]

Once the samples were collected, we tested the applicability of both the linear model and the scaling factor to the relationship of Ni-63 and Co-60 activities (see Sect. 2.1).

The bivariate dispersion diagram of the pair Ni-63/Co-60 is shown in Figure 2. Two linear models were tested. In black is the regression line without intercept ($\beta_0 = 0$ in Eq. (2)). In red the regression line obtained with intercept. The amount of explained variance is 88% and 59% for the models with and without intercept respectively.

The estimation of the average and total activity of Ni-63 was also performed calculating the scaling factors as the mean and the geometric mean (see Eq. (3)) of the scaling factors for each pair Ni-63/Co-60, according to equation (1). Summary statistics of the scaling factor are given in Table 9. Table 10 shows the comparison of Ni-63 activities calculated using the different methods.

4.3.2 Analysis and discussion

Table 10 compares the values of specific and total activity of Ni-63 using 5 different methods. These methods can be separated into 2 classes:

- Authoritative and stratified sampling allow us to estimate an average content of Ni-63, which is identical for each single package of the batch.
- Geometric SF, mean SF and linear model allow us to estimate the specific activity of Ni-63 in each package, scaled by the activity of Co-60.

Authoritative and stratified sampling (as applied in the present study) are conservative methods because they tend to overestimate the concentration of Ni-63 either via measurements of high-dose judgemental samples (authoritative case) or sampling in the strata with higher Co-60 total-activity (stratified sampling).

The use of the geometric SF, the mean SF or the linear model depends on the distribution of the ratios of Ni-63/Co-60 activity. As a general rule, geometric SF should be preferred for right-skewed distributions and mean SF for approximately normal distributions.

With the exception of authoritative sampling, the methods suggested to estimate the activity of Ni-63 give a concentration which is within 1 standard deviation from the mean calculated over the 5 estimations. The concentration obtained from authoritative sampling lies within 2 standard deviations of the mean. A similar conclusion is reached using medians and inter-quartile ranges. The maximum relative error of Ni-63 concentration is found between the estimations from authoritative sampling and geometric SF (35%). Excluding authoritative sampling, the relative errors calculated between the considered methods is below 16% and, if the standard error is calculated at $k=2$, the confidence intervals include all the central tendency estimators of Ni-63.

Due to the variability of the authoritative samples the SE is very high in this last case. The confidence interval of the total activity of Ni-63 from authoritative sampling includes the confidence intervals obtained by applying any other of the methods considered here.

Judgemental sampling is a powerful method for DTM estimation when statistical sampling cannot be performed. Sampling with a non-probabilistic approach can be fast, cheap and a good indicator of extreme values. In the present study, the Ni-63 concentration estimated by authoritative sampling can be used as a conservative content estimator. If conservative values (such as the one obtained when sampling high dose rate samples) are available, it is possible to compare them with limits from regulations for waste elimination. A common practice in waste characterization consists of comparing extreme values with limits from waste management authorities. If extreme values respect these limits, it can be inferred that the entire population respects the limits.

The results obtained from stratified sampling show that the content of Ni-63 is very close to the concentrations obtained applying scaling factors and regression. The present study suggests that, when only a limited number of samples can be withdrawn from a population, it is possible to concentrate the sampling effort on the strata of the population with higher variability and activity. This technique also allows the selection of samples that can be used for scaling factor calculations.

Finally, to test the validity of the methods discussed, 24 complementary random samples of copper were collected on the left-over population. Summary statistics of Ni-63 activity from quality assurance samples are given in Table 11. The statistics of interest are calculated twice, with and without values below detection limits.

Amongst the 24 samples collected, 7 have a specific activity of Ni-63 below the detection limit (~ 0.2 Bq/g). The average concentrations of Ni-63 from test samples are below the activity of Ni-63 estimated by using both probabilistic and non-probabilistic techniques. This result was expected because the left-over population is characterized by a low γ activity.

Table 11. Summary statistics of Ni-63 activity in 24 random samples from the left-over copper population.

	With DL	Without DL
Mean $a_{\text{Ni-63}}$ in Bq/g	0.57	0.75
SE ($k=1$) in Bq/g	0.11	1.95
Median activity in Bq/g	0.32	0.54
I.Q. range in Bq/g	0.73	0.79
Minimum in Bq/g	<0.082	0.19
Maximum in Bq/g	1.95	1.95

Stratification and sampling in the strata of high activity are robust techniques to estimate mean activity values of DTMs and to avoid the collection of samples which have an activity below the detection limit. We recall here that 2.5% of the stratified samples were below DL and that $\sim 30\%$ of the samples withdrawn from the left-over population were below the DL.

4.4 Comparison of calculations and experiments

The summary statistics of the theoretical scaling factors obtained by Actiwiz calculation are shown in Table 5. The calculations can be compared with the experimental SFs from campaign 2. The estimated values from calculations are below the experimental scaling factors. This is mainly due to the large number of scenarios considered for the calculations. These scenarios include decay times from 1 up to 30 years.

When the decay time increases the ratio of the Ni-63/Co-60 activities also increases. This is a consequence of the long half-life of Ni-63 (~ 100 years) compared to the half-life of Co-60. For instance, after a 10 year decay time $\sim 25\%$ activity of Co-60 is left whilst $\sim 93\%$ activity of Ni-63 is still present. This is equivalent to a scaling factor ~ 3.6 times bigger.

Calculating mean scaling factors for decay times equal or above 10 and 20 years we obtain respectively $SF_{T_c \geq 10y} = 6.33$ and $SF_{T_c \geq 20y} = 9.61$. The result obtained for decay times above or equal to 20 years are in very good agreement with the experimental results from campaign 2. All other factors being equal, it is possible to use simulations to identify potential decay time for historical waste by comparing theoretical and experimental scaling factors.

Theoretical studies, such as the ones proposed in the present study, can be used to predict the order of magnitude of the confidence interval of scaling factors and to quantify average statistics of theoretical scaling factors.

5 Conclusion

In the present study, we tested and compared different techniques to sample historical waste. These methods were used to estimate the concentration of Ni-63 in copper after studying the correlation between a key nuclide and the DTM nuclide of interest. To estimate the specific activity of

Ni-63 we used linear regression, the scaling factor method and the so-called “mean activity method”. Among the statistical techniques available to sample materials, we discussed simple random and systematic sampling, census, 2-stage stratified and authoritative sampling and we introduced the use of the bootstrap for DTM activity estimation.

We used as an example 2 waste populations of copper from cables activated at CERN. The waste populations are respectively called campaign 1 and campaign 2.

For campaign 1, we chose simple random and systematic sampling when selecting samples. The bootstrap, which is a resampling technique with repetition, is used to estimate distributions of the concentration of Ni-63 around an average value. The Ni-63 activities obtained were compared with census data.

The present study results suggest that the bootstrap is a robust tool to estimate average activity of DTM nuclides from samples and that this estimation is more precise when the number of repetition and samples increases. It is also found that the estimation of the activity of Ni-63 is more precise when values below the detection limit are excluded. The results of the simulations performed are in very good agreement with results from census data. When the number of resampling is above 500, the relative error of Ni-63 concentration from bootstrap with respect to census data is below 25%.

Systematic sampling performs better when estimating Ni-63 with respect to random sampling. However this result cannot be generalized since it is due to a specific set of random numbers and the use of different seeds can generate a different score. The bootstrap can be used as a complement to these strategies since it can easily evaluate the distribution of statistics instead of simple statistics such as the mean or the median. In practice, we can collect samples using random or systematic sampling and process the results using the bootstrap.

For practical reasons we suggest the use of systematic sampling because it is easy to implement in an industrial process in which waste packages are routinely produced. Care must be taken however because systematic sampling depends on the file order and can be affected by a periodic or repetitive structure of the waste flow.

A correlation between the activities of Ni-63 and Co-60 was not found for campaign 1. The mean activity method was applied to estimate the content of Ni-63 in the batch. This method consists of calculating the average activity from all the samples collected – including values below detection limits – and attributing the average value of the DTM to each single package of the batch.

For the population of campaign 2 samples are taken using 2-stage stratified sampling. This sampling method allows us to concentrate the sampling effort on the strata where the γ -activity is higher. We used authoritative sampling to estimate preliminary standard deviations of the activities on the strata and the Neyman allocation to identify primary units for sampling. A 2-stage sampling was chosen due to the unknown heterogeneity of the activity on shredded copper.

For campaign 2, the activities of Ni-63 and Co-60 are correlated. We applied linear regression and the scaling factor method to estimate the content of Ni-63 in each drum

of the population. The relative error affecting the average specific activity of Ni-63, calculated via stratified sampling, linear regression and the scaling factor method – either geometric or mean scaling factor – is below 16%.

The choice among mean and geometric mean scaling factor depends on the experimental distribution of the SF. For symmetric unimodal distributions a large difference from the 2 calculations is not expected.

We also calculated the average specific activity of Ni-63 from authoritative, high-dose, samples. As expected, the activity so calculated is biased towards higher values since the samples were chosen to be conservative in terms of γ activity. Carefully chosen, judgemental samples can be used to estimate higher bounds of activities for a batch of waste.

Finally, we compared the results from theoretical with experimental scaling factors. The results obtained with both methods are within one order of magnitude and can be improved if we consider realistic decay times for a given campaign. We showed that, for the waste family of campaign 2, decay times of 20 years or more explain the difference between simulations and experiments.

The present study shows how various existing sampling methods can be applied to sample historical waste produced at CERN. Each technique should be adapted to the needs of the waste producer. The sampling techniques introduced, combined with linear regression, scaling factors and mean activity methods are a robust set of tools that can be used to characterize historical waste in research centres and nuclear installations.

The authors wish to thank Bertrand Cellier, Nick Walter and Thijs Wijnands for the sample collection and the CERN HSE-RP Group for its support.

References

1. International Organization for Standardization, *ISO 21238. Nuclear energy – Nuclear fuel technology – Scaling factor method to determine the radioactivity of low- and intermediate-level radioactive waste packages generated at nuclear power plants* (ISO, Geneva, Switzerland, 2007)
2. IAEA Nuclear Energy Series, No. NW-T-1.18. *Determination and use of scaling factors for waste characterization in nuclear power plants* (IAEA, Vienna, Austria, 2009)
3. W.G. Cochran, *Sampling techniques* (John Wiley and Sons, New York, USA, 1977)
4. R.O. Gilbert, *Statistical methods for environmental pollution monitoring* (Van Nostrand Reinhold Company, New York, USA, 1987)
5. ASTM D6311 98, *Standard guide for generation of environmental data related to waste management activities: selection and optimization of sampling design* (ASTM, West Conshohocken, USA, 2014)
6. B. Efron, R. Tibshirani, *An introduction to the bootstrap* (Springer, New York, USA, 1993)
7. T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning: data mining, inference and prediction* (Springer-Verlag, New York, USA, 2009)
8. G.F. Knoll, *Radiation detection and measurements* (John Wiley and Sons, New York, USA, 2010)

9. G. Gilmore, *Practical γ -ray spectrometry* (John Wiley and Sons, New York, USA, 2008)
10. M.F. L'Annunziata, *Handbook of radioactivity analysis* (Academic Press, San Diego, USA, 1998)
11. D. Tattam, L. Keightley, *Radiometric non-destructive assay. Measurement good practice guide No 34 (Issue 2)* (National Physical Laboratory, Teddington, UK, 2012)
12. J.A. Dean, *Analytical chemistry handbook* (McGraw-Hill, Columbus, USA, 1995)
13. H. Vincke, C. Theis, Actiwiz – optimizing your nuclide inventory at proton accelerators with a computer code, in *Proceedings of the ICRS12 Conference. Progress in Nuclear Science and Technology, Nara, Japan* (2012), Vol. 4, p. 228
14. C. Theis, H. Vincke, The use of Actiwiz in operational radiation protection, in *Proceedings of the SATIF-12 Conference. Nuclear Science Report NEA/NSC/R 3, Fermilab, USA* (2014)
15. G. Battistoni, F. Cerutti, A. Fasso', A. Ferrari, S. Muraro, J. Ranft, S. Roesler, P.R. Sala, The FLUKA code: Description and benchmarking, in *Proceedings of the Hadronic Shower Simulation Workshop 2006, Fermilab, USA, 6–8 September 2006*, edited by M. Albrow, R. Raja, AIP Conf. Proc. **896**, 31 (2007)
16. A. Ferrari, P.R. Sala, A. Fasso', J. Ranft, FLUKA: A multi-particle transport code. CERN-2005-10 (2005), INFN/TC-05/11, SLAC-R-773, 2005
17. R. Froeschl, C. Theis, F. La Torre, H. Vincke, N. Walter, A. Sgobba, Radiological Hazard classification of material in CERN's accelerators. CERN-DGS-2012-003-RP-IR; EDMS Id: 1184236, Geneva, Switzerland, 2012
18. IAEA, *Strategy and methodology for radioactive waste characterization* (IAEA-TECDOC-1537, Vienna, Austria, 2007)

Cite this article as: Biagio Zaffora, Matteo Magistris, Gilbert Saporta, Francesco Paolo La Torre, Statistical sampling applied to the radiological characterization of historical waste, EPJ Nuclear Sci. Technol. **2**, 34 (2016)