



Analysis of the comparison of in situ measurements made on biological shielding of the BR3 nuclear reactor

Marielle Crozet¹ · Séverine Demeyer² · Valérie Lourenço³ · Margarita Herranz⁴ · Sven Boden⁵

Received: 17 January 2022 / Accepted: 17 March 2022 / Published online: 5 May 2022
© The Author(s) 2022

Abstract

INSIDER (Improved Nuclear Site characterization for waste minimization in D&D operations under constrained EnviRonment) was a European project funded under the H2020-EURATOM programme and launched in June 2017. The project was coordinated by the French Commissariat à l'énergie atomique et aux énergies alternatives (CEA), it had a total duration of 4 years and covered a budget of 4 M€. INSIDER's work was performed by 5 technical working groups (WG) which brought together 18 institutions from 10 countries, leading to a total of 68 participating researchers. The objective of the project was to optimise the radiological characterisation of nuclear installations in constrained environments in order to obtain an accurate estimate of the content of contaminated materials as well as to optimise the quantity of contaminated materials to be treated as waste. The focus of this paper is on the statistical analysis of an interteam comparison of measurement results (dose rate, total gamma measurement, and gamma spectrometry) made in situ at the BR3 reactor, Belgium.

Keywords INSIDER project · BR3 reactor · Proficiency test · Analysis of variance

Abbreviations

INSIDER	Improved Nuclear Site characterization for waste minimization in DD operations under constrained EnviRonment
CETAMA	Commission for the Establishment of Analytical Methods
BR3	SCK-CEN nuclear reactor

Introduction

Having been built in the middle of the twentieth century, civil nuclear installations across the world are entering a period when their first installations are reaching the end of their projected lifetimes. The stakes are high: the aim is to show that, by using optimized methods, it is possible to cleanly close down the life cycle of nuclear facilities and leave cleaned up sites, at a controlled cost. The nuclear industry already has a solid experience in the remediation and dismantling techniques that will be required to meet this challenge [1].

In terms of dismantling operations, one can distinguish between the dismantling of reactors, that of cycle facilities, and post-accident dismantling. Whilst each of these cases presents unique specificities, the dismantling and decommissioning (D&D) process of a site or an installation is always the same: site or installation characterization, elaboration of a scenario for the operations, preliminary decontamination of surfaces, cutting and dismantling operations, and finally management of waste and effluents. The characterization of site or installation [2–6] to be dismantled is a crucial step in the process: it is essential to define the actions to be carried out. The INSIDER (Improved Nuclear Site characterization for waste minimization in D&D operations under constrained EnviRonment) project [7] aimed to develop and

✉ Marielle Crozet
marielle.crozet@cea.fr

¹ CEA, DES, ISEC, DMRC, University of Montpellier, Marcoule, France

² Laboratoire national de métrologie et d'essais (LNE), Trappes, France

³ CEA, List, Laboratoire National Henri Becquerel (LNE-LNHB), Université Paris-Saclay, 91120 Palaiseau, France

⁴ Nuclear Engineering Department, University of the Basque Country (UPV/EHU), Pza. Ingeniero. Torres Quevedo, 1, 48013 Bilbao, Spain

⁵ SCK CEN, DDW, Boeretang 200, 2400 Mol, Belgium

validate a new and improved integrated characterisation methodology and strategy during nuclear D&D operations of nuclear power plants, post accidental land remediation, or nuclear facilities under constrained environments. In this project, three use cases are studied, corresponding to the three categories mentioned above [8]. SCK-CEN provided the case study with regards to the dismantling of nuclear reactors: characterization of the biological shield of the BR3 reactor, made of irradiated heavy concrete called UC2 [9]. The Horizon H2020 INSIDER project is composed of 7 workpackages [7]. The workpackage 6, in charge of estimating the performance of measurement methods used and assessing the measurement uncertainty [10], organised several interlaboratory or interteam comparisons [11–13] in compliance with ISO 17043:2010 [14]. Amongst these comparisons, a comparison of non-destructive in situ measurements [15] was organized at the BR3 reactor in collaboration with workpackage 5 (on site analysis) and with the use case supplier SCK-CEN. Each team came to measure on site each in turn at previously defined points [16]. The results of this in situ measurement interteam comparison are presented in this paper.

In situ measurement comparison at BR3: presentation

The results of the comparisons are of prime importance in the INSIDER project. They will allow both to compare the different methods applied and to assess result uncertainties from comparisons performed on real cases (benchmarking analysis) on one side and on synthetic certified reference materials (CRMs, Interlaboratory comparisons) on the other side (RM were produced by WP4 in compliance with [17, 18]). The choices made in the project allowing to cover the most important and common use cases in D&D operations are presented in Table 1.

This in situ benchmarking exercise can be used as a proficiency test where each team has to give a measurement value with an associated expanded uncertainty. The aim was

also to perform measurements in conditions as similar as possible to routine ones. The results were used to perform a variance analysis (ANOVA) to determine the contribution of the different sources of uncertainty to the measurement uncertainty. Therefore, repeatability, intermediate precision by varying certain experimental conditions (eg: repositioning the measurement equipment, comparing different types of detectors used ...), and reproducibility need to be assessed by performing multiple measurements (only when short measurement times are possible).

The BR3 interteam comparison was the test scheme chosen to highlight non-destructive in situ measurements of contamination into/on solid surfaces [16]. Three measurands were chosen for the scheme and a common measurement protocol (approved by all participants) was applied in order to perform the measurements. The aim of this comparison was to allow the participants to compare their results with those of the other participants (using identical or different methods), always with the ultimate aim of improving their analyses. Where possible, the outcome of the comparison was also to derive an estimate of measurand uncertainties for each measurand [19, 20].

The UC2 benchmarking exercise occurred at the end of 2018 and consisted in several characterizations of the biological shield of the BR3 reactor—located in the SCK CEN (Mol, Belgium)—in terms of dose-rate and total gamma emission as well as a trial for the use of high resolution gamma spectrometry measurements at an early characterization stage [21]. Today, most of the installations in this pressurized water reactor have been dismantled. The biological shield characterized consisted in high-density concrete which was neutron activated during the reactor's 25 years of operation. A more detailed description of the measurement methods and site can be found in Herranz et al. [22, 23].

The aim of this in situ benchmarking comparison focused on the measurement of gamma emissions by using a number of commercially available portable area dosimeters to carry out dose rate and total gamma measurements. The measurements allowed to evaluate the different dosimeter's equivalence in providing comparable

Table 1 Summary of the 2 benchmarking exercises and the 4 interlaboratory comparisons (ILC) organized within the frame of the INSIDER project

Use case number	1	2
Object of D&D	Back/end fuel cycle and/or research facility	Nuclear reactor
Location	Ispra (Italy)	Mol (Belgium)
Theme	Liquid waste storage facility	Biological shield of a nuclear reactor
Benchmarking analysis: in situ NDA	Yes	Yes (analysis in this article)
In lab analysis via ILC on Reference Materials (being Certified during the project)	ILC1 on CRM1 (synthetic liquid)	ILC2 on CRM2 (spiked non irradiated real concrete)
In lab analysis via ILC on real samples	No	Yes (2 RM with 2 activity levels)

results. Three measurement locations were chosen, with emission values ranging from natural background level to several $\mu\text{Sv/h}$, in order to cover low dose rates and activity measurements. The radiations measured were composed of beta/gamma emitters, mainly activation products Eu-152, Ba-133, Co-60, and Cs-137, leading to a measurement range of about 30 to 1400 keV. For both dose rate and total gamma measurements at the three locations (named A, B and C, with increasing dose rate or total gamma activities), five sets of five consecutive measurements each were carried out in order to estimate the repeatability of the measurements. The repeated measurements also allowed to evaluate the reproducibility of repositioning by removing the measurement device from the spot after each set of 5 measurements. This resulted in a total of 25 single dose rate and total gamma measurements for each of the three locations (A, B and C) [16].

Measurements were performed in direct contact with the surface to characterize without any collimation or shielding. To help standardize the various measurement probes, calibration was performed in terms of ambient gamma dose rate equivalent $H^*(10)$ (in $\mu\text{Sv/h}$) using a Cs-137 reference source. For total-gamma measurement devices, a Cs-137 point source provided by the SCK CEN was used to compare the probes. In situ gamma spectrometry measurements were performed at location C (highest dose rate, 30 cm away from the surface) using shielded detectors and background correction enabled through repeated measurements with or without collimation. All teams used their own equipment and calibration procedure. Modelling was necessary and the SCK CEN provided the chemical composition and density of concrete, but also the standard activation profile parameters such that the teams could assess several characteristics: depths, Eu-152/Eu-154 activity ratios, and Cs-137 surface activity.

Usually, the aim of proficiency tests is to compare a measurement result with an assigned certified value. However, in the present comparison exercise the purpose was rather to estimate the equivalence of common measurement devices with a particular interest in assessing dose rates and total gamma surface activities. To do so, the assigned value for the proficiency assessment was chosen to be the robust mean from the results reported by participants. Standardized performance statistics such as the z - and ζ -scores were favoured to compare the measurement devices or teams (in accordance with NF ISO 13528 [24]).

Precision decomposition was performed through an analysis of variances (ANOVA) using a model, which takes into account two main factors: the team and/or type of devices used and the repositioning effect. The significance of these effects was estimated using a F test at a risk level of 5%.

Mathematical models

Proficiency test

The aim of proficiency tests (PT) is to compare a result on a proficiency test item with an assigned value, where a result is the average of all the measurements from a participant on the test item. In this study, the assigned value x_{pt} for the proficiency test item was defined as the robust mean from the participants' reported results and standardized performance statistics (z -score, z' -score, and ζ -score) are considered. The idea behind performance statistics is to compare the difference $D_i = x_i - x_{\text{pt}}$ (x_i is the result of laboratory code (i) with an allowance for measurement errors ([24] Sect. 8.1.1) described as standard uncertainties or standard deviations (as discussed below).

The difference D_i may be expressed as a relative difference:

$$D_i\% = 100(x_i - x_{\text{pt}})/x_{\text{pt}}\%$$

In the present study, it is assumed that all the consensus estimates (of means and standard deviations) used in the performance statistics are obtained exclusively from the analysis of the participants' results obtained in the same round of the PT scheme. The standard uncertainty of the assigned value $u(x_{\text{pt}})$ reflects the confidence in the assigned value i.e. in the process that led to that value. When x_{pt} is derived as a robust mean, the standard uncertainty of the assigned value is estimated as [24]:

$$u(x_{\text{pt}}) = 1.25 \times \frac{s^*}{\sqrt{p}}$$

where s^* is the robust standard deviation of the results. The factor 1.25 is based on the standard uncertainty of the median in a large set of results drawn from a normal distribution. The standard deviation for proficiency assessment σ_{pt} characterizes the dispersion of results around a central value and can be computed, for instance from the data obtained in the PT, as the robust estimate of the standard deviation of the results s^* . Procedures to estimate the robust mean and the robust standard deviation of results can be found in [24]. The performance statistics considered in this study are the z -score and the ζ -score, computed from participant results $x_i, i = 1, \dots, p$.

The z -score for a proficiency test result x_i is calculated as:

$$z_i = \frac{x_i - x_{\text{pt}}}{\sigma_{\text{pt}}}$$

The ζ -score can be used instead of the z -score when there is an effective system in operation for validating laboratories' own estimates of the standard uncertainties of their results

$$\zeta_i = \frac{x_i - x_{pt}}{\sqrt{u_i^2 + u^2(x_{pt})}}$$

where u_i^2 is the laboratory's own estimate of the standard uncertainty of its result x_i .

Remarks

- Although the robust mean x_{pt} is correlated with each x_i , it is possible that performance statistics do not take this correlation into account. As a consequence, ζ_i would be under-estimated, but according to [24], this under-estimation is not significant if the uncertainty of the assigned value is small, allowing the ζ -score to be used with consensus statistics without prior adjustment for correlation.
- In practice, the uncertainty of the assigned value should be small in comparison with σ_{pt} ; a recommendation from [24] is to ensure $u(x_{pt}) < 0.3\sigma_{pt}$. Otherwise, participants could receive action and warning signals due to the inaccuracy in the determination of the assigned value. If $u(x_{pt}) \geq 0.3\sigma_{pt}$, [24] recommends to take into account $u^2(x_{pt})$ by using the z' -score

$$z'_i = \frac{x_i - x_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}}$$

The interpretation of z -scores is the following [24]:

- A result that gives $|z| \leq 2.0$ is considered to be acceptable.
- A result that gives $2.0 \leq |z| \leq 3.0$ is considered as a warning signal.
- A result that gives $3.0 \leq |z|$ is considered to be unacceptable (or action signal).

A similar interpretation holds for z' -scores and ζ -scores.

Measurement uncertainty estimation (ANOVA)

The measurement result is modeled by a random effects model (model 1):

$$x_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}, \quad i = 1, \dots, a \quad j = 1, \dots, b \quad k = 1, \dots, n$$

where μ is the overall mean response, α_i is the effect of level i of the *team* factor, $\beta_{j(i)}$ is the effect of level j of the *repositioning* factor for the level i of the *team* factor, ε_{ijk} is a random error term.

The analysis of variance (anova) of model 1 is performed under the following hypotheses $\alpha_i \sim \text{iid } N(0, \sigma_{\text{team}}^2)$,¹

¹ iid: independent, identically distributed.

Table 2 Anova table: sum of squares decomposition and variance components for a nested anova with 2 random factors (B nested within A, A: teams and B: repositioning) under the assumptions of model 1

Factor	d.f	MS	F-ratio	Variance	p-value
A	$a - 1$	MS_A	$\frac{MS_A}{MS_{B'(A)}}$	$\frac{MS_A - MS_{B'(A)}}{nb}$	$P(F > F_{\text{crit}})$
B'(A)	$(b - 1)a$	$MS_{B'(A)}$	$\frac{MS_{B'(A)}}{MS_{\text{resid}}}$	$\frac{MS_{B'(A)} - MS_{\text{resid}}}{n}$	$P(F > F_{\text{crit}})$
Residual	$(n - 1)ba$	MS_{resid}			

$\beta_{j(i)} \sim \text{iid } N(0, \sigma_{\text{repos}}^2)$, $\varepsilon_{ijk} \sim \text{iid } N(0, \sigma^2)$ (homoscedasticity), $\alpha_i, \beta_{j(i)}, \varepsilon_{ijk}$ are pairwise independent. Significance testing of factors is performed with p -values obtained as the probability $P(F > F_{\text{crit}})$ where F is the value of a test statistic (F ratio in Table 2 estimated on the data and F_{crit} is the tabulated value corresponding associated with a risk level $\alpha = 5\%$. A p -value less than $\alpha = 5\%$ indicates a significant effect. Under these assumptions (normality, homoscedasticity and independence), variance components can be obtained from the anova sum of squares decomposition according to the formulas shown in Table 2.

To express that the residual variance differs among teams (heteroscedasticity), the error term ε_{ijk} can explicitly depend on the levels of the *team* factor, for instance $\varepsilon_{ijk} \sim N(0, \sigma_i^2)$. The estimation of parameters, among which the variance components σ_{team}^2 and σ_{repos}^2 , requires dedicated software (for example the R package nlme) using iterative algorithms such as ML (maximum likelihood) or REML (restricted maximum likelihood) [25]. Significance testing of factors is performed by model comparison. In order to test the significance of the random effect Repositioning (B'(A)), we can fit a new model with only the team factor (model $x_{ijk} = \mu + \alpha_i + \varepsilon_{ijk}$, $\varepsilon_{ijk} \sim N(0, \sigma_i^2)$) and test the significance of the likelihood ratio. If the p -value of the test is less than 0.05, the factor has a significant effect at the level 5%. The significance of the random factor team is tested against the model $x_{ijk} = \mu + \varepsilon_{ijk}$, $\varepsilon_{ijk} \sim N(0, \sigma_i^2)$. The intermediate precision variance is given by

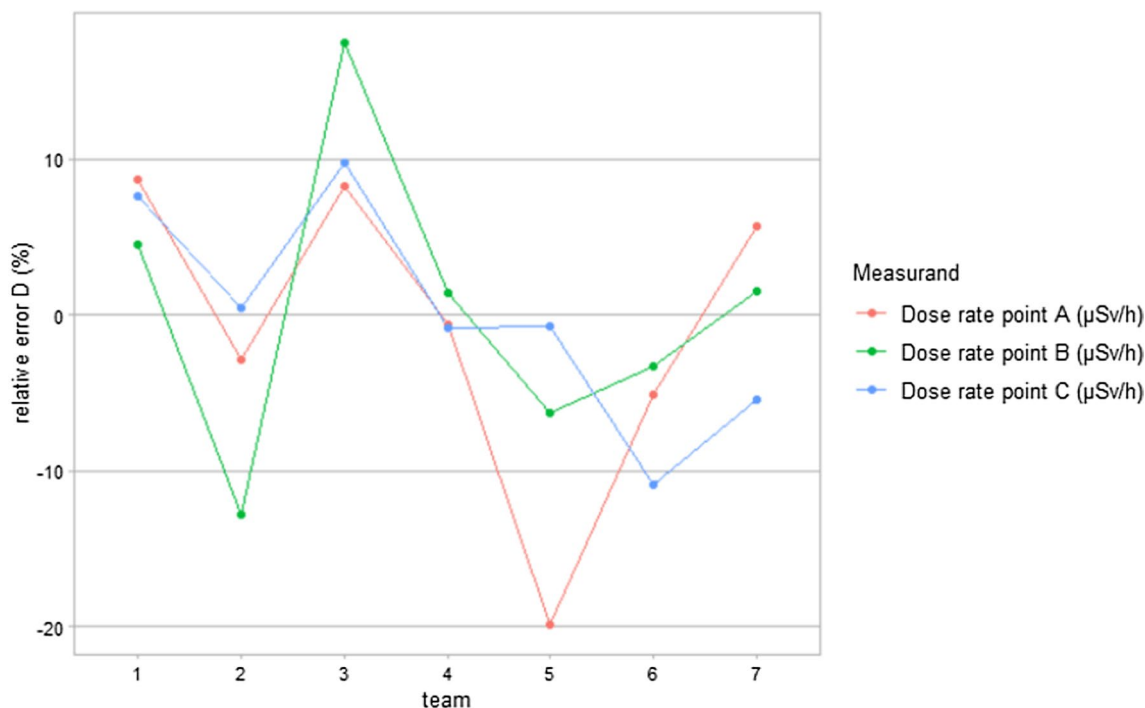
$$s_f^2 = \hat{\sigma}_{\text{team}}^2 + \hat{\sigma}_{\text{repos}}^2,$$

and can be used as part an uncertainty budget to characterize uncertainty originating from the measurement procedure (operator, measurement device, repositioning, ...). It is important to notice that s_f^2 does not account for measurement trueness or spatial variability.

Table 3 Consensus estimates for dose rate

	Point A	Point B	Point C
x_{pt}	0.383	1.027	3.06
$u(x_{pt})$	0.017	0.048	0.12
σ_{pt}	0.036	0.10	0.25

evaluation of performance using proficiency tests was based on z' -scores in this study. All the individual team results, augmented with values of z -scores and ζ -scores, are presented in the appendix. For clarity, only summary results based on z' -scores are displayed for each measurand in the main body of this paper.

**Fig. 1** Relative deviation from consensual value for dose rate PT

Comparison results

Proficiency test

For Dose rate and Total Gamma measurements, team uncertainty estimates associated with dose rate measurements are only based on type A uncertainty evaluation (no calibration is taken into account, for example) such that ζ -scores are computed only for information, not for performance evaluation (more details on the measurements made in [22, 23]). For Gamma spectrometry, the ζ -score is not computed since the uncertainty (estimated from the peak area) is missing for many teams.

For Dose rate, Total Gamma and Gamma spectrometry measurement, the ratio $u(x_{pt})/\sigma_{pt}$ exceeds the threshold 0.3. According to [24], this means that the uncertainty of the consensus estimate $u(x_{pt})$ is too large with respect to the estimate of the dispersion of measurements σ_{pt} . In this case, the z' -score should be used instead of the z -score. The z -score is computed for information in all cases. The

Dose rate

Commercially available radiation detection instruments can be categorized as a function of their detector type: gas counters, scintillation counters, and solid state detectors. It is interesting to note that these three types of detectors are capable of measuring the main types of radiation produced by radioactive decay (alpha, beta, X, or gamma), which makes them particularly suitable for the initial mapping stage. Consensus estimates for dose rate are presented in Table 3. Performance statistics such as relative deviations and z' scores are plotted in Figs. 1 and 2.

Regarding performance statistics: Team 5 has a z' -score very close to 2, the warning signal limit. At location B, team 3 (organic scintillator) had the lowest relative standard deviation, which contributed to a ζ score in the unacceptable range (resulting in a warning signal). At location C, the dose rate measured reached the calibration range of detector 3 but the relative standard deviation is still the lowest, again

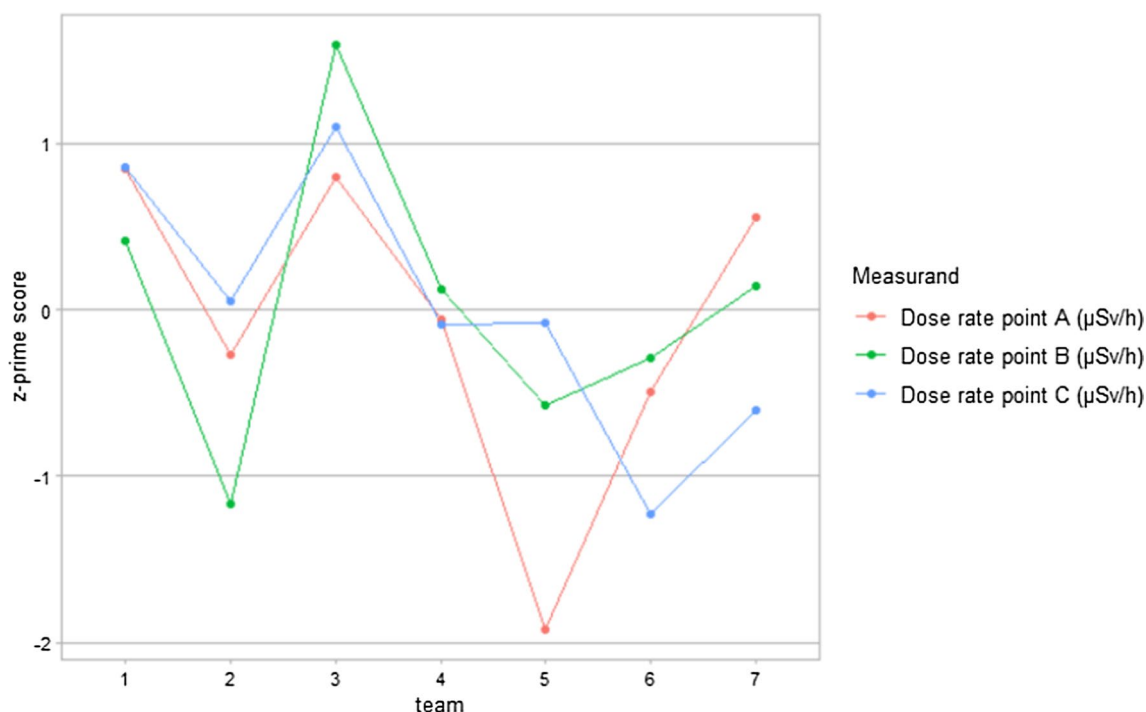


Fig. 2 z' score for dose rate PT

resulting in a warning signal for ζ score performance as well as resulting in the highest z' performance score, which is still acceptable. It should be noted that the uncertainties were only derived from the measurements' repeatability and reproducibility (i.e. the 25 measurements per location). As such, the uncertainties can be assumed to be under-estimated as calibration or correction factors were not taken into account. Whatever the location, the uncertainties reported by team 5 were at least 20% higher (3 times higher on average) than those of the other participants.

Total gamma

Despite a common calibration with a point source of Cs-137 provided by the SCK-CEN, a large dispersion of the absolute surface activities measured by the different devices was observed. This is probably due both to the fact that the main radiation measured during the test is far from the Cs-137 signal (662 keV) and the fact that several of the detectors used have non-flat energy responses. Indeed, the majority of the radiation in the studied surfaces is due to the presence of Ba-133 (< 400 keV) and to some extent Eu-152 (120–1400 keV). In order to normalise these differences, it was decided to study the ratios between two measurement points. Consensus estimates for total gamma measurement ratios calculated for point A over point B and point A over point C are presented in Table 4. Relative deviations are plotted in Fig. 3, and z' scores are shown in Fig. 4.

Relative deviations show that teams 4, 5, and 6 tended to over-estimate total gamma surface activity. On the contrary, the relative deviations of teams 7, 2, and 8 show a similar under-estimation, greater than 10%.

Gamma spectrometry

The gamma spectrometry measurements carried out, together with monte carlo modelling, were intended to estimate mainly two parameters: the depth limit up to which the activated concrete should be considered as waste to be treated, and the surface activity of contaminants.

Table 5 summarises the consensus values for the five measurands. Relative deviations are plotted in Fig. 5 and z' scores are presented in Fig. 6.

It is important to note that all depth limit consensus values displayed standard uncertainties close to 5% whilst PT standard deviations usually lie below 10%. This shows that the results of this study can be taken to be mostly in suitable agreement irrespective of the type of the detector and calculations used. For isotopic ratios of Cs surface activity,

Table 4 Consensus estimates for total gamma

	A/B	A/C
x_{pt}	0.324	0.1128
$u(x_{pt})$	0.022	0.0093
σ_{pt}	0.050	0.021

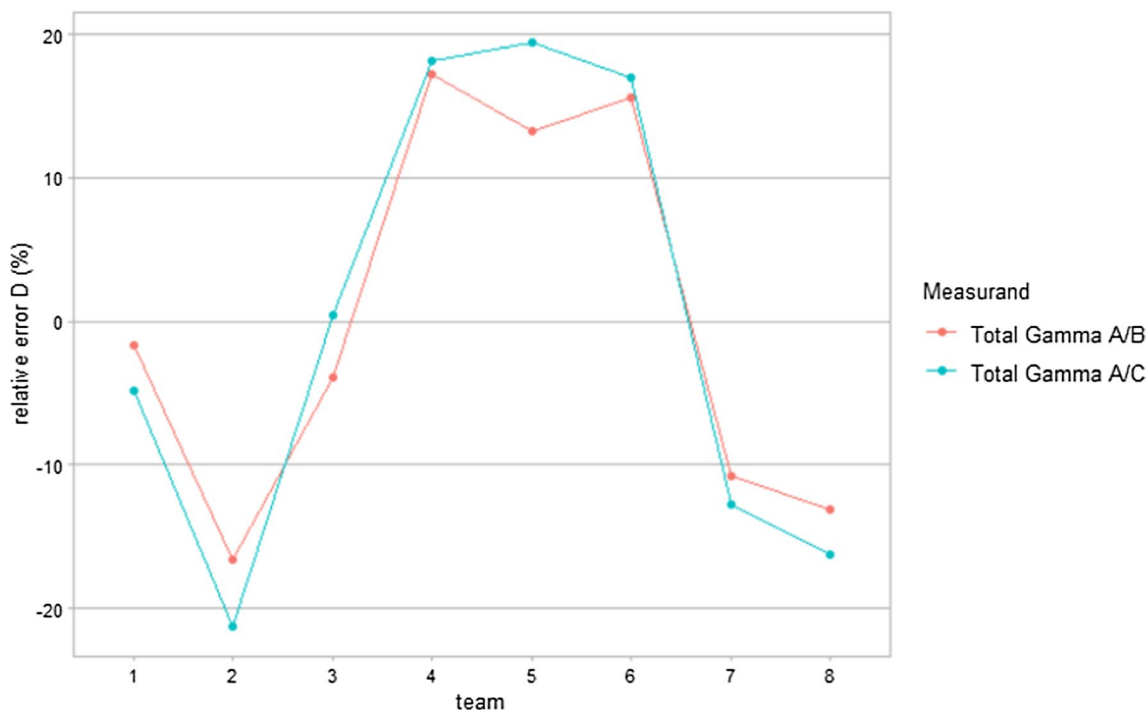


Fig. 3 Relative deviation from consensual value for total gamma measurement ratios

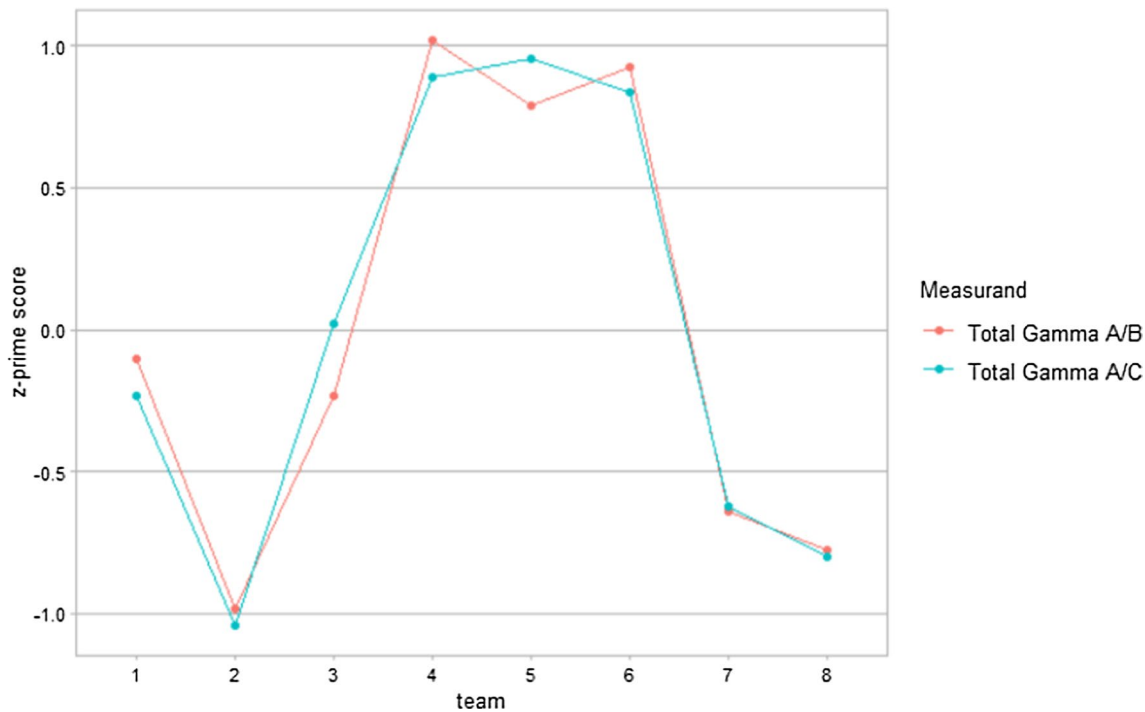


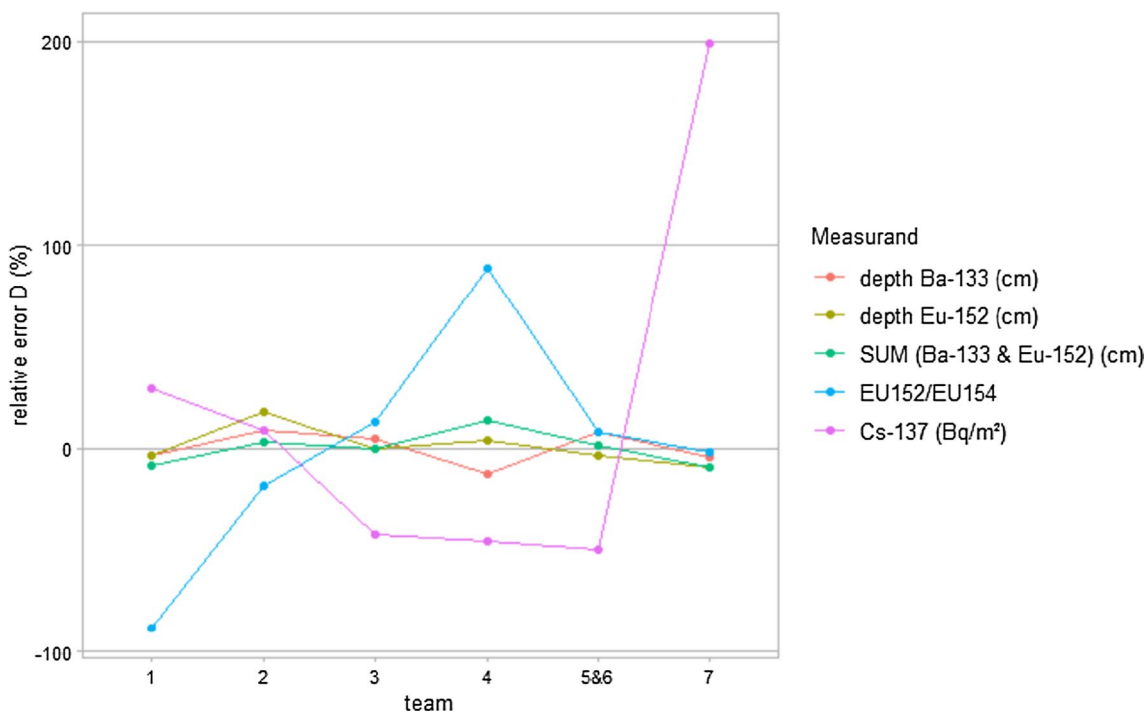
Fig. 4 z' scores from consensual value for total gamma measurement ratios

the results were influenced by discrepant data. This effect resulted in consensus values with standard uncertainties around 33% and PT standard deviations twice as high.

Team 4 submitted the highest uncertainties and its depth limit estimate for Ba-133 appears to be the most underestimated. Team 1 also submitted large uncertainties associated

Table 5 Results for the five measurands

	Depth Ba-133 (cm)	Depth Eu-152 (cm)	SUM (Ba-133 and Eu 152) (cm)	Eu-152/Eu-154	Cs-137 (Bq/m ²)
x_{pt}	52.6	38.2	57.2	15.9	6818
$u(x_{pt})$	2.58	1.77	2.81	5.25	2317
σ_{pt}	5.06	3.47	5.50	10.3	4541

**Fig. 5** Values of relative error D (%) for each team and each measurand

with a depth limit estimate that appears underestimated. Team 2 provided the highest limiting depth for Ba-133. Finally, teams 3 and 5&6 submitted very similar results.

Regarding the depth limit estimates for Eu-152; team 4 submitted the highest uncertainties but this time its depth limit estimate was the highest. Team 1 submitted large uncertainties and, similarly as for Ba-133, their depth limit estimate for Eu-152 appears underestimated. Team 2 which provided the highest limiting depth for Ba-133 also provided the highest estimate for Eu-152. Finally, teams 3 and 5&6 continued to show results similar to each other.

The results for the combination of Ba-133 and Eu-152 depth limits are very similar to those for depth limit estimates for Eu-152 which have been previously presented in this paper.

In estimating the ratio of the two isotopes of europium, the detector that provided the lowest value remains that used by team 1. The relative uncertainty provided was the highest (around 30%). Team 2 provided an estimate close to the consensus value, with large uncertainties. Team 4 provided the

highest estimate of the ratio, including a very large relative uncertainty (~27%). Again, team 3 and team 5&6 submitted comparable results.

For the measurement of the Cs-137 surface activity, the results submitted by all participants were very scattered. Teams 3 and 5&6 submitted results which were very low but close to each other. Team 4 supplied a low value with large uncertainties. Teams 1 and 2 reported similar results both in terms of absolute value and in relative uncertainties.

Measurement uncertainty estimation

Dose rate results for each team at each of the three sites (A, B, and C) are displayed in Fig. 7 using diamond plots. As indicated by the graphs and confirmed by Levene's Test of variance homogeneity, the hypothesis of homogeneity of variance is rejected for all sites. This means that anova decomposition as previously described in Table 2 cannot be applied. Team variance parameters σ_i^2 are thus introduced to model the residual variance per team.

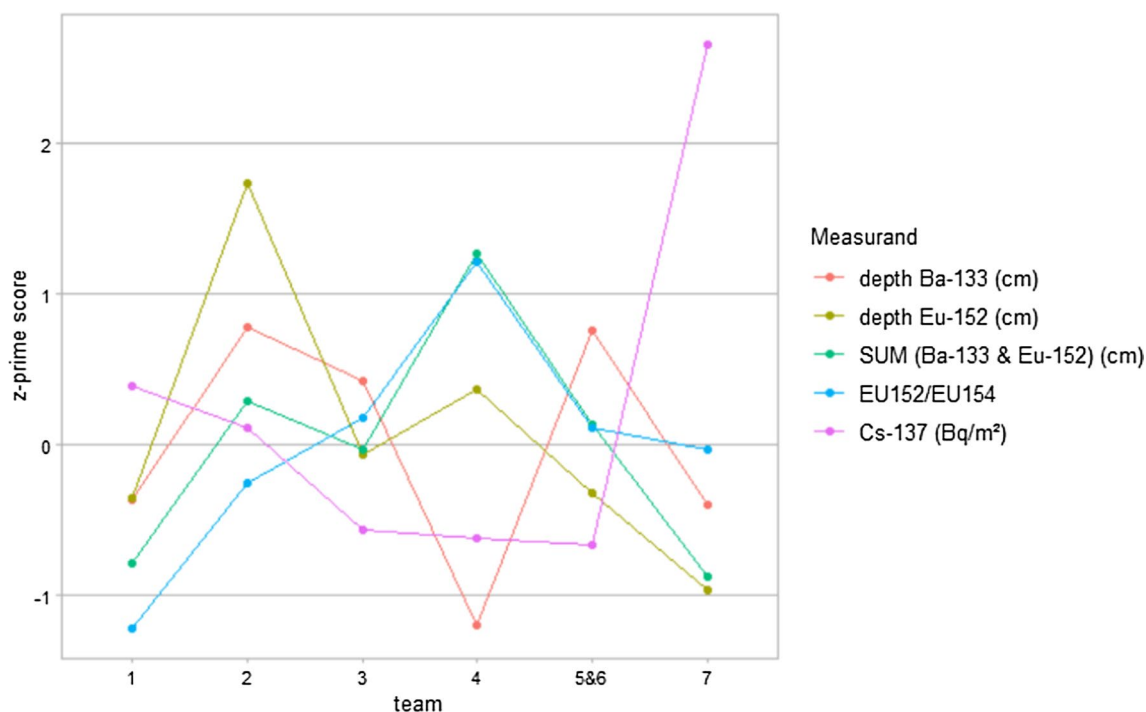


Fig. 6 Values of the z' score for each team and each measurand

The utility of the introduction of group variance parameters σ_i^2 in order to model random residual error at a team scale was verified by plotting residuals under both hypotheses (a) $\varepsilon_{ijk} \sim N(0, \sigma^2)$ and (b) $\varepsilon_{ijk} \sim N(0, \sigma_i^2)$, as seen in Fig. 8. Under hypothesis (a) residuals for site A are plotted against their corresponding fitted values and higher fitted values seem to have lower residuals. Under hypothesis (b), residuals are plotted for each team against the corresponding team mean, $\hat{\alpha}_i$, and have a comparable dispersion.

The model comparison approach shows not only that the Repositioning factor has a significant effect for measurements performed at sites B and C but also that the team factor has a significant effect for measurements performed at all three sites. Table 6 displays, for each site, the estimates of the overall mean $\hat{\mu}$, of the variance components $\hat{\sigma}_{\text{team}}^2$ and $\hat{\sigma}_{\text{reposit}}^2$ obtained with the *R* nlme package and the corresponding measurement variance components S_f^2 . The quantities $\hat{\sigma}_{\text{team}}$, $\hat{\sigma}_{\text{reposit}}$, and S_f are also presented and expressed in terms of relative standard deviation $\hat{\sigma}/\hat{\mu}$ or relative uncertainty $S_f/\hat{\mu}$ in parentheses.

Total gamma

Estimation of the variance components due to the team effect by analysing the Total Gamma ratios A/B and A/C

Figures 9 and 10 display the dispersion of ratios for Total Gamma measurements between teams. The Levene test does not reject the hypothesis of homogeneity of variances at the level $\alpha = 5\%$ (p -value = 0.16 for A/B ratio and p -value = 0.25 for A/C ratio).

A one-way random effect anova is used to estimate the components of variance due to team when analysing the Total Gamma ratios A/B and A/C, whose results are displayed in Table 7.

Estimation of the repositioning effect for each team by analysing direct measurements at points A, B and C

Results obtained with one-way random effect anova for each team are displayed for each site A, B, and C in Tables 8, 9 and 10, respectively. For all teams at each site, the Levene Test does not reject the hypothesis of homogeneity of variances across repositioning.

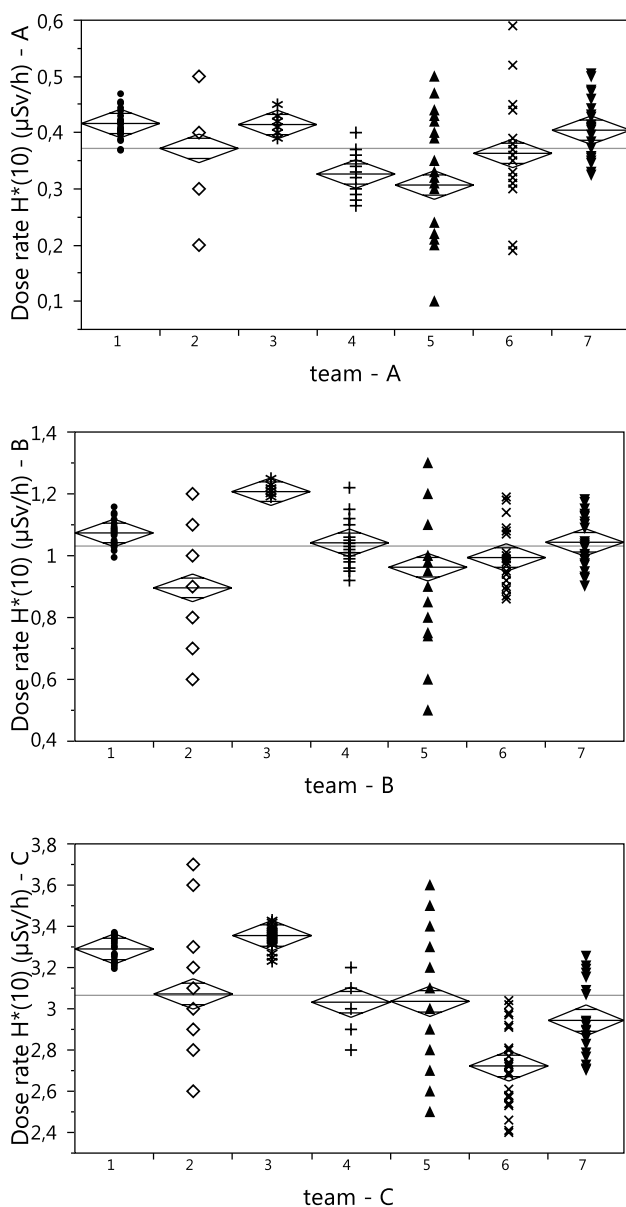


Fig. 7 Diamond plots displaying, at each point (A, B, C), the dose rate measurements performed by each team (1–2–3–4–5–6–7) during the 5 series of measurements (Repositioning). Middle horizontal line: mean, lower and upper horizontal lines: 95% coverage interval. The horizontal line is the overall mean of measurements

Site A

See Fig. 11

Site B

See Fig. 12

Site C

See Fig. 13

Discussion

Dose rate

Amongst the seven participating teams, four different types of detector were utilised: one ionization chamber, one proportional counter, two Geiger-Müller probes and three scintillators. The team factor is thus highly interfered by the detection principle of the device used. Indeed, ionization chambers, proportional counters, and Geiger-Müller devices are gas filled detectors, whereas scintillators have higher sensor densities (from 1 g/cm^3 for organic scintillator, 4.5 g/cm^3 for CsI to nearly 7 g/cm^3 for BGO crystals) as well as higher mean atomic number. These advantages mean that scintillator detectors can be smaller devices whilst maintaining a good performance for more sensitive measurements.

It is worth noting that for measurements at the lowest dose rate (location A), about half the natural background, the density of the detection medium can be seen to have a clear effect on the measurements. Namely, ionization chamber, proportional counter, and Geiger-Müller devices are gas filled detectors which show lower dose rate estimates than solid scintillator based probes. At location B, although at a lesser scale, this tendency can still be observed with gas probes showing lower dose rate estimates. It is nonetheless important to specify that for the two locations A and B, the standard deviation reported for the measurements by each team was also high and some of detectors were out of their measurement (team 2) or calibration (teams 3 and 4) ranges. Furthermore, whilst the detector used by team 5 showed the lowest dose rate estimate it is believed that this could be due to the fact that the detector was calibrated in terms of photon equivalent dose H_x , instead of Ambient dose equivalent ($H^*(10)$), potentially resulting in up to 5% underestimation of the measured dose rate. Team 7 detector may have the highest response time mentioned (about 60 s to be compared with the 30 s measurement time allowed for each data point) and is also the one measuring the highest dose rate but at this measurement spot A (team 7). The higher variability observed in the measurements of team 5 can be attributed to the higher response time for their detector, even when small signal variations are measured, especially if the necessary stabilisation time was not respected before reporting measurement values during consecutive measurements.

From the ANOVA analysis, it was possible to see that the most important component of the team factor arises from the variability of the detector used rather than the dexterity of the team operator. The apparent negligible effect of the

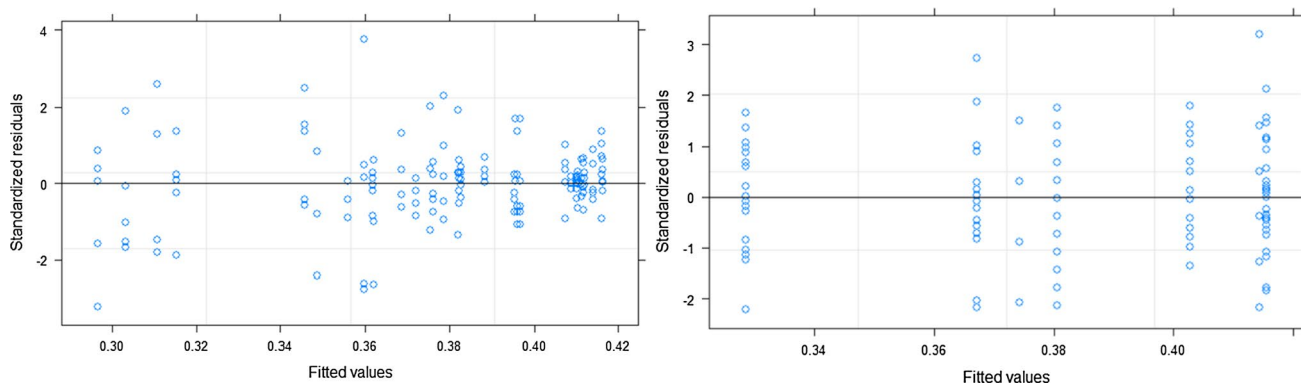


Fig. 8 Plot of residuals against fitted values under $\epsilon_{ijk} \sim N(0, \sigma^2)$ (left) and $\epsilon_{ijk} \sim N(0, \sigma_i^2)$ (right), for site A

Table 6 Estimates of variance components $\hat{\sigma}_{team}^2$, $\hat{\sigma}_{repos}^2$ and precision variance S_f^2 for each site

	Site A		Site B		Site C	
$\hat{\mu}$	0.383		1.034		3.06	
$\hat{\sigma}_{team}^2$	0.0011		0.0087		0.0469	
$\hat{\sigma}_{team}$	0.033	(8.6%)	0.093	(9.0%)	0.22	(7.1%)
$\hat{\sigma}_{repos}^2$	–	–				
$\hat{\sigma}_{repos}$	–	–	0.055	(5.3%)	0.038	(1.2%)
S_f^2	0.0011		0.011		0.048	
S_f	0.033	(8.6%)	0.11	(10%)	0.22	(7.2%)

Their squared root $\hat{\sigma}_{team}$, $\hat{\sigma}_{repos}$ and S_f are followed by their relative estimates with respect to the overall mean estimate $\hat{\mu}$, in parentheses. Dashes are displayed in place of estimates when the effect of the factor repositioning is not significant

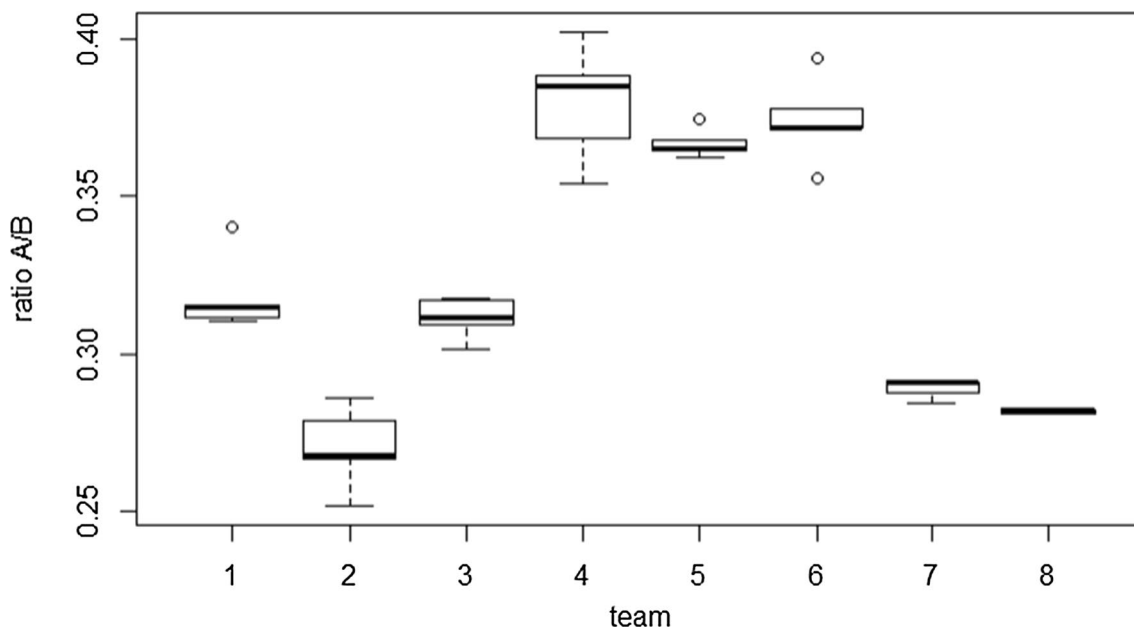


Fig. 9 Boxplot of the Total Gamma ratios A/B for each team

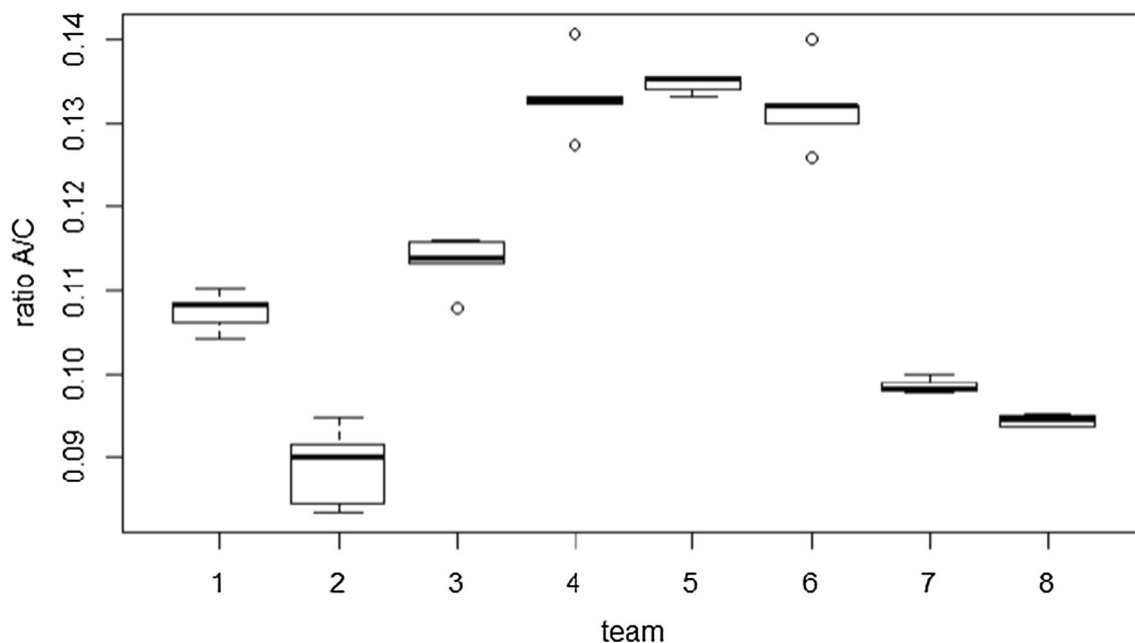


Fig. 10 Boxplot of the Total Gamma ratios A/C for each team

Table 7 Estimates of variance components $\hat{\sigma}_{\text{team}}^2$ for the ratios A/B and A/C

	A/B		A/C	
$\hat{\mu}$	0.324		0.113	
$\hat{\sigma}_{\text{team}}^2$	0.001897		0.000342	
$\hat{\sigma}_{\text{team}}$	0.044	(13%)	0.018	(16%)

repositioning factor at the lowest dose rate can be attributed to the fact that at such a low dose rates this factor is hidden by the higher relative dispersion of measurements, close to the measurement capabilities of the devices. The calculation of the relative deviation from consensus estimates for all teams further confirms the previously observed importance of the nature of the detection medium. Indeed, teams 1, 3, and—to some extent—7, tended to overestimate the contact dose rates. All three teams had scintillation detectors, i.e. a solid detection medium, which possesses a better performance and higher sensitivity than gas filled detectors at low energies.

Total gamma measurements

Relative deviations (Fig. 3) show that the two gas filled detectors (teams 4 and 6) tend to over-estimate total gamma surface activity. The relative deviations of scintillation detectors of similar density (team 7, NaI(Tl), $\sim 3.7 \text{ g/cm}^3$; team 2, ZnS, $\sim 4 \text{ g/cm}^3$; team 8, LaBr3, $\sim 5 \text{ g/cm}^3$) show comparable under-estimations, greater than 10%. As the energy response

of gas detectors is rather flat at all energies, it is possible that the apparent over-estimation is due to an overcorrection of the high sensitivity of low energy solid scintillation detectors. In addition to this, it is possible that the low energy photon field of the surface or even beta emissions may also have influenced the larger proportional counters, resulting in their overestimation of the total gamma count rates.

Whatever the chosen ratio (A/B or A/C), the reported values from team 5 are the highest. This is likely due to the fact that the detector was not shielded, resulting in a higher influence of the surrounding background to the total gamma radiation measured.

Three teams used surface contamination monitors (teams 2, 3, and 5). The sensitive medium of these monitors is a rectangular scintillator with a surface area of more than 300 cm^2 . These detectors are characterised by the fact that they are also suitable for the measurement of alpha and beta radiation (with the latter being particularly important in terms of the present study). According to ISO 8769, these surface contamination monitors must be calibrated using a—usually square shaped—surface source with an area of at least 100 cm^2 . As such, the calibration performed in the present study wherein a point source was used in contact with a large-surface detection medium can be seen as a source of measurement bias. In addition to this, the Cs-137 standard source also had a beta particle component (mostly of energy below 200 keV), whereas the surface to be measured contained a charged particle emission component mostly below 50 keV (Ba-133) but also a higher energy component (up to 550 keV for Eu-152). All these possible biases make it

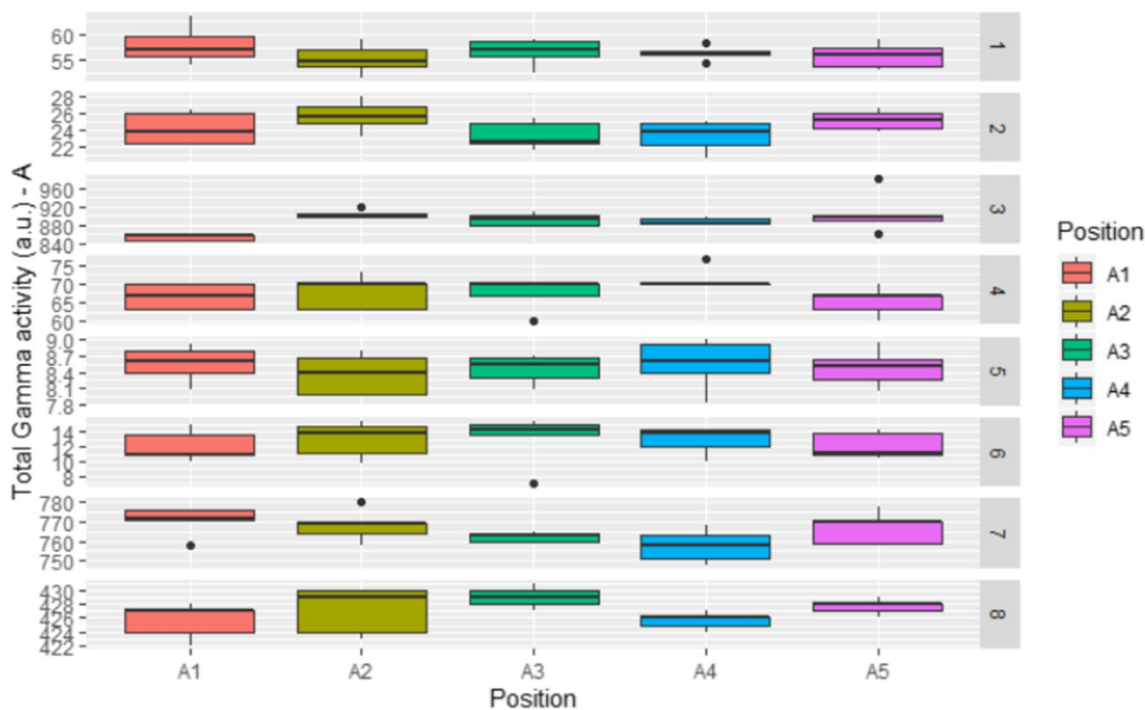


Fig. 11 Boxplots of Total Gamma measurements for each team at each random position A1–A2–A3–A4–A5 for site A

Table 8 Estimates of the variance components due to the repositioning effect for each team at site A

Team	1	2	3	4	5	6	7	8
$\hat{\mu}$	56.4	24.28	888	67.6	8.5	12.7	765.2	427.0
$\hat{\sigma}_{\text{reposit}}^2$	–	0.5976	349.7	2.079	–	–	16.22	1.156
$\hat{\sigma}_{\text{reposit}}$	–	0.77	19	1.4	–	–	4.0	1.1

Dashes are displayed in place of estimates when the effect of the factor repositioning is not significant

difficult to compare and interpret the results produced by surface contamination monitors.

Gamma spectrometry

Amongst the six teams, two types of detectors were used: a solid scintillator (LaBr₃(Ce), team 4) and semiconductor detectors. Of the semiconductor detector users it is worth noting that team 1 used a CZT detector, which is denser (~5.8 g/cm³) but significantly smaller in size (1 cm³) than the hyper-pure germanium detectors used by the other teams. Another important point is that, unlike the other detectors, the detector of team 3 was not collimated.

Aside from the volume and density of the detection materials used, the main bias between detectors was their energy resolution. The resolution of a detector is its ability to discriminate between peaks of similar energy. The resolution of HPGe detectors is usually expressed as the full width at half maximum (FWHM) of the 1332.5 keV energy peak of Co-60, whereas the energy resolution for a

scintillation detector is taken to be the percentage of the relative efficiency of the energy peak from Cs-137 (around 4% for LaBr₃, for instance). The resolution of HPGe detectors can be three times better than the resolution of other inorganic scintillators. However, this advantage is accompanied by greater constraints during field use as, unlike scintillators that can be used at room temperature, HPGe detectors must be cooled with liquid nitrogen. CZT detectors are considered to have a medium resolution, lying between those of inorganic scintillators and HPGe.

The uncertainties provided by the study's participants are underestimated as they only take into account the uncertainty in determining the area of the photoelectric peaks. As such, the better resolved detectors have significantly lower uncertainties than others. This is likely to bias the statistical performance criteria where measurement uncertainty is involved.

This effect can be seen from the results provided by team 4 which, using the lowest resolution detector (LaBr₃(Ce)), submitted the highest uncertainties. Team 4's depth limit

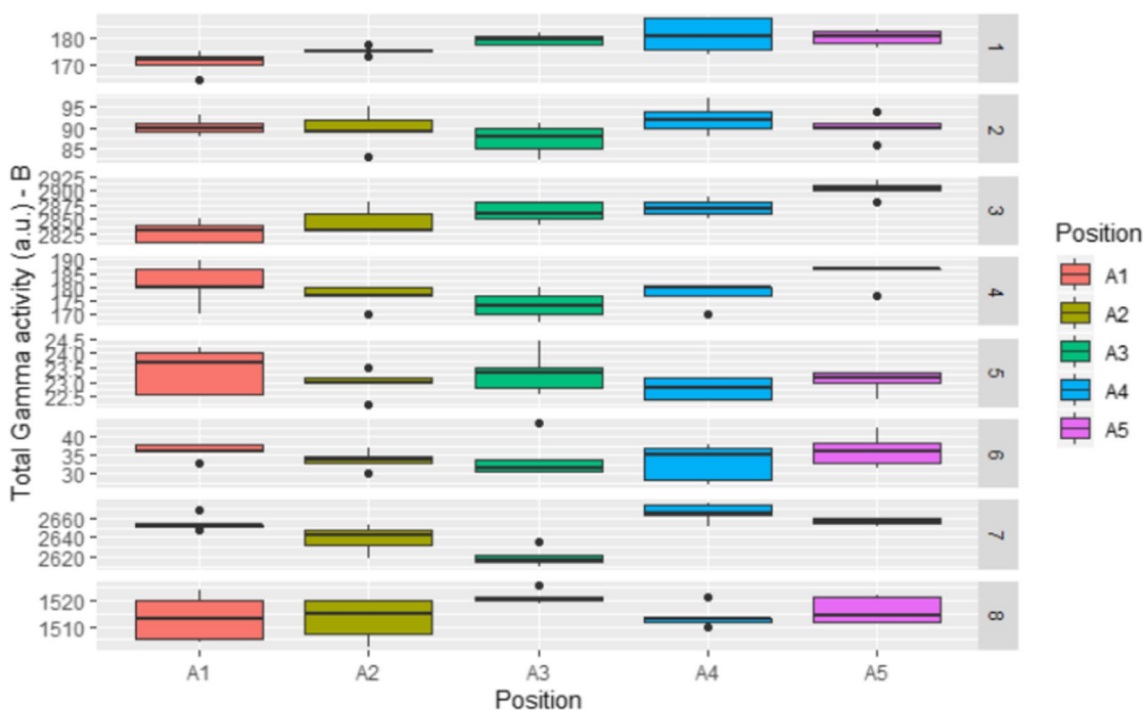


Fig. 12 Boxplots of Total Gamma measurements for each team at each random position A1–A2–A3–A4–A5 for site B

Table 9 Estimates of the variance components due to the repositioning effect for each team at site B

Team	1	2	3	4	5	6	7	8
$\hat{\mu}$	177.5	89.88	2862	178.4	23.108	34	2647	1515.4
$\hat{\sigma}_{\text{reposit}}^2$	16.06	0.908	713.4	13.66	0.0069	–	307.0	4.228
$\hat{\sigma}_{\text{reposit}}$	4.0	0.95	26	3.7	0.083	–	18	2.1

Dashes are displayed in place of estimates when the effect of the factor repositioning is not significant

estimate for Ba-133 (a low energy emitter) appears to be the most underestimated, this may be due to its lower sensitivity to low energy emissions (X-ray photons and electrons). Team 1, using the very small CZT detector, submitted larger uncertainties. Similarly to team 4, team 1's depth limit estimate for Ba-133 appears to be underestimated. This may be due to the very small size of the sensor. The results of team 7 using the unshielded HPGe detector are difficult to explain. It is interesting to note that the HPGe detector that provides the highest limiting depth for Ba-133 is the HPGe GL2020 detector optimised for low energy measurements. Finally, the two HPGe detectors with the closest dimensions (team 3 and team 5&6) submitted similar results.

Regarding the depth limit estimates for Eu-152, team 4, using the lowest resolution detector (LaBr3(Ce)), provided the highest uncertainties but this time its depth limit estimate is the highest. Team 1, using the very small CZT detector, also submitted large uncertainties and, similarly as for Ba-133, their depth limit estimate for Eu-152 appears underestimated. This may still be due to the very small size

of their sensor. The results of team 7 using the unshielded HPGe detector are once again difficult to explain, contrary to the previous low energy emitter depth limit estimation, team 7's estimate was the lowest of those for Eu-152. It is interesting to note that the HPGe GL2020 detector optimised for low energy measurements (team 2) that provided the highest limiting depth for Ba-133, also provided the highest estimate for Eu-152. Finally, the two HPGe detectors with the closest dimensions (team 3 and team 5&6) submitted similar results, confirming that they provide comparable results.

The results for the combination of Ba-133 and Eu-152 depth limits are very similar to those for depth limit estimates presented for Eu-152.

In estimating the ratio of the two isotopes of Eu, it is important to note that their emissions are different but sufficiently similar for spectral interferences to occur, especially when Co-60 is also present. The better resolved detectors are thus expected to provide better results. The detector that provides the lowest value remains the CZT (team 1), which is very small despite its medium resolution. The relative

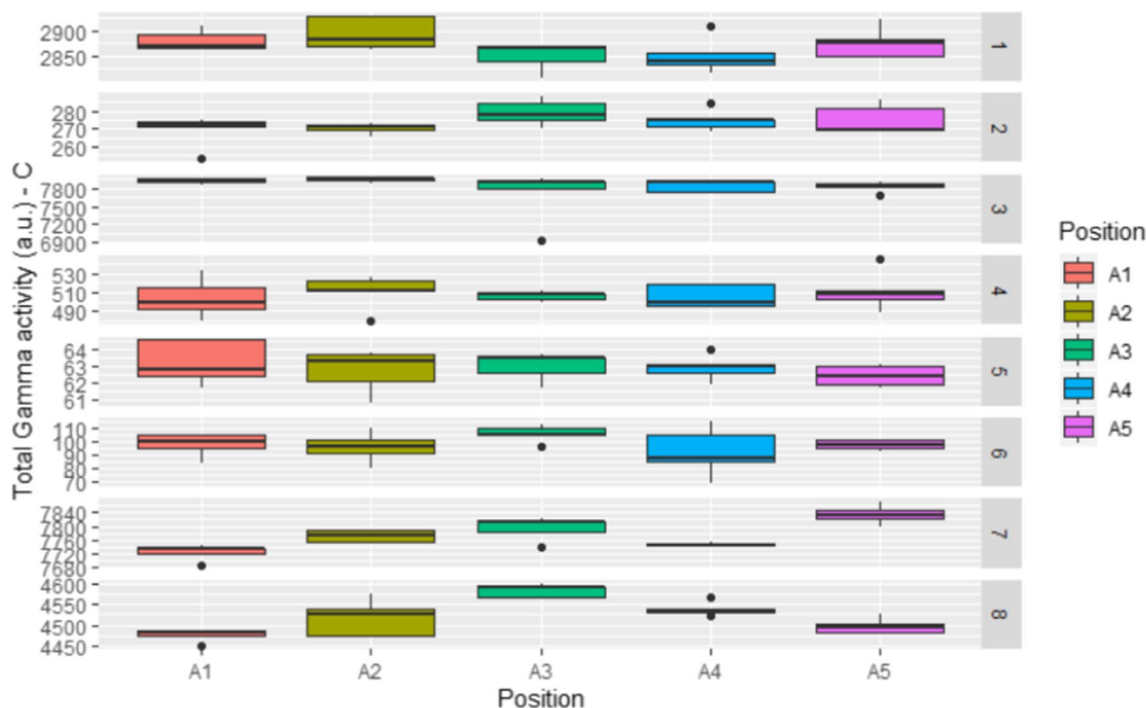


Fig. 13 Boxplots of Total Gamma measurements for each team at each random position A1–A2–A3–A4–A5 for site C

Table 10 Estimates of the variance components due to the repositioning effect for each team at site C

Team	1	2	3	4	5	6	7	8
$\hat{\mu}$	2870	273.5	7841	507.8	62.9	97.8	7774	4522
$\hat{\sigma}_{\text{repos}}^2$	210.1	6.064	877.6	–	–	1.736	1706	1515
$\hat{\sigma}_{\text{repos}}$	14	2.5	30	–	–	1.3	41	39

Dashes are displayed in place of estimates when the effect of the factor Repositioning is not significant

uncertainty provided is the highest (around 30%). Team 2, with its low-energy sensitive HPGe detector, provides an estimate close to the consensus value, with larger uncertainties than detectors of a similar type. Probably for this reason team 4, with its less resolved scintillation detector, provides the highest estimate of the ratio with a very large relative uncertainty ($\sim 27\%$). Again, team 3 and team 5&6 submit comparable results, using their similar detectors. The results of team 7 using the unshielded HPGe detector are close to the consensus value, but again it is difficult to comment.

For the measurement of Cs-137 surface activity, the results submitted by all participants are very scattered. Teams 3 and 5&6 submitted very low results which remained close to each other. Team 4 with its inorganic scintillator submitted a low value with large uncertainties, probably due to the low counting rate. Teams 1 and 2 submitted similar results both in terms of absolute values and relative uncertainties. As for the non-collimated detector, its estimate of surface activity was the highest, probably interfered by the

ambient radiation coming from the biological shield as a whole.

Conclusion

Despite the wide variety of detectors used, the majority of the parameters measured show consistent results. A limiting point in the statistical interpretation of the submitted results concerns the reported uncertainties, as they are known to be heavily underestimated. The low-level contamination measurements show the limitations of on-site measurements in a high noise environment wherein, notwithstanding collimations, the noise will often disturb the detector measurements.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10967-022-08319-9>.

Acknowledgements The authors would like to thank the participants in the comparison: Fernando Legarda and Raquel Idoeta from Basque Country University (UPV/EHU Spain), Jesús Ruiz from TECNATOM

(Spain), SCK-CEN (Belgium), Peter Volgyesi from MTA-EK (Hungary), Polina Otiougova from PSI (Switzerland), Lousai Leong from MIRION (France), Axel Klix from KIT (Germany).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

References

- IAEA (2004) International atomic energy agency, application of the concept of exclusion, exemption and clearance – safety guide. IAEA Safety Standards Series No. RS-G-1.7, Vienna
- NEA (2017) Radiological characterisation from a waste and materials end-state perspective: practices and experience
- NEA (2013) Radiological characterisation for decommissioning of nuclear installations
- IAEA (1998a) Radiological characterization of shut down nuclear reactors for decommissioning purposes. International Atomic Energy Agency, Vienna. IAEA technical report series n. 389
- IAEA (1998b) Characterization of radioactively contaminated sites for remediation purposes. International Atomic Energy Agency, Vienna. IAEA-TECDOC-1017
- IAEA (2007) Strategy and methodology for radioactive waste characterization. IAEA-TECDOC-1537 International Atomic Energy Agency, Vienna
- INSIDER (2017) The INSIDER Project. <http://insider-h2020.eu/>
- Aspe F, Idoeta R, Auge G, Herranz M (2020) Classification and categorization of the constrained environments in nuclear/radiological installations under decommissioning and dismantling processes. *Prog Nucl Energy*. <https://doi.org/10.1016/j.pnucene.2020.103347>
- Boden S, Cantrel E (2007) Pre-decommissioning radiological characterization of concrete. In: Proceedings of the 11th international conference on environmental remediation and radioactive waste management ICEM 2007, vol 1. Bruges, Belgium
- JCGM 200 (2012) International vocabulary of metrology – basic and general concepts and associated terms (VIM), 3rd edn, 2008 version with minor corrections
- Lourenço V, Crozet M, Demeyer S (2018) Interlaboratory comparisons organisation. INSIDER deliverable 6.2, <https://app.lgi-consulting.org/ecm/insider-ecm>
- Crozet M, Boden S, Herranz M, Branger T, Peerani P, Russel B (2018) Benchmarking (ILC on real sample) organization, INSIDER deliverable 6.1, <https://app.lgi-consulting.org/ecm/insider-ecm>
- Peerani P, Boden S, Crozet M, Zanolello F, Herranz M, Serrano R, Roudil D (2018) Design of the benchmarking exercise. Deliverable 2.5. <http://www.insider-h2020.eu>
- ISO 17043 (2010) Conformity assessment – general requirements for proficiency testing. International Organization for Standardization, Geneva, Switzerland
- Herranz M, Aspe F, Amgarou K (2020) Recommended in situ measurement techniques for each constrained environment, INSIDER deliverable 5.3, <https://app.lgi-consulting.org/ecm/insider-ecm>
- Boden S, Rogiers B, Verstrepen G, Mangelschots N, Broeckx W (2019) WP3 sampling strategy - sampling plan for use case 2: BR3 bioshield. Deliverable 3.5. http://insider-h2020.eu/wp-content/uploads/2019/04/INSIDER-D3.5_Sampling-plans-for-use-case-2.pdf
- ISO 17034 (2016) General requirements for the competence of reference materials producers. International Organization for Standardization, Geneva, Switzerland
- ISO Guide 35 (2006) Reference materials – general and statistical principles for certification. International Organization for Standardization, Geneva, Switzerland
- JCGM 100 (2008) Evaluation of measurement data — guide to the expression of uncertainty in measurement. GUM 1995 with minor corrections
- ISO 21748 (2017) Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation
- Boden S, Rogiers B, Dadoumont J, Noynaert L, Garcia-Miranda M, Roudil D (2018) R&D within the BR3 PWR decommissioning project, DEM 2018, Avignon, France
- Herranz M, et al (2021) Progress in nuclear energy. 137:103740
- Herranz M, Boden S, Abbas K, Idoeta R, Broeckx W, Ruiz J, Aspe F, Leong L, Otiougova P, Klix A, Volgyesi P, Peerani P, Legarda F (2020) Reporting on in situ measurement campaign, INSIDER deliverable 5.4, <https://app.lgi-consulting.org/ecm/insider-ecm>
- ISO 13528 (2015) ISO/TC69/SC6, statistical methods for use in proficiency testing by interlaboratory comparison. International Organization for Standardization, Geneva, Switzerland
- Knoll GF (2010) Radiation detection and measurement, 4th edn. p 864

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.