ICEM2011-59159
Waste Characterization

A Proposed Figure of Merit for Evaluating the Performance of Radiation Waste Measurement Systems

Ronald M. Keyser[1], Timothy R. Twomey[1], Neil A. Webster[2],
and Michael D. Belbot[2]

[1]ORTEC, 801 South Illinois Avenue, Oak Ridge, TN 37831 USA
[2]Thermo Fisher, One Thermo Fisher Way, Oakwood Village, OH 44146 USA

Keywords: FOM, Figure of Merit, detection limit

Abstract:

The instruments used to monitor the radioactive content of waste materials, either in container monitors, hand held radiation detectors, or mobile analysis systems, are constructed in widely different ways with widely varying detector materials and analysis software. However, within the various groupings (e.g., automatic waste monitors), all instruments are expected to solve the same problem, that is, to identify and quantify any radioactive material present according to the prescribed investigation methods. The best way to compare the performance of different instruments is with a numerical score or Figure of Merit (FOM). The FOM must quantify the performance of the instrument with respect to true positives (TP), false positives (FP), and false negatives (FN). The minimization of FN for certain radionuclides (e.g., uranium, plutonium, or other SNM) is more important than the minimization of FN for non-threat nuclides (e.g., low NORM). Likewise, the minimization of FP for SNM is also more important than falsely reporting the common NORM nuclides, which are of minimal concern in waste. The performance depends on the details of the testing, so the analysis conditions must also be included in the statement of the FOM. A FOM has been developed based on the number of true positives (TP), the number of false important positives (FIP), the number of FP, the number of true positives for SNM (TPSNM), and the number of false positives for SNM (FPSNM). This formula rates the overall performance with extra weight given to FP and FN for SNM. Examples will be shown for testing of test measurement systems with different types of detectors.

INTRODUCTION

The need for improved characterization of waste has resulted in the development of many different solutions designed to solve the same problem, that is, detect and quantify small amounts of specific nuclides in a measurement situation with wide variation in conditions, but with high reliability and minimal delay in the operations. In some cases the time for measurements needs to be short because of the large number of containers. Mixtures of nuclides are expected as well as variations in the sample matrix especially in decommissioning. For the majority of the potential users of these assay systems, it is

not possible to perform testing to determine the efficacy of a specific system in the expected conditions at the measurement point. To aid the users in the evaluation of systems, others have developed scoring methods for reporting performance [1]. In this work, we expand on the description of a scoring system (Figure of Merit or FOM) [2, 3] which is intended to result in a simple numerical score for each system in a prescribed test by comparing systems using longer counting times. The examples below include results with TP and FP in the FOM score.

Each test is the data collection, analysis, and reporting that results in either a negative (no activity) or a positive (activity) result. Once identified, the activity can then be calculated. There is one set of results per test which is the list, if any, of nuclides found during the test. For large waste containers, a test could consist of repeated measurements and the results reported as the average (or sum) activity of all the measurements. In this work, each test is only one measurement.

The result of the test consists of True Positives (TPs), False Positives (FPs), and False Negatives (FNs). A TP (at a specified confidence level) is used to characterize the waste assay system. Different thresholds define the different levels. The term "nuisance alarm" has been used in homeland security applications to describe the reporting of any nuclide that is actually present, but whose presence is not of concern. An example is ^{40}K, which is naturally occurring in many materials. Another possible report is "irrelevant TP" where the nuclide is present in the background and may be reported by some sensitive devices or long counting times and not reported in other conditions.

A TP is the reporting of a nuclide being detected when the nuclide is actually present in the test situation. A FP is the reporting of a nuclide being detected when the nuclide is not present. A FN is failure to report a nuclide, when the nuclide is actually present in the test situation. However, some TPs and FPs are more important than others. For example, a positive report of plutonium detection is more serious than a report of ^{40}K or other Naturally Occurring Radioactive Material (NORM).

In this FOM formula, the nuclides are divided into categories: Special Nuclear Material (SNM), Important, and Other. Different weighting factors are assigned to each category.

The main purpose of the instrument is to correctly identify and quantify the materials, that is, maximize the number of TPs. In any monitoring situation, FPs, meaning the presence of a nuclide, must be minimized because each FP can cause the mischaracterization of the container. In this calculation of the FOM, the FPs for some nuclides are given more weight than FPs for other nuclides. Likewise, FNs must also be minimized because this is a failure of the monitoring system to perform its primary function, which can also mischaracterize the waste, but in a different way.

The FOM is normalized by dividing by the value of the FOM for a perfect score and scaled by 10. That is, the weighted score for all the possible TPs in the test. This gives the best score as 10, but with no lower limit as FPs and FNs have negative factors. The exception is the null test with no activity in the container, where a perfect score is 0.

The intent is that a high FOM will mean that the system is performing better in the stated test conditions than one with a lower FOM. Here "better" includes all factors, including mischaracterizations due to FPs and FNs as well as the ability to detect the nuclides present in a variety of conditions. Many standards and test plans only require the determination of the identification of the nuclide without regard to the

number of false indications. In waste measurements, any false indication can mean either more expensive disposal or the undesired release of material into the "free release" waste stream.

Two different special situations are important: mixtures of nuclides, where a mixture of nuclides gives a response in the detector that mimics another nuclide, which may be of high interest, and masking of one nuclide by a large amount of another nuclide or nuclides. The performance when multiple nuclides are in the test mixture is accounted for in the scoring as the combination of TPs, FPs, and FNs. The masking situation is different in that the masking nuclide(s) could be NORM, which is often detected in many common materials. To alert the user to a possible masking situation, specific nuclides, such as NORM, can have multiple levels of concern: one level for low amounts and a higher level for masking situations. This will allow small quantities of the nuclide to pass without identification (not counted as either TP or FN), but will be identified if the level is above the higher threshold (should be TP). If the test condition contains a high level of a "masking nuclide", the result is a TP if the detected level of the nuclide is above a limit and a FN if the nuclide is not reported (i.e., no identification). For these tests, "high NORM" is defined as the emissions from 3 tonnes of potassium chloride (KCl) stacked uniformly on the entire bottom of a single large waste container.

To properly compare the performance of different systems, the details of the test conditions must be included with the FOM. For example, the length of time for data collection should be included in the results description.

Table I shows the categories and the nuclides in each category. The exact nuclides and classification is not important, but must be the same for all scoring in order to compare the results of different tests. In the situation where a table entry (e.g., elevated uranium) is reported as a result of the detection of a specific isotope (i.e., ^{235}U) only one result is included in the FOM. In this example, however, if elevated uranium was reported with no other uranium isotope reported, it would be counted as a TP or FP according to the test situation.

Table I

SNM	**Important**	**Other**
^{233}U	^{57}Co	^{18}F
^{235}U	^{60}Co	^{67}Ga
^{237}Np	^{133}Ba	^{51}Cr
Pu (^{239}Pu)	^{137}Cs	^{75}Se
Elevated uranium	^{192}Ir	^{89}Sr
U^{232}/Th^{228}	^{204}Tl	^{99m}Tc
SNM	^{241}Am	^{103}Pd
		^{111}In
		^{123}I
		^{125}I
		^{131}I
		^{153}Sm
		^{201}Tl
		^{133}Xe
	High NORM	Natural (NORM)
		^{40}K
		*^{226}Ra
		^{232}Th and daughters
		^{238}U and daughters

Table II shows the weighting factors for the different results.

Table II Weighting Factors					
SNM		**Important**		**Other**	
TPs	FPs & FNs	TPi	FPi & FNi	TPo	FPo & FNo
4	-4	3	-3	1	-1

The FOM is calculated in three components as shown in the equations below:

$$FOM(TP) = 10 * \frac{\sum_{SNM\ found} TPs + \sum_{Imp\ found} TPi + \sum_{Other\ found} TPo}{\sum_{SNM\ Actual} TPs + \sum_{Imp\ actual} TPi + \sum_{O\ actual} TPo}$$

$$FOM(FP) = 10 * \frac{\sum_{SNM\ not\ actual} FPs + \sum_{Imp\ not\ actual} FPi + \sum_{Other\ not\ actual} FPo}{\sum_{SNM\ Actual} TPs + \sum_{Imp\ actual} TPi + \sum_{O\ actual} TPo}$$

$$FOM(FN) = 10 * \frac{\sum_{SNM\ actual\ missed} FNs + \sum_{Important\ actual\ missed} FNi + \sum_{Other\ actual\ missed} FNo}{\sum_{SNM\ Actual} TPs + \sum_{Imp\ actual} TPi + \sum_{O\ actual} TPo}$$

With the total FOM as the sum of these three:

$$FOM = FOM(TP) + FOM\ (FP) + FOM(FN)$$

EQUIPMENT AND SETUP

The test results shown are based on measurements from two systems: one using germanium (HPGe) detectors and one using sodium iodide (NaI). The HPGe system consisted of two columns of 4 detectors each on opposite sides of the container. The NaI system consists of two columns each with 8 large volume NaI detectors each on opposite sides of the container. The detectors and arrangement are described in [2, 3, 4].

Figure 1 Example system with HPGe detectors.

An example of a system using HPGe detectors is shown in Fig. 1. Only 2 detectors on each side are used in this system. This is a commercial system designed for waste container characterization installed at three SOGIN decommissioning sites located at Caorso, Trino and Latina, Italy and in operation since April, 2010 [5].

To compare the two detector types, tests were made using ^{133}Ba and ^{60}Co sources mounted, individually, inside the center of the cargo section of a standard 8 m box body truck. Measurements were made with and without shielding (to simulate the absorbing matrix in the container) and NORM (to simulate the expected background in many containers). The truck was stationary with the source centered horizontally and vertically in the plane of the detectors. The two systems were tested sequentially. The shielding was solid steel plates to simulate a high-density matrix in the waste container.

Source and attenuators

For the examples here, only ^{133}Ba was selected, both with and without NORM. The ^{133}Ba source closely resembles ^{239}Pu, an SNM material. The source is a point source and was mounted in a wooden frame. The attenuators were 3.2 mm thick steel plates which could be combined to give total thicknesses of 0.32 to 10 cm. The attenuators were placed so as to completely shield the detectors from the source. The ^{133}Ba source was 5.4 MBq.

The NORM was placed evenly on the floor of the cargo section of the truck to give an increased background, but was piled low enough so that it did not block the direct path for gamma rays between the sources and the detectors. The NORM was commercial water softener salt substitute or KCl and zirconium sand. About 3 tonne of KCl in 18 kg bags and 713 kg of sand in 23 kg bags on wooden pallets were in the truck.

The ^{133}Ba source was placed in the container and positioned so that the source was in the middle of the detection zone. Two measurements were made for each system and shielding-NORM combination: 2 minutes and 5 minutes counting time. The shielding thicknesses were from 1.27 to 4.5 cm. The nuclides detected were recorded and the FOM calculated for each test using the nuclides in Table I and the factors in Table II.

RESULTS

To show the efficacy of this proposed FOM method vs just the true positives, that is, only the detection of ^{133}Ba, the two calculations are shown in Fig. 2 for the case of no NORM and 120 second count. The FOM (TP) and FOM (TP +FP) are plotted as a function of steel shielding thickness for both detectors. The reporting of TP only is a requirement in many standards and specifications currently in use. In this case, only ^{133}Ba is accepted as a TP Important.

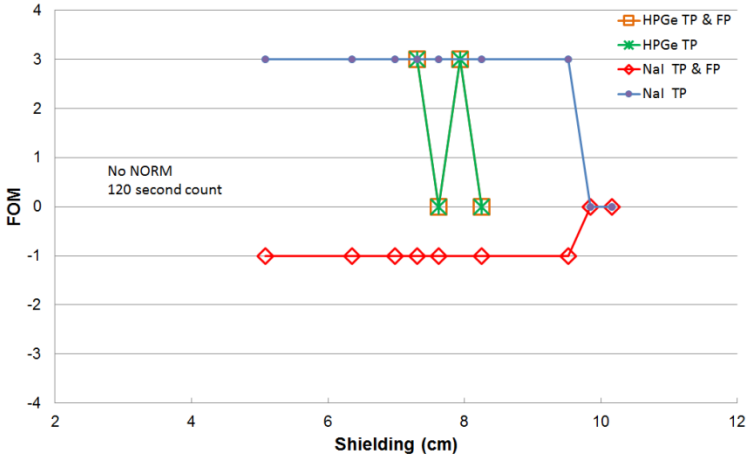

Figure 2. Plot of ^{133}Ba FOM vs Steel Shielding Thickness for Two Counters Comparing Results for True Positives only and True Positives + False Positives.

Note that the HPGe detector system does not report false positives in this case, so the TP only and TP + FP results are the same. The NaI system reports 133Ba with more shielding, but also reports FPs resulting in a lower FOM (TP+FP).

Figure 3 shows the advantage of high resolution in the identification of nuclides in the presence of background radiation. In the high NORM case, the HPGe detector system is able to identify 133Ba at shielding thicknesses up to 6 cm in 120 seconds.

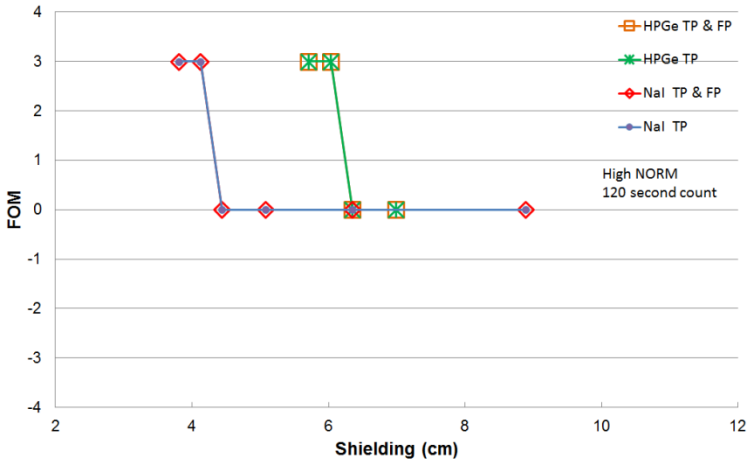

Figure 3. Plot of ^{133}Ba FOM vs Steel Shielding Thickness with High NORM for Two Counters Comparing Results for True Positives only and True Positives + False Positives.

Waste containers are often counted for longer time periods to improve the detection ability. Figure 4 shows the same situation as Fig. 2, but with a count time of 300 seconds. The increased counting time improves the detection ability of both systems, but does not change the general character of the response except for the detection of uranium in the HPGe system.

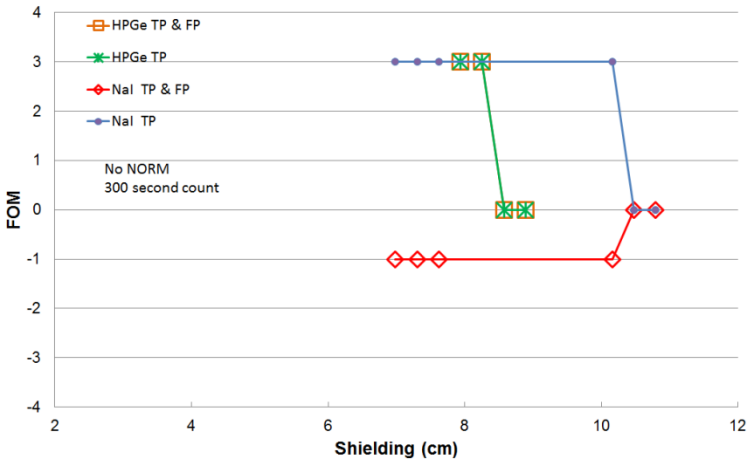

Figure 4. Plot of ^{133}Ba FOM vs Steel Shielding Thickness for Two Portal Monitors Showing Results for the FOM Including TPs and FPs with Longer Count Times.

Figure 5 shows the results for the 300 second count with high NORM. In this case the response has clearly changed. The high NORM has increased the background counts for the NaI system to effectively mask the 133Ba signal when the shielding is increased above 4.2 cm. The HPGe system detection limit is reduced from 10 cm to 6.7 cm.

DISCUSSION

The measurement of waste materials is generally limited to the detection and quantification of a specific list of nuclides. The list is generally small and is based on the known sources of the waste and the processes involved at the production facility. The classification of waste can be based on the total activity in the container or on the activity of each nuclide in the list. If a nuclide of interest is mistakenly reported as present (FP) and an activity reported, the characterization of the waste could be incorrect causing it to be dispositioned incorrectly.

The NaI system shown here exhibits a good ability to identify the test nuclide with increasing shielding (simulating denser matrices), however the poorer resolution gives a high number of FPs causing the FOM to be reduced. Increasing the counting time also increases the performance. This is due to the increased signal and the reduction in uncertainty in the spectrum channel counts. The response to increased background is a reduction in detection ability due to the decrease in signal-to-noise ratio.

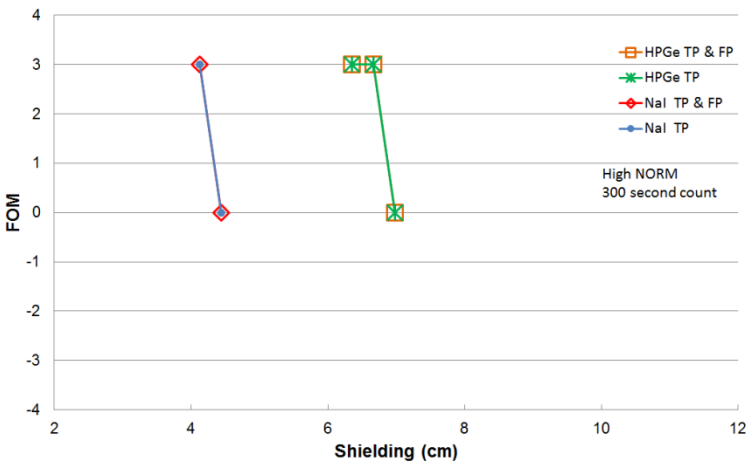

Figure 5. Plot of ^{133}Ba FOM vs Steel Shielding Thickness for Two Portal Monitors Showing Results for the FOM Including TPs and FPs with Longer Count Times and NORM.

The HPGe system shows a lower shielding thickness limit in the ability to identify the test nuclide, but the very good resolution reduces the number of FPs to almost none. This combination gives a higher FOM up to the limit where both systems fail to report any nuclides in the sample. The increasing count time improves the performance. The one FP (Fig. 4) reported for HPGe was due to reporting uranium in the 300 second count time spectrum. The increased background also reduces the signal-to-noise ratio in the spectrum, but the reduction in detection ability is not as much as in the NaI system due to the better resolution of the HPGe detectors.

Increasing the threshold for detection was not considered a valid way to improve performance because this would increase the detection limit, thus increasing the activity reported for each container.

CONCLUSIONS

The FOM described in this paper has been shown to provide data that assists in the choice of instrumentation to meet measurement requirements, including the identification of nuclides present and considering the impact of false reporting of nuclides not present. This FOM accounts not only for the positive detection of nuclides present, but also the false reporting of nuclides not present, making it useful in the comparison of differing systems designed to solve the same problem. It also provides a useful indication of the extent to which the efficacy of a given system will be impacted by FPs, which can cause increased disposal costs. This FOM applies to any monitoring test comparison, because it only uses the results reported by the instrument. In using the FOM, the complete details of the test situation must be given and be as consistent as possible for the different instruments. Future work will include work with different radionuclides and measurement scenarios. These data will assist in the correct choice of instruments for different performance needs.

REFERENCES

[1] Leonid Kagan and Andrei Stavrov, "Methodology of Testing Vehicle Radiation Portal Monitors Considering Field Operation Conditions," INMM Annual Meeting Proceedings, July 2010

[2] Ronald M. Keyser, Neil A. Webster, Michael D. Belbot, and Timothy R. Twomey, "Figure of Merit for Evaluating the Performance of Radionuclide Identification in Portal Monitors and Handheld Devices," Proceedings of the 33rd ESARDA Conference, May 2011, Budapest, Hungary

[3] Ronald M. Keyser, Neil A. Webster, Michael D. Belbot, and Timothy R. Twomey, "A Figure of Merit for Comparing the Performance of Different Radiation Identification and Detection Systems Used in Illicit Trafficking Control," INMM Annual Meeting Proceedings, July 2011

[4] Ronald M. Keyser and Timothy R. Twomey, "Performance of a Radiation Portal Freight Monitor based on Integrated Germanium Detector Modules," INMM Annual Meeting Proceedings, July 2010

[5] Timothy R. Twomey, Ronald M. Keyser, Davide Sacchi , Giuseppe Merlino, Paolo Grignani, Gianluigi Migliore, Nunzia Radatti, and Sabrina Romani, "Radiological Performance of an Automated HPGe Assay System for Bulk Containers of Decommissioning Waste Intended for Free Release," INMM Annual Meeting Proceedings, July 2010