SAND2016-2854R

Unclassified

# Algorithm Improvement Program

## Nuclide Identification Algorithm Scoring Criteria And Scoring Application

Domestic Nuclear Detection Office

*February 2016*

Homeland Security

Sandia National Laboratories

Unclassified

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

U.S. DEPARTMENT OF ENERGY

THIS PAGE INTENTIONALLY LEFT BLANK

**This document was prepared for the**

**U.S. Department of Homeland Security (DHS)**

**Domestic Nuclear Detection Office (DNDO)**

**Under Contract HSHQDC12-X-000354**

THIS PAGE INTENTIONALLY LEFT BLANK

# Approval

## Submitted

| | | | |
|---|---|---|---|
| Michael Enghauser | Date | Name | Date |
| Principal Member of Technical Staff | | Title | |
| Sandia National Laboratories | | Office | |

## Reviewed

| | | | |
|---|---|---|---|
| Gregory C. Slovik | Date | Name | Date |
| Technical Director | | Title | |
| DNDO | | Office | |

## Approved

| | | | |
|---|---|---|---|
| David Chu | Date | Name | Date |
| Program Manager | | Title | |
| DNDO | | Office | |

# Record of Changes

| Version | Date | Modified By | A = Add.<br>M = Mod.<br>D = Del. | Change Description:<br>Page, Table, Figure, Paragraph |
|---|---|---|---|---|
| 0.00 | 2/9/2015 | | | Initial Version |
| 1.00 | 2/2/2016 | Michael Enghauser | A | Section 2: Added “Paired observations” definition |
| 1.00 | 2/2/2016 | Michael Enghauser | M | Table 17 modified |
| 1.00 | 2/2/2016 | Michael Enghauser | A | Section 2.5: Added discussion on how the application interprets multiple nuclide decay chain equivalencies |
| 1.00 | 2/2/2016 | Michael Enghauser | M | Section 3: Modified to reflect that the scoring application has been updated to consider nuclide identification confidence indices |
| 1.00 | 2/2/2016 | Michael Enghauser | M | Section 3.2: Replaced “Scoring Application Nonparametric Statistical Comparisons” section with “Scoring Application Algorithm Performance Comparisons” section |
| 1.00 | 2/2/2016 | Michael Enghauser | A | References: Added applicable references |
| 1.00 | 2/2/2016 | Michael Enghauser | A | Appendix A: Added appendix for Percentile Bootstrap Confidence Interval Determinations |

# Executive Summary

The goal of the Domestic Nuclear Detection Office (DNDO) Algorithm Improvement Program (AIP) is to facilitate gamma-radiation detector nuclide identification algorithm development, improvement, and validation. Accordingly, scoring criteria have been developed to objectively assess the performance of nuclide identification algorithms. In addition, a Microsoft Excel spreadsheet application for automated nuclide identification scoring has been developed.

This report provides an overview of the equations, nuclide weighting factors, nuclide equivalencies, and configuration weighting factors used by the application for scoring nuclide identification algorithm performance. Furthermore, this report presents a general overview of the nuclide identification algorithm scoring application including illustrative examples.

# Contents



# Figures



# Tables

# 1 Introduction

The goal of the Domestic Nuclear Detection Office (DNDO) Algorithm Improvement Program (AIP) is to facilitate gamma-radiation detector nuclide identification algorithm development, improvement, and validation. Accordingly, scoring criteria have been developed to objectively assess the performance of nuclide identification algorithms. In addition, a Microsoft Excel spreadsheet application for automated nuclide identification scoring has been developed.

This report provides an overview of the equations, nuclide weighting factors, nuclide equivalencies, and configuration weighting factors used by the application for scoring nuclide identification algorithm performance. Furthermore, this report presents a general overview of the nuclide identification algorithm scoring application including illustrative examples.

# 2 Scoring Application and Definitions

To assist with understanding the scoring application equations presented in this report, the following definitions are provided:

- True Positive (tp): Nuclide reported by the algorithm that is present.
- False Positive (fp): Nuclide reported by the algorithm that is not present.
- False Negative (fn): Nuclide not reported by the algorithm that is present.
- Paired observations: Helsel [1] describes paired observations for two-groups or two data sets as “both groups have the same number of observations, and the first observation in the group is linked to the first observation in the second group. Similarly, the second observation in the first group is linked to the second observation in the second group, the third with the third, and so on.”

## 2.1 Scoring Application Equations

The fundamental equation used to evaluate nuclide identification algorithm performance is based on F-scores. F-scores are a statistical method for determining accuracy by utilizing precision (p) and recall (r). For more detailed information on F-scores, please see the reference “The truth of the F-measure” [2].

In general terms, precision is the fraction of nuclides reported by an algorithm that should have been reported. For example, if a nuclide identification algorithm has a calculated precision of 0.9, then 90% of the nuclides reported by the algorithm were correct. Eq. (1) illustrates the computation of precision.

$$p = \frac{tp}{tp+fp} \quad (1)$$

Similarly, recall is related to the fraction of nuclides not reported by an algorithm that should have been reported. For example, if a nuclide identification algorithm has a calculated recall of 0.8, then 0.2 (1.0 minus 0.8) or 20% of the nuclides were not reported by the algorithm that should have been reported. Eq. (2) illustrates the computation of recall.

$$r = \frac{tp}{tp+fn} \tag{2}$$

Once precision and recall are determined, the F-score (F) is determined by calculating the harmonic mean of precision and recall, Eq. (3).

$$F = 2 * \frac{p*r}{p+r} \tag{3}$$

For evaluating nuclide identification algorithm performance, the harmonic mean is used since it provides a more accurate representation than the arithmetic mean when averaging rates such as precision and recall [2]. To demonstrate the suitability of using the harmonic mean for evaluating algorithm performance, the following example is provided.

**Table 1. Harmonic Mean Versus Arithmetic Mean Example**

| | Algorithm Results |
|---|---|
| Precision | 0.05 |
| Recall | 1.00 |
| F-score (Arithmetic mean) | 0.53 |
| F-score (Harmonic mean) | 0.10 |

In the example, the algorithm liberally reports numerous nuclides resulting in very low precision, due to a high fraction of false positives, and very high recall, due to a low fraction of false negatives. Intuitively, the performance of the algorithm should be very low since nearly all of the nuclides reported are incorrect rendering the algorithm practically useless. However, the F-score arithmetic mean is an unrealistic value of 0.53 while the F-score harmonic mean is an appropriate value of 0.10. For more detailed information on the harmonic mean, please see the reference “The truth of the F-measure” [2].

## 2.2 Scoring Application Nuclide Weighting Factors

To utilize a scoring scale of zero to 100, the equations for precision and recall were multiplied by 100. In addition, the traditional F-score formula was augmented to allow the use of nuclide weighting factors (WF) based on nuclide importance. Although default scoring application nuclide weighting factors have been assigned (see Table 2), the scoring application has been programmed to allow default nuclide weighting factors to be changed easily to meet the goals and objectives of a given test campaign.

**Table 2. Default Weighting Factors (WF) for High, Medium, and Low Importance Nuclides**

| Nuclide Category | Example | tp WF | fp WF | fn WF |
|---|---|---|---|---|
| High Importance | U-235 in HEU | 4 | 2 | 4 |
| Medium Importance | Shielded Ir-192 | 2 | 1 | 2 |
| Low Importance | K-40 in Fertilizer | 1 | 1 | 1 |

As shown in Table 2, false negatives (failures to correctly identify nuclides present) are deemed more serious than false positives (reporting nuclides that are not present) for medium and high importance nuclides. Accordingly, the assigned default false positive weighting factor is lower than the assigned default true positive and false negative weighting factors.

In addition to the nuclide weighting factors presented in Table 2, special categories were assigned for algorithms that provide uranium and plutonium type determinations and algorithms that correctly identify difficult to detect nuclides present in trace quantities. For example, if plutonium is present, a nuclide identification algorithm that correctly identifies the plutonium type (WGPu or RGPu) is preferable to one that does not. Similarly, a nuclide identification algorithm that identifies the presence of trace nuclides is preferable to one that does not. Consequently, the default "rewards" and "penalties" shown in Table 3 have been assigned for detection of trace nuclides and for algorithms that report uranium and plutonium material types.

**Table 3. Default Weighting Factors for Reported Uranium and Plutonium Type Determinations and Trace Nuclides**

| Reward (+) or Penalty (-) | Example | tp WF | fp WF | fn WF |
|---|---|---|---|---|
| (+) | WGPu or RGPu correctly identified | 0.5 | 0 | 0 |
| (-) | WGPu or RGPu incorrectly identified | 0 | 0.5 | 0 |
| (+) | Ir-194m$^2$ with Ir-192 | 0.5 | 0 | 0 |

*Note: Ir-194m$^2$ represents the second metastable state of Ir-194.*

To further define how the "reward" and "penalty" system is used for detection of trace nuclides and for algorithms that report uranium and plutonium material types, applicable examples are provided in the following section, 2.3 Scoring Application Examples.

An additional category for "nuclides" that are "not applicable" was also assigned. As shown in Table 4, this category assigns a value of zero to each of the weighting factors which effectively removes the reported "nuclide" from scoring. This is currently assigned to "Annihilation" when nuclides have readily identifiable gamma emissions in addition to annihilation radiation (e.g., Na-22, Ge-68/Ga-68, and Sr-82/Rb-82). For example, if a nuclide identification algorithm reports "Annihilation", "Eu-154", and "Na-22" when Na-22 is present, the scoring application will remove "Annihilation" and score the algorithm considering only "Eu-154" and "Na-22" as reported nuclides.

**Table 4. Default Weighting Factors for Nuclides That Are Not Applicable**

| Nuclide Importance | Example | tp WF | fp WF | fn WF |
|---|---|---|---|---|
| Not Applicable | Annihilation | 0 | 0 | 0 |

## 2.3 Scoring Application Examples

To assist in understanding the formulas and principles used by the scoring application, four general examples are presented for illustrative purposes. Information relevant to the four examples is supplied in Tables 5 and 6.

**Table 5. Nuclide Weighting Factors**

| Nuclide | Importance | tp WF | fp WF | fn WF |
|---|---|---|---|---|
| Cs-137 | Low | 1 | 1 | 1 |
| Ga-67 | Low | 1 | 1 | 1 |
| Np-237 | High | 4 | 2 | 4 |

**Table 6. Nuclides Present and Reported by the Identification Algorithm**

| Example | Nuclides Present | Nuclide(s) Reported |
|---|---|---|
| 1 | Ga-67, Cs-137 | Np-237, Ga-67 |
| 2 | Ga-67, Cs-137 | Cs-137 |
| 3 | Np-237, Cs-137 | Np-237, Ga-67 |
| 4 | Np-237, Cs-137 | Cs-137 |

For Example 1 shown in Table 6, the nuclide identification algorithm correctly reported Ga-67, incorrectly reported Np-237, and did not report Cs-137. Using Table 5, the appropriate weighting factors for Example 1 are: tp WF = 1 for Ga-67; fp WF = 2 for Np-237; and fn WF = 1 for Cs-137. Accordingly, the precision, recall, and F-score are calculated on a scoring scale of zero to 100 as follows:

$$100 * p = 100 * \left(\frac{tp}{tp+fp}\right) = 100 * \left(\frac{1}{1+2}\right) = 33.3 \quad (4)$$

$$100 * r = 100 * \left(\frac{tp}{tp+fn}\right) = 100 * \left(\frac{1}{1+1}\right) = 50.0 \quad (5)$$

$$F = 2 * \left(\frac{p*r}{p+r}\right) = 2 * \left(\frac{33.3*50.0}{33.3+50.0}\right) = 40.0 \quad (6)$$

In a similar fashion, precision, recall, and F-scores were calculated for Examples 2 through 4 with results summarized in Table 7.

**Table 7. Calculated Precision, Recall, and F-scores**

| Example | Nuclides Present | Nuclide(s) Reported | Precision | Recall | F-score |
|---|---|---|---|---|---|
| 1 | Ga-67, Cs-137 | Np-237, Ga-67 | 33.3 | 50.0 | 40.0 |
| 2 | Ga-67, Cs-137 | Cs-137 | 100.0 | 50.0 | 66.7 |
| 3 | Np-237, Cs-137 | Np-237, Ga-67 | 80.0 | 80.0 | 80.0 |
| 4 | Np-237, Cs-137 | Cs-137 | 100.0 | 20.0 | 33.3 |

A review of Table 7 shows the impact of incorrectly and correctly identifying Np-237, a Special Nuclear Material (SNM) nuclide with high importance.

An additional two examples are provided to aid in the understanding on how to compute the "reward" and "penalty" for the detection of trace nuclides. Information relevant to these two examples is supplied in Tables 8 and 9.

**Table 8. Nuclide Weighting Factors**

| Nuclide | Importance | tp WF | fp WF | fn WF |
|---|---|---|---|---|
| Ir-192 | Medium | 2 | 1 | 2 |
| Ir-194m$^2$ | Trace (+) | 0.5 | 0 | 0 |
| K-40 | Low | 1 | 1 | 1 |

**Table 9. Nuclides Present and Reported by the Identification Algorithm**

| Example | Nuclides Present | Nuclides Reported |
|---|---|---|
| 5 | Ir-192, Ir-194m$^2$ (Trace) | Ir-192, Ir-194m$^2$, K-40 |
| 6 | Ir-192, Ir-194m$^2$ (Trace) | Ir-192, K-40 |

For Example 5 shown in Table 9, the nuclide identification algorithm correctly reported Ir-192 and Ir-194m$^2$ and incorrectly reported K-40. Using Table 8, the appropriate weighting factors for Example 5 are: tp WF = 2 for Ir-192, tp WF = 0.5 for Ir-194m$^2$; and fp WF = 1 for K-40. Accordingly, the precision, recall, and F-score are calculated on a scoring scale of zero to 100 as follows:

$$100 * p = 100 * \left(\frac{tp}{tp+fp}\right) = 100 * \left(\frac{2+0.5}{(2+0.5)+1}\right) = 71.4 \quad (7)$$

$$100 * r = 100 * \left(\frac{tp}{tp+fn}\right) = 100 * \left(\frac{2+0.5}{(2+0.5)+0}\right) = 100.0 \quad (8)$$

$$F = 2 * \left(\frac{p*r}{p+r}\right) = 2 * \left(\frac{71.4*100.0}{71.4+100.0}\right) = 83.3 \quad (9)$$

In a similar manner, the precision, recall, and F-score were calculated for Example 6 with the results summarized in Table 10.

**Table 10. Calculated Precision, Recall, and F-scores When a Trace Nuclide is Present**

| Example | Nuclides Present | Nuclide(s) Reported | Precision | Recall | F-score |
|---|---|---|---|---|---|
| 5 | Ir-192, Ir-194m2 (Trace) | Ir-192, Ir-194m2, K-40 | 71.4 | 100.0 | 83.3 |
| 6 | Ir-192, Ir-194m2 (Trace) | Ir-192, K-40 | 66.7 | 100.0 | 80.0 |

As shown in Table 10, the algorithm that correctly identified the trace nuclide, Ir-194m[2], is "rewarded" with a better F-score.

Three final examples are provided to assist in understanding the "reward" and "penalty" system for algorithms that report uranium and plutonium material types. Information relevant to the three examples is displayed in Tables 11 and 12.

**Table 11. Nuclide Weighting Factors**

| Nuclide | Importance | tp WF | fp WF | fn WF |
|---|---|---|---|---|
| Pu-239 | High | 4 | 2 | 4 |
| Am-241 | Low | 1 | 1 | 1 |
| Pu-241 | Low | 1 | 1 | 1 |
| K-40 | Low | 1 | 1 | 1 |
| WGPu | (+) | 0.5 | 0 | 0 |
| RGPu | (-) | 0 | 0.5 | 0 |

**Table 12. Nuclides Present and Reported by the Identification Algorithm**

| Example | Nuclides Present | Nuclides Reported |
|---|---|---|
| 7 | Pu-239(H), Pu-241(L), Am-241(L), WGPu(+) | Pu-239, Pu-241, Am-241, RGPu, K-40 |
| 8 | Pu-239(H), Pu-241(L), Am-241(L), WGPu(+) | Pu-239, Pu-241, Am-241, K-40 |
| 9 | Pu-239(H), Pu-241(L), Am-241(L), WGPu(+) | Pu-239, Pu-241, Am-241, WGPu, K-40 |

For Example 7 shown in Table 12, the nuclide identification algorithm correctly reported Pu-239, Pu-241, and Am-241 and incorrectly reported K-40 and RGPu. Using Table 11, the appropriate weighting factors for Example 7 are: tp WF = 4, 1, and 1 for Pu-239, Pu-241, and Am-241, respectively; and fp WF = 1 and 0.5 for K-40 and RGPu, respectively. Accordingly, the precision, recall, and F-score are calculated on a scoring scale of zero to 100 as follows:

$$100 * p = 100 * \left(\frac{tp}{tp+fp}\right) = 100 * \left(\frac{4+1+1}{(4+1+1)+(1+0.5)}\right) = 80.0 \qquad (10)$$

$$100 * r = 100 * \left(\frac{tp}{tp+fn}\right) = 100 * \left(\frac{4+1+1}{(4+1+1)+0}\right) = 100.0 \quad (11)$$

$$F = 2 * \left(\frac{p*r}{p+r}\right) = 2 * \left(\frac{80.0*100.0}{80.0+100.0}\right) = 88.9 \quad (12)$$

In a similar manner, precision, recall, and F-scores were calculated for Examples 8 and 9, shown in Table 12, with results summarized in Table 13.

**Table 13. Calculated Precision, Recall, and F-scores With Plutonium Type Reported**

| Example | Nuclides Present | Nuclide(s) Reported | Precision | Recall | F-score |
|---|---|---|---|---|---|
| 7 | Pu-239, Pu-241, Am-241, WGPu | Pu-239, Pu-241, Am-241, RGPu, K-40 | 80.0 | 100.0 | 88.9 |
| 8 | Pu-239, Pu-241, Am-241, WGPu | Pu-239, Pu-241, Am-241, K-40 | 85.7 | 100.0 | 92.3 |
| 9 | Pu-239, Pu-241, Am-241, WGPu | Pu-239, Pu-241, Am-241, WGPu, K-40 | 86.7 | 100.0 | 92.9 |

As displayed in Table 13, the algorithm that identified WGPu is “rewarded” and the algorithm that incorrectly identified RGPu is “penalized”. This results in a better F-score for the algorithm that correctly identified WGPu.

## 2.4 Scoring Application Configuration Weighting Factors

Similar to nuclide weighting factors, the scoring application allows the use of weighting factors based on configuration importance and allows default configuration weighting factors to be changed easily to meet the goals and objectives of a given test campaign. In general, both the frequency of observation and the associated consequence with non-detection are used as the basis for assigning configuration weighting factors. For example, SNM and Radiological Dispersion Device (RDD) configurations, which are observed with low frequency, would be assigned a high configuration weighting factor due to the consequence associated with non-detection. Default scoring application weighting factors based on configuration importance are presented in Table 14.

**Table 14. Default Weighting Factors for High, Medium, and Low Importance Configurations**

| Configuration Importance | Example | Configuration Weighting Factor (WFc) |
|---|---|---|
| High | High frequency of observation and/or high consequence associated with non-detection | 3 |
| Medium | Medium frequency of observation and/or medium consequence associated with non-detection | 2 |
| Low | Low frequency of observation and low consequence associated with non-detection | 1 |

To illustrate how the scoring application utilizes configuration weighting factors, the following example is provided which compares grouped F-scores calculated with and without applying configuration weighting factors. Information applicable to the example is supplied in Table 15.

**Table 15. Default Weighting Factors for High, Medium, and Low Importance Configurations**

| Configuration ID | Configuration Importance | Configuration Importance WF | F-Score |
|---|---|---|---|
| A | High | 3 | 15.0 |
| B | Medium | 2 | 90.0 |
| C | Low | 1 | 95.0 |
| D | Low | 1 | 85.0 |

$$F(unweighted) = \left( \frac{\sum_{i=1}^{n} F}{\sum_{i=1}^{n} n} \right) = \left( \frac{(15.0+90.0+95.0+85.0)}{4} \right) = 71.3 \quad (13)$$

$$F(weighted) = \left( \frac{\sum_{i=1}^{n} F \times WFc}{\sum_{i=1}^{n} WFc} \right) = \left( \frac{(15.0*3)+(90.0*2)+(95.0*1)+(85.0*1)}{(3+2+1+1)} \right) = 57.9 \quad (14)$$

As shown in the example, the grouped F-score calculated with configuration weighting factors applied places additional emphasis on the high importance configuration which scored poorly. Accordingly, the utilization of configuration weighting factors provides a better representation of overall operational nuclide identification algorithm performance.

## 2.5 Scoring Application Nuclide Reporting Conventions and Nuclide Equivalencies

Since some nuclides are reported differently by nuclide identification algorithms, it is necessary for the application to interpret reported nuclides correctly for accurate scoring. To minimize the complexity of interpreting reported nuclides, ANSI N42.42-2011 [3] nuclide reporting conventions will be specified as the standard to be used by nuclide identification algorithms. Specifically, use of ANSI N42.42-2011 nuclide reporting conventions for "other radiation sources" (reproduced directly from ANSI N42.42-2011 in Table 16 below) is necessary.

**Table 16. ANSI N42.42-2011 Name Format for Other Radiation Sources**

| Name | Definition |
|---|---|
| Annihilation | The 511 keV annihilation peak. Such photopeak can be produced by positron emission tomography (PET) sources; examples of such sources are: $^{11}$C, $^{13}$N, $^{15}$O, $^{18}$F. |
| Bremsstrahlung | The signature of bremsstrahlung radiation has been observed. Bremsstrahlung is produced when fast electrons interact with the Coulombic field of the nucleus or when the fast electrons are decelerated when interacting with a metal target. |
| DU | Depleted Uranium is uranium with lower than natural abundance of $^{235}$U. Approximate abundance: 99.799% $^{238}$U, 0.2% $^{235}$U, 0.001% $^{234}$U. |
| HEU | Highly Enriched Uranium is uranium with high abundance of $^{235}$U. The $^{235}$U abundance is higher than 20%. |
| LEU | Low Enriched Uranium is uranium with an abundance of $^{235}$U of approximately 3% to 20%. |
| *N*(*reaction*) | Nuclear reactions are indicated by the chemical element or nuclide name (*N*) followed by the reaction notation (*reaction*). Reaction notations include:<br>• n,g<br>• n,n'g<br>• a,n<br>• n,2n<br>Examples are: "H(n,g)", "Fe(n,g)", and "O-18(a,n)". |
| Plutonium | If the radiation measurement instrument cannot discriminate between the different levels of plutonium enrichments (RGPu and WGPu), then they should all be indicated as "Plutonium". |
| *N*-xray | X-rays are indicated by the element name followed by "-xray". Examples: "U-xray", "Pb-xray". |
| Radium | Naturally occurring radium (Ra-226) decay chain in equilibrium. |
| Refined U | Natural uranium chemically processed to be separated from daughters ($^{234}$Th and $^{234m}$Pa being short lived daughters of $^{238}$U are still present). |
| RGPu | Reactor Grade Plutonium is plutonium with > 7% $^{240}$Pu |
| Shielded Source | The signature of a shielded radioactive source that cannot be fully identified due to the presence of shielding material. |
| Thorium | Naturally occurring thorium (Th-232) decay chain in equilibrium. |
| Unknown | Sources not identified because radionuclides are not listed in the radiation measurement instrument library or because the energy spectrum is distorted due, for example, to the presence of masking or shielding material. |

| Name | Definition |
|---|---|
| U-natural | Uranium natural is equivalent to uranium-ore; that is, uranium in natural abundance and in secular equilibrium with an abundance of 99.2745% $^{238}$U, 0.72% $^{235}$U, and 0.0055% $^{234}$U. |
| Uranium | If the radiation measurement instrument cannot discriminate between the different levels of uranium enrichments (DU, LEU, HEU and Refined U), then they should all be indicated as “Uranium”. |
| WGPu | Weapons Grade Plutonium is plutonium with ≤ 7% $^{240}$Pu. |

Table 17 presents example nuclide equivalences currently assigned by the scoring application for reported “nuclides” that are not unique. The most common use of nuclide equivalences applies to decay chain nuclides associated with a parent decay chain nuclide. For example, if an algorithm reports Bi-214, a decay product of the Ra-226 decay chain, the scoring application converts Bi-214 to Ra-226 prior to scoring. If other nuclide equivalencies are needed to properly interpret a reported “nuclide”, additional nuclide equivalencies can be readily added to the application for accurate scoring.

**Table 17. Example Nuclide Equivalences Currently Assigned by the Scoring Application**

| Assigned Nuclide | Nuclide Equivalencies |
|---|---|
| Annihilation | Annihilation, F-18, Positron Emitter |
| Background | Background, None |
| Bi-213 | Bi-213, Tl-209 |
| Bremsstrahlung | Beta, Bremsstrahlung, Sr-90, P-32, Y-90, Beta Emitter |
| Cf-252 | Cf-252, Cf-249 |
| Ge-68/Ga-68 | Ge-68, Ga-68, Ge-68/Ga-68 |
| Neutrons | Neutrons, Neutron, H(n,g), Fe(n,g), Neutrons On Fe, Neutrons On Hydrogen |
| Np-237 | Np-237, Pa-233 |
| Pu-239 | Pu-239, Plutonium |
| Pu-241 | Pu-241, U-237 |
| Ra-226 | Ra-226, Radium, Bi-214, Pb-214 |
| Sr-82/Rb-82 | Sr-82, Rb-82, Sr-82/Rb-82 |
| Sr-85/Kr-85 | Kr-85, Sr-85, Sr-85/Kr-85 |
| Th-232 | Th-232, Thorium, Ac-228 |
| U-232/Th-228 | U-232, Th-228, Bi-212, Pb-212, Tl-208 |
| Zr-95 | Zr-95, Nb-95 |

Similar to nuclide equivalences, it is occasionally necessary for the application to interpret and convert a reported “nuclide” to multiple nuclides for accurate scoring. For example, if a nuclide identification algorithm reports U-ore or U-natural, the scoring application will convert the reported nuclides to U-238 and Ra-226 prior to scoring. This is needed since uranium ore that has not been chemically processed to remove all decay chain products from the uranium will contain all decay products from the U-238 decay chain including all decay products from the Ra-226 decay chain. Alternatively, uranium that has been chemically processed will not contain Ra-

226 and its decay chain products due to the long half-lives of U-234 and Th-230, which precede Ra-226 in the U-238 decay chain, and effectively eliminates in-growth of Ra-226.

Table 18 presents example nuclide equivalences currently assigned by the scoring application for reported “nuclides” that require conversion to multiple nuclides for accurate scoring. As needed to meet the goals and objectives of a given test campaign, default assigned nuclide equivalencies are editable and can be easily changed. Lastly, if other nuclide conversions are needed to properly interpret a reported “nuclide”, then additional nuclide conversions can be readily added to the application for accurate scoring.

**Table 18. Example Assigned Nuclide Equivalencies For Reported “Nuclides”**

| Reported Nuclide | Assigned Nuclide Equivalencies |
|---|---|
| U-Ore | U-238 + Ra-226 |
| U-natural | U-238 + Ra-226 |
| HEU | U-235 + U-enr |
| LEU | U-235 + U-enr |
| DU | U-238 + U-dep |
| RefinedU | U-238 + U-nat |
| WGPu | Pu-239 + WGPu |
| RGPu | Pu-239 + RGPu |

Lastly, it is sometimes necessary for the application to properly interpret multiple nuclide decay chain equivalencies. To illustrate how multiple decay chain equivalencies are interpreted by the application, an example utilizing the Th-229, Ac-225, and Bi-213 decay chains (see Figure 1) will be provided. To further assist with understanding the example, the gamma emitting nuclides most likely to be detected for these three decay chains is listed for reference purposes in Table 19.

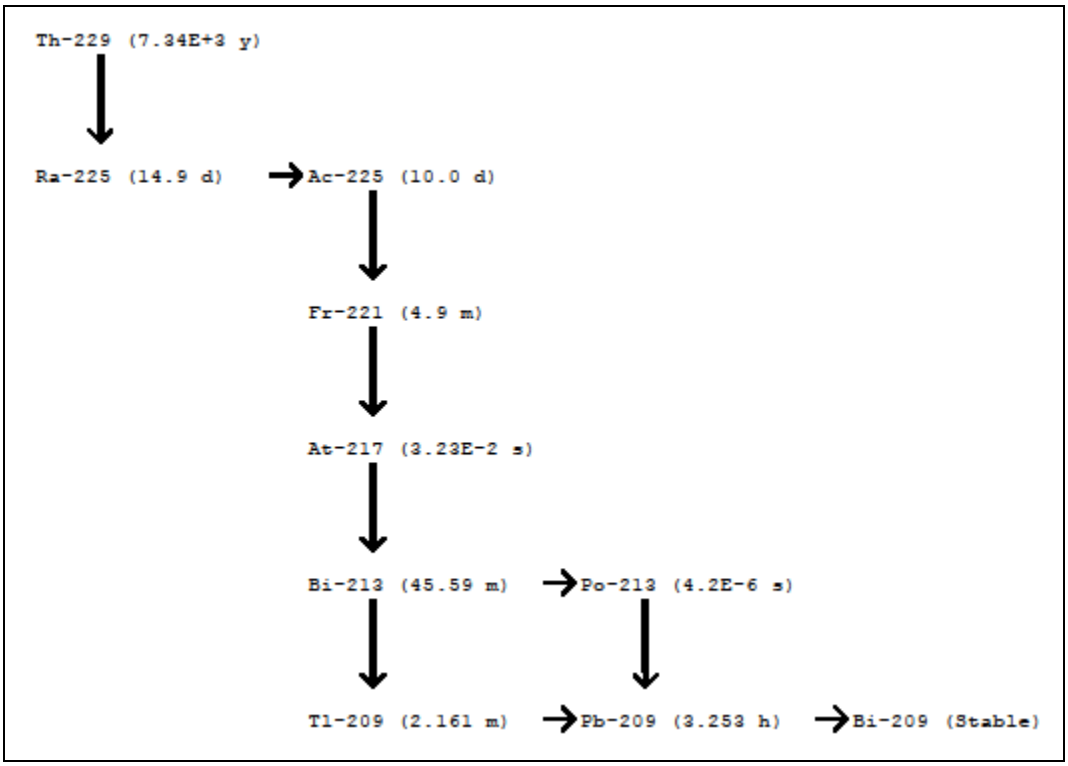


**Figure 1. Th-229 Decay Chain.**

**Table 19. Gamma Emitting Decay Chain Nuclides Most Likely to be Detected from the Th-229, Ac-225, and Bi-213 Decay Chains**

| Decay Chain | Gamma Emitting Nuclides Most Likely to be Detected |
|---|---|
| Th-229-DC | Ac-225, Fr-221, Bi-213, Tl-209 |
| Ac-225-DC | Ac-225, Fr-221, Bi-213, Tl-209 |
| Bi-213-DC | Bi-213, Tl-209 |

A hypothetical sample of Th-229 contains both the Ac-225 decay chain and the Bi-213 decay chain. Accordingly, Ac-225 decay chain nuclides (Ac-225, Fr-221) and Bi-213 decay chain nuclides (Bi-213, Tl-209) can each contribute gamma signatures to a measured spectrum of the Th-229 sample. Since algorithms may interpret these signatures as evidence of one or both of these nuclides, products from both nuclide decay chains (Ac-225 and Bi-213) are valid answers.

Similarly, since a sample of Ac-225 contains the Bi-213 decay chain, products from both nuclide decay chains (Ac-225 and Bi-213) are valid answers.

However, since a sample of Bi-213 will not contain the Ac-225 gamma signature (Ac-225, Fr-221), only products from the Bi-213 decay chain (Bi-213, Tl-209) are valid.

Hence, to accurately assign the reported nuclide to the correct decay chain, the application initially converts the reported nuclide to the immediate parent nuclide within the decay chain. If that parent is actually present in the sample, it is converted by the application to the parent nuclide decay chain for scoring.

To enhance comprehension, two general examples are presented in Table 20.

**Table 20. Multiple Nuclide Decay Chain Equivalency Examples**

| Example | Nuclide(s) Present | Nuclide(s) Reported Initial | Decay Chain Conversion Initial | Decay Chain Conversion Final | Nuclide(s) Reported Final |
|---|---|---|---|---|---|
| 11 | Ac-225-DC, Cs-137 | Tl-209, Bi-213, Cs-137 | Bi-213-DC | Ac-225-DC | Ac-225-DC, Cs-137 |
| 12 | Tl-201, Tl-202 | Tl-201, Bi-213 | Bi-213-DC | Bi-213-DC | Tl-201, Bi-213-DC |

In Example 11, the identification algorithm reports the presence of Bi-213, a decay product of Ac-225 decay chain, which is actually present. Therefore, the identification of Bi-213 is converted to the parent Ac-225 nuclide decay chain for scoring. In Example 12, Ac-225 is not present, so Bi-213 is not subsequently converted.

# 3 Nuclide Identification Confidence Indices

Many nuclide identification algorithms use confidence indices designed to evaluate the confidence of the nuclide identification. Currently, the scoring application has been programmed to consider nuclide identification confidence indices if reporting scales of 0 to 10 (integer values only) or High, Medium, Low are used.

To illustrate how the scoring application utilizes nuclide identification confidence indices, the following example is provided. Information relevant to the example is displayed in Tables 21 through 23.

**Table 21. Default Weighting Factors for High, Medium, and Low Nuclide ID Confidence Indices**

| Nuclide ID Confidence Index | Nuclide ID Confidence Index Acronym | Weighting Factor |
|---|---|---|
| High | (H) | 3/3 |
| Medium | (M) | 2/3 |
| Low | (L) | 1/3 |

**Table 22. Nuclides Present and Reported by the Identification Algorithms**

| Algorithm | Nuclides Present | Nuclides and Confidence Indices Reported |
|---|---|---|
| A | Ga-67, Sm-153, Eu-154 | Ga-67(H), Np-237(L), Sm-153(H) |
| B | Ga-67, Sm-153, Eu-154 | Ga-67(H), Np-237(H), Sm-153(L) |

**Table 23. Nuclide Weighting Factors**

| Nuclide | Nuclide Importance | tp WF | fp WF | fn WF |
|---|---|---|---|---|
| Ga-67 | Low | 1 | 1 | 1 |
| Sm-153 | Low | 1 | 1 | 1 |
| Eu-154 | Low | 1 | 1 | 1 |
| Np-237 | High | 4 | 2 | 4 |

As shown in Table 22, nuclide identification algorithm A correctly reported Ga-67 and Sm-153 with high confidence, incorrectly reported Np-237 with low confidence, and did not report Eu-154. Using the nuclide identification confidence index weighting factors presented in Table 21, the precision, recall, and F-score are calculated as follows:

$$100 * p = 100 * \left(\frac{tp}{tp+fp}\right) = 100 * \frac{\left[\left(1*\frac{3}{3}+1*\frac{3}{3}\right)\right]}{\left[(1+1)+\left(2*\frac{1}{3}\right)\right]} = 75.0 \quad (15)$$

$$100 * r = 100 * \left(\frac{tp}{tp+fn}\right) = 100 * \frac{\left[\left(1*\frac{3}{3}+1*\frac{3}{3}\right)\right]}{[(1+1)+(1)]} = 66.7 \quad (16)$$

$$F\ Score = 2 * \left(\frac{p*r}{p+r}\right) = 2 * \left(\frac{7.50*66.7}{7.50+66.7}\right) = 70.6 \quad (17)$$

In a similar manner, precision, recall, and F-scores were calculated using the nuclides and confidence indices reported by algorithm B. Lastly, precision, recall, and F-scores were calculated without considering nuclide identification confidence indices. A comparison of the results is presented in Table 24.

**Table 24. Calculated Precision, Recall, and F-scores**

| Nuclide ID Confidence Indices | Precision | Recall | F-score |
|---|---|---|---|
| Algorithm A | 75.0 | 66.7 | 70.6 |
| Algorithm B | 33.3 | 44.4 | 38.1 |
| Not Considered | 50.0 | 66.7 | 57.1 |

By considering nuclide identification confidence indices, Algorithm A is not fully penalized for reporting Np-237 with low confidence which provides a better assessment of nuclide identification algorithm performance. Likewise, Algorithm B demonstrates the impact of misplaced confidence with Sm-153 being correctly reported but with low confidence and Np-237 incorrectly reported with high confidence.

# 4 Scoring Application Bar Chart and Histogram Generation

The scoring application can be used to generate summary bar charts and histograms to evaluate nuclide identification algorithm performance. Currently, summary bar charts and histograms are automatically generated for the following categories.

- Radionuclide (Natural, medical, industrial, and threat)
- Count time (User selectable)
- Source strength or standard deviations above background (User selectable)
- Unshielded versus shielded

- Areal density of shielding (User selectable)
- Configuration importance
- Detector type

Figures 21 and 22 present example bar charts and histograms, respectively, which compare nuclide identification algorithm performance for unshielded and shielded configurations.

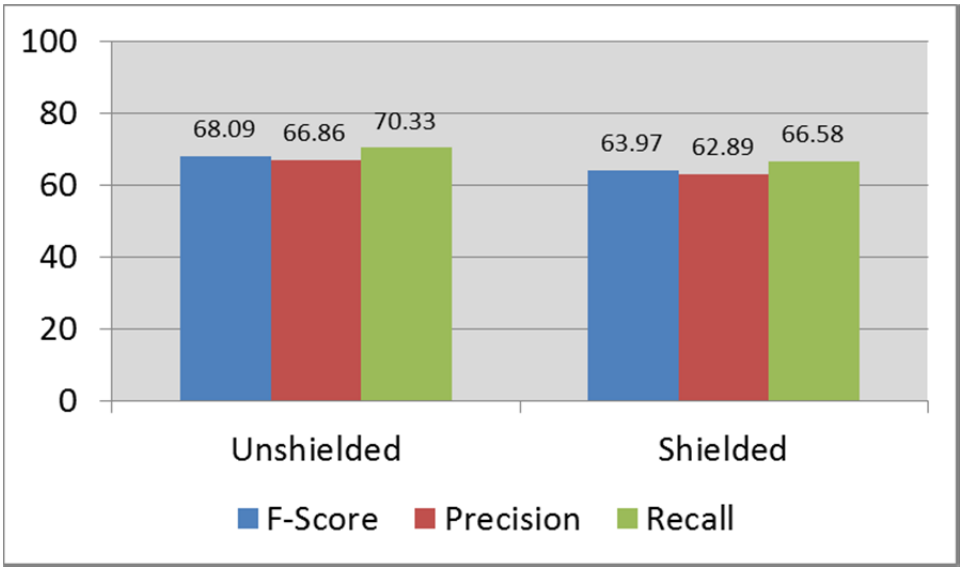


**Figure 2. F-Score, Precision, and Recall Barchart Example**

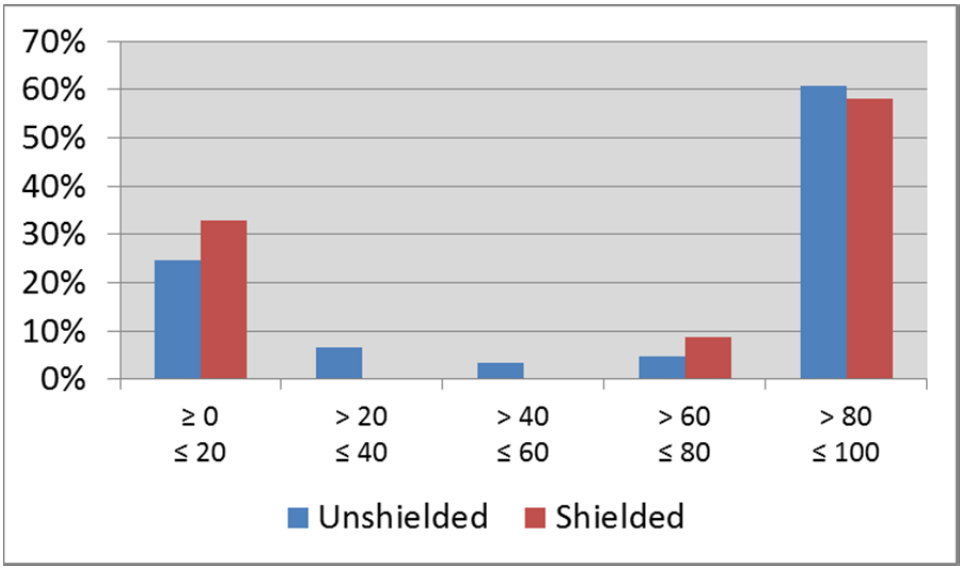


**Figure 3. F-Score Histogram Example**

# 5 Scoring Application Algorithm Performance Comparisons

In addition to category comparisons, the scoring application can also be used to compare the performance of two nuclide identification algorithms with paired observations [1]. Along with comparing average algorithm F-score, precision, and recall values (Figure 4), the scoring application uses the two analysis methods described below to evaluate differences in nuclide identification algorithm results.

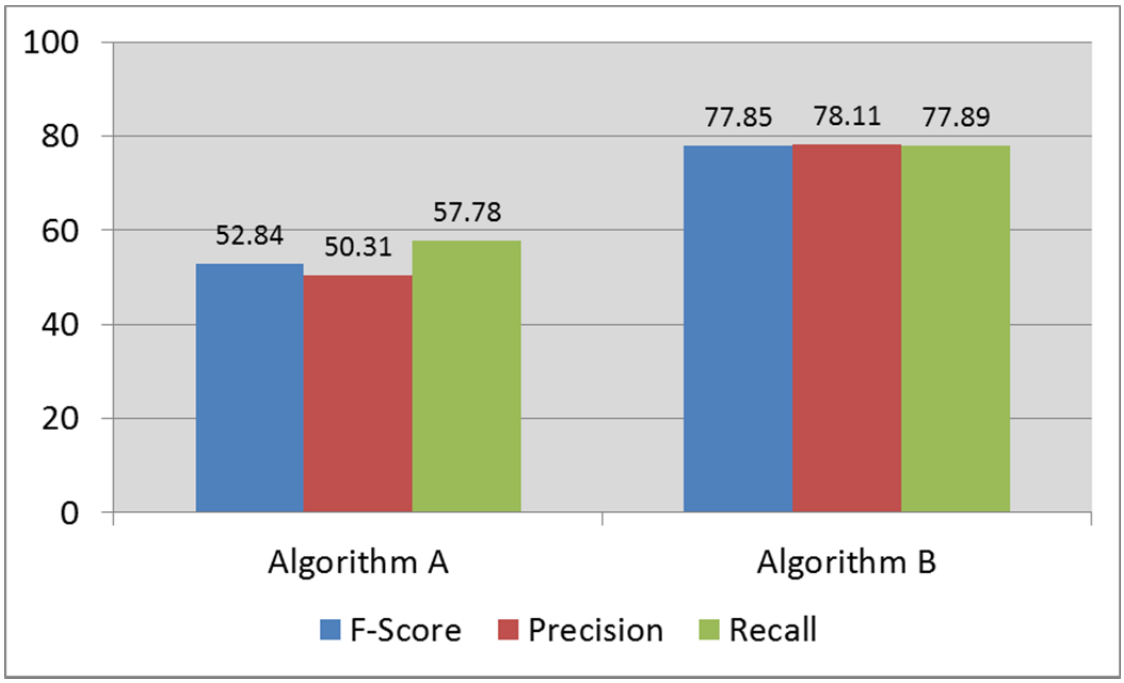


**Figure 4. F-Score, Precision, and Recall Algorithm Comparison Barchart Example**

Difference of means

If nuclide identification algorithm results for measured spectral data pairs are sufficiently alike, the mean difference for the paired nuclide identification algorithm scores should approximate zero. Figure 5 presents an example histogram for the mean differences between Algorithm A and Algorithm B paired nuclide identification algorithm scores. Since the histogram is heavily skewed to the left, Algorithm B scores higher than Algorithm A.

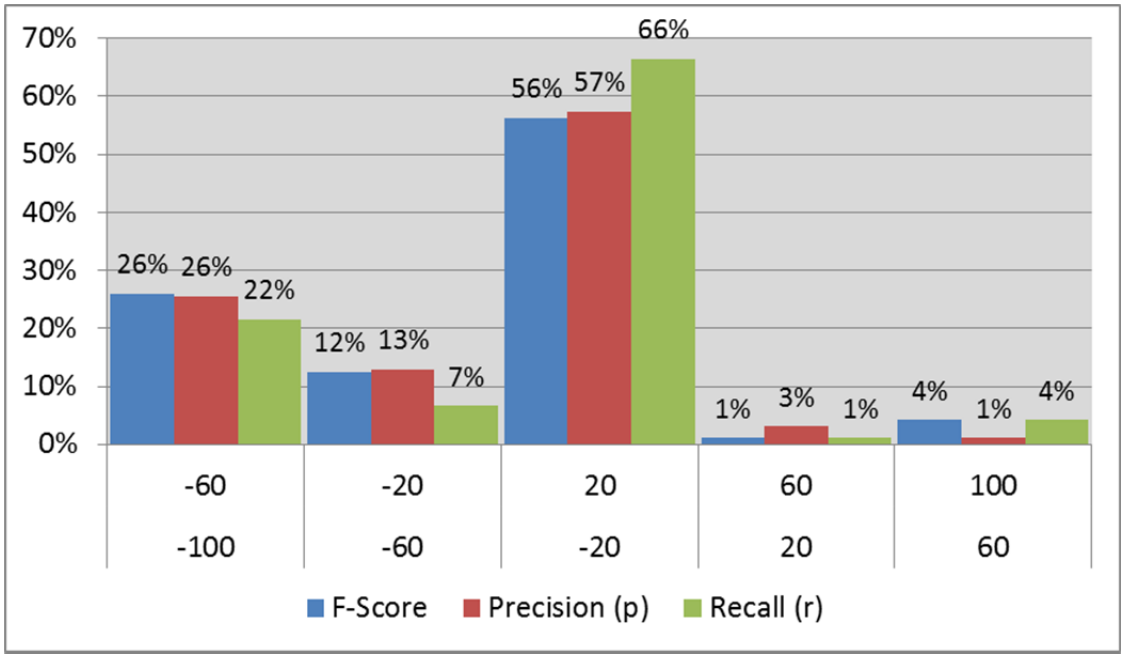


**Figure 5. F-Score, Precision, and Recall Paired Data Mean Difference Histogram (Algorithm A - Algorithm B)**

Difference of proportions

If nuclide identification algorithm results for spectral data pairs are suitably similar, the proportion of times that each algorithm scores better than the other algorithm should be roughly equal. Or stated differently, the mean difference between these proportions should approximate zero. Figure 6 presents an example histogram for the mean proportion differences between Algorithm A and Algorithm B paired nuclide identification algorithm scores. Since the histogram is heavily skewed to the left, Algorithm B scores higher than Algorithm A.

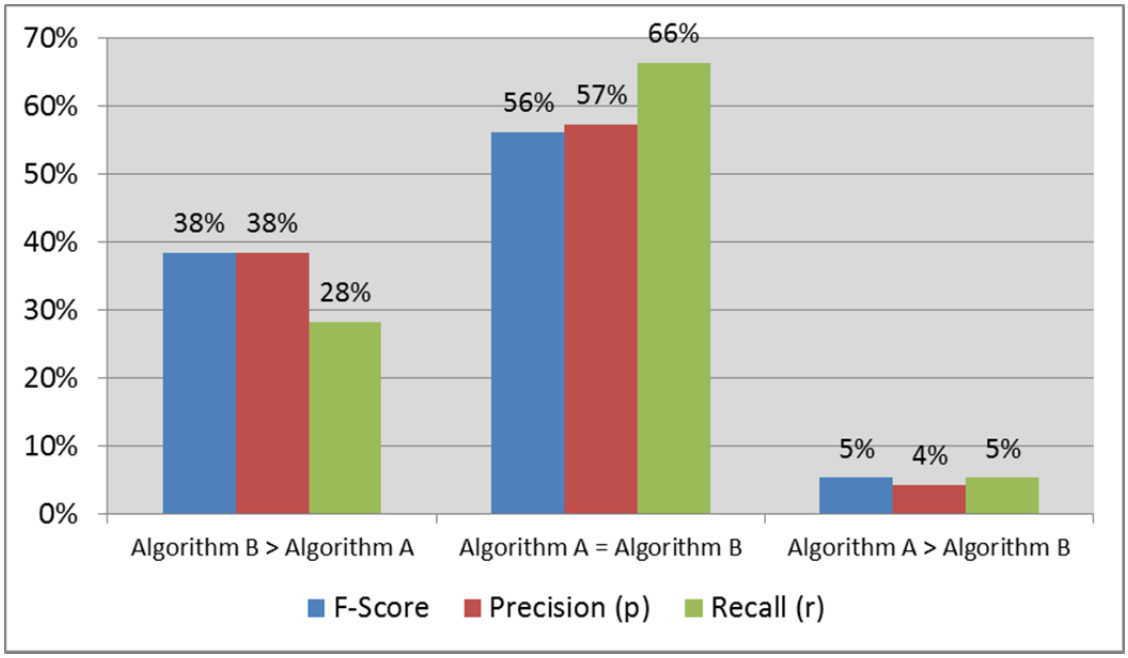


**Figure 6. F-Score, Precision, and Recall Paired Data Proportions**

Confidence Interval Determinations and Recommended Method

Although not calculated by the application due to file size limitations, the recommended method for determining confidence intervals on the mean differences described above is the percentile bootstrap method. Detailed information on how to calculate percentile bootstrap confidence intervals on the mean differences is presented in Appendix A.

Statistical Comparison Interpretation

If the derived confidence interval contains zero, there is insufficient evidence to conclude that nuclide identification algorithm results are different. If the derived confidence interval does not contain zero, there is sufficient evidence to conclude that nuclide identification algorithm results are different.

Example: The 95% confidence interval for the mean algorithm score difference between Algorithm A and Algorithm B is (0.5, 4.1). As such, we are 95% confident that the mean algorithm score difference between Algorithm B and Algorithm A is between 0.5 and 4.1 with Algorithm B scoring higher than Algorithm A.

# Summary

Scoring criteria and an associated Microsoft Excel application have been developed to objectively assess the performance of nuclide identification algorithms. As described in this report, the fundamental equation used to evaluate nuclide identification algorithm performance is F-scores which have been modified using nuclide weighting factors.

Additionally, this report discusses the importance of nuclide reporting conventions and outlines the use of nuclide equivalencies. Lastly, use of scoring application bar charts, histograms, and methods for evaluating and comparing nuclide identification algorithm performance is discussed.

# References (or Related Documents)

[1] D. R. Helsel and R. M. Hirsch, "Statistical Methods in Water Resources, U.S. Geological Survey, Techniques of Water-Resources Investigation, Book 4, Chapter A3".

[2] Y. Sasaki, "The Truth of the F-measure," October 2007. [Online]. Available: http://www.cs.odu.edu/~mukka/cs795sum10dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf.

[3] "ANSI N42.42-2011, Data Format Standard for Radiation Detectors Used for Homeland Security".

[4] Wellner, et al., "Rapidly Retargetable approaches to De-identification in Medical Records," *Journal of the American Medical Informatics Association*, pp. 14:564-573, 2007.

[5] K. T. Chess and B. Chess, "Fortify Software, A Metric for Evaluating Static Analysis Tools," [Online]. Available: www.securitymetrics.org/attachments/Metricon-1-Chess-Software.ppt.

[6] D. R. Helsel, Nondetects and Data Analysis, Statistics for Censored Environmental Data, NJ, USA: John Wiley and Sons, 2005.

[7] J. Lane, Interviewee, *Personal Communication with Jonathan Lane, Statistics and Surveillance Assessment Department.* [Interview]. February 2015.

[8] *AIP Nuclide Identification Algorithm Scoring Application*, SNL_NucIDAlgScrApp_V001.xlsx.

# Acronyms and Abbreviations

| ACRONYM | TERM |
|---|---|
| AIP | Algorithm Improvement Program |
| DHS | Department Of Homeland Security |
| DNDO | Domestic Nuclear Detection Office |
| DU | Depleted Uranium |
| fn | False Negative |
| fp | False Positive |
| HEU | Highly Enriched Uranium |
| p | Precision |
| r | Recall |
| RDD | Radiological Dispersion Device |
| RGPu | Reactor Grade Plutonium |
| SNM | Special Nuclear Material |
| tp | True Positive |
| U-dep | Depleted Uranium |
| U-enr | Enriched Uranium |
| U-nat | Natural Uranium |
| WF | Weighting Factor |
| WGPu | Weapons Grade Plutonium |

# Appendix A. Percentile Bootstrap Confidence Interval Determinations

Difference of means

1. Randomly sample, with replacement, n pairs from the data set {(x1,y1),(x2,y2),…,(xn,yn)}.
2. Calculate m_i as the mean of xi – yi for the n pairs of data randomly sampled.
3. Repeat 1-2 a large number of times (2,000 replications were used for this report).
4. Define the 95% bootstrap confidence interval as having the lower bound equal to the 2.5 percentile of {m1, m2, …, mn} and the upper bound equal to the 97.5 percentile of that set.

Difference of proportions

1. Randomly sample, with replacement, n pairs from the data set {(x1,y1),(x2,y2),…,(xn,yn)}.
2. Calculate the proportion, p_a, of times that xi > yi and the proportion, p_b, of times that yi > xi for the n pairs of data randomly sampled.
3. Define d_i as p_a – p_b. That is, d_i is the difference of the proportions for the ith replication.
4. Repeat 1-3 a large number of times (2,000 replications were used for this report).
5. Define the 95% bootstrap confidence interval as having the lower bound equal to the 2.5 percentile of {d1, d2, …, dn} and the upper bound equal to the 97.5 percentile of that set.

THIS PAGE INTENTIONALLY LEFT BLANK