OPEN-SOURCE IMPLEMENTATION OF AN ALGORITHM FOR PHOTOPEAKS SEARCH AND ANALYSIS IN GAMMA-RAY SPECTROMETRY WITH SEMICONDUCTOR DETECTORS

Marcelo F. Máduar[1] and Brigitte R. S. Pecequilo[1]

[1] Instituto de Pesquisas Energéticas e Nucleares, IPEN - CNEN/SP
Av. Professor Lineu Prestes 2242
05508-000 São Paulo, SP
mmaduar@ipen.br

Keywords: Gamma-ray spectrometry, HPGe detectors, spectrum analysis, open-source code.

ABSTRACT

Radioactivity quantification of gamma-ray emitter radionuclides in samples measured by HPGe gamma spectrometers relies on the analysis of the photopeaks present in the spectra, especially on the accurate determination of their net areas. This paper presents a methodology and an algorithm description for the peak search and analysis in order to obtain the relevant peaks parameters and their uncertainties. The procedure is performed on a three step approach: a) a preliminary search is done by using the second-difference method; b) experimental peaks widths are assessed in order to obtain a width vs. channel relationship and to define regions with single or overlapping peaks; c) a non-linear fit is applied to each region of the spectrum with candidate peaks. The final target function is in the form $G(x) = B(x) + F(x)$, where $B(x)$ is the baseline composed by a sum of a weighed left-side $B_L(x)$ and right-side $B_R(x)$ base-line quadratic functions and the photopeaks term $F(x)$ is a sum of Gaussian functions. The computational implementation is released entirely in open-source license. The code was developed in C++ language and the interface was developed with Qt GUI software toolkit. GNU scientific library, GSL, was employed to perform linear and non-linear fitting procedures as needed. Spectra previously generated at our laboratories were analyzed with the presented methodology and with the commercial software package WinnerGamma. Results obtained are consistent with those obtained with the aforementioned package, suggesting that it could be safely used in general-purpose gamma-ray spectrometry.

1. INTRODUCTION

Radioactivity quantification of gamma-ray emitter radionuclides in samples measured by HPGe gamma spectrometers relies on the analysis of the photopeaks present in the spectra, especially on the accurate determination of their net areas [1]. This task is usually performed with proprietary software tools [2,3,4,5]. Such tools, although usually offering a wide range of graphical and user-friendly software interface capabilities, are highly costly and commonly do not allow user adaptations nor modifications to one's laboratory particularities.

As a result, the developing of in-house software can emerge as an attractive option, both for the immediate costs advantages as well as the availability to adapt the tool for the user needs, especially when such tools are released under an open-source license [6]. This is the case, for instance, of the recently developed software SAANI [7], at IPEN, with modern-style interface and planned to replace an older solution, VISPECT, based on systems with primitive graphical capabilities.

This paper presents a methodology and an algorithm description for the peak search and analysis in order to obtain the relevant peaks parameters and their uncertainties. The computational implementation is released entirely in open-source license [6] for the main code and with the use of open software packages for interface design and mathematical libraries. The present work is part of a comprehensive work under progress aiming to develop a complete software solution, called OpenGamma, for gamma-ray spectrometry.

The procedure to obtain peaks areas is performed on a three step approach. Firstly a preliminary search is done by using the second-difference method. In the second step, the experimental peaks widths are assessed in order to obtain a FWHM vs. channel relationship. Well-formed and isolated peaks are chosen for this fit. Lastly, a non-linear fit is again applied to each multiplet region of the spectrum to separate the peaks and obtain their net areas.

2. COMPUTATIONAL METHODS

2.1. Peak Preliminary Search

The initial search is done by using the now classical second-difference method firstly described by Mariscotti [8], consisting in the generation of a derived spectrum $d(i)$ by applying the expression

$$d(i) = s(i-1) + s(i+1) - 2\,s(i) \tag{1}$$

where $s(i)$ is the original spectrum count in each channel i. The candidate peaks are set in channels i where $d(i)$ obeys the criterion

$$|\,d(i)\,| < \mathrm{k}\,.\,u(i) \tag{2}$$

where $u(i)$ is the uncertainty in $s(i)$ and k is a user defined parameter to set the search sensitivity.

For the preliminary search we used an order-2 binomial filter applied to the spectra [9], prior to the derivation, in order to minimize misidentifications caused by statistical fluctuations.

2.2. Peak Width vs. Channel Fit

In the second step, the experimental peaks widths are assessed in order to obtain the commonly used full width at half maximum parameter (FWHM). The FWHM vs. channel relationship is obtained to be applied afterward. In the continuous domain, the second difference $d(i)$ corresponds to the second derivative of the Gaussian function of the peak, $F(x)$, where x is the continuous variable corresponding to the discrete one i. Thus a fit is done around each candidate peak p, by using the target function

$$F''(x) = 2\,a_p\,k_p\,[2\,k_p\,(x\text{-}c_p)^2 - 1]\,.\,\exp[-k_p\,(x-c_p)^2] \tag{3}$$

and employing the Levenberg-Marquardt minimization method for non-linear fitting [10]. The parameters to be obtained are a_p, k_p and c_p, where a_p is the Gaussian height, k_p the exponent factor related to the peak FWHM, now called w, by the expression

$$w = 2 \, (\ln 2)^{1/2} \, k^{-1/2} \tag{4}$$

and c_p is the centroid channel.

Once the peaks' widths are obtained, "good" peaks, i.e., well-formed and isolated ones, are chosen by their low uncertainty in w, and applied to a linear fit with a 2nd order polynomial in order to obtain $w_f(x)$, the fitted values of w as a function of channel x. The fitted values of k, called $k_f(x)$, are also generated.

2.3. Multiplets separation and peak height fit

In the last step, a non-linear fit is again applied to each multiplet region of the spectrum with m candidate peaks. For each candidate c_p, other peaks c_q are assumed to be in the same multiple when they are close enough to c_i to satisfy the relationship

$$|c_p - c_q| < n \, . \, w_f(c_p) \tag{5}$$

where n is a user-set minimum number of FWHMs accepted to consider peaks to be separated. In this step the target function $G(x)$ is composed by the baseline term $B(x)$ and the photopeaks term $F(x)$. Function $B(x)$ is a sum of a left-side $B_L(x)$ and right-side $B_R(x)$ base-line quadratic functions, each of them weighed by a summation of m logistic functions normalized in the form

$$\tfrac{1}{2} \, \{1 + \mathrm{erf}[(k_f)_p^{1/2} \, (x - c_p)]\} \tag{6}$$

Photopeaks term $F(x)$ is a summation of m Gaussian functions in the form

$$a_p \exp \, [-(k_f)_p \, (x - c_p)^2] \tag{7}$$

The target parameters in this step are those of the baseline function and 2 parameters, c_p and a_p, for each peak in the multiplet. The fitted net area A under the peak is directly obtained by the expression

$$A = \tfrac{1}{2} \, (\pi/\ln 2)^{1/2} \, a_p \, w_p \tag{8}$$

3. SOFTWARE IMPLEMENTATION

The code was developed in C++ language with the use of object-oriented codification [11]. The the interface was developed with Qt GUI software toolkit [12]. The mathematical software GNU scientific library (GSL) [13], was employed to perform linear and non-linear fitting procedures as needed. The present code is part of the package OpenGamma, which is hosted and registered in the public software repository SourceForge [14].

4. RESULTS AND DISCUSSION

Spectra previously generated at our laboratories were analyzed by using the presented methodology and by the commercial software package WinnerGamma. In order to compare the results from both codes, a procedure outlined in the AIEA document TEC-DOC-1011 [15] was adopted. This document describes a software intercomparison exercise performed with several software packages, including InterWinner; the performance of each package in determining peaks parameters was evaluated by computing the z-scores for each relevant variable calculated by the packages in spectra sets, according to the expression

$$z_{rep} = \frac{A_{rep} - A_{ref}}{\sqrt{\sigma_{rep}^2 + \sigma_{ref}^2}} \qquad (9)$$

In the previous expression, A and σ are the value and uncertainty, respectively, of the variable. The index *rep* refers to the result reported by the software and index *ref*, to the corresponding reference value. At the end, all evaluated packages were considered adequate for use in gamma-ray spectrometry [15].

Therefore, in this paper, as the net area is the variable of interest, this one was computed for all the set of peaks found in the spectra analyzed, where *rep* refers to the results obtained in this work and *ref*, to the corresponding ones from InterWinner.

In order to identify potential issues according to spectral characteristics, spectra were grouped according to their general features, as follows:

- Group A: low counting rate and long counting intervals (2 days or more), resulting in both adequate statistics, low base-line and low peak distortion; 29 spectra from the Brazilian Intercomparison National Program (PNI) were analyzed in this category.
- Group B: low count rate at short intervals, resulting in poor counting statistics; 22 spectra, mainly from effluents with low concentrations and some neutron activation analysis samples.
- Group C: High counting rate from a few peaks in each spectrum, resulting in good counting statistics, but significant base-line from radiation scattering; 24 spectra from radioactive gaseous effluents were analyzed, with significant ^{131}I contents.
- Group D: Several radionuclides (10 or more) in the same sample, resulting in numerous and large spectrum regions with 3 or more overlapping peaks. 16 spectra of process samples from IEA-R1 Reactor at IPEN were analyzed.

The same user defined parameters (peak search sensitivity and width of side base-lines) was employed for all the spectra sets, both in the present procedure and in InterWinner. The fit function did not take into account incomplete charge collection, as left-side tailing in the peaks was virtually not observed in none of the spectra sets.

Table 1 summarizes the results obtained for each group. The number of peaks corresponds to all matches from InterWinner and by the procedure described in this work. The variable χ^2 was calculated by

$$\chi^2 = \sum_{i=1}^{N} z_i^2 \qquad (10)$$

In this expression, N is the total number of peaks in each spectra set and z_i is z_{rep} calculated by Eq. 9, for each peak i in the set.

**Table 1. Peak sets and corresponding χ^2 values in each group.
Variable studied was the calculated net area.**

Group	Number of peaks in the group	χ^2
A	898	6829,5
B	624	4804,8
C	212	284,5
D	435	612,8

Individual z_i values were plotted as dots in dispersion graphics, against the corresponding gamma-ray energy, in order to identify potential dependencies on the regions of the spectrum. For better visualization, y-scale from –20 to 20 was used for groups A and B (Fig. 1), and -10 to 10 for groups C and D (Fig. 2). Furthermore, group B energy scale is limited to 1800 keV as the detector employed was a low-energy detector.
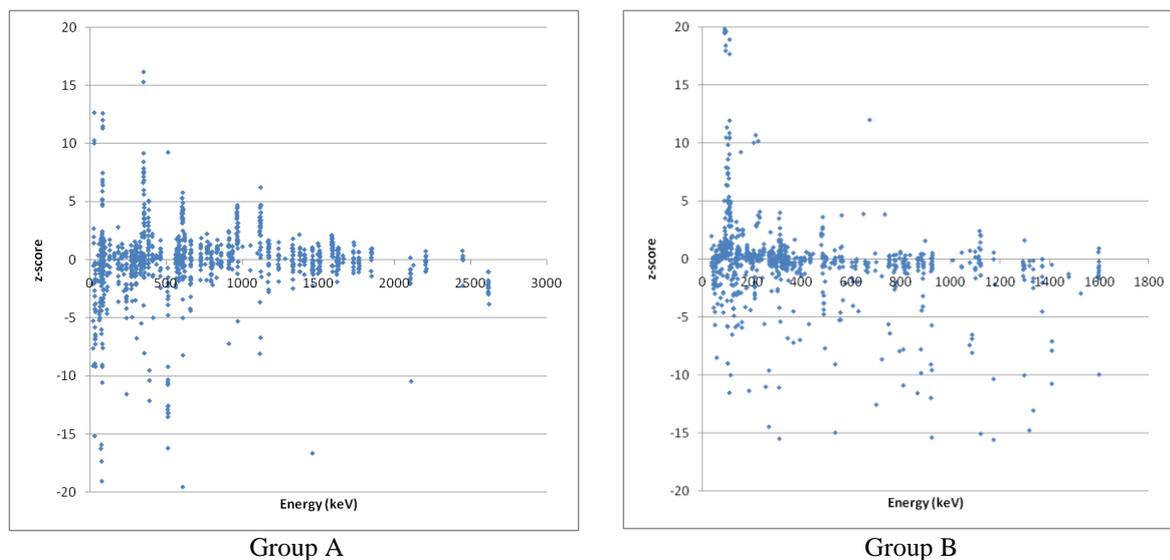

Group A   Group B

Figure 1. z-scores for spectra groups A and B for peak net areas.
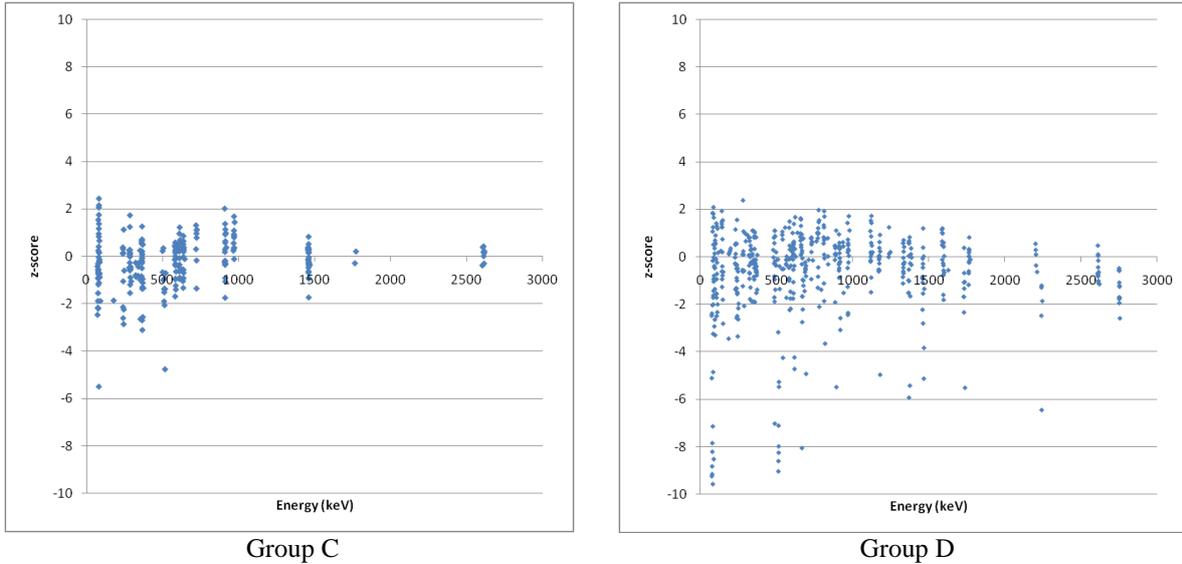
Group C

Group D

Figure 2. *z*-scores for spectra groups C and D for peak net areas.

In order to graphically show the capabilities of the presented procedure, Fig. 3a shows a complex peak multiplet successfully resolved. In Fig. 3b, a somewhat complicated multiplet led to a poorer fit, mainly because of the irregular baseline at the left side.
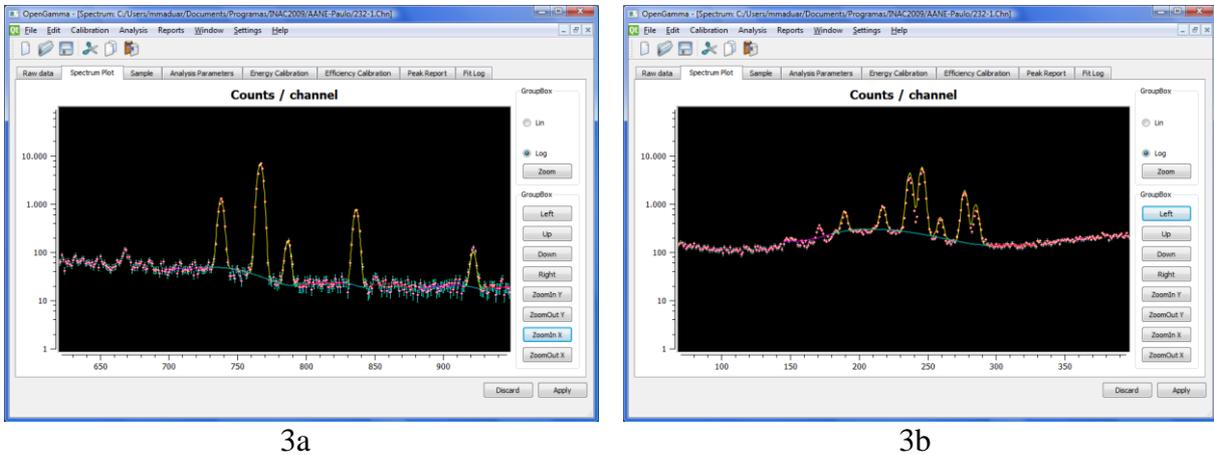

3a

3b

Figure 3. Spectra excerpts analyzed by using the procedure described in this work, part of OpenGamma software under development.

5. CONCLUSIONS

Results obtained with the present code were consistent with those obtained with the aforementioned package, by using the comparison criterion expressed in Eq. 9. No significant trend was found related to the groups of spectra as well. However, a somewhat systematic

trend to get outliers lower than zero is noticeable in all spectra sets. Also, data dispersion is more prominent in the left-side half of spectra; however, the overall behavior or z-scores does not appear to depend on energy peak.

Nonetheless, one have to bear in mind that results from software tools must always be carefully examined by the spectrum analyst, since misfits can occur, as illustrated in Fig. 3b. In this respect, the main advantage to have a promptly available analysis result in graphical form is to allow, even to the moderately skilled analyst, to visually identify problematic spectra.

The results suggest that the present methodology and the corresponding computational code could be safely used in general-purpose gamma-ray spectrometry.

Related work is under progress, including the integration of nuclide libraries, efficiency calibration curves and interfaces for entry of samples parameters, in order to allow obtaining results directly in units of total activity and activity concentrations in samples.

ACKNOWLEDGMENTS

REFERENCES

1. G. F. Knoll, "*Radiation detection and measurement*", 3rd ed., Wiley, New York (2000).
2. P. A. Aarnio, M. T. Nikkinen, J. T. Routti, "UNISAMPO, comprehensive software for gamma-spectrum processing", *J. Radional. Nucl. Ch.*, **248(2)**, pp.371-375 (2001).
3. "GammaVision®-32 V6 Gamma-ray spectroscopy software", http://www.ortec-online.com/pdf/a66.pdf (2005).
4. "InterWinnerTM 5 Gamma and alpha spectroscopy software", http://www.ortec-online.com/pdf/interwinner.pdf (2005).
5. "Genie 2000 Gamma Analysis Software", http://www.canberra.com/pdf/Products/SW_pdf/5gamma.pdf (2009).
6. "The GNU General Public License", http://www.gnu.org/licenses/gpl.html (2007).
7. S. R. de Lúcia, "Desenvolvimento de um software de espectrometria gama para análises por ativação com nêutrons utilizando o conceito de código livre", MSc. dissertation, IPEN, São Paulo, 2008.
8. M. A. J. Mariscotti, "A method for automatic identification of peaks in the presence of background and its application to spectrum analysis", *Nucl. Instr. and Methods*, **50**, pp.309-320 (1967).
9. P. Marchand, L. Marmet, "Binomial smoothing filter: A way to avoid some pitfalls of least-squares polynomial smoothing", *Rev. Sci. Instrum.*, **54**, pp.1034-1041 (1983).
10. P. R. Gill, W. Murray, M. H. Wright, "The Levenberg-Marquardt Method", In: "*Practical Optimization*", Academic, London (1981).
11. M. Fowler, "*UML essencial: um breve guia para a linguagem-padrão de modelagem de objetos*", 3ª ed., Bookman, Porto Alegre (2005).

12. J. Blanchette, M. Summerfield, "*C++ GUI programming with Qt 4*", Prentice Hall (2006).

13. M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, F. Rossi, "*GNU scientific library reference manual, 2nd ed.*", http://www.gnu.org/software/gsl (2007).

14. "SourceForge.net: Find and build open source software", http://sourceforge.net (2009).

15. IAEA, "Intercomparison of gamma ray analysis software packages", IAEA-TECDOC-1011, IAEA, Vienna, 1998.