Peak detection using fuzzy inference systems in gamma-ray spectrometry

Bruno Burini Robles Arine, Jose Arnaldo Frutuoso Roveda,
Sandra Regina Monteiro Masalskiene Roveda, and
Antonio Cesar Germano Martins

UNESP: Campus de Sorocaba,
Av. Três de Março, 511, Sorocaba - SP, Brazil
barine@posgrad.sorocaba.unesp.br,
{roveda,sandra,amartins}@sorocaba.unesp.br
http://www.sorocaba.unesp.br

Abstract. Gamma-ray spectrometry is a well-established analytical method for radionuclide concentration assessment. Given the state of modern computer technology, analysis workflow is conveniently automatized. Nonetheless, automated analysis algorithms still return a large number of false positives and false negatives, mainly when dealing with samples whose radioactivity is close to the minimum detectable value. To counter this issue, an attempt was made to implement a better peak search algorithm based on fuzzy inference systems (FIS). Compared to the classical solution, the FIS-based algorithm returned less false negatives and significantly less false positives. Peak area calculation performances of both solutions were similar, although the FIS-based algorithm had slightly lesser deviation. This paper shows that a simple set of fuzzy rules can improve both precision and sensitivity of gamma-ray analysis algorithms.

Keywords: gamma-ray, spectrometry, fuzzy, peak detection, signal processing

1 Introduction

Gamma-ray spectrometry is a widely adopted analytic method for radioactivity analysis in a number of scenarios such as nuclear medicine, work safety, and environmental monitoring[1][2]. The spectrometry system is composed of a detector made of either a semi-conductive crystal or a scintillator, converting incident gamma-rays into electrical current which is amplified and read by appropriate hardware. During the dawn of gamma-ray spectrometry, the resulting spectra were interpreted by a capable technician, who would visually detect all present peaks and compare their energy centroids with a reference library of nuclides. Optionally, the technician might evaluate the area of the peaks in the spectra, adjust them according to the detector efficiency curve, and evaluate the concentration of radionuclides in the sample. Lately, the process has become automated by commercial software packages, which are able to replace much of the manual labor that were by the technician[3].

However, automation came with the cost of sometimes unreliable results when dealing with low levels of radioactivity. In this context, peak detection processes based purely on statistical methods return a high rate of false positives and false negatives.

To tackle this problem, we developed a method of automated peak searching using fuzzy inference system (FIS), which relies on peak shape patterns defined by empirical expertise through a set of fuzzy rules instead of comparing signals with arbitrary one-dimensional thresholds. Signal processing techniques based on fuzzy systems are not novel. It has been used extensively in several kinds of spectra such as doppler [4], oximeter pulse signals[5][6], and radars[7]. Alshennawy & Aly[8] developed a very promising method of edge detection in images using fuzzy inference systems. Their novel approach yielded significantly better results than some algebra-oriented filters. The work of Alshennawy & Aly served as inspiration to our work, that transports the idea of edge detection in images to peak detection in spectra.

This paper describes the mechanism by which our algorithm works and tests both the new proposed method and the classical method of peak detection on artificially generated spectra.

1.1 Gamma-ray spectrometry functionality

Gamma-ray spectrometry is a direct analysis technique which can be used to measure activity concentration of some radioactive isotopes in different samples. The detectable energy in usual gamma-ray spectrometry systems covers the range from 30 to 3000 keV[9].

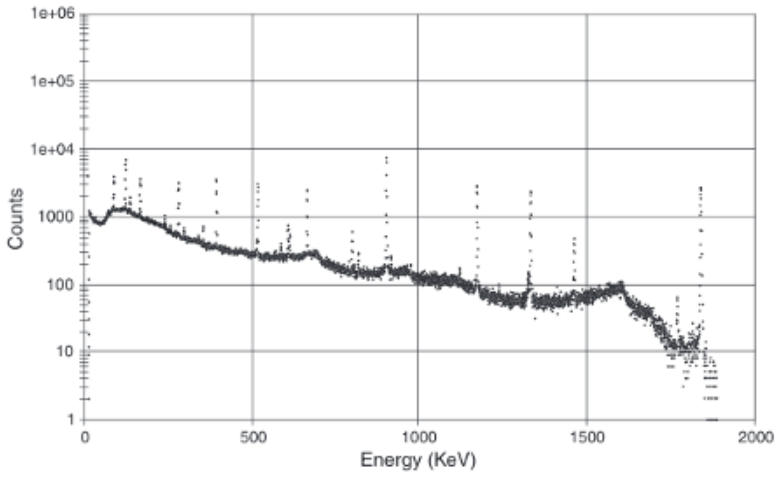

Fig. 1. Example of a typical spectrum acquired by high-purity germanium (HPGe) gamma-ray detectors.[11]

Each radioactive isotope may produce one or more gamma rays bearing different, characteristic energy intensities. When a gamma ray interects with the material that composes the detector, an electrical pulse is produced and converted into digital signal through a spectrometry system. After a reasonable time, the system produces a gamma-ray spectrum, which is the stored histogram of observed emissions by the detector in function of energy or channels[10].

Due to technical limits, spectrometry systems represent emission energies as peaks rather than single-valued lines in the gamma-ray spectrum. Therefore, the first step to the calculation of activity concentration is the identification of peaks in the spectrum. The classical method of peak detection in software packages is based on derivative calculation to identify sudden rises in the spectrum baseline above a certain threshold. Once the peak region is marked, the next step in the detection process is the peak area calculation either by integration of a fitted curve or by direct summing of counts in each channel.

1.2 Fuzzy Inference Systems

Fuzzy inference systems are a group of production rules that use linguistic variables to execute a decision-making process. It can be built around actual data (like measurements and observations), or empirical knowledge suggested by specialists in the field of interest. Therefore, the knowledge base of a fuzzy inference system is formed by a group of conditional statements and which is put into practice through a three-step procedure: fuzzification, inference and defuzzification.[12] During fuzzification, the domain of the input variables is mapped into qualitative descriptions spaces defined by membership functions. The inference step combines the fuzzified domains according to a set of if-then rules, and then defuzzifies the obtained result into crisp numbers as output, normally in the range [0,1]. Nonetheless, fuzzy systems output can be mapped to any suitable domain range.

2 Methods

2.1 Input normalization and scaling

Photopeaks in gamma-ray spectra may have unpredictably different heights due to different sample activities and counting times. Therefore, input data must be normalized to ensure that the same set of fuzzy rules covers all possible cases of signal-to-noise ratios.

The transformation of raw channel counting values into a meaningful variable for the fuzzy system was inspired by outlier detection methods. The most common used form of outlier detection would be normalizing the data in terms of mean and standard deviation, and setting an arbitrary threshold k[14]. Therefore, normalized x_i is given by

$$\frac{x_i - \bar{x}}{\sigma} \, . \tag{1}$$

Thus, every value not in the range $[\bar{x} - k\sigma, \bar{x} + k\sigma]$ would be considered an outlier. Empirically, $k = 2$ is commonly chosen as threshold, so that the 5% most statistically distant points from the center tendency are classified as outliers, assuming the population distribution is approximately normal.

However, the mean becomes an inappropriate measure of central tendency when one or more outliers have extreme values that are much larger than the population average[15]. That would be the case in spectra with peaks that are considerably larger than others. If the average was used as the measure of central tendency, it would be displaced accordingly and minor peaks would be discarded, even though they are sufficiently above the background continuum.

In this regard, some authors[16] recommend the use of median and median absolute deviation (MAD) as a simple and robust tool for outlier detection. Therefore, equation 1 can be adapted to use median instead of mean as central tendency estimator, and MAD instead of standard deviation as a robust measure of the variability of a univariate sample of quantitative data. However, in order to use the MAD in place of the standard deviation σ, a scale factor K must accounted to ensure equivalency between these features[17]. Hence, the equivalent median deviation estimator $\hat{\sigma}$ is given by

$$\hat{\sigma} = K \cdot \mathrm{MAD} \ . \tag{2}$$

Assuming a normal distribution, K is given by $1/(\Phi^{-1}(3/4)) \approx 1.4826$, where Φ^{-1} is the inverse of the cumulative distribution function for the standard normal distribution. Therefore, spectra normalization is done by transforming the count in each channel c_i in the input spectrum by a function f as shown:

$$f(c_i) = \frac{c_i - \mathrm{median}(C)}{1.4826 \times \mathrm{median}_i(|c_i - \mathrm{median}_j(c_j)|)} \ . \tag{3}$$

Consequently, if the resulting value for c_i is so that $f(c_i) > 1$, the channel should be taken as a relevant *signal* instead of background *noise*. This rule is the core of the membership functions as the normalized input spectrum is passed to the FIS.

It was empirically observed that $f(c_i)$ values larger than 2 are generally a relevant signal, usually associated to a peak. Likewise, all values below zero will most certainly stand for background noise. Therefore, to ensure that all input points lie within the fuzzy rules domain, the resulting normalized spectrum is scaled so that negative values are set to zero and values larger than 2 are capped down to 2.

2.2 The algorithm

The algorithm was implemented on MATLAB 7.12.0. The routine uses input spectra in ASCII format, in which each line represents each channel of the spectrum. It then scans the spectrum with a floating window of 3 channels, passing over their normalized values to the FIS.

2.3 Fuzzy Inference System

The FIS was implemented in MATLAB's Fuzzy toolbox using a Mamdani approach. Each channel from the 3-channel window (left, center, and right channel) is classified as either *signal* or *noise*, based on the normalized value calculated by Eq. 3 and a pair of membership functions (Figure 2).

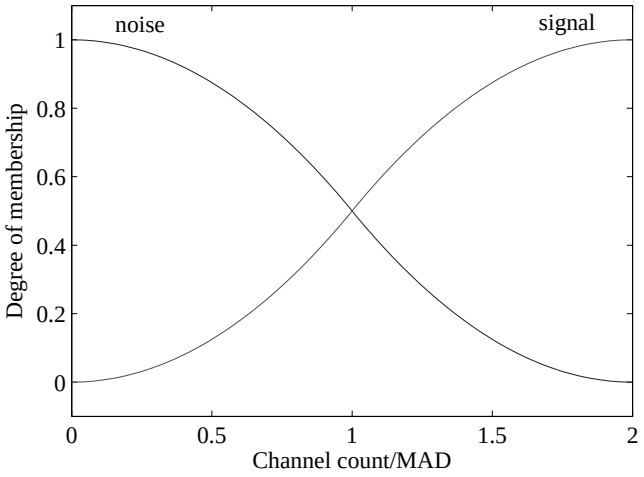

Fig. 2. Channel classification according to median-normalized value. The three input variables (left, center, and right channel) share the same membership functions in this picture.

It can be noticed that when the normalized count value is 1, the membership functions attribute equal membership values (0.5) to both signal and noise, thus softening the arbitrariety of the deterministic threshold used in classical approaches.

Table 1. Rules for channel classification

Rule #	**Left**	**Center**	**Right**	**Output**
1	noise	noise	noise	background
2	noise	noise	signal	edge
3	noise	signal	noise	background
4	noise	signal	signal	peak
5	signal	noise	noise	edge
6	signal	noise	signal	peak
7	signal	signal	noise	peak
8	signal	signal	signal	peak

The fuzzified input is passed over to a Mamdani-type set of rules, which classifies the central channel based on a visual pattern as shown in Table 1. The choice of each rule was based on empirical observation of peak behavior.

The reasoning behind Rules #3 and #6 in table 1 are that fluctuations along the spectra may occur. A lone relevant signal among background noise is most likely not to be part of a peak, whereas a slightly below threshold signal among relevant signals is most likely to be a faulty signal which would still fit as a peak.

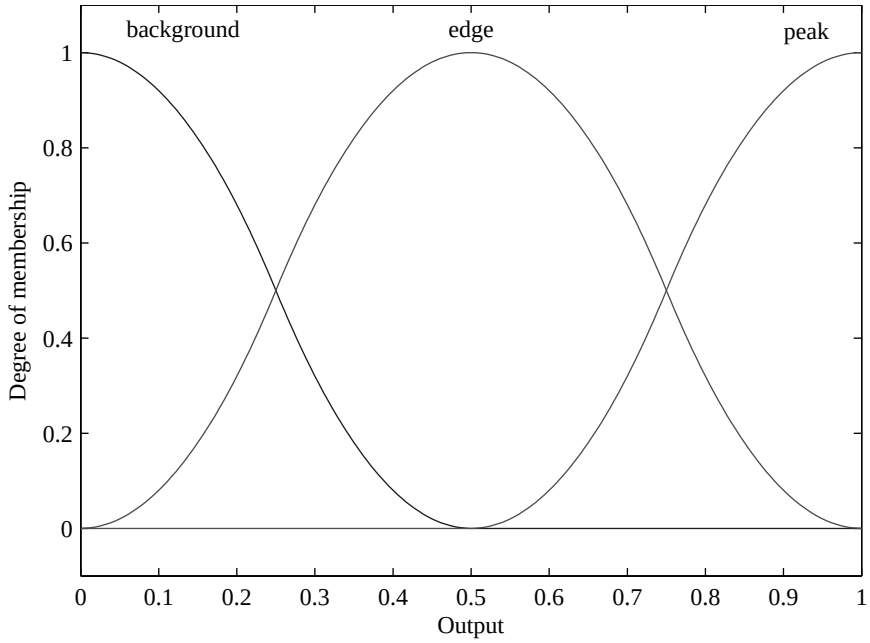

Fig. 3. Channel classification output.

The remaining rules seek to find peak edges as they stand out from the continuum, and afterwards, blend with the continuum as the floating 3-channel window moves to the right. After all relevant channels are marked as part of a peak, the values of all channels between the two peak edges are multiplied by a weight vector composed by their respective FIS defuzzified ouput. Defuzzification is done with the Largest of Maximum (LOM) method. The peak area is then calculated by summing the weighted channels and subtracting the baseline region under the peak.

2.4 Testing

The algorithm was benchmarked against the generalized second difference algorithm[3], a de facto standard peak detection method in several kinds of spectrometry[18]. The test was made with 1000 artificially generated spectra simulating

an energy calibration of 0.5 keV per channel. From this dataset, 50% spectra contained pure background noise, and the other 50% contained background noise added to a randomly generated gaussian peak. The pure background spectra contained 16 counts per channel in average. In the background spectra with added peak, the peak was generated based on a FWHM (full width at half-measure) of 1.0 keV, and its true count value is a random number between 0.5 and 1.5 times the theoretical limit of detection (LD)[19], where:

$$\mathrm{LD} = 2.71 + 4.65 \times \sqrt{B} \tag{4}$$

B represents the background continuum beneath the peak. A value of 100 was used, which is close to the evaluated background continuum.

Both the second difference and the fuzzy algorithm had their peak detection performance evaluated by area under ROC (receiver operating characteristic) curve, and area calculation was evaluated by coefficient of determination (r^2), root-mean-squared error (RMSE), and mean relative deviation.

3 Results

The results of the first part of the performance test, in which the algorithms should detect the presence of the artificial peak in nearly half the 1000 generated spectra, is shown in Figure 4.

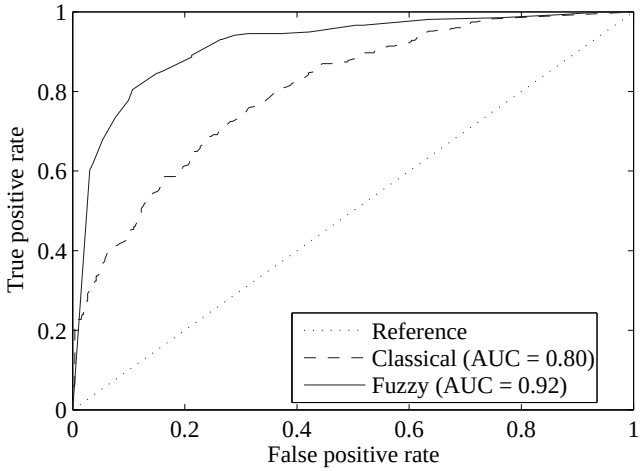

Fig. 4. ROC curve of both methods.

ROC curves are used to evaluate performance of binary classifier systems as their discrimination thresholds are changed[20]. The curves are created by plotting the false positive rate (1 - specificity) against the true positive rate

(sensitivity) of each model at various threshold settings. Performance then is numerically expressed as the area under curve (AUC). The dashed line in the diagonal is equivalent to the ROC curve of a random predictor with AUC of 0.5. The random predictor is commonly used as a baseline to see how useful the models are compared to pure chance. Therefore, a good model shall approach 1.0 whereas nearly random models shall get closer to 0.5. Given that peak areas were as near as possible to the theoretical limit of detection, both classical and fuzzy approaches showed reasonable performance, although the fuzzy model yielded a significantly better result (AUC = 0.80 and 0.92 respectively).

In the second part of the performance test, in which the algorithms should estimate the actual count of the generated peak, performance difference was less pronounceable, as shown in Table 2.

Table 2. Peak count estimation performance

	Classical	**Fuzzy**
r^2	0.16	0.33
RMSE	15.64	9.63

Both coefficient of determination (r^2) and root-mean-squared error (RMSE) were calculated using only true positive events to give a better perspective of the algorithms functioning. Either marks were numerically close to each other, although the fuzzy approach still returns better figures.

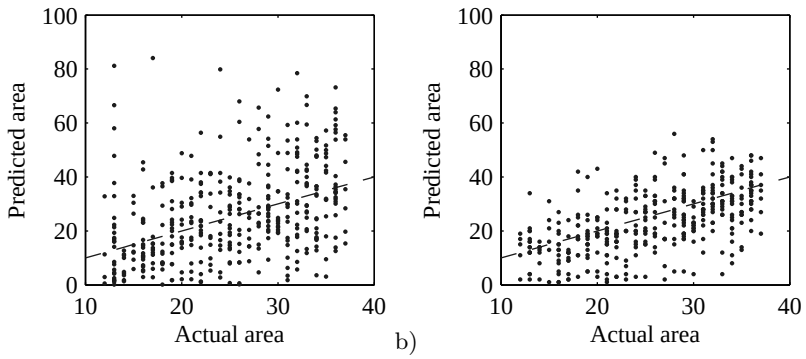

Fig. 5. Scatter plots depicting actual peak area versus calculated peak area by both (a) second difference algorithm and (b) proposed fuzzy algorithm. The diagonal dashed line represents the optimal scenario where estimated and actual peak count would be equal.

As seen in Figure 5a, the classical approach deviates from non-uniformity around the optimal solution (near to the diagonal line). Also, there is a noticeable

trend of the classical algorithm to overshoot estimated counts when peak size is close to zero. This effect can be seen more clearly in Figure 6. Nonetheless, the fuzzy approach (Figure 5b) yielded a more linear correlation between actual and estimated counts ($r^2 = 0.16$ for classical approach and 0.33 for fuzzy approach).

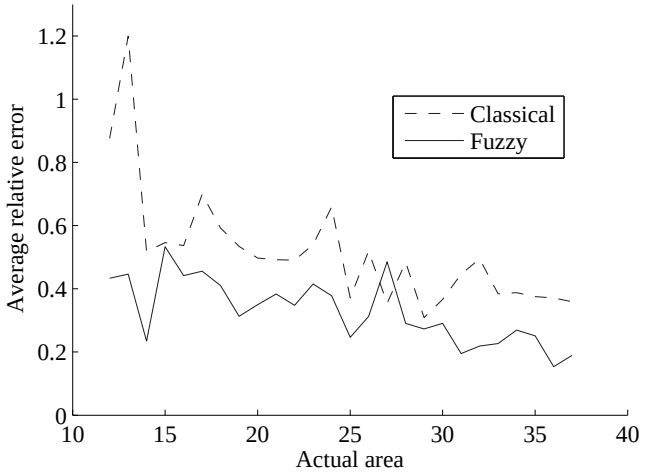

Fig. 6. Average relative error (mean absolute deviation divided by peak area value) for each possible value.

4 Conclusion

This work shows how a minimalist fuzzy system can improve manifold the rate of false negatives and false positives when dealing with low-activity peaks in typical gamma-ray spectra. The use of fuzzy thresholds, along with empirical knowledge from technicians who are able to visually establish patterns of peak shape, shows that classical methods on signal detection and novel computational approaches can work together to yield better results in the field of gamma-ray spectrometry. New problems arise, though, as it is the case where multiplets are present in the spectrum. As of yet, the implemented fuzzy model is unable to differentiate convoluted peaks. However, the solution is straightforward and could be implemented in future works. Also, further studies must be made with densely populated spectra, as well as peaks with more intense activity.

References

1. L'Annunziata, M.F., ed.: Handbook of radioactivity analysis. 2nd ed edn. Academic Press, San Diego (2003)

2. Gilmore, G.: Practical gamma-ray spectrometry. 2nd ed edn. Wiley, Chichester, England ; Hoboken, NJ (2008)
3. Canberra: Genie 2000 Spectroscopy Software: Customization Tools. Canberra Industries, Meriden (2006)
4. Cornman, L.B., Goodrich, R.K., Morse, C.S., Ecklund, W.L.: A Fuzzy Logic Method for Improved Moment Estimation from Doppler Spectra. Journal of Atmospheric and Oceanic Technology **15**(6) (December 1998) 1287–1305
5. Stetson, P.: Determining heart rate from noisy pulse oximeter signals using fuzzy logic. In: The 12th IEEE International Conference on Fuzzy Systems, 2003. FUZZ '03. Volume 2. (May 2003) 1053–1058 vol.2
6. Ho, K., Tsuchiya, N., Nakajima, H., Kuramoto, K., Kobashi, S., Hata, Y.: Fuzzy logic approach to respiration detection by air pressure sensor. In: IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE 2009. (August 2009) 911–915
7. Gader, P.D., Nelson, B.N., Frigui, H., Vaillette, G., Keller, J.M.: Fuzzy logic detection of landmines with ground penetrating radar. Signal Processing **80**(6) (June 2000) 1069–1084
8. Alshennawy, A.A., Aly, A.A.: Edge detection in digital images using fuzzy logic technique. World Academy of science, engineering and technology **51** (2009) 178–186
9. Rosa, M.M., Custódio, L.G., Bonifácio, R.L., Taddei, M.H.: Validação do Ensaio de Medidas por Espectrometria Gama de Radionuclídeos em Matrizes de Alimentos. In: IX Latin American IRPA Regional Congress on Radiation Protection and Safety-IRPA 2013. (2013)
10. Reilly, D., Ensslin, N., Smith Jr, H., Kreiner, S.: Passive nondestructive assay of nuclear materials. Technical report, Nuclear Regulatory Commission, Washington, DC (United States). Office of Nuclear Regulatory Research; Los Alamos National Lab., NM (United States) (1991)
11. Canberra: Gamma and X-Ray Detection (2014)
12. Rabêlo, R.d.A.L., Carneiro, A.A.d.F.M., Fernandes, R.A.S., Braga, R.T.V.: Uma abordagem baseada em sistemas de inferência Fuzzy Takagi-Sugeno aplicada ao planejamento da operação de sistemas hidrotérmicos de geração. Sba: Controle & Automação Sociedade Brasileira de Automatica **22**(1) (February 2011) 49–64
13. Zadeh, L.A.: Fuzzy sets. Information and control **8**(3) (1965) 338–353
14. Hodge, V.J., Austin, J.: A survey of outlier detection methodologies. Artificial Intelligence Review **22**(2) (2004) 85–126
15. Brookner, R.J.: A Note on the Mean as a Poor Estimate of Central Tendency. Journal of the American Statistical Association **36**(215) (September 1941) 410–412
16. Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L.: Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology **49**(4) (July 2013) 764–766
17. Rousseeuw, P.J., Croux, C.: Alternatives to the Median Absolute Deviation. Journal of the American Statistical Association **88**(424) (December 1993) 1273–1283
18. Yang, C., He, Z., Yu, W.: Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. BMC Bioinformatics **10** (January 2009) 4
19. Currie, L.A.: Detection and quantification limits: origins and historical overview. Analytica Chimica Acta **391**(2) (1999) 127–134

20. Beck, J.R., Shultz, E.K.: The use of relative operating characteristic (ROC) curves in test performance evaluation. Archives of pathology and laboratory medicine **110**(1) (1986) 13–20