

A Procedure to Fit Nuclear Decay Data With Proper Treatment of Outliers

Guilherme Soares Zahn^{*†}

IPEN-CNEN - São Paulo

E-mail: gzahn@ipen.br

Vitor C. Gonçalves

IPEN-CNEN - São Paulo

E-mail: vitorc.goncalves@hotmail.com

Frederico Antonio Genezini

IPEN-CNEN - São Paulo

E-mail: fredzini@ipen.br

In this work, a variation of the Normalized Residuals [1] robust average technique is proposed where a regular least-squares fit is performed, and then points with large residuals are either rejected or have their uncertainties adjusted; this process is repeated until all residuals are in a reasonable range, and then the result of the last iteration should be the best estimate for the fit parameters, with minimum interference from the outlier (or near-outlier) values. This technique was then tested in the process of fitting the half-lives of the ^{57}Co and ^{60}Co isotopes using data from daily detector verifications performed at the LAN-CRPq-IPEN laboratories, which records are subject to many types of mistakes, therefore they could be used as a good testbench for the outlier treatment proposed.

XXXIV edition of the Brazilian Workshop on Nuclear Physics

5-10 June 2011

Foz de Iguaçu, Parana state, Brasil

^{*}Speaker.

[†]The authors would like to thank Dr. Mitiko Saiki and Regina Ticianelli from LAN/IPEN for the detector verification data used in this work.

1. Introduction

In the process of analysing experimental results, especially in situations when the experiment can't be fully controlled, one faces the problem of identifying and treating outlier results which can be a result of faulty equipment, experimental mistakes, and so on. To address the issue of identifying outliers, many statistical tools have been proposed [2]; nevertheless, not only this is still a matter of dispute but the problem of how to deal with these outlier data is an open question.

When the task is to take an average of different measurements for the same physical parameter, with the proper evaluation of all different values, of their uncertainties, and the identification and treatment of outlier values with the aim of reaching the best possible estimate for both the measured value and its uncertainty, the *Rajeval Technique* or the *Normalized Residuals Mean* [1], for instance, are strong tools for making a good estimate with low uncertainty and a good treatment of outlier data. When the task is to fit data to a function, though, these techniques don't apply directly and there aren't tools to perform this task other than the regular fitting procedures, which don't take care of possible outliers.

1.1 The Proposed Technique

In order to address the problem of fitting a function to the data while properly treating outliers, a generalization of the *Normalized Residuals* (NR) technique is proposed. The original NR procedure (as quoted in [1]) can be summarized as follows:

1. Given an original "dirty" set of results $R[y, \sigma(y)]$, an unweighted mean is taken and the normalized residuals for each data point (r_i) are calculated;
2. A limiting value for this normalized residual, r_0 , is also calculated in terms of the number of data in the set;
3. Measurements with $|r_i| > r_0$ have their uncertainties enlarged so that $|r_i| = r_0$;
4. A *weighted* mean of the new set of results $R[y, \sigma'(y)]$ is taken, the normalized residuals are recalculated and the step 3 is repeated;
5. Step 4 is repeated until a predetermined convergence rule is fulfilled.
6. The resulting \bar{y} and $\bar{\sigma}$ are considered as the best estimates for the measured parameter and its uncertainty, respectively.

When the problem is fitting a function to the data, some changes are required to the NR procedure. First of all, the normalized residual, defined as:

$$r_i = \sqrt{\frac{w_i \cdot W}{W - w_i}} \cdot (y_i - \bar{y}) \quad (1.1)$$

(where y_i and w_i are the individual values and weights, and \bar{y} and W the average and total weight) can't be used because in some cases it may lead to a square root of a negative number (keeping in mind that in such a fit a sensible translation would be $w_i = \sigma_i^{-2}$ and $W = \bar{\sigma}^{-2}$); therefore, the

regular residual ($r_i = (y_i - \bar{y})/\sqrt{\sigma_i^2 + \bar{\sigma}^2}$) was used instead. Also, due to this change, in a first approach the limiting residual was chosen as 3 – i.e., data outside a 99.7% probability interval are considered as outliers and have the uncertainties adjusted. Moreover, data with $|r_i| > 5$ (i.e., outside a 99.9999% probability interval) were considered as gross outliers and discarded. In this way, the new procedure can be outlined as follows:

1. Given the original “dirty” set of results $R[x, y, \sigma(y)]$, an unweighted fit is performed, the predicted results for each value of x are interpolated (with their respective uncertainties) and the residuals for each data point (r_i) are calculated;
2. Measurements with $|r_i| > 5$ are discarded;
3. A *weighted* fit of the new set of results $R'[x, y, \sigma(y)]$ is performed, the predicted results for each value of x are interpolated again (with their respective uncertainties) and the residuals for each data point (r_i) are recalculated;
4. Measurements with $|r_i| > 5$ are discarded and the ones with $|r_i| > r_0$ have their uncertainties enlarged so that $|r_i| = r_0$;
5. Steps 3 and 4 are repeated until a predetermined convergence rule is fulfilled (in the present case it is $\chi_{i-1}^2 - \chi_i^2 \leq 0.01$), where χ^2 for N datapoints is defined as:

$$\chi^2 = \frac{1}{N} \sum_{j=1}^N \frac{(y_{fit,j} - y_j)^2}{\bar{\sigma}_j^2 + \sigma_j^2} \quad (1.2)$$

2. Experimental Procedure

In order to test the proposed procedure, data from daily detector verification from IPEN Neutron Activation Laboratory was used to determine the half-lives of the isotopes used in these verifications (^{57}Co and ^{60}Co). The dataset used was taken in an almost daily basis with a 25% Canberra beryllium window HPGe detector over a period of about 2 years. This verification is performed by the first person to use the detector on each day by placing a source comprised by a ^{57}Co source glued over a ^{60}Co source at a standard source holder at 9 cm from the face of the detector and counting for 300s (Live Time). The spectrum is then processed using the in-house developed software VISPECT to obtain the *counts per second* for each peak and the results for the 122keV transition of ^{57}Co and for the 1333 keV transition from ^{60}Co are printed and manually transcribed into a digital spreadsheet. This dataset is a good testbench for the proposed procedure since it is very prone to errors, from transcription mistakes to mispositioning of the source and such. The results were then fitted to an exponential decay (Eq. 2.1) using a covariant fitting procedure implemented in the MatLab environment.

$$Y = A_0 \cdot e^{-\ln(2) \cdot t / T_{1/2}} \quad (2.1)$$

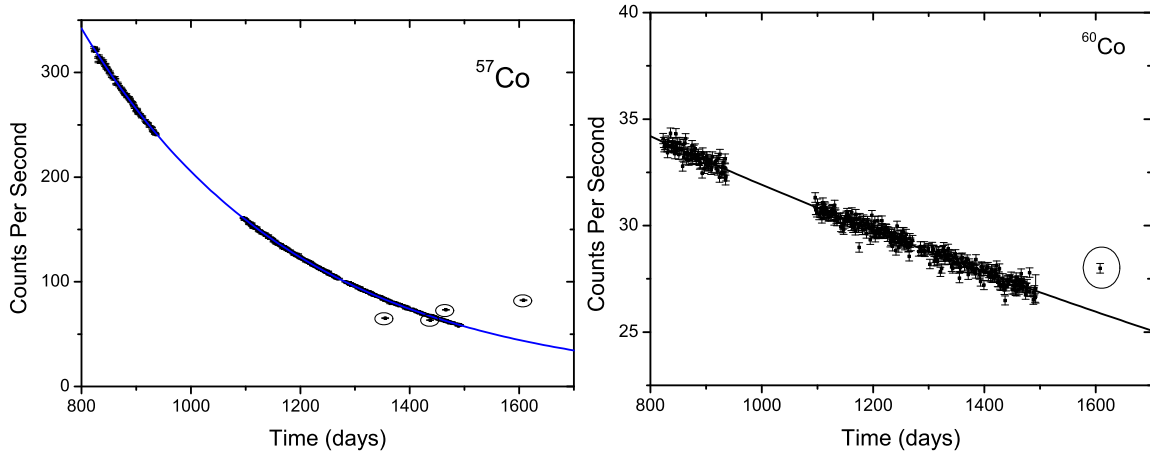


Figure 1: Experimental data used in this work for the ^{57}Co (left) and ^{60}Co (right) decays; the continuous lines show the curves obtained using the proposed method and the circles show the datapoints that were automatically rejected by the procedure as “gross outliers”.

3. Results and Discussion

The results obtained using the proposed technique are shown in table 1, together with the results obtained using a regular σ^{-2} -weighted least squares fit and the values found in the literature for the half-lives. The experimental data for both decays are shown in Fig. 1, together with the curves obtained using the proposed fitting procedure.

	^{57}Co			^{60}Co		
	A_0 (cps)	$T_{1/2}$ (days)	χ^2	A_0 (cps)	$T_{1/2}$ (days)	χ^2
This Method	2633.6(25)	271.73(9)	2.1	45.05(10)	2012(11)	1.3
Regular Fit	2826.6(25)	271.97(9)	36.3	44.97(10)	2020(11)	1.6
Literature	*	271.74(6)[3]	*	*	1925.28(14)[4]	*

* Does not apply.

Table 1: Results obtained with the proposed method compared to a regular σ^{-2} -weighted fit and, in the case of the half-lives, to the tabulated value from ENSDF [3, 4].

The results for both decays show that the half-life values obtained using the proposed procedure are in better agreement with the values found in the literature than the results obtained using a regular σ^{-2} -weighted fit. In the case of ^{57}Co , the proposed procedure yielded a result that is very close to the tabulated value and has an uncertainty of the same magnitude, while a regular fit would give the same uncertainty, but lead to an overestimated value. In the case of ^{60}Co , though, while the value obtained using the proposed procedure is closer to the tabulated value than the one from the regular fit, both results are considerably higher than the tabulated value – in fact, the experiment duration of about 2 years is too short to allow for a good half-life estimate for this isotope, so one shouldn’t really expect to reach a good value in this test. In both cases the gross outliers were efficiently rejected, too. It also should be mentioned that convergence was reached after 3 full

iterations (i.e., three consecutive weighted fits) in both cases, and that both datasets were initially comprised of 381 values.

4. Conclusion

The proposed fitting procedure proved to be efficient in fitting nuclear decay data, eliminating gross outliers present in the two datasets used in the test, reducing the χ^2 of the fit and yielding results that were closer to the expected values than a regular σ^{-2} -weighted fit. Convergence was reached after only a few iterations (three in both cases) for a quite large dataset (381 values), thus proving that it can be efficiently used in routine data fitting.

References

- [1] M. U. Rajput, and T. D. MacMahon, *Nucl. Instrum. Meth. A* **312**, 289–295 (1992).
- [2] P. M. S. Oliveira, C. S. Munita, and R. Hazenfratz, *J. Radioanal. Nucl. Ch.* **283**, 433–437 (2010).
- [3] M. R. Bhat, *Nucl. Data Sheets* **85**, 415–536 (1998).
- [4] J. K. Tuli, *Nucl. Data Sheets* **100**, 347–481 (2003).