

A New Data Smoother Applied to Low-resolution Sodium-Iodide Detectors

Thomas L. Burr, CCS-6;
Nicholas W. Hengartner, CCS-3;
Steven C. Myers, N-2

The energy spectra of gamma-rays emitted by radioisotopes act as fingerprints that enable identification of the source. Such identification from low-resolution sodium iodide (NaI) detectors over short time periods is challenging for several reasons, including the Poisson fluctuations in the recorded counts. Smoothing the data over neighboring energy bins can reduce noise in the raw counts, at the cost of introducing a bias that de-emphasizes the peaks and valleys of the spectrum. This note describes a new two-stage smoothing procedure that uses a multiplicative bias correction for adjusting initial smoothed spectra. The benefits of this new method are illustrated on real spectra.

Performance of radioisotope identification (RIID) algorithms using gamma-ray spectroscopy is a subject of increasing importance. For example, sensors deployed at locations that screen for illicit nuclear material rely on isotope identification to resolve innocent nuisance alarms arising from naturally occurring radioactive material (NORM) [1-3]. Studies comparing the energy spectra measured by low, medium, and high resolution detectors suggest that relatively low-resolution detectors, such as sodium iodide (NaI) detectors that typically count the number of gamma photons in 512 or 1024 calibrated energy channels, will continue to play a key role in NORM identification [1]. However, RIID performance from the raw counts collected over short time intervals is strongly impacted by intrinsic variations from Poisson count statistics.

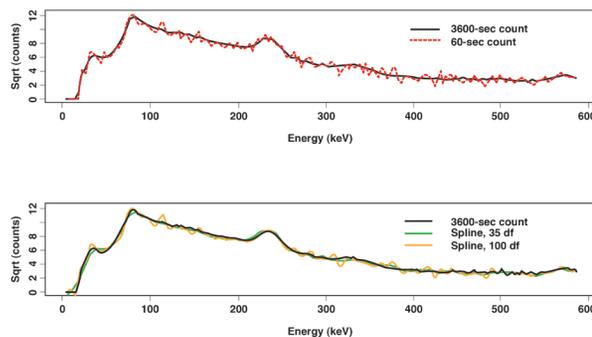
Various strategies for improving the RIID by reducing the noise in the data are being considered. One noise-reduction strategy is to smooth the spectral data over neighboring energy bins. This smoothing decreases the variance of the raw counts at the cost of introducing some bias that de-emphasizes the peaks and valleys of the spectrum. Most smoothers also distort peak shape by broadening the base and midsection. Preliminary results in [4] suggest, but do not prove, that substituting the raw NaI spectral counts with their smoothed values can lead to improved performance of existing RIID algorithms. Even better performance can be expected if one can reduce the bias of the smoothed counts without substantially increasing their variance.

Reference [5] describes a new two-stage smoother that uses a multiplicative bias correction (MBC) for adjusting initial smoothed energy spectra, thereby correcting some of the deficiencies of traditional smoothers. Reference [6] describes the effect of the MBC-smoother in the context of detecting injected point-like threat isotopes onto each of many real vehicle profiles.

The bias of main concern in smoothing spectral data is the tendency by all smoothers (local polynomial, local kernel, wavelet, etc.) to underestimate the peaks and overestimate the valleys of the true spectrum $f(x)$. The true count rate as a function of energy x is denoted $f(x)$, and the observed count rate in a given spectrum is denoted $y = f(x) + e$ (see Fig. 1). High-degree-of-freedom fits, such as the 100-degree-of-freedom spline in Fig. 1, will have smaller bias in the peaks, but larger variance in the off peaks. As a result, the peaks are fit very well, but the smoother tends to generate spurious smaller peaks. The opposite is true for low-degree-of-freedom fits, in which the smoother fails to fully catch the peaks but does not tend to introduce additional spurious peaks.

Finding a smoother that minimizes the root mean squared error (RMSE) does not by itself guarantee that the resulting smoother has desirable properties at the peaks while not generating spurious ones. This leaves open the question of selecting the smoothing parameter. Instead of attempting to optimize the bandwidth h for the RIID application, we focus on estimating the bias and correcting the initial smoother using that estimate. This is not a new idea, but the application is new, and most of the approaches described in the literature focus on additive corrections of the bias. Such corrections may not be desirable in the present context, as they can lead to negative-valued energy spectra that are undesirable as input into existing RIID algorithms.

Fig. 1. Example ^{232}Th spectrum for one minute, for sixty minutes (top), and two example spline fits (bottom). The 100 df spline over-fits the off-peak channels. The 35 df spline over-smooths the peaks and valleys.



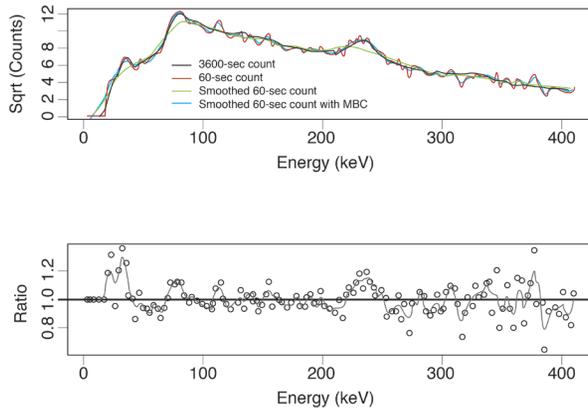


Fig. 2. Example ²³²Th spectrum for 3600 seconds, for 60 seconds, and a lokerns fit with and without the multiplicative bias correction; (bottom) The ratio $R_i = \frac{y_i}{\hat{y}_{1,i}}$

A recent attractive alternative [5 is to apply an MBC to an initial smoother. The asymptotic analysis and small sample simulations in [7] show that this method is effective at reducing the bias of the initial smoother without increasing the variance. As a result, one obtains estimates for the smoothed spectra that better fit the peaks and valleys but do not introduce additional spurious peaks or under-smooth the flat regions. For good performance, the MBC approach needs the initial smoother to not under-smooth. Over-smoothing by the initial smoother is acceptable. Therefore, we used cross-validation to select the bandwidth for the initial smoother. Cross-validation [8] involves

leaving out x,y pairs while calculating the smooth function at trial bandwidths and then testing the performance of the trial bandwidths on the held-out pairs.

For completeness, we describe the MBC smoother here. Starting with an initial kernel smoother for $\sqrt{f(x)}$, with weight function $w_{1,i}(x, h_1)$ and bandwidth h , $\hat{g}(x) = \sum_{i=1}^n w_{1,i}(x, h) \sqrt{y_i}$ we denote by $\hat{y}_{1,i} = \hat{g}(x_i)^2$ the predicted spectrum in bin i , and construct the auxiliary variables $R_i = \frac{y_i}{\hat{y}_{1,i}}$.

Each value R_i is an unbiased, but noisy, estimate of the relative bias in the i th energy bin. Smoothing the latter gives rise to a multiplicative correction factor $\hat{R}(x) = \sum_{i=1}^n w_{2,i}(x, h_2) R_i$ for the initial smoother.

Multiplying the predicted spectrum by the latter leads to the estimate $\hat{f}_2(x) = \hat{R}(x) \hat{g}(x)^2$ for the energy spectra $f(x)$, and in particular the estimate $\hat{y}_{2,i} = \hat{f}_2(x_i)$ for the expected counts of gamma-photons in the i th bin. Neither the bandwidth nor the smoother for the correction factor need be the same as for the initial smoother.

Figure 2 illustrates the MBC for the same ²³²Th source as in Fig. 1. Notice again at least three distinct peaks and notice the trend in $R_i = \frac{y_i}{\hat{y}_{1,i}}$ which suggests that peaks will be better estimated using the corrected values $\hat{y}_{2,i}$ rather than predictions from the initial smoother $\hat{y}_{1,i}$.

In NaI spectra, peak width varies as a function of energy because detector resolution varies with energy. Adaptive smoothing is therefore preferred. There are two main options for the adaptive smoothing of

$y = f(x) + e$. One option is to adjust the local bandwidth according to the local density of the x data. This option is not relevant for our context because of equally spaced x values, regardless of whether one regards x to be energy or channel number. A second option is to identify regions of rapid change in y , such as peak regions, and modify the bandwidth accordingly. In effect, the MBC applies the second option. However, note that the MBC does not require one to identify the peak regions. This is a key advantage of the MBC, which provides adaptive smoothing, by “smoothing less” in the peaks. Smoothing less in the peaks occurs because the ratio $R_i = \frac{y_i}{\hat{y}_{1,i}}$ used in the MBC tends to pull

the first smooth $\hat{y}_{1,i}$ toward y_i and hence leads to “smoothing less,” as desired in the peaks.

We have found that both smoothing splines and a local kernel smoother (such as lokern) work well as a second smoother, although that choice is not critical.

[1] Stromswold, D. et al., IEEE Nucl Sci Symp, <http://ieeexplore.ieee.org/iel5/9892/31433/01462180.pdf> (2004).
 [2] Burr, T. and M.S. Hamada, *Algorithms* **2**, 339 (2009).
 [3] Burr, T. et al., *Qual Reliab Eng Int* **25**, 899 (2009).
 [4] Myers, S. et al., “Choosing the Number of Repeat Measurements in Radioisotope Identification Performance Testing,” Los Alamos National Laboratory Unrestricted Release Report LA-UR 08-1764 (2008).
 [5] Burr, T. et al., *IEEE Trans Nucl Sci* **57**, 2831 (2010).
 [6] Burr, T. et al., *Appl Radiat Isot* **69**, 1436 (2011).
 [7] Hengartner, N. et al., “Multiplicative Bias Corrected Nonparametric Smoothers with Application to Nuclear Energy Spectrum Estimation,” <http://arxiv.org/abs/0908.0128v1> (2009).
 [8] Hastie, T. et al., *The Elements of Statistical Learning*, Springer, New York (2001).

Funding Acknowledgments
 Department of Homeland Security