New prospects for automated particle identification

A.B. McIntosh, S. Wuenschel, K. Hagel, S.J. Yennello, S. He,[1] and J. Huang[1]
[1]*Department of Statistics, Texas A&M University, College Station, Texas*

A new method for analysis of raw experimental data to obtain PID (particle identification) in an automated way is being explored. This report contains a brief description of the goals of PID and the drawbacks of current methods. This report focuses on the description of the new method in development, and the successes and current status of the new method. A much more detailed description will be submitted to a statistics journal in the coming months.

Particle identification in nuclear physics experiments has become incredibly sophisticated over the last few decades. Detector arrays have become larger and have achieved new records in particle isotopic resolution. This allows new physical phenomena to be explored, but comes at a price: the particle identification is incredibly labor intensive. For example, in the NIMROD array, nearly 200 detector telescopes employ the ΔE-E technique for charged particle identification for Z>2 and pulse shape discrimination for lighter ions. Each 2D spectrum contains a series of visually discernible bands (see Fig. 1), which correspond to unique particle types. The goal of the PID process is to develop an algorithm that

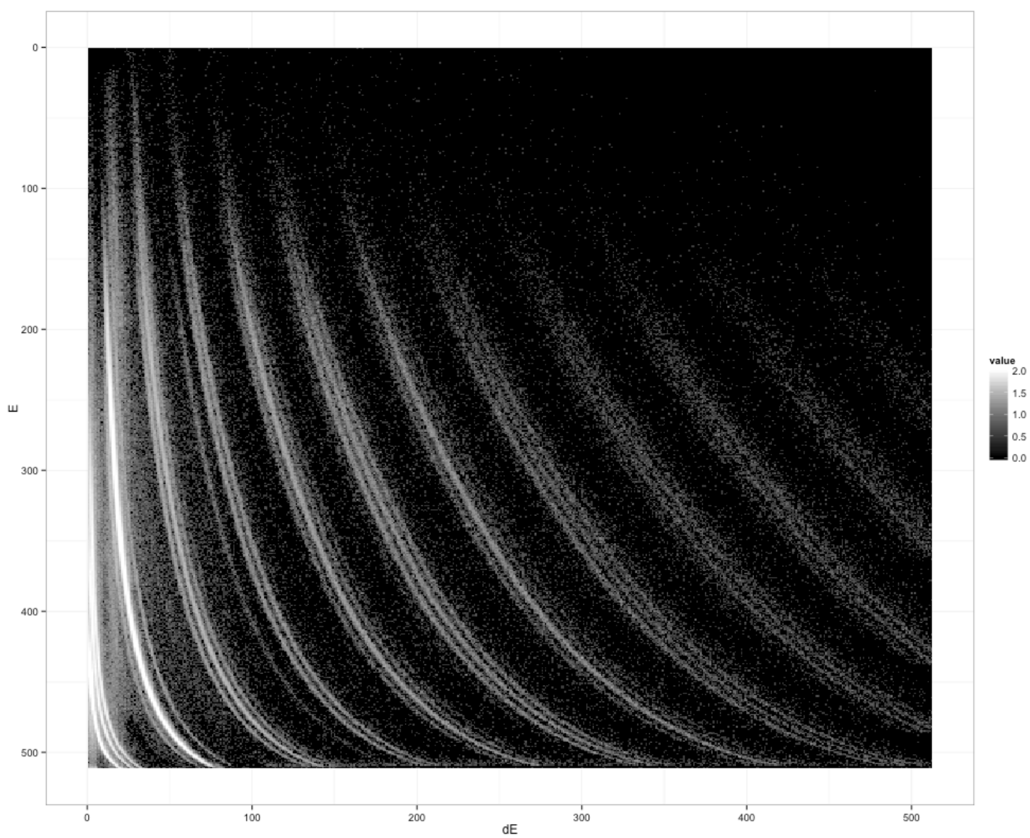

FIG. 1. Raw E vs ΔE histogram for a detector of the NIMROD array showing isotopic resolution. The reaction producing the charged particles is Kr+Ni@35AMeV.

can sift through the hundreds of millions of events and automatically determine what type of particle each datum corresponds to. To date, this process relies on humans to generate a large series of curves along the bands of data by manually defining points through which they desire the curve to pass. This labor-intensive step is measurable in man-years. The goal of the present investigations is automation through statistical methods.

The new method shows much promise already at automatically finding curves to describe the ridges in the data. The first step in the process is to bin the data appropriately and apply a smoothing function to minimize the effects of statistical fluctuations. Fig. 2 shows a smoothed version of the data in Fig. 1. The isotopic resolution is still clearly discernible. The smoothing uses a wavelet method. Briefly, this takes the two-dimensional surface and applies a local transformation akin to a Fourier transform to filter out undesirable frequency components.

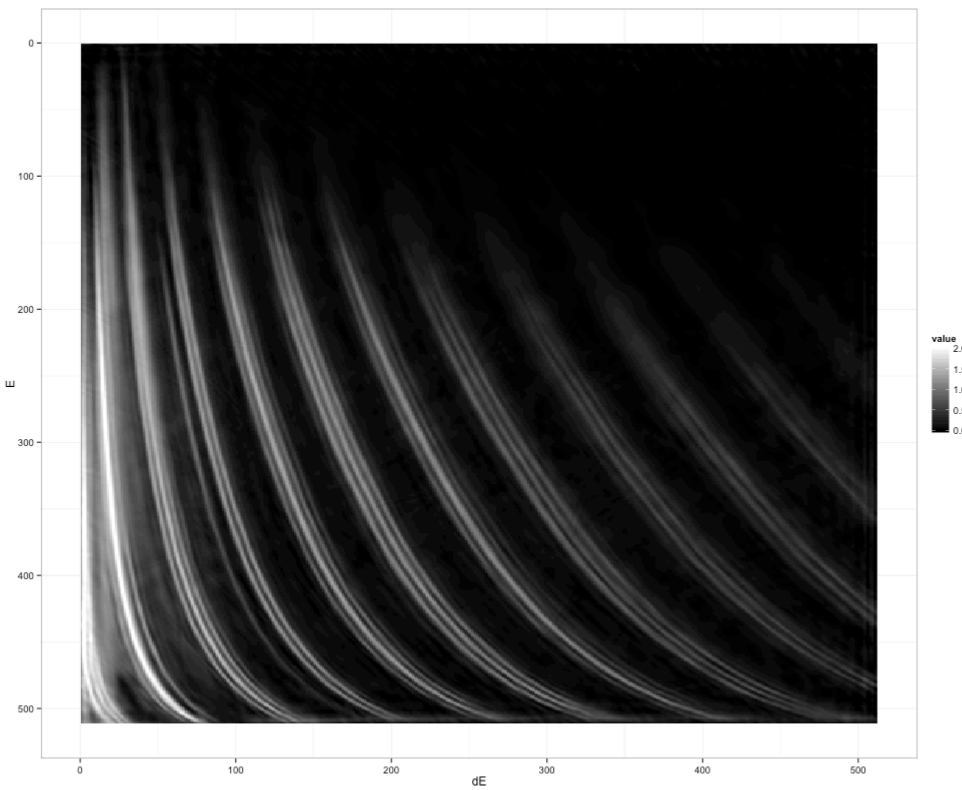

FIG. 2. A smoothed version of the data in figure 1. See text for explanation.

The new method then examines slices of this smoothed two-dimensional surface to find local maxima. The results of this are shown in Fig. 3. The gray data points show the same raw data represented in Figs 1 and 2. The red points are the automatically obtained local maxima. The local maxima describe an impressive range of isotopic loci from Z=1 to Z=11. A nearest neighbor algorithm is then able to group these maxima together to allow each series of point to be fit with a spline that follows the ridge in the original data.

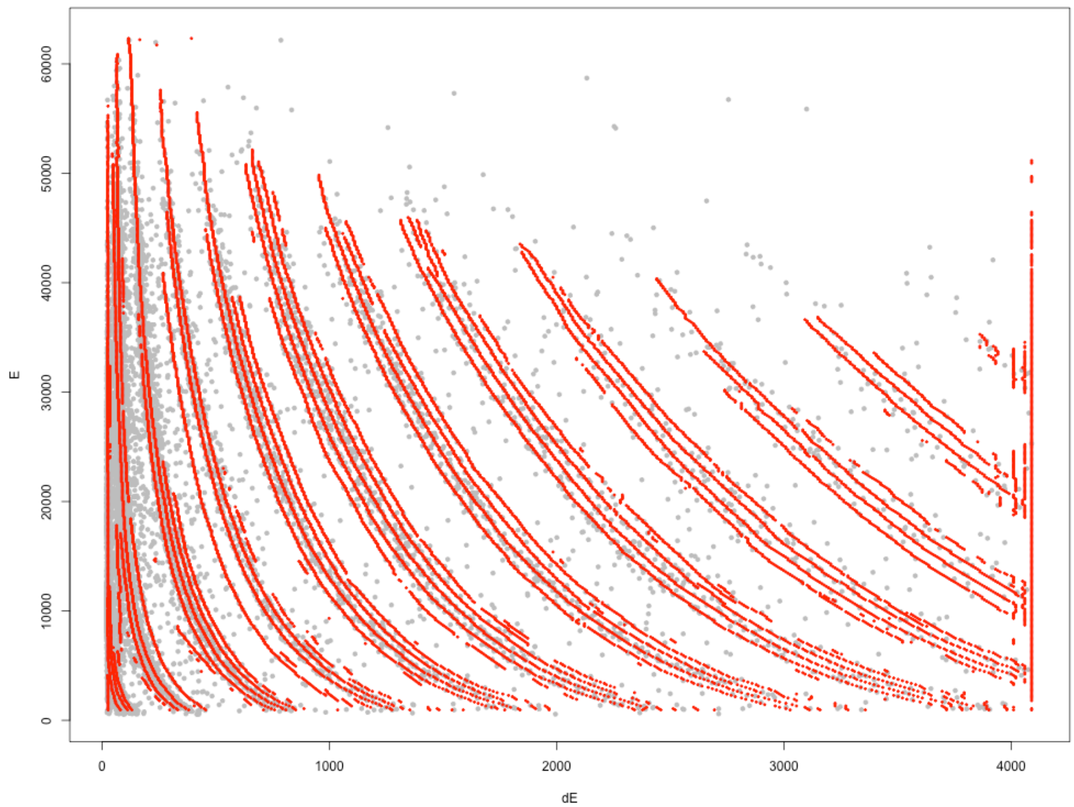

FIG. 3. Same raw data (gray) as for Figs 1 and 2. The red points are obtained automatically by finding the local maxima in Fig. 2.

The algorithm is being testing on a broader range of detectors to test is robustness. The ease with which these local maxima are obtained already suggests that this algorithm and derivatives of it may find widespread use in particle identification and similar problems in nuclear science.

This work is part of a broad initiative to advance scientific output in “Big Data” experiments through collaboration between scientists and statisticians. This work is supported by the Department of Energy (DE-FG02-93ER40773, Cyclotron Inst.), the Welch Foundation (A-1266, Chemistry Dept.), and the National Science Foundation (DMS-1208952, Statistics Dept.).