EPJ Plus

EPJ.org
your physics journal

Eur. Phys. J. Plus (2014) **129**: 239 DOI 10.1140/epjp/i2014-14239-3

# Detection of nuclear sources in search survey using dynamic quantum clustering of gamma-ray spectral data

Marvin Weinstein, Alexander Heifetz and Raymond Klann

Società Italiana di Fisica

Springer

Eur. Phys. J. Plus (2014) **129**: 239
DOI 10.1140/epjp/i2014-14239-3

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

# Detection of nuclear sources in search survey using dynamic quantum clustering of gamma-ray spectral data

Marvin Weinstein[1,2], Alexander Heifetz[3,a], and Raymond Klann[3,b]

[1] Quantum Insights LLC, 3845 Nathan Way, Palo Alto, CA, 94303, USA
[2] Stanford Linear Accelerator Center (Emeritus), 2575 Sand Hill Road, Menlo Park, CA 94025, USA
[3] Nuclear Engineering Division, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

Received: 18 June 2014 / Revised: 19 September 2014
Published online: 7 November 2014 – 

**Abstract.** In a search scenario, nuclear background spectra are continuously measured in short acquisition intervals with a mobile detector-spectrometer. Detecting sources from measured data is difficult because of low signal-to-noise ratio ($S/N$) of spectra, large and highly varying background due to naturally occurring radioactive material (NORM), and line broadening due to limited spectral resolution of nuclear detector. We have invented a method for detection of sources using clustering of spectral data. Our method takes advantage of the physical fact that a source not only produces counts in the region of its spectral emission, but also has the effect on the entire detector spectrum via Compton continuum. This allows characterizing the low $S/N$ spectrum without distinct isotopic lines using multiple data features. We have shown that noisy spectra with low $S/N$ can be grouped by overall spectral shape similarity using a data clustering technique called Dynamic Quantum Clustering (DQC). The spectra in the same cluster can then be averaged to enhance $S/N$ of the isotopic spectral line. This would allow for increased accuracy of isotopic identification and lower false alarm rate. Our method was validated in a proof-of-principle study using a data set of spectra measured in one-second intervals with sodium iodide detector. The data set consisted of over 7000 spectra obtained in urban background measurements, and approximately 70 measurements of $^{137}$Cs and $^{60}$Co sources. Using DQC analysis, we have observed that all spectra containing $^{137}$Cs and $^{60}$Co signal cluster away from the background.

## 1 Introduction

In a search scenario, nuclear background spectra are continuously measured in short acquisition intervals (usually one second) with a mobile detector-spectrometer (*i.e.* in a vehicle or carried in a backpack) [1–3]. In principle, sources can be detected and identified from the measured data by their unique spectral lines. Detecting sources from data measured in a search is difficult because of low signal-to-noise ratio ($S/N$) of spectra, large and highly varying background due to naturally occurring radioactive material (NORM), and line broadening due to the limited spectral resolution of radiation detectors.

Data analysis is typically performed in a post-processing mode, and consists of saliency and anomaly searches in a data set containing tens of thousands of one-second spectra. Visual inspection of data by a human analyst is not efficient because of the large volume of data. Existing computational tools utilize a total counts threshold-triggered alarm, and radiation isotope identification (RIID) algorithms. Using a threshold-based alarm results in a large number of false alarms due to large background fluctuations in an urban search, and may result in missed detection of a weak source with sub-threshold total counts. RIID algorithms applied to one-second spectra produce a large number of false alarms because of statistical background fluctuations. The number of false alarms can be decreased if one-second spectra are averaged over some time window to improve the $S/N$. However, the size of the window is not known *a priori*. Averaging over the wrong window would decrease the $S/N$ of a source by washing out spectral lines with random counts. In general, the spectra containing measurements of the same source might not be sequential in time.

[a] e-mail: aheifetz@anl.gov
[b] *Current address*: National Security Directorate, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA, 99352.

This can happen if the detector approaches the same sources at different times during the search, or if the source has been moved. Since there is no *a priori* information, one can average over randomly selected combinations of one-second spectra. Given the large size of the data, the solution of such a combinatorial problem would require enormous computational resources.

We have developed a method for the detection of sources in search applications using a data clustering technique called Dynamic Quantum Clustering (DQC) [4–7]. Our approach provides the scientific basis for grouping one-second spectra of the same source, irrespective of time stamps of these spectra. The physical idea behind this approach is that the presence of a source would not only impact a narrow spectral band corresponding to emission line, but also has the effect on the entire detector spectrum via the Compton continuum [8,9]. This allows characterizing spectra, including the ones without distinct isotopic lines, using multiple data features. DQC selects the spectra according to their overall shape similarity. The cluster or clusters possibly containing a source can be identified in the feature space by their separation from the rest of the data. Then spectra in the same cluster can be averaged to enhance the $S/N$ of the isotopic spectral line. This would allow for increased accuracy of isotopic identification and a lower false alarm rate.

## 2 Challenges in nuclear source detection and identification in search surveys

Detection of radioactive isotopes during a gamma-ray spectrometry screening campaign consists of performing both saliency and anomaly searches in large amounts of recorded spectral data. In a typical drive-through search scenario background or environmental gamma-ray spectra are continuously measured over short acquisition time intervals (*e.g.*, one second), with a fast response nuclear detector-spectrometer, such as thallium-doped sodium iodide (NaI(Tl)) scintillator. The gamma-ray spectra of energy range of up to 3 MeV are typically recorded by means of a 1024-channel multichannel analyzer, which can be set up to 512 or 256 channels. In principle, the presence of a particular isotope can be inferred from the measured gamma spectrum either by an experienced spectroscopist, or by using a digital computer running an automated radioisotope identification (RIID) algorithm. The challenges of nuclear spectroscopy in a search scenario, which set it apart from the conventional spectroscopy, consist of (1) the low signal-to-noise ($S/R$) ratio of gamma-ray spectral lines; (2) large and highly varying background; and (3) line broadening due to the limited spectral resolution of the nuclear detector.

The strength of the signal in nuclear detection depends on a number of factors, including i) the activity of the source (*e.g.* 26 MBq–560 GBq point source $^{60}$Co, $^{137}$Cs, $^{192}$Ir, $^{99m}$Tc were tested for identification search by car-borne and air-borne gamma-ray spectrometry in [10]; ii) the source-to-detector distance (it is known that the gamma photon flux from an isotropic point source decreases as the square of the distance); iii) the presence of shielding material (*e.g.* for 600 keV photons the mean free path in air is approximately 100 m); iv) detector size and efficiency (*e.g.* 16 liter of modular NaI(Tl) detector was used for airborne gamma-ray spectrometry measurements in [11], and the theoretically optimal choice for detector size is described in [12]). In the case of a low number of counts, $S/N$ can be improved by increasing the integration time. However, in a wide area search scenario, the spectral data is typically acquired by a moving detector in a sequence of one-second measurement intervals. The choice of acquisition interval depends on the tradeoff between the spatial resolution and signal-to-noise ratio ($S/R$) of the recorded spectrum (*e.g.*, see the discussion in [10]). Because of the probabilistic nature of the source to detector distance and the short acquisition time during screening, most measured spectra have low signal-to-noise ($S/R$) ratios. The signals of interest are weak even for sources without shielding, and might not exhibit distinctive isotopic spectral lines. Within the framework of a particular radioactive background, nuclear detection is associated with measurement uncertainties. For example, in each energy bin in the spectrum, gamma count as a function of time has a Poisson distribution, or degrades to a Binomial distribution for very small sample rates. Because there is limited measurement time available in a search scenario, there will be a large temporal variance in the amplitude of the detected signal spectrum. Thus the potential increase in the count rate for a particular energy bin —due to the presence of special nuclear material (SNM) or other radioactive materials— can be smaller than the variance of the background count in that bin. Current approaches to this problem include, for example, using a noise adjusted singular value decomposition analysis method for orphan sources identification [13].

Another complication for nuclear search stems from the presence of a fairly strong and highly varying nuclear background due to emission by naturally occurring radioactive materials (NORM). The radioactive background changes randomly due to a number of natural and man-made causes. These include but are not limited to the presence of building materials (functioning both as source and shielding), variation in concentration of NORM isotopes in the soil, increasing flux of the cosmic background radiation with increasing elevation above the sea level, and seasonal and nocturnal variations. The nuclear background spectrum in the energy bandwidth from 0 to 3 MeV and measured with a NaI(Tl) detector is dominated by gamma emission of primordial radionuclide $^{40}$K, and $^{238}$U, $^{232}$Th decay chain progenies. While these primordial isotopes are found in virtually all surroundings on Earth, urban building materials typically contain high concentrations of NORM. This frequently causes large random fluctuations in the background count in an urban search. In addition to random background variations and noise, isotopic detection is

further complicated by potential shielding. If a shielded SNM source is present at the scene, unknown shielding would result in unpredictable attenuation and distortion of the isotopic spectra. Thus, the anomaly to search for in the data is not well defined ahead of time. Current approaches to solving this problem include, for example, applying filtering of a "rolling" average background spectra to improve the signal-to-background ratio [14].

Additional challenges in nuclear spectrum analysis come from line broadening due to the limited spectral resolution of nuclear detectors. NaI(Tl) is one of the most commonly used gamma-ray detectors because of a high response speed, relatively low cost, and technological ability to produce large-volume NaI(Tl) crystals. However, a NaI(Tl) detector has a spectral resolution of approximately 6–7% at 662 keV ($\sim 46$ keV FWHM). Thus, spectra lines that are theoretically expected to be narrow appear as broad Gaussian-shaped peaks when measured with NaI(Tl) detector. This is because of the statistical fluctuations in the number of free charge carriers excited per monoenergetic interaction event in the scintillator crystal. Moreover, the natural background gives rise to a triplet near 662 keV ($^{137}$Cs) with 583 keV ($^{208}$Tl) and 609 keV ($^{214}$Bi), which makes it even more difficult to resolve the individual isotopic lines. In addition, non-zero counts, with a particularly high number contribution in the low-energy range, are registered below spectral peaks. These counts are referred to as the Compton continuum. They appear due to scattering events taking place in the scintillator crystal where only partial energy deposition occurs because the scattered gamma-ray escapes the detector crystal. Current approaches to mitigating this problem include, for example, using full spectrum analysis to resolve the contribution from $^{137}$Cs in environmental spectra [15].

## 3 Current approaches to nuclear search data analysis

A screening campaign usually lasts for several days, with 8–12 hour shifts. During each day, there are approximately 30000 one-second spectra per detector platform to analyze. Current practices employed by radiological response teams consist of monitoring the count rate levels for large anomalies in real time (*i.e.* threshold alarms) and analyzing nuclear background data collected during the work day in off-line mode at the end of the shift. The usual approach to data analysis consists of examining time series of spectrum-integrated total counts. If during visual inspection of time series the analyst discovers what he or she believes to be an anomaly, spectral content of the suspicious data is subjected to further examination. Most current nuclear detection systems are coupled with GPS receivers. If an anomaly is found in the spectral data, then one would return to the corresponding geographical location to make additional measurements. Although real-time detection of sources is the ultimate goal, with the current state of technology, data analysis in a post-processing mode results in more thorough review of the data and higher accuracy of detection.

### 3.1 Computational methods using total counts threshold alarm

Because there is a high probability of human error in analyzing a large volume of nuclear background data, utilizing computational tools has the potential of increasing the accuracy of detection. In addition, using a computer-based method brings the possibility of extending background data analysis to real-time mode. Most commercial systems use an algorithm which produces an alarm if the total counts, or the counts within a spectral region of interest, exceed some threshold value. If an alarm is triggered, spectral content of the data is examined [16].

The problem with using an alarm trigger —based on some threshold value of total counts— is that it results in large number of false alarms and missed detections. This is because abrupt changes in the measured background rate, such as when a mobile platform detector moves between buildings, might accidentally trigger a false alarm. On the other hand, a source that does not produce enough counts above the threshold would not trigger an alarm.

To illustrate the problems that occur, let us consider a specific example. Figure 1 shows the gamma exposure rate along the trajectory of a drive through a portion of downtown Chicago with an Exploranium GR-460 commercial NaI(Tl) crystal-based 512-channel detector. The detector was a $2 \times 4 \times 16$ inch NaI(Tl) crystal —128 cubic inches or 2.1 liter background spectra were measured every second for a total of approximately 5000 spectra. The plot shows that the gamma exposure rate due to background varies significantly depending on location. (It also shows an intermittent GPS signal due to urban canyons.) The time dependence of total gamma counts per second (CPS) for the example in fig. 1 is shown in fig. 2. The count rate varies widely from a low of about 1000 cps in the park areas with minimal structures to as high as 5000 cps near many of the buildings. These significant changes are all due to background NORM materials.

### 3.2 Computational methods using RIID algorithms

The operational principle of most radiation isotope identification (RIID) algorithms consists of matching either spectral peaks or entire measured spectra to known templates in a spectral library of isotopes, which may also include attenuated and distorted spectra of shielded sources.

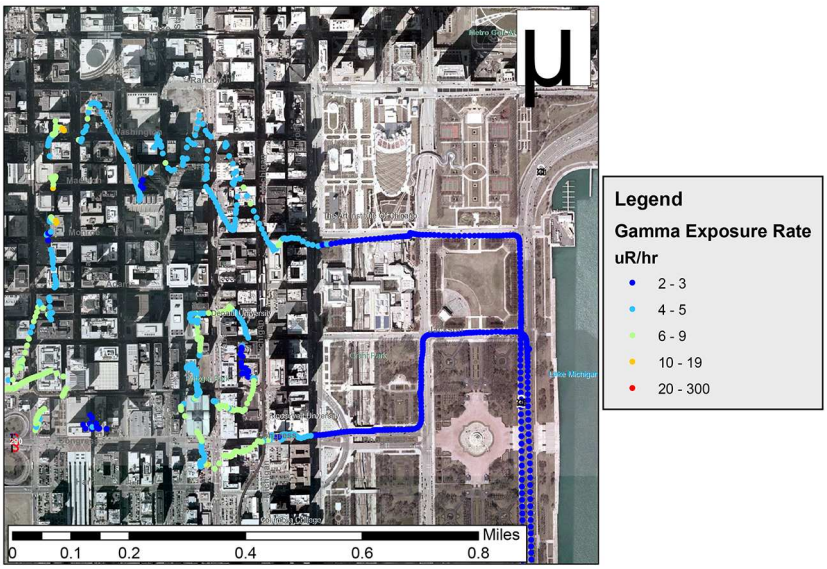


**Fig. 1.** Gamma exposure rate ($\mu$R/hr) in a section of Chicago. Also shown is the intermittent GPS signal due to urban canyons.

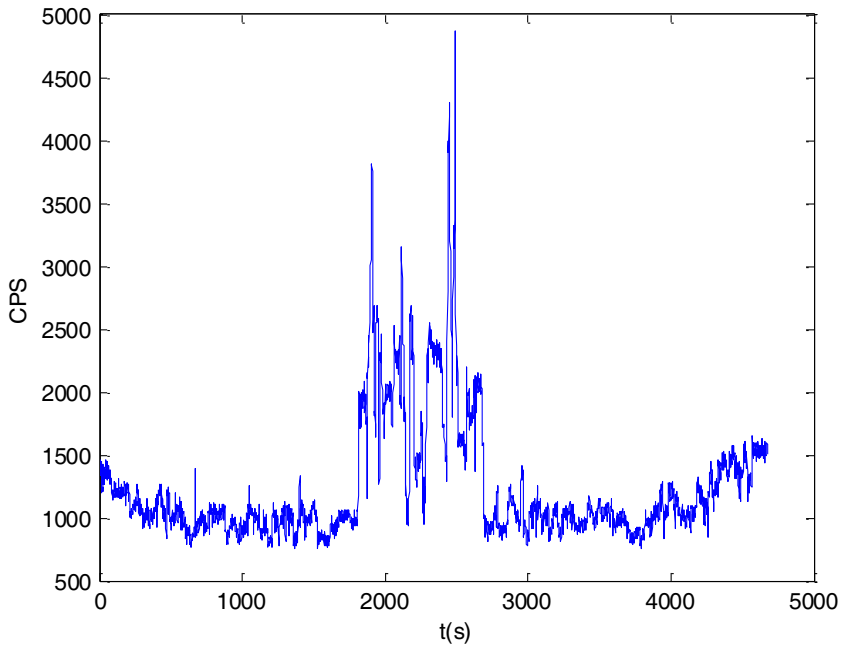


**Fig. 2.** Time dependence of total gamma counts per second (CPS) of the example in fig. 1. Significant fluctuations of CPS are all due to variability in background NORM materials.

Unless the source registers enough counts to display a clear spectral line, RIID algorithms fail. Spectra measured during one-second time intervals are, in general, very noisy, and do not exhibit clear spectral lines. Thus, any RIID algorithm applied to noisy and largely featureless one-second spectra would yield inconclusive results. In our prior studies, applying the multiple regression method of Gamma Detector Response and Analysis Software (GADRAS) [17], Maximum likelihood and fuzzy logic RIID algorithms [1] to one-second spectra produces a large number of false alarms and missed detections. The cause for large numbers of false alarms was statistical background fluctuations, which produced random peaks in spectral channels corresponding to emission lines of isotopes of interest. As an illustration, in fig. 3 we plot four one-second spectra from the nuclear background data set measurements in downtown Chicago shown in figs. 1 and 2. The four spectra correspond to four tallest spikes in the time series in fig. 2. The spectra are noisy with poorly visible spectral lines. One can recognize $^{40}$K line at 1460 keV, but the rest of the spectral lines are blurry.

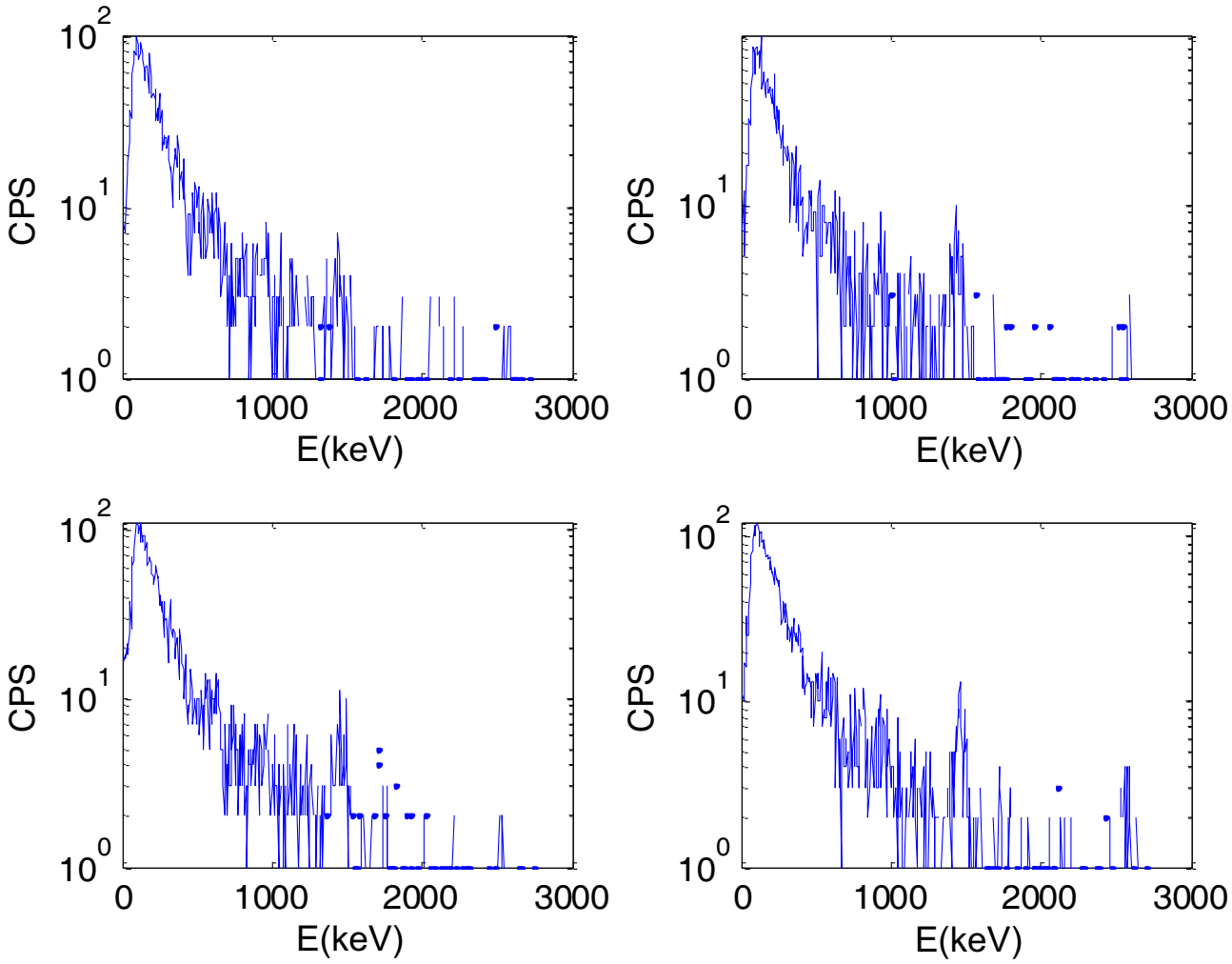


**Fig. 3.** Four one-second spectra of the nuclear background measurements corresponding to the four tallest spikes in the time series in fig. 2.

RIID algorithm performance can be improved if spectral data $S/N$ is increased by averaging over multiple time frames containing one-second spectra. As an illustration, in fig. 4 we plot search-time averaged number of gamma counts $\langle N_\gamma \rangle_T$ as function of energy $E$ for the data in figs. 1 and 2. The result is a typical background spectrum measured with NaI(Tl). Energy windows used to measure gamma rays from the decay of $^{40}$K, and $^{238}$U, and $^{232}$Th decay chain progenies are shown by highlighting. The plot in fig. 4 shows clear isotopic spectral lines. The fundamental problem with this approach is that the number of time frames to use in averaging is not known. Averaging the spectra over the entire search time will increase $S/N$ of the spectral lines of the most common NORM isotopes encountered during the search. On the other hand, a source is not commonly observed repeatedly during the entire search. Thus, source spectral lines will be washed out by averaging with a large number of time frames containing random background fluctuations in the same spectral bands.

One possibility to address the spectrum averaging problem is to use a moving window average of different lengths. This approach is, generally, incorrect because there is a possibility that the same source could be measured by the detector in frames that are not sequential in time. This can happen if during search the detector platform approaches the same source from different directions, or if the source has been moving during the screening campaign. There is no guidance on the limits of the averaging window. If one is averaging over small number of randomly selected time frames, statistical fluctuations of the background may randomly produce what appears to be a resemblance of a source spectral line. On the other hand, averaging over large number of one-second frames can lead to washing out of source spectral lines. Since there is no *a priori* information, one can average over randomly selected combinations of one-second spectra. Given the large size of the data, the solution of such a combinatorial problem would require enormous computational resources.

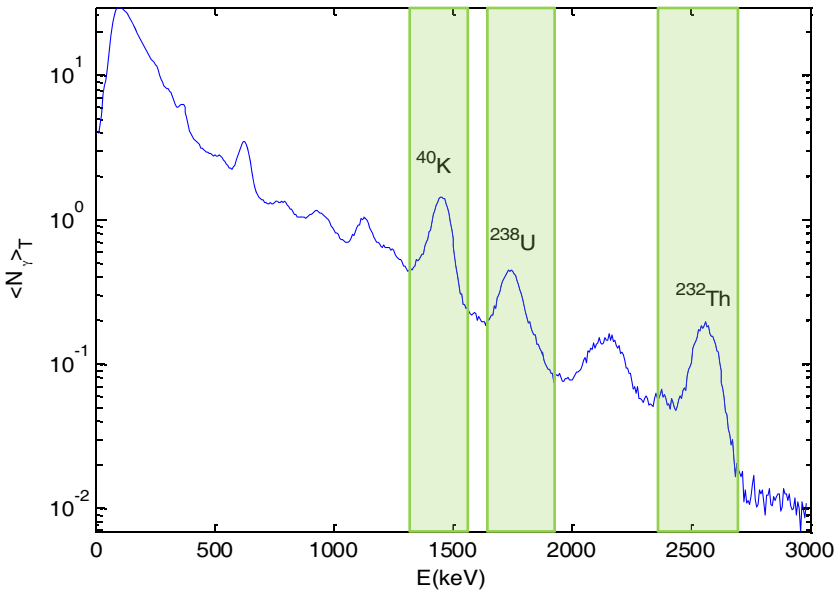


**Fig. 4.** Search-time averaged number of gamma counts $\langle N_\gamma \rangle_T$ as a function of energy $E$ for the data in figs. 1 and 2. The result is a typical background spectrum measured with NaI(Tl). Energy windows used to measure gamma rays from the decay of $^{40}$K, $^{238}$U, and $^{232}$Th decay chain progenies are shown by highlighting.

## 4 Data analysis using dynamic quantum clustering

Our approach provides the scientific basis for selecting time frames containing one-second spectra that can be subsequently averaged without losing information about the source. The physical idea behind this approach is that the presence of a source would not only impact a narrow spectral band by corresponding to emission spectral line, but would also affect the entire detector spectrum via the Compton continuum and detector response function (DRF). This allows characterizing the spectrum using multiple data features, as opposed to the single feature of a spectral line. Implementation of our method is based on the application of a data clustering technique called Dynamic Quantum Clustering (DQC) [4–7]. The spectra that end up in the same cluster —as determined by their overall shape— can be averaged to enhance $S/N$ of the isotopic spectral line. This would allow for increased accuracy of isotopic identification.

DQC uses ideas borrowed from quantum mechanics to solve the problem of clustering data —or equivalently, revealing hidden correlations among the many features being measured. The method works by creating a quantum potential that serves as a faithful proxy of the density of the data in feature space. It then uses the mathematics of quantum evolution in Hilbert space to reveal the correlated subsets of the data as simple (point-like) clusters, or as extended, parameterizable shapes. The outcome of a DQC analysis is a movie that shows how and why sets of data points are eventually classified as members of simple clusters or as members of extended structures. Prior studies have shown that DQC has much higher sensitivity-to-density variation in the data compared to other clustering methods, such as the Parzen window estimator, the support vector machine, or K-means clustering [4,6].

The validity of the proposed approach was demonstrated with a proof-of-concept numerical experiment based on the example case listed above. In this study, DQC was applied to a database of 7675 one-second spectra collected with NaI(Tl) detector with 512 spectral channels in a drive-through campaign in the Chicago area. A subset of this data was displayed in figs. 1 and 2. The detection system uses multiple-peak gain stabilization that exploits the naturally occurring isotopes of U, K and Th. Automatic spectral stabilization is achieved using computer algorithms and the spectral signatures of these isotopes. The technique provides for a fast stabilization at start-up as well as maintaining stabilization during operation. This continual stabilization allows for us to maintain the energy calibration so that spectra can continue to be compared over disparate time intervals.

Approximately 70 one-second spectra in the total dataset consisted of measurements with a $^{137}$Cs and $^{60}$Co radioactive sources with various levels of $S/N$. $^{137}$Cs isotope has a characteristic line at 662 keV, while $^{60}$Co has two characteristic lines at 1173 keV and 1332 keV. Search time-averaged number of gamma counts $\langle N_\gamma \rangle_T$ for all 7675 spectral measurements in the data set is plotted in fig. 5. Note that the isotopic spectrum is essentially the same as that in fig. 4. There are no visible $^{137}$Cs or $^{60}$Co lines. This is highlighted by drawing vertical lines in fig. 5 to indicate spectral locations of where the centers of $^{137}$Cs and $^{60}$Co lines are expected to be. In addition, there are spectral lines of NORM $^{214}$Bi centered at 609 keV, and $^{208}$Tl centered at 583 keV, which can potentially obscure the $^{137}$Cs line.

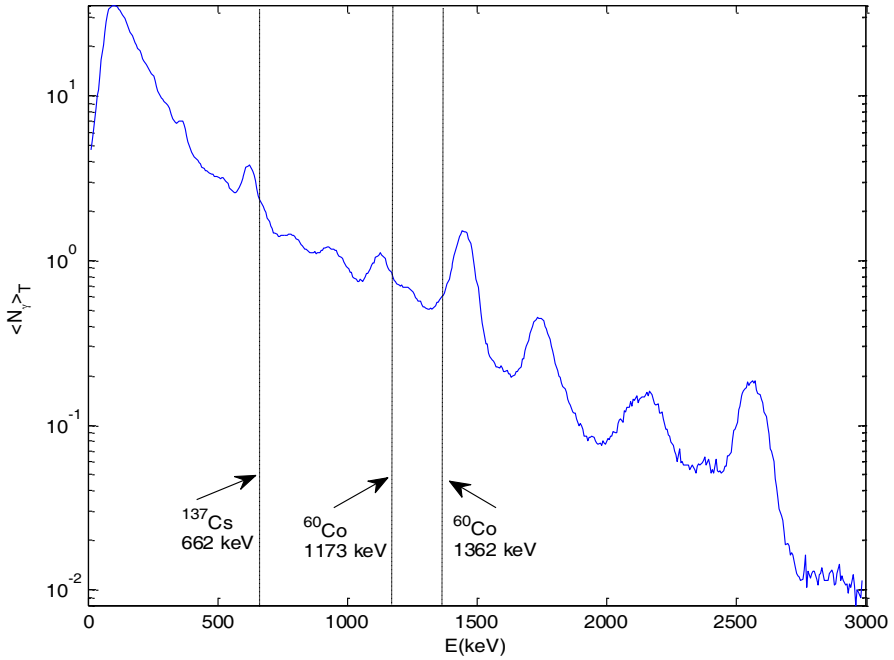


**Fig. 5.** Time-averaged number of gamma counts $\langle N_\gamma \rangle_T$ for 7675 one-second spectra consisting of background measurements and approximately 70 one-second spectra of $^{137}$Cs and $^{60}$Co source measurements. Vertical lines indicate where the centers of $^{137}$Cs and $^{60}$Co spectral lines are expected to be.

In this study, we have 7675 one-second spectra, each with 512 channels, so the data matrix $M$ is $512 \times 7675$. It turned out that five detector channels had zero counts at all times, and indeed a computer check revealed that $rank(M) = 507$. We performed a normalization procedure so that the largest value in each of 7675 spectra was scaled to one. This normalization procedure ensures that clustering would not be sensitive to fluctuation in total counts. As described in our previous publications, data dimensionality reduction speeds up clustering computations. Utilizing singular-value decomposition (SVD), we obtain $M = U\Sigma V^T$, where $U$ is a $512 \times 512$ matrix containing the orthonormal basis for the column space of $M$. The columns of $U$ are also known as the principle components (PC's). We take the first $d$ PC's to form a $d$-dimensional subspace of the 512-dimensional space. The criteria for selecting the value of $d$ is that the reconstructed matrix obtained from the first $d$ singular values of $\Sigma$, preserves the information contained in the original matrix $M$. In this study, we used the criteria that reconstructed matrix should retain spectral signatures of $^{137}$Cs and $^{60}$Co. We determined that setting $d = 120$ preserved the desired information in reconstruction. Setting $d$ to lower values, *e.g.*, $d = 110$, erased any traces of $^{137}$Cs and $^{60}$Co in the reconstructed matrix. To proceed with DQC analysis, we associate with each of $n = 7675$ data points $\vec{x}_i$ in a Euclidean space of $d = 120$ dimensions a Gaussian wave function $\psi_i(\vec{x}) = \exp(-(\vec{x} - \vec{x}_i)^2/2\sigma^2)$. In this study, we set $\sigma = 0.34$. Using the DQC formalism described in our previous publications, we study the temporal behavior of the expectation values, or centers of Gaussians, $\langle \vec{x}_i(t) \rangle$ for all $i$.

We have observed that two point-like clusters, consisting of 24 and 44 points, immediately separated from the rest of clusters in the feature space. These stand-out clusters represent anomalies in the data. Upon inspection of the two anomaly clusters, we have observed that both clusters consist of spectra containing the $^{137}$Cs and $^{60}$Co signals and no background-only spectra. Also, we have confirmed that all spectra in the original set which could contain $^{137}$Cs or $^{60}$Co or both signals have clustered into either one of the anomaly clusters. Figures 6(a)–(c) illustrate the results of clustering in the space of first three SVD components. The anomaly clusters are circled red dots, while the rest of the data is displayed as aqua-marine dots. Frame captures in figs. 6(a), (b) and (c) show that as the field of view is rotated, anomaly clusters 1 and 2 are moving relative to the background points. One anomaly cluster contains 24 spectra and has a strong Cs and Co peaks in every spectrum. The second anomaly cluster contains 44 spectra and shows modest Cs peaks, or arguably Cs peaks that are not visible. Some of the spectra in the second cluster also show modest $^{60}$Co peaks. One-second spectra in clusters 1 and 2 are shown in figs. 7 and 8, respectively. Each subplot has log-linear axes, where the $x$-axis is the energy $E$ in units of keV, and the $y$-axis is gamma counts per second. The energy range was limited to 2000 keV because the counts above that threshold value were negligible. Central locations of $^{137}$Cs and $^{60}$Co spectral peaks are marked by vertical lines.

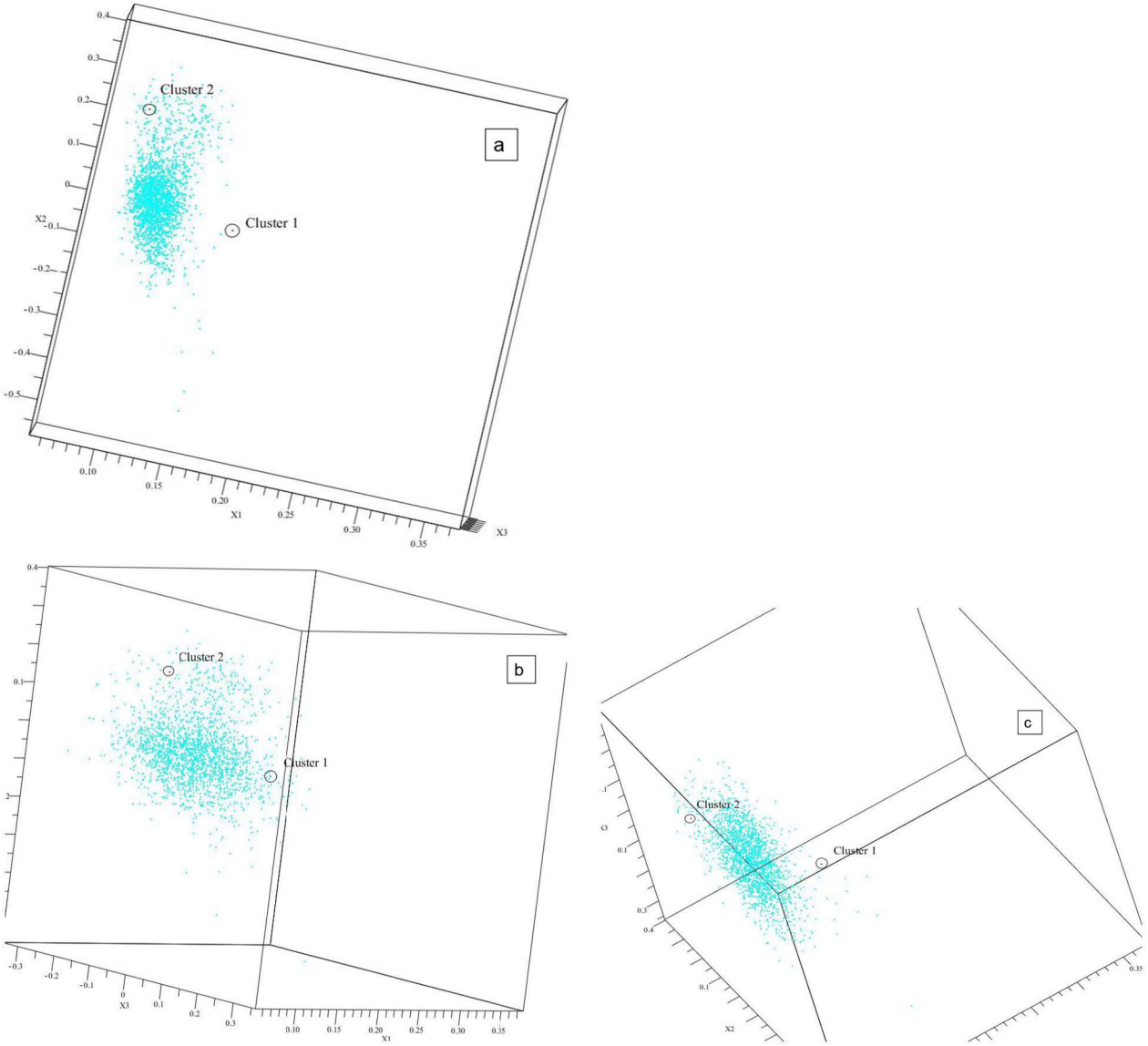


**Fig. 6.** Separation of two clusters containing $^{137}$Cs and $^{60}$Co spectra (circled red dots) from background data (aqua-marine dots) in the feature space of the first three SVD components. Frame captures in (a), (b) and (c) show that as the field of view is rotated, anomaly clusters 1 and 2 are moving relative to the background points.

Seven spectra in the second cluster each have total gamma counts that are less than 4000. These spectra might not be registered by threshold-triggered alarm in urban search, where the total count of the background fluctuates between 1000 and 5000 total gamma counts per second (according to fig. 2). In fig. 8, the spectra with total counts less than 4000 are enclosed by red boxes. It should also be noted that some of one-second spectra in the second cluster with total counts exceeding 4000 do not display clear isotopic lines. For example, spectrum #39 in fig. 8 (9th row, 3rd column) has over 6000 total counts but no clear isotopic spectral lines.

As hypothesized previously, the data in anomaly clusters can be integrated in time to improve source(s) $S/N$ without losing spectral information in the averaging process. Figure 9 shows log-linear plots of cluster-averaged spectra in the first and second clusters. The average spectrum of the first anomaly cluster in fig. 9(a) shows a very strong $^{137}$Cs peak at 662 keV, and strong $^{60}$Co peaks at 1173 keV and 1332 keV. The average spectrum of the second anomaly cluster in fig. 9(b) shows a good $^{137}$Cs peak and weak $^{60}$Co peaks.

In the present study, one-second spectra of $^{137}$Cs and $^{60}$Co sources were measured in a time sequence. Since the DQC analysis uses no information about the time at which a spectrum is taken, it will be insensitive to temporal correlations among cluster members. Thus the same results would be obtained if the one-second spectra with $^{137}$Cs and $^{60}$Co sources were randomly distributed among the entire database of 7675 spectra. This feature of DQC allows using the method for detecting and tracking sources in real time.

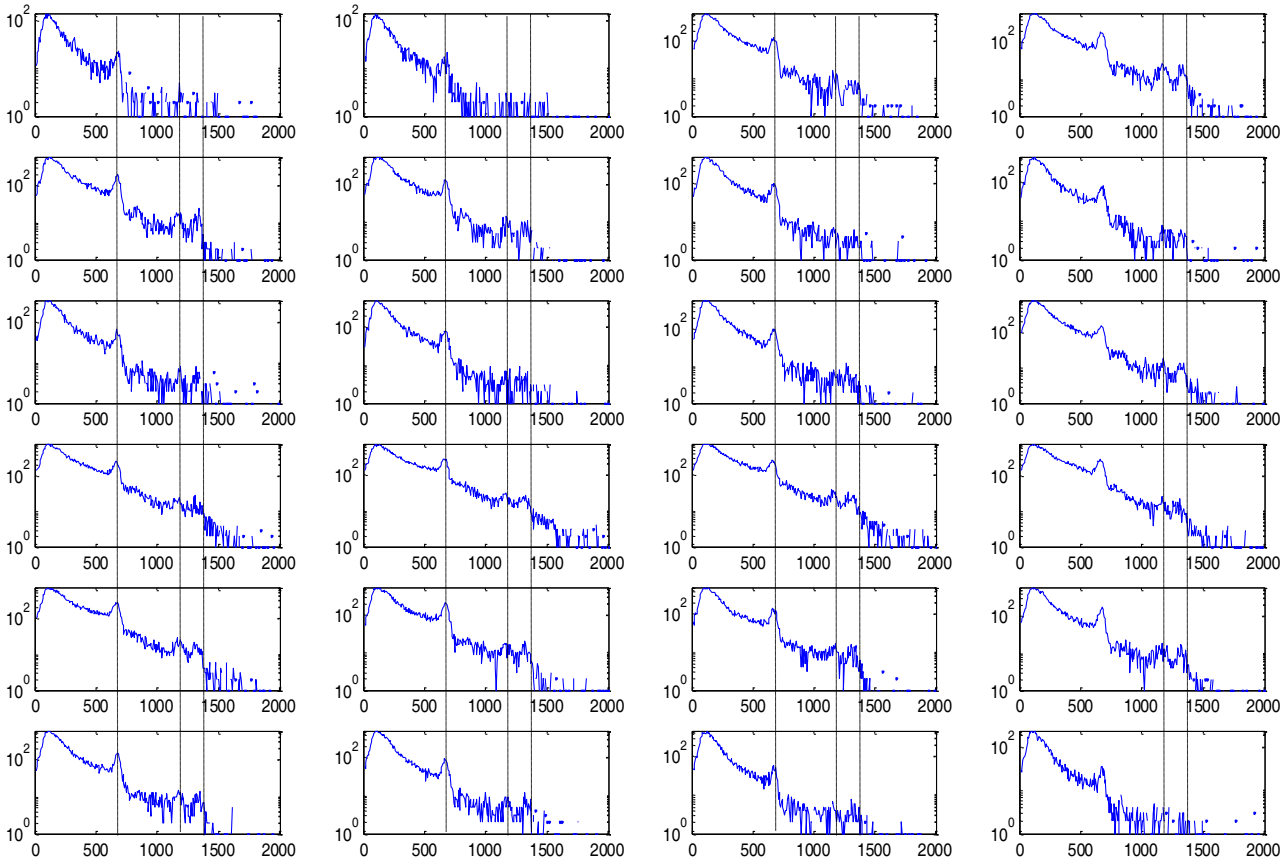

**Fig. 7.** Log-linear plots of 24 one-second spectra displaying strong $^{137}$Cs peaks at 662 keV and good $^{60}$Co spectra at 1173 keV and 1332 keV in cluster 1. Centers of the $^{137}$Cs and $^{60}$Co peaks are marked by vertical lines.

## 5 Conclusion

We have designed a method for detecting nuclear sources in search applications using clustering of spectral data. Our method allows for increased accuracy of isotopic identification and lower false alarm rate. The validity of our approach has been demonstrated in a proof-of-principle study using the Dynamic Quantum Clustering (DQC) method. The validation study consisted of a data set of spectra measured in one-second intervals with Sodium Iodide detector. The data set of over 7000 spectra consisted mostly of urban background measurements, and approximately 70 measurements of $^{137}$Cs and $^{60}$Co sources. Using DQC analysis, we observed that all spectra containing the $^{137}$Cs and $^{60}$Co signal clustered away from the background, producing two point-like clusters. One cluster contained spectra with strong Cs and Co peaks in every spectrum. The second cluster contained spectra with modest Cs and Co peaks, or arguably Cs and Co peaks that are not visible. Averages of spectra in either cluster reveal clear $^{137}$Cs and $^{60}$Co spectral lines.

Development of a robust DQC-based technique for detection of sources will require further algorithm optimization studies and large-scale validation and verification experiments. It is necessary to investigate the relationship between sensitivity of DQC feature identification in the spectrum to the detector response function (DRF) of NaI(Tl). This would provide a better understanding of the limits of the source detection capability of a DQC-based technique. It is expected that DQC performance will depend on such factors as source spectrum and signal strength, background isotopic composition and variability with time. Such studies will also provide an indication if the detector spectral channels can be ranked in order of importance for DQC analysis. Channels with least importance could be excluded from data to reduce its size, and hence increase the speed of analysis. In the present study, we did not take advantage of data sparsity. In fact, 57% of entries are zeros in the $512 \times 7675$ data matrix $M$. Data sparsity results primarily from the fact that detector channels corresponding to energies above 2000 keV contain virtually no counts. In the preliminary study, DQC analysis was performed in 120-dimensional feature space. We plan to study the dependence of DQC's performance on the dimension of the feature space for different sources and backgrounds. Also, we plan to investigate DQC performance for different values of the width of the Gaussian wave function, $\sigma$. In the present study, we considered clustering of spectral data taken at different times. In the future work, we will investigate clustering of temporal data in the energy domain, and compare the information content of either approach.

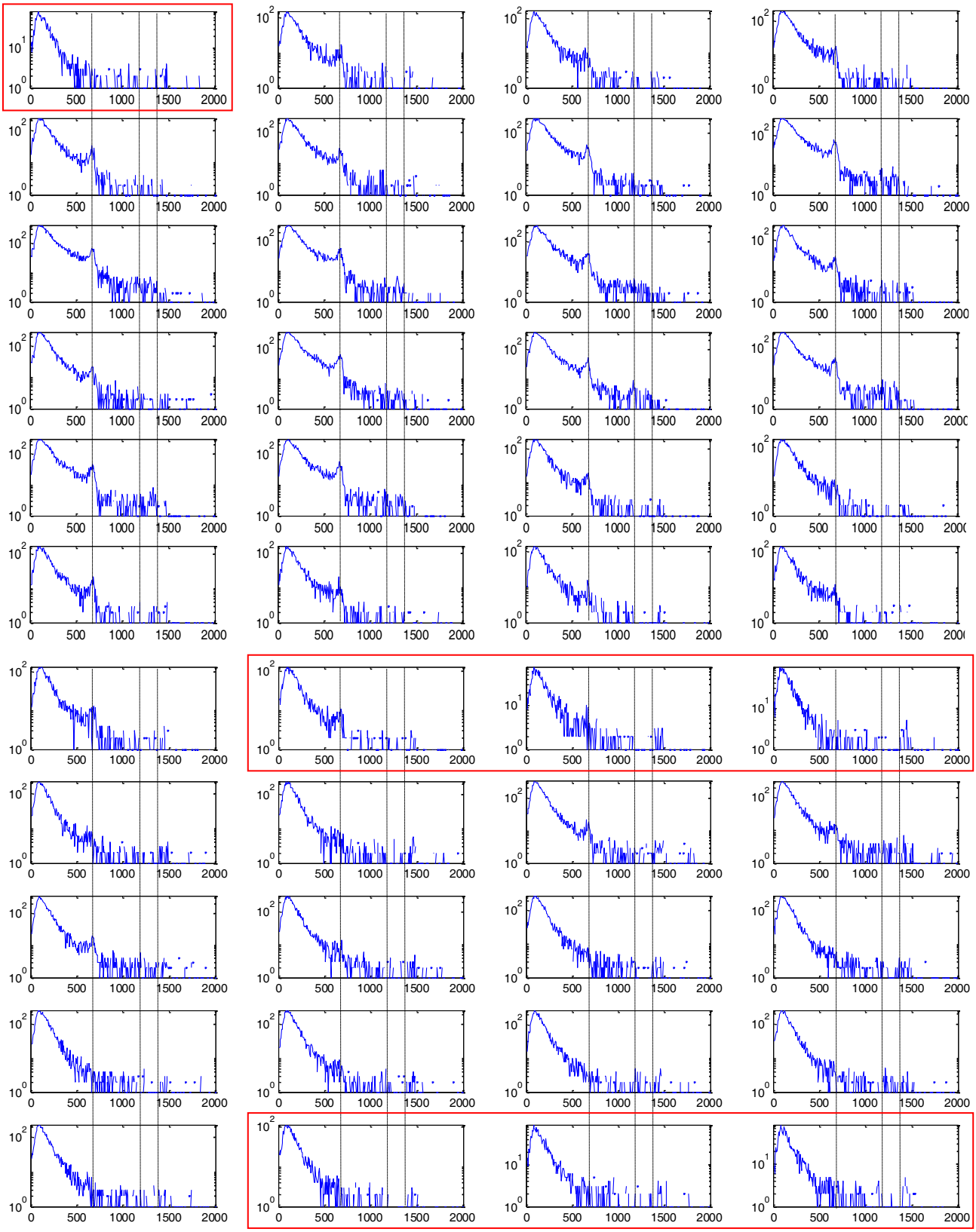

**Fig. 8.** Log-linear plots of 44 one-second spectra containing modest to indistinct $^{137}$Cs peaks at 662 keV and $^{60}$Co spectra at 1173 keV and 1332 keV in cluster 2. Centers of the isotopic peaks of $^{137}$Cs and $^{60}$Co are marked by the vertical lines. Spectra with total counts less than 4000 are enclosed by red boxes.

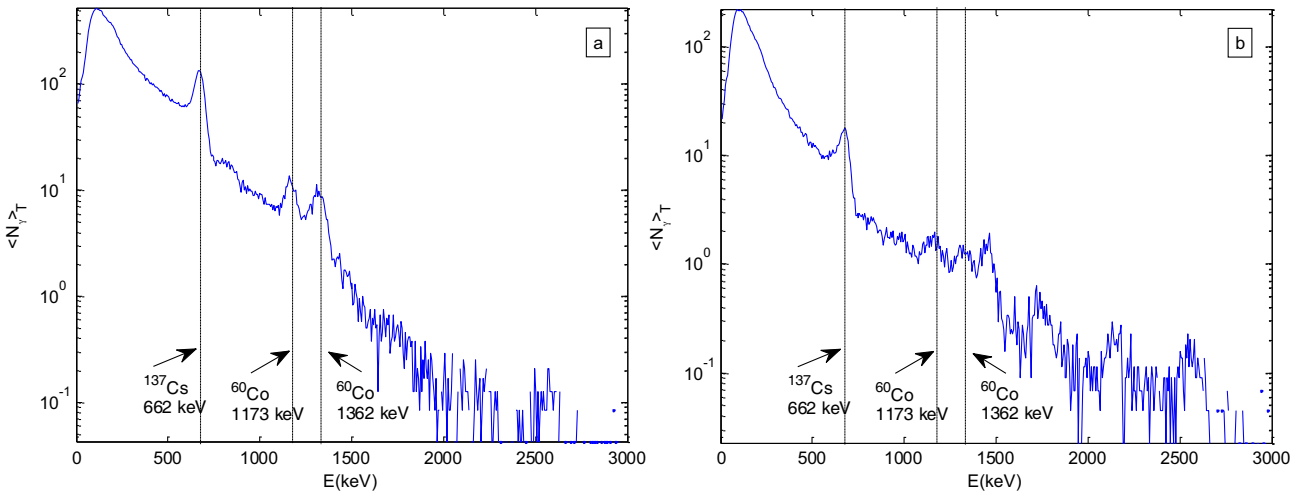


**Fig. 9.** (a) Cluster-averaged number of gamma counts $\langle N_\gamma \rangle_T$ of 24 spectra in cluster 1. (b) Cluster-averaged number of gamma counts $\langle N_\gamma \rangle_T$ of 44 spectra in cluster 2. In either (a) or (b), there is a clearly identifiable $^{137}$Cs peak at 662 keV. The spectrum in (a) shows clear $^{60}$Co peaks at 1173 keV and 1332 keV, while the spectrum in (b) shows weak $^{60}$Co peaks.

We will also investigate the possibilities of expediting DQC performance. For example, it has been shown recently that DQC can be parallelized and efficiently executed using graphics processors [8]. We will investigate if faster algorithm execution can be achieved by clustering of spectral data in time or clustering of temporal data in the energy domain by introducing appropriate simplifications in either approach. DQC's performance on spectral data collected with other medium resolution detectors used in search applications will be considered as well. Also, we will investigate the possibilities for developing a technique using DQC for real-time source detection and tracking combined with an estimation approach, *i.e.* building on the data set as it is being collected.

This work was supported in part by the National Science Foundation Cyber-Physical Systems grant CNS-1329657. The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.

## References

1. M. Alamaniotis, A. Heifetz, A.C. Raptis, L.H. Tsoukalas, IEEE Trans. Nucl. Sci. **60**, 3014 (2013).
2. E. Bai, A. Heifetz, S. Dasgupta, R. Mudumbai, submitted to IEEE Trans. Nucl. Sci. (2014).
3. T. Burr, M. Hamada, Algorithms **2**, 339 (2009).
4. M. Weinstein, D. Horn, Phys. Rev. E **80**, 066117 (2009).
5. M. Weinstein, F. Meirer, A. Hume, Ph. Sciau, G. Shaked, R. Hofstetter, E. Persi, A. Mehta, D. Horn, arXiv:1310.2700v2 (2013).
6. D. Horn, A. Gottlieb, Phys. Rev. Lett. **88**, 018702 (2001).
7. P. Wittek, J. Comput. Phys. **233**, 262 (2013).
8. K.S. Krane, *Introductory Nuclear Physics*, 3rd edition (Wiley, New York, 1987).
9. G.F. Knoll, *Radiation Detection and Measurement*, 4th edition (Wiley, New York, 2010).
10. J. Hovgaard (Editor), RESUME 95: Rapid Environmental Surveying Using Mobile Equipment, Nordic Nuclear Safety Research (NKS) Copenhagen, Denmark (1997).
11. E. Guastaldi *et al.*, Remote Sens. Environ. **137**, 1 (2013).
12. K.-P. Ziock, K.P. Nelson, Nucl. Instrum. Methods A **579**, 357 (2007).
13. H.K. Aage, U. Korsbech, Appl. Radiat. Isot. **58**, 103 (2003).
14. A.J. Cresswell, D.C.W. Sanderson, Nucl. Instrum. Methods A **607**, 685 (2009).
15. A. Caciolli *et al.*, Sci. Total Environ. **414**, 639 (2012).
16. K.D. Jarman, R.C. Runkle, K.A. Anderson, D.M. Pfund, Appl. Radiat. Isot. **66**, 362 (2008).
17. D.J. Mitchell, H.M. Sanger, K.W. Marlow, Nucl. Instrum. Methods Phys. Res. A **276**, 547 (1989).