Back to BDC Data Competition (/)

Hello, **cbritt2**

MUSE Data Competition

Overview

(/competition/1/)

Data and Details

(/competition/1/download)

Submit Results

(/competition/1/upload)

Leaderboard

(/competition/1/leaderboard)

Team Dashboard

(/competition/1/team/9/)

Files for Download

The runs are separated into two sets: a training set for which an answers file with the correct answers for each run (source type, time at which the detector was closest to the source) is also provided, and a test set for which you will populate and submit an answers file for online scoring.

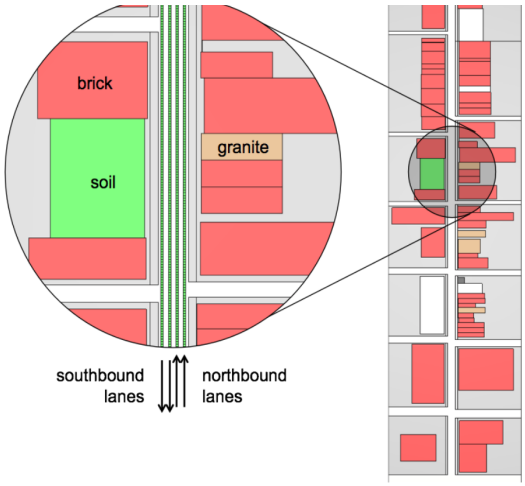
- testData.zip (/competition/1/file/1) - Test set — 15,923 runs (6.8 GB)
- trainingData.zip (/competition/1/file/2) - Training set — 9,800 runs (4.4 GB)
- answers.csv (/competition/1/file/3) - Training set answer file (149 kB)
- answerTemplate.csv (/competition/1/file/4) - Blank test set answer file (172 kB)
- SourceInfo.zip (/competition/1/file/5) - Templates for sources 1-5 (266 kB)

Data Description

For the competition we have generated data from thousands of runs that mimic what would be acquired by a 2”×4”×16” NaI(Tl) detector moving down a simplified street in a mid-sized U.S. city. Each run has been designed so that no source is present within the first 30 seconds of measurements.

In a real urban search scenario, you would be able to see the general layout of the street as the detector moves along. A notional diagram is shown at the right, though the exact street geometry for each run will be different (and unspecified). Buildings made of different materials (red, beige, and dark gray) occupy most of the space along the

covered park areas (green). The main street (also white) has four travel lanes, two in each direction, and several cross streets. Travel lanes are each ten feet wide and there is at least 10 feet of sidewalk (gray) between the outer travel lane and the building facades. Detector paths are all within one of the four main traffic lanes – with no lane switching and no turning onto the cross streets.



The set of data for a given run mimics what a detector would see in real-time: a list of pairs, each comprising the time at which a gamma ray interacted within the radiation detector and the corresponding energy the gamma ray deposited in the detector. Times are in microseconds (μsec, millionths of a second) since the last event, and the deposited energy is in units of kilo-electronvolts (keV). The time recorded for the first event in each run will be zero. Algorithms can bin the data into any time and energy structure you choose.

For each run, several characteristics are fixed a priori:

- The lane and direction of travel.
- The speed of travel, a constant between 1 and 13.4 m/s.
- The street geometry (e.g., what buildings are present where).

If a source is present, the following characteristics are also fixed:

- The type of source (see below for the list of types and their SourceID labels).
- The strength of the source.
- Whether the source is shielded.
- The location of the source along the street.

Although the source characteristics will not be revealed in the test data set, they do impact the data recorded in each file. For instance, fewer source-induced events will be recorded if the detector is moving at a faster speed, and more events will be recorded if the source has a higher strength.

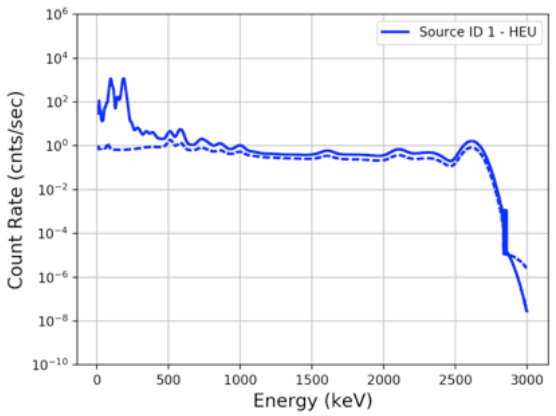
Source Types

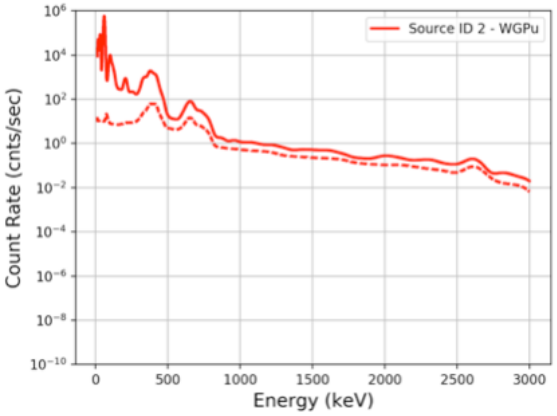
If a source is present in a run, it will be one of six types:

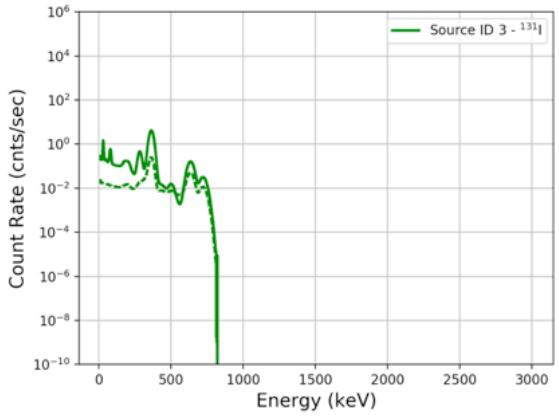
SourceID	Source Type
1	HEU: Highly enriched uranium
2	WGPu: Weapons grade plutonium
3	^{131}I: Iodine, a medical isotope
4	^{60}Co: Cobalt, an industrial isotope
5	^{99m}Tc: Technetium, a medical isotope
6	A combination of ^{99m}Tc and HEU

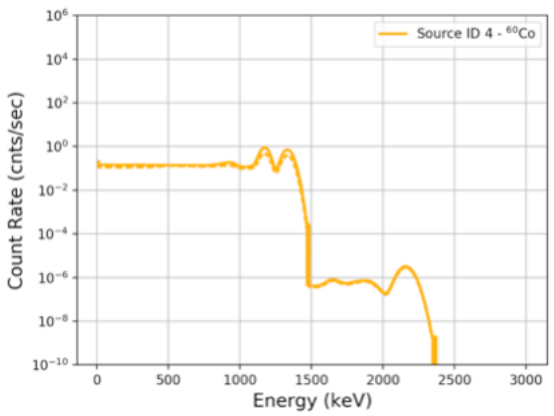
If no source is present, the SourceID field should be entered as 0. For all runs, there will be no source within the first 30 seconds of measurements. That is, you're guaranteed at least 30 seconds of background-only measurements in each run.

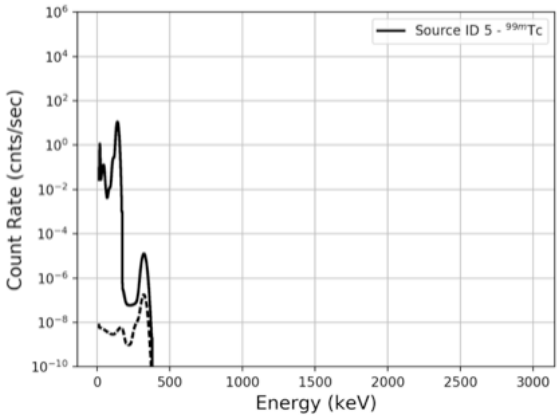
Energy spectra for each source type are shown below for a significant quantity [1] of source types 1 and 2 and 1 microcurie (μCi) for source types 3-5. In each figure, the solid curve shows the unshielded spectrum and the dashed curve, the spectrum with 1 cm of lead shielding. The plots show sources 1 meter away from the center of











The data that generated these figures can be downloaded here (/competition/1/file/5).

File Formats

Files for this competition are standard ASCII and comma delimited (CSV format). They include run files that contain the radiation detection event data for each run and answer files that contain the source type and source time for all the runs in either the training or the test set.

Run Files

unique run ID number. Each run file contains a list of radiation detection events, one event per row. Each row gives the time in microseconds since the previous recorded event and the deposited energies of this event in keV. For the first event in each run file, the time since the last event is recorded as 0.

Example Run File (excerpt from Training File 100001.csv):

```
0,69.0
985,154.7
757,55.5
908,1106.9
1105,356.1
391,116.9
…
```

In this example, the first row records the first event, which had a deposited energy of 69 keV. The second row describes the 2nd event, which occurred 985 μsec after the first with a deposited energy of 154.7 keV. The third row records an event that occurred 757 μsec after the second event (985 +757 = 1742 μsec since the beginning of the run) with a deposited energy of 55.5 keV. The event descriptions continue, row-by-row until the end of the run.

Answers Files

For the training set, the answers file contains the correct answer for each run file in CSV format. Each row of the training set answers file gives an identifier for the run (RunID), an identifier for the source type (SourceID), and the time in seconds (reported to the nearest 1/100 of a second) at which the detector was closest to the source (SourceTime), sometimes referred to as the time of closest approach. For runs without an extraneous source (background only), SourceID and SourceTime are both 0.

Example Training Set Answers File:

```
RunID,SourceID,SourceTime
100001,0,0.00,
100002,0,0.00,
100003,0,0.00,
100004,1,65.58,
100005,0,0.00,
100006,6,36.19,
100007,1,44.10, …
…
```

In the example training file, the first row is for run number 100001, which, like runs 100002 and 100003, had no extraneous source present (0 for SourceID and 0 for SourceTime), while the fourth row (after the header) is for run number 100004, which had source 1 (HEU) present at a location that was nearest the detector 65.58 seconds after the start of that run.

The format for the test set answers file is the same as for the training set answers file, except that the user may optionally add ASCII text comments into a fourth column that the competition hosts will be able to see. You will fill in the SourceID and SourceTime fields of the answers file for each run in the test set and submit the entire file for scoring (/competition/1/upload)