SANDIA REPORT
SAND2023-11060
Printed October 2023

Sandia National Laboratories

Questionnaire for Radioisotope Identification and Estimation from Gamma Spectra using PyRIID v2

Tyler J. Morrow (tmorro@sandia.gov)

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@osti.gov
Online ordering:	http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.gov
Online order:	https://classic.ntis.gov/help/order-methods

ABSTRACT

Accurate targeting of radioisotope classifiers and estimators requires an understanding of the target problem space. In order to facilitate clear communication on expected model behavior and performance between practitioners and stakeholders on their problems, this questionnaire was created. Stakeholder responses form the basis of a trained model as well as the start of usage requirements for the model as it is integrated with analysis processes or detection systems. This questionnaire may also be useful to machine learning practitioners and gamma spectroscopists developing new algorithms as a starting point for characterizing their problem space, especially if they are using PyRIID [5].

ACKNOWLEDGMENT

This work was funded by the U.S. Department of Energy, National Nuclear Security Administration, Office of Defense Nuclear Nonproliferation Research and Development (DNN R&D).

CONTENTS

NOMENCLATURE

DRF Detector Response Function

GADRAS Gamma Detector Response and Analysis Software

OOD Out-of-Distribution

PyRIID Open-Source Python package for ML-based RIID-related software utilities

SI International System of Units

SNR Signal-to-Noise Ratio

1. INTRODUCTION

This questionnaire collects the information needed to construct one or more radioisotope classifiers or estimators for your detection system. Whether targeting classification or estimation, the model or models produced will be simple neural networks and can be delivered in a variety of file formats: Keras HDF, ONNX, TFLite, and more [1] [2] [3]. In addition, model files will be accompanied by an auxiliary file describing their input, output, and assumptions which can be easily parsed. Please put your responses in a separate document and send back in whatever environment is appropriate.

Completing this questionnaire provides several benefits:

- the problem you want to solve is clearly established
- the usage requirements of your model(s) are known up front
- detection system requirements can be reconciled with model usage requirements up front
- the scope of a test plan for a measurement campaign to provide further validation naturally follows

Note that using a model outside its target problem space (i.e., using it on out-of-distribution (OOD) data) often leads to unexpected, inexplicable, or incorrect results. Ways in which one may obtain OOD data include, but are not limited to, the following:

- change in detector response
  - change in *true* energy calibration
  - change in resolution
- change in environment
  - change in scattering
  - change in background composition or rate
- out-of-distribution sources
- detection threshold set lower than the minimum supported signal-to-noise ratio (SNR)

The degree to which detection systems can support OOD detection, by maintaining stability of detector response or environment, varies. While not maintaining detection system assumptions does not necessarily disqualify your system from using this technology, it does introduce a greater risk of processing data which is not supported. As such, unless otherwise noted by model creators, the degree to which you should trust this technology is fundamentally tethered to the engineered stability of your detection system and its ability to hold the assumptions represented by your responses to this questionnaire.

This page left blank

2. REQUIRED INFORMATION

Information here is the minimum required to create a viable model. If you are unsure about any of it, we may be able to provide guidance upon request.

2.1. Handling guidance

How should the model(s) we create be handled and delivered to you?

Why we ask: by default, model training is performed using the public version of GADRAS (available via RSICC) and our Open Source Python package, PyRIID [4] [5]. However, the specific model created for you will be distinct from any model we have made and handled in accordance with the guidance you provide here.

2.2. Detector

Which detector will you be using with the model(s) we create?

To provide info for this section, you can attach your own GADRAS Detector Response Function (DRF) based on your detector or specify the path from the GADRAS Detector folder. Certain subsequent sections will ask about source distance and height, but if that info is stored (and accurate) within the DRF, simply say so in your responses below. If you instead need DRF characterization performed for your detector, note that below.

Why we ask: the DRF provides us with a mandatory baseline understanding of the response your detector will have when it encounters a variety of sources and environments. We use this baseline characterization to vary certain parameters that explore what sources of interest would look like under alternative environmental conditions you might expect.

2.3. Deployment environment

This section seeks to establish the range of environmental conditions your detector should expect. Note that measurements from your actual deployment environment are welcome and particularly helpful.

2.3.1. Surroundings

In what types of locations will your detector be placed, e.g., inside vs. outside, big room vs. small room?

Why we ask: info about your detector's surroundings gives us a clue about scattering.

2.3.2. Background composition

What is the composition of background radiation in your target environment in terms of Potassium (K), Uranium (U), Thorium (T), cosmic, and intrinsic sources. One or more background measurements from the target environment with your detector is most ideal for answering this question, but we recognize it may not be possible. When a measurement is not available, any information about the nature of environmental materials is useful, such as dirt, concrete, etc. GPS location, or even just altitude above sea level, is also useful.

Why we ask: this helps us capture representative background signature assumptions.

2.3.3. Background rate

What range of background rates (in counts per second) do you anticipate from all background sources including K, U, T, cosmic, and intrinsic?

Why we ask: after we synthesize various background compositions, this information is one part of how we bound the expected signal-to-noise (SNR) ratios.

2.3.4. Distance to source

What range of distances to the source do you anticipate?

Why we ask: sometimes the distance to a source during measurement varies and that may significantly affect scattering. This information helps us try to capture that effect.

2.3.5. Detector height above ground

What range of detector heights above the ground do you anticipate?

Why we ask: sometimes detectors are sat on the ground, on tables, or suspended in the air, all of which may significantly affect scattering. This information helps us try to capture that effect.

2.3.6. Event duration

What range of event durations (i.e., time, in seconds, that source is present) do you anticipate?

Why we ask: sometimes it is not known how long a source will be present. This information helps us vary event duration to account for such variation in signal-to-noise ratio calculations.

2.4. Model type

This section captures the type and behavior of the model you would like.

2.4.1. Model input(s)

What is the exact form of the gamma spectrum your detection system will provide to the model?

The main goal is to understand the pre-processing your detection system can and will perform, if any.

Examples of pre-processing:

- integration of spectra over event duration
- binning of list-mode data
- energy calibration determination and reinterpolation
- down-binning
- spectrum normalization (L1-norm, L2-norm, etc.)
- dimensionality reduction (PCA, NMF, etc.)

If you have neutron information accompanying your gamma spectrum, please note that. Generally speaking, gamma or neutron counts (or transformations of them) is all we will use, whereas engineered features will not.

Why we ask: we want the trained model to take as input exactly what your detection system provides.

2.4.2. Model output(s)

What model output would you like?

For classification problems, the options are:

- Radiation source (think isotope or named mixture of isotopes) + configuration (shielding)
- Radiation source only
- Arbitrary category

For regression problems, the options are:

- Proportion estimates: the proportion of counts in the spectrum associated with each source; this is for the full spectrum, not peak.
- Activity estimates: this is just proportion estimation plus a post-processing step to convert proportion estimates into activity by incorporating detection system and source assumptions.

After noting your problem type, please include a list of the radiation sources and shielding expectations (materials and thickness ranges) you are targeting.

This page left blank

3. OPTIONAL INFORMATION

Information here is not required to create a useful model, but when available, it helps specialize your model even more. When this info is not provided, or you are unsure about it, we may be able to provide guidance upon request.

3.1. Duration of deployment

For how long will your detector be deployed (in months, years, etc.)?

Why we ask: the performance of a model may degrade as the detector ages and deviates from its original DRF. This info gives us an opportunity to catch that up front and potentially create models for various detector ages.

3.2. Sources not-of-interest

What sources, if any, should the model still be aware of even though they are not of interest?

Why we ask: this info is useful for understanding true negatives, false positives, or false negatives. It gives the model the opportunity to have better explanations for measurements you might ask it to analyze.

3.3. Additional test data

Would you like to provide a dataset of your own on which to further test the model(s)?

If so, please attach or link to it, and include any requirements or documentation for it.

Why we ask: datasets captured here for this question can be measured from the target environment, but can also be synthetic datasets that would yield more experimental results to reveal even more about model behavior.

3.4. Source strength

What range of signal-to-noise ratios (SNR) or source activities do you anticipate?

The SNR to which we refer is the full spectrum net counts divided by the square-root of background counts.

Why we ask: this knowledge may allow us to shrink or expand the problem space in terms of SNR to better prepare your model.

3.5. Additional information

Is there any other relevant info about your detection system or deployment environment? If so, please let us know.

REFERENCES

[1] Junjie Bai, Fang Lu, Ke Zhang, et al. *ONNX: Open Neural Network Exchange*. https://github.com/onnx/onnx. 2019.

[2] François Chollet et al. *Keras*. https://keras.io. 2015.

[3] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[4] Dean J. Mitchell et al. *GADRAS Detector Response Function*. Tech. rep. Sandia National Laboratories, Albuquerque, NM (United States), 2014.

[5] Tyler Morrow, Nathan Price, and Travis McGuire. *PyRIID*. Apr. 2021. DOI: 10.11578/dc.20221017.2.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov

This page left blank

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.