SAND2016-7489D

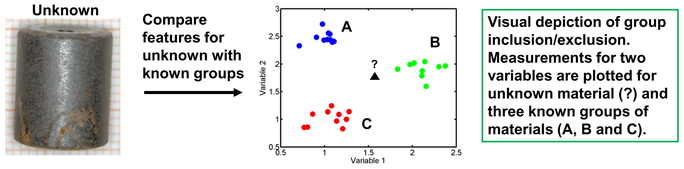
Dynamic Analysis Environment (DAE) – An Interactive Test Bed for Nuclear Forensics

Christopher Stork[1], Matthew Clopeck[1], David Stuart[1], Chad Ummel[2], Susan Bodily[1]

[1]Sandia National Laboratories*, [2]University of California, Berkeley

Introduction

Dynamic Analysis Environment (DAE) is a graphical, user-interactive environment to facilitate group inclusion/exclusion method testing, evaluation and comparison for pre-detonation nuclear forensics applications. Employing DAE, the multivariate signatures of a questioned/test material can be compared to the signatures for different, known groups, enabling the linking of the questioned material to its potential process, location, or fabrication facility. Advantages of using DAE for group inclusion/exclusion include built-in query tools for retrieving data of interest from a database, the recording and documentation of all analysis steps, a clear visualization of the analysis steps intelligible to a non-expert, and the ability to integrate analysis tools developed in different programming languages. Two group inclusion/exclusion methods are implemented in DAE: (1) principal component analysis (PCA), and (2) K nearest neighbors (KNN). In 2016, three new data visualization tools have been incorporated into DAE to help an analyst graphically compare questioned materials with known groups of materials, facilitating group inclusion/exclusion.

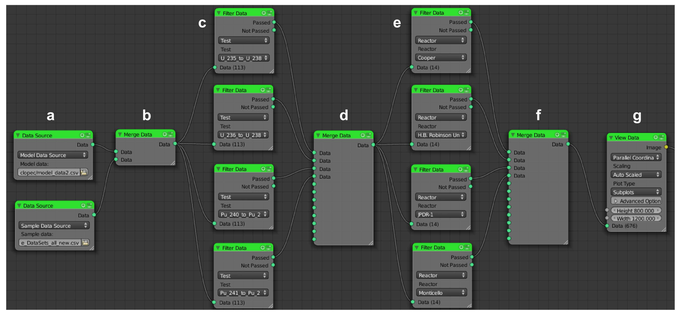


Visual depiction of group inclusion/exclusion. Measurements for two variables are plotted for unknown material (?) and three known groups of materials (A, B and C).

DAE – Basic Function and Capabilities

DAE is an interactive, configurable environment that facilitates the analysis of nuclear forensics data sets through the use of predefined analytical modules, called nodes. These nodes can be arranged into an analytical chain of data processing steps. Available nodes include:

1. Data Source – Extracts known and questioned/test data from files.
2. Augment Data – Extends source data by invoking analytical processes via dropdown selections. Modules exist for constructing PCA and KNN models.
3. Filter Data – Splits data into two separate flow paths based on selected filter criteria. Modules exist for filtering groups/reactors and variables.
4. Merge Data – Combines data streams (e.g., groups/reactors or variables).
5. Relate Data – Relates test data to PCA or KNN model.
6. View Data – Displays data at any point in the analysis chain.



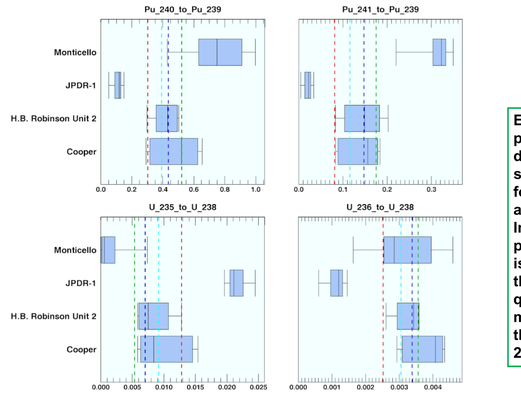
An analysis chain consisting of (a) two Data Source nodes, (b) a Merge Data node, (c) four Filter Data nodes filtering on isotopic variables, (d) a second Merge Data node, (e) four Filter Data nodes filtering on individual reactors, (f) a third Merge Data node, and (g) a View Data node to graphically compare the test data to the model data from the four filtered reactors.

Nuclear Forensics Data - SFCOMPO

Spent Fuel Isotopic Composition (SFCOMPO), an open source international database of isotopic compositions for Spent Nuclear Fuels (SNF) obtained through post-irradiation experiments, is used to demonstrate the utility of DAE in addressing the group inclusion/exclusion problem. SFCOMPO consists of SNF compositions for 14 reactors in 4 countries. Simulated SNF samples were generated by Argonne National Lab and treated as questioned/test samples.

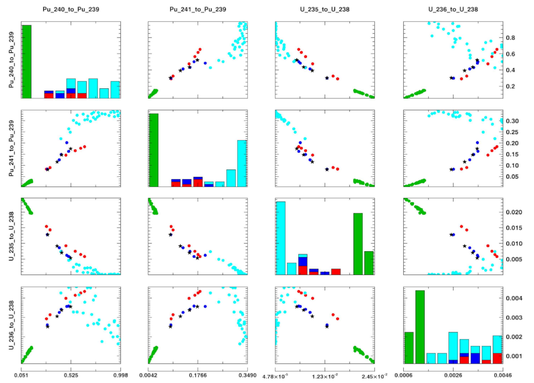
New DAE Data Visualization Tools – Box, Pairwise and Parallel Coordinates Plots

Three new data visualization tools, namely (1) a univariate box plot utility, (2) a bivariate pairwise plot utility, and (3) a multivariate parallel coordinates plot utility, have been incorporated into DAE to help an analyst compare questioned materials with known groups of nuclear materials, facilitating group inclusion/exclusion. The figures below demonstrate the box plot, pairwise plot, and parallel coordinates plot utilities included in DAE using data extracted from SFCOMPO. For this example, $^{240}Pu/^{239}Pu$, $^{241}Pu/^{239}Pu$, , $^{235}U/^{238}U$, and $^{236}U/^{238}U$ data for four questioned samples are compared to samples from four reactors.



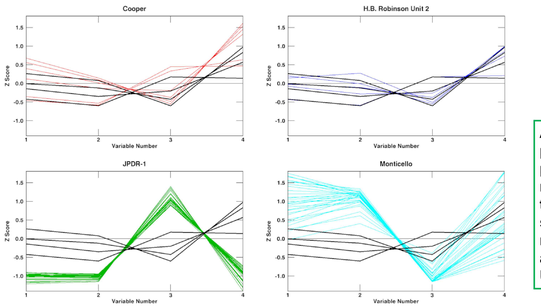
Each box plot provides a graphical depiction of the statistical distribution for a given reactor for a given isotopic ratio. Inspection of the box plots for the four isotopic ratios reveals that the four questioned samples most closely resemble the H.B. Robinson Unit 2 reactor.

A separate box plot is displayed for each of the four reactors for each of the four isotopic ratios. For each isotopic ratio, the values for the four questioned samples are represented by the vertical, dashed red, cyan, blue, and green lines.



An inspection of the pairwise scatter plots reveals that the four questioned samples most closely resemble the H.B. Robinson Unit 2 reactor.

In the pairwise plot, each off diagonal element shows the scatter plot for the four reactors and four questioned samples for a given pair of isotopic ratios.

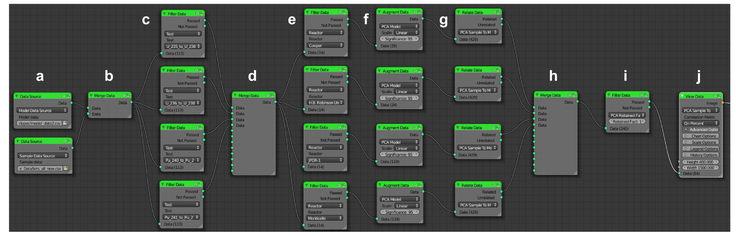


An inspection of the parallel coordinates plots for the four reactors reveals that the four questioned samples most closely resemble the Cooper and H.B. Robinson Unit 2 reactors.

Reactor-specific parallel coordinates plots are displayed that compare the four isotopic ratio values for the four questioned samples to each given reactor.

DAE Group Inclusion/Exclusion Methods – Principal Component Analysis and K Nearest Neighbors

Two group inclusion/exclusion methods, principal component analysis (PCA) and K nearest neighbors (KNN), are fully integrated in DAE. Integration of a third group inclusion/exclusion method, T^2 Fullrank, into DAE is near completion. The figures below demonstrate the process of constructing PCA and KNN models in DAE using training data composed of four reactors from SFCOMPO, and classifying four questioned/test samples using these models.



A PCA analysis chain consisting of (a) two Data Source nodes, (b) a Merge Data node, (c) four Filter Data nodes filtering on isotopic variables, (d) a second Merge Data node, (e) four Filter Data nodes filtering on individual reactors, (f) four Augment Data nodes used to construct a PCA model for each filtered reactor, (g) four Relate Data nodes to statistically compare the test data to the four reactor-specific PCA models, (h) a third Merge Data node, (i) a Filter Data node to retain only the first PC in each PCA model, and (j) a View Data node to visualize the group inclusion/exclusion results for the test samples. A similar analysis chain was constructed for KNN.

PCA Sample To Model Correlation - Qn

Model	RF	1	2	3	4
Cooper	1	0.0000	0.0000	0.0000	0.0000
H.B. Robinson Unit 2	1	0.4254	0.5361	0.4587	0.5510
JPDR-1	1	0.0000	0.0000	0.0000	0.0000
Monticello	1	0.0000	0.0000	0.0000	0.0000

PCA Sample to Model Correlation table output by DAE using a View Data node, summarizing the PCA group inclusion/exclusion results for the four test samples. The probability values for the H.B. Robinson Unit 2 model all exceed 0.25, suggesting that the test samples all originated from this reactor.

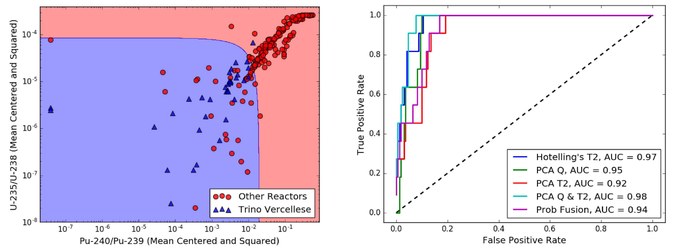
KNN Sample To Training Classification

Training Class	K	1	2	3	4
H.B. Robinson Unit 2	3	NaN	-1.28501	-1.05530	-0.736983
JPDR-1	3	NaN	NaN	NaN	NaN
Monticello	3	NaN	NaN	NaN	NaN
Cooper	3	NaN	NaN	NaN	NaN

KNN Sample to Training table output by DAE using a View Data node, summarizing the KNN group inclusion/exclusion results for the four test samples. Three of the four test samples are deemed as originating from the H.B. Robinson Unit 2 reactor. NaN signifies a test sample is inconsistent with a given reactor.

Planned DAE Enhancements

- Incorporate additional group inclusion/exclusion methods such as probabilistic fusion and one-class support vector machines (SVM).
- Incorporate group inclusion/exclusion performance evaluation metrics, such as accuracy and precision, to enable the quantitative comparison of methods.



Left Figure: Decision boundaries differentiating Trino Vercellese materials (blue regions) from other reactors (red regions) generated by probabilistic fusion model. Right Figure: Receiver Operating Characteristic (ROC) curves summarizing the performance of five group inclusion/exclusion models for data extracted from SFCOMPO. Both of these functionalities are incorporated in an expert, Python-based test bed being developed at Sandia.

*Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

This work was performed by Sandia National Laboratories for Department of Homeland Security, Domestic Nuclear Detection Office, National Technical Nuclear Forensics Center.