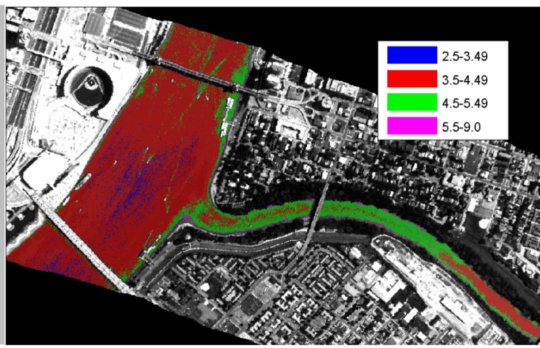
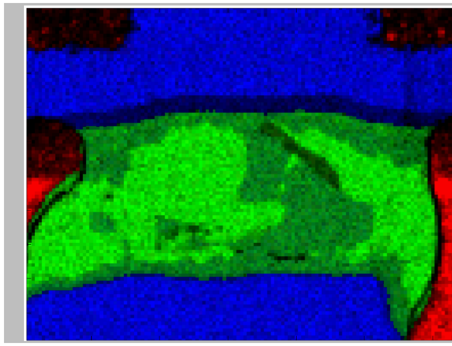
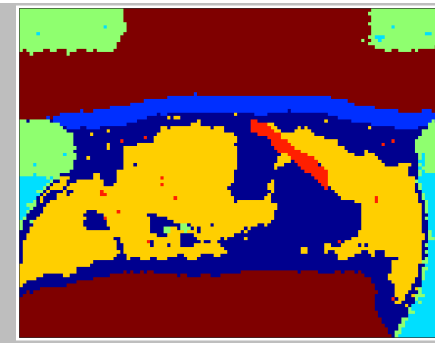
SAND2015-10520PE

Exceptional service in the national interest

Sandia National Laboratories







Machine Learning Capabilities

Chris Stork, Dept. 1814

December 9, 2015

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Machine Learning Overview

- Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Machine learning algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions.

Machine Learning Methods

Sandia National Laboratories

<u>Method</u>	<u>Use</u>
Clustering	Identify intrinsic groups with little or no prior knowledge
Constrained factor analysis (CFA)	Extract spectral and concentration profiles
Group inclusion/exclusion	Quantify similarity of unknown material to known groups
Multivariate regression	Predict chemical quantity using sensor response
Multivariate statistical process control (MSPC)	Detect process/sensor changes

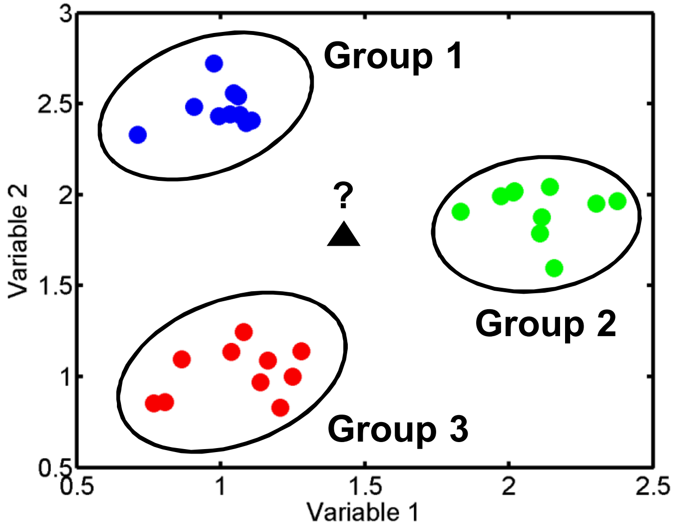
Group Inclusion and Exclusion of Nuclear Forensics Data

- **Developing principal component analysis and nearest neighbor based classification algorithms to enable rapid, accurate, and credible identification of origins of questioned nuclear materials**
  - **Compare multivariate signatures (isotopics, trace elements) for questioned nuclear material to signatures for materials from known groups (reactors, processes), enabling group inclusion and exclusion**

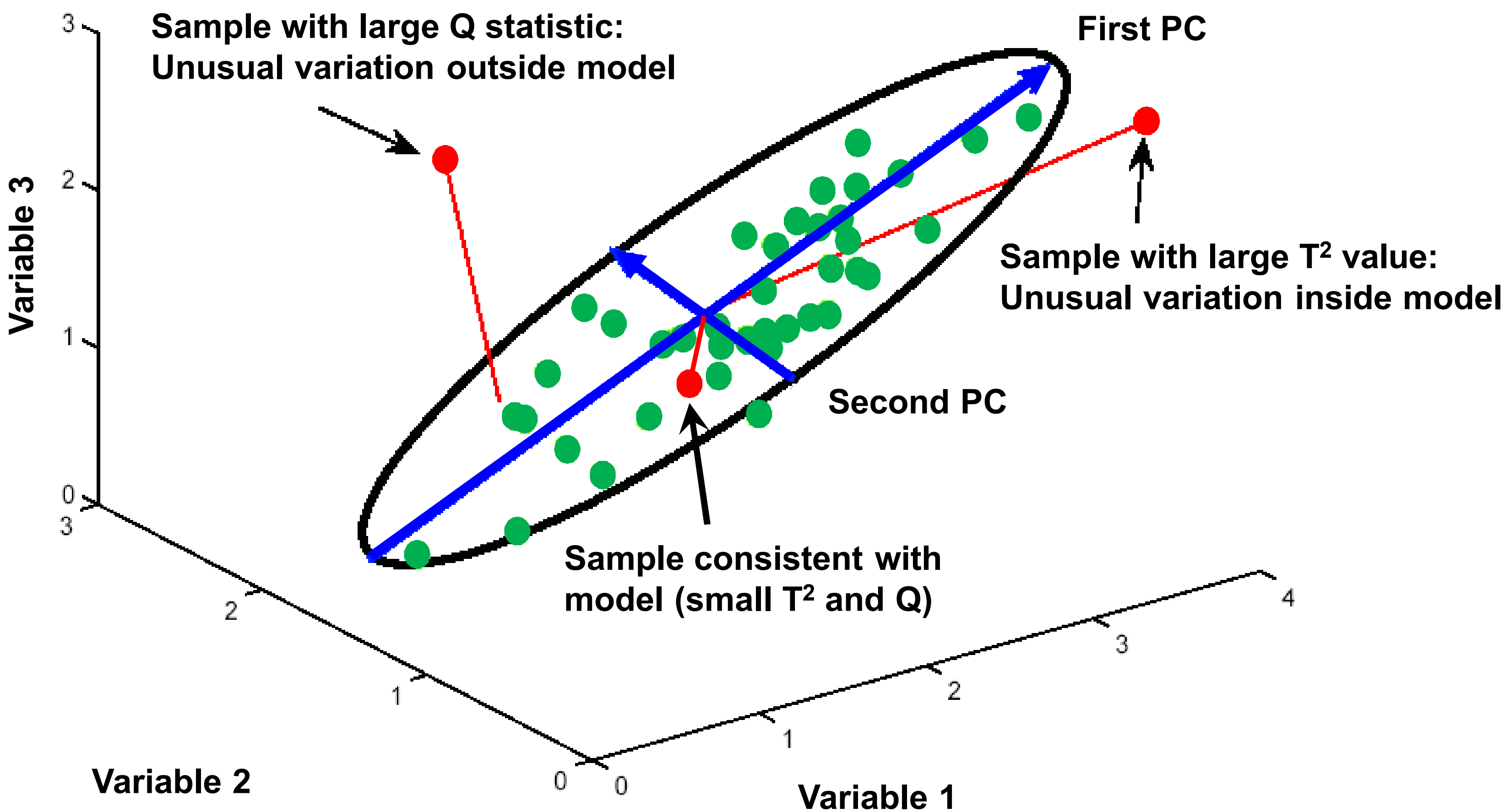
Questioned material

Compare signatures for material with known groups

Exclusion - questioned material inconsistent with known groups



Technical Approach: Address Group Inclusion/Exclusion Problems using PCA



Spent Fuel Isotopic Composition (SFCOMPO) Database

- Composed of isotopic composition data for spent nuclear fuel (SNF) samples from 14 reactors in 4 countries
  - Includes U, Pu, Am, Cm and several fission products (Nd, Cs, Sr)
  - 7 boiling water reactors (BWRs) and 7 pressurized water reactors (PWRs)

Gundremmingen/Germany

Tsuruga Plant/Japan

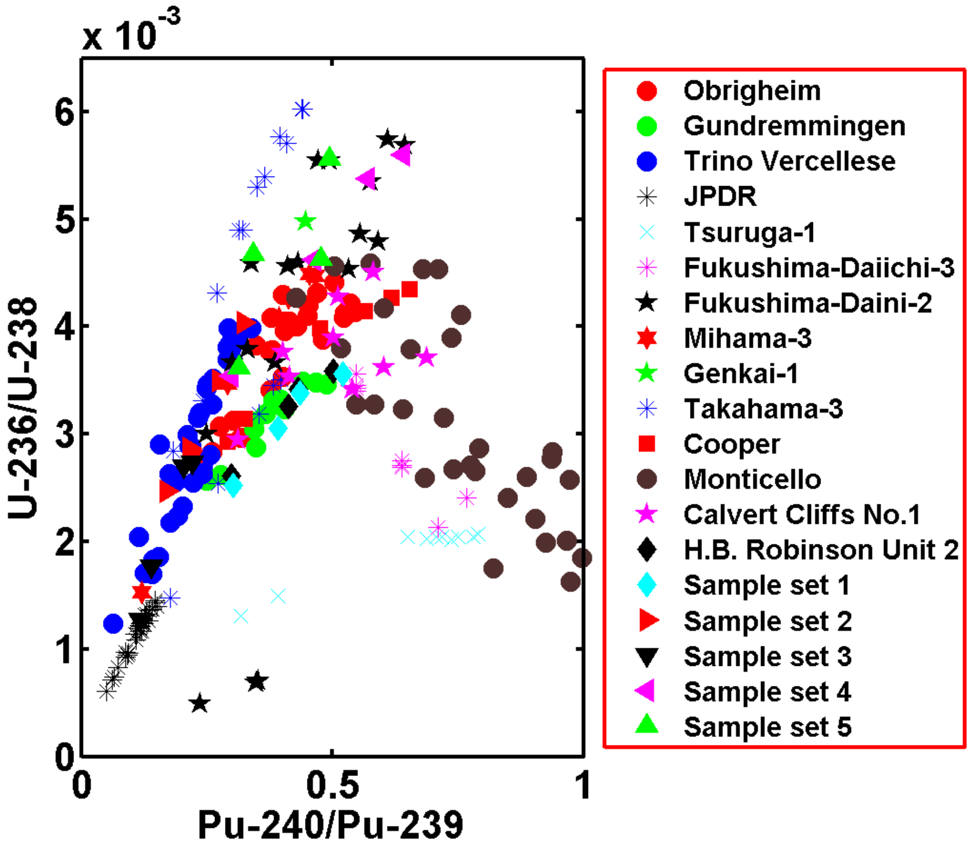
Calvert Cliffs No. 1/USA

Group Inclusion/Exclusion Analysis using SFCOMPO Database

- Separate PCA model constructed for each reactor or related set of reactors within SFCOMPO
- In blind study, five simulated SNF sample sets from Argonne National Laboratory (ANL) treated as questioned samples and statistically compared to each PCA group model
  - Each simulated sample set originated from reactor in SFCOMPO
  - Four samples within each set
- Four indices used for group inclusion and exclusion:
  - Q statistic
  - Hotelling's T^2 statistic
  - Hawkins' T^2_H statistic
  - Mahalanobis distance
- Assessment and ranking of these indices based on their group inclusion/exclusion performance

Group Inclusion/Exclusion Analysis of Simulated SNF Sample Sets

- 5 isotopic ratios used in statistical comparisons:
  - $^{240}Pu/^{239}Pu$, $^{241}Pu/^{239}Pu$, $^{242}Pu/^{239}Pu$, $^{235}U/^{238}U$, $^{236}U/^{238}U$



- **Sample set 1 overlaps H.B. Robinson Unit 2 and Gundremmingen reactor samples**
- **Sample sets 2 and 3 overlap Trino Vercellese and Mihama-3 reactor samples**
- **Sample sets 4 and 5 overlap Mihama-3 and Fukushima-Daini-2 reactor samples**

Q Statistic Probabilities Calculated for Sets 1-5 for 14 PCA Models

PCA Model	Set 1	Set 2	Set 3	Set 4	Set 5
Obrigheim	0.002-0.007	*0.015-0.093*	*0-0.076*	0.000-0.004	0.000-0.005
Gundremmingen	0.000-.001	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.000
Trino Vercellese	0.000-0.000	*<u>0.100-0.343</u>*	*<u>0.013-0.124</u>*	0.000-0.000	0.000-0.000
JPDR	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.000
Tsuruga-1	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.000
Fukushima-Daiichi-3	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.002	0.000-0.004
Fukushima-Daini-2	0.000-0.000	*0.078-0.155*	*0.001-0.145*	*<u>0.410-0.669</u>*	*<u>0.122-0.678</u>*
Mihama-3/ Genkai-1	0.000-0.000	*0.007-0.506*	*0.027-0.284*	0.000-0.005	0.000-0.006
Takahama-3	0.000-0.000	*0.024-0.055*	*0.004-0.065*	0.000-0.004	*0.000-0.014*
Cooper	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.000
Monticello	0.000-0.001	0.000-0.004	0.000-0.000	0.000-0.002	0.000-0.001
Calvert Cliffs No. 1/H.B. Robinson Unit 2	*0.247-0.508*	0.000-0.003	0.000-0.000	0.000-0.001	0.000-0.003
Calvert Cliffs No. 1	*0.052-0.154*	0.000-0.002	0.000-0.000	0.000-0.002	0.000-0.008
H.B. Robinson Unit 2	*<u>0.033-0.460</u>*	0.000-0.000	0.000-0.000	0.000-0.000	0.000-0.000

Q statistic performs best:

- **Correctly excludes large percentage of reactor groups for all 5 sample sets**
- **Correctly includes true reactor of origination for all 5 sample sets**

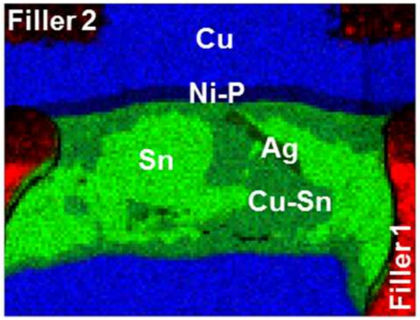
Entries above exclusion threshold of 0.01 highlighted in *red italics*. True reactor of origination for each sample set is <u>underlined</u>.

Material Characterization Through Cluster Analysis of Hyperspectral Images

- **Identification of intermetallic phases in material samples through cluster analysis**

Fuzzy clustering identifies Cu-Sn intermetallic phase in EDX solder bump data set

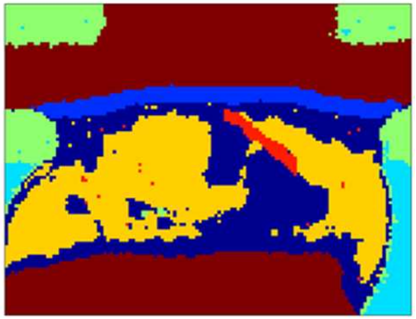
RGB composite EDX image



Clustering, D = CS^T
(Fuzzy c-means, m = 1.3)

Clustering estimates:
- S, spectrum for each group
- C, group assignment for each pixel

Pixel group assignments

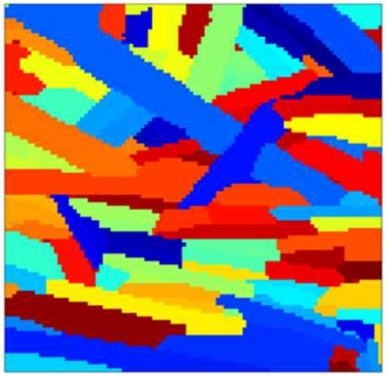


- **Phase identification in complex materials using fast hierarchical clustering algorithm**

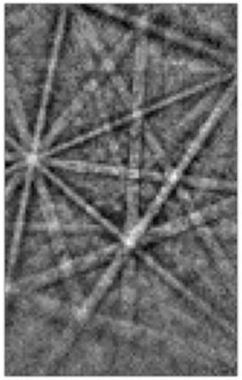
Two US patents issued (8,280,887 and 8,554,771)

Application of hierarchical clustering to bismuth titanate (BTO) diffraction data reveals 120 readily interpretable phases

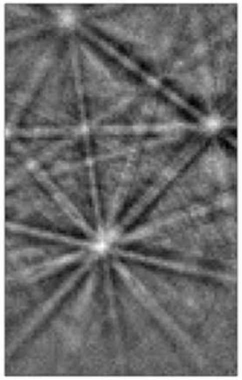
Pixel Assignments
(120 Clusters)



Diffraction Pattern
(Cluster 1)



Diffraction Pattern
(Cluster 2)



Potential Areas of Collaboration

- Summer internship opportunities at Sandia for undergraduate and graduate students.
- Materials Data Science collaboration opportunities:
  - Sandia has extensive materials databases that could be used to facilitate development of new/improved materials.
  - Sandia has machine learning expertise that can be extended to model relationships between materials processes, structures and properties.