ADAPD:
Advanced Data Analytics for Proliferation Detection

DSI Workshop – August 8, 2018
Eddy Banks

LLNL – Jim Brase
LLNL – Eddy Banks
LANL – Jim Smith
SNL – Danny Rintoul
ORNL – Phillip Bingham
PNNL – Mark Greaves
ADAPD – Goal: Early Detection of Low-Profile Proliferation

ADAPD will build science-based capabilities to predict and detect proliferation observables and characterize associated activities.

Data Science to the rescue:

- Emerging analytic methods that integrate physics-based simulation models with experimental data have demonstrated significant increases in prediction performance for complex physical modeling applications.
- ADAPD will draw upon and extend these emerging methods to integrate physics and subject-matter-expert models with DOE/NNSA testbed data to improve prediction of proliferation activities.
- These methods will provide new capabilities for detection of low-profile proliferation observables in noisy, limited data environments with quantified uncertainties.
Advances in data science enable new approaches to low-profile proliferation detection

Knowledge, models, testbeds and activities

Testbed ground truth data

Testbed process simulations

“Learn the physics”

Train with combined facility data and SME-built simulations

Re-train the top layers using limited observable data

Prediction of observables in the new environment

Query the model with new process state and environmental inputs

Advances in data science enable new approaches to low-profile proliferation detection

Knowledge, models, testbeds and activities

Testbed ground truth data

Testbed process simulations
### ADAPD – R&D Objectives

<table>
<thead>
<tr>
<th>U Isotope Separation</th>
<th>WGPu Production</th>
<th>Buildup to Testing</th>
<th>Component Manufacturing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive forward models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-based multi-phenomenology detection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer and validation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### R&D Objectives

- Subject Matter Expertise
- Data
- Testbeds
Approach: Predictive models of proliferation observables driving model-based detection

Objective 1 – Combine SME-built with data-driven models for predicting multiple observables

Objective 2 – Multi-phenomenology detection for complex/varied environments

Objective 3 – Model validation and transfer to new Low-Profile environments

DOE/NNSA testbeds
Predictive models integrated into multisource analytics are essential to limited data applications

- Testbeds
  - SPE
  - REDC/HIFR
  - …

- Data Streams
  - Acoustic, seismic
  - Optical
  - Facility power
  - Materials procurement and shipping patterns
  - Patterns of life
  - …

- Analytic Capabilities
  - Integrating direct and indirect data in the context of a predictive model will enable detection in the noisy, limited data environments and extend capabilities for quantitative estimates of rates and uncertainties
To detect Low-Profile Proliferation, ADAPD will address key data science gaps

**Predictive models integrating physics and data-driven approaches**

- On-line machine learning
- Experimental data
- The next experiment?
- Simulation ensemble pipeline
- Model predictions, uncertainties
- The next simulation?

Collaboration with DP efforts of LLNL and LANL

**Multi-source data integration**

- Deep multimodal learning methods
- Sparse and noisy data – tensor models
- SME model-based transfer learning

**Scalable network and graph analytics**

- 2:00AM 5:00AM 12:00PM
- Aggregated trip counts between Manhattan census tracts
- November 2 2013 (marathon day)

- Pattern of life analytics
- Temporal activity learning
- Rare event and anomaly learning
- Learning from sparse data

**Uncertainty quantification in data-driven systems**

- Models learned from data
- High dimensional wave forms, spectra, images
- Accountability from inference back to data

Uncertainty Quantification (UQ) in concept

- System Config
- Computer Model
- Model Error
- Model Output
- Uncertain parameters
- Compare to learn about θ and δ
- Experimental Observations

UQ is learning about θ and δ, and using probability distributions to make predictions.

Lawrence Livermore National Laboratory
LLNL-PRES-754931
ADAPD will be driven by a set of science questions

**Objective 1: Predictive modeling** – Can we predict the broad range of observables associated with Low-Profile Proliferation?
- Can we combine physics, SME, and data-driven models to accurately predict a broad range of observables?
- What observations and experiments on existing testbeds are needed to develop and validate predictive models?

**Objective 2: Multi-phenomenology detection and characterization** – Can we use multiple observables to detect and characterize proliferation activities?
- Can we use predictive models to improve performance with noisy and sparse data?
- Can we quantify uncertainty in detection results that are based on multiple uncertain observables?

**Objective 3: Validation and extension to new environments** – Can we transfer a learned model to new geographical locations, different infrastructures, ....
- What tests should be performed on existing testbeds to best validate and calibrate models?
- Can we combine models for proliferation activities involving multiple facilities and sites?
Industry and university partnerships are an important component of the ADAPD plan

**Universities**
- Basic R&D in machine learning algorithms and methods
- Much of our data can be shared
- Output is prototype algorithms and – most importantly - students

**Industry**
- Industry-standard opensource computing frameworks and tools
- Innovative computing technology and architectures
- Large open data sets

**ADAPD**
- Application to nonproliferation
- Work with full spectrum of data and models
- Full access to testbeds and activities
- R&D in targeted nonproliferation modeling and analytics
The ADAPD R&D Roadmap focuses on building cross-cutting and sustainable capabilities

<table>
<thead>
<tr>
<th>Year</th>
<th>Predictive modeling</th>
<th>Multi-phenomenology detection</th>
<th>Transfer to new environments</th>
<th>Testbed and Experiment Partnerships</th>
<th>Data Management and curation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Predictive model development – baseline scenarios</td>
<td>Detection and characterization development – baseline methods</td>
<td>Develop synthetic data approaches to enhanced model generalization</td>
<td>Collaborative testbed exp 1 training set development</td>
<td>Assess existing data sets • Catalog and curate existing data sets</td>
</tr>
<tr>
<td>2</td>
<td>Predictive model development – expanded observables – both direct and indirect</td>
<td>Detection and characterization development – inverse inference methods</td>
<td>Test and evaluate transferability of learned models to new data sets on the same testbed</td>
<td>• Collaborative experiment 2</td>
<td>Catalog and curate new data sets</td>
</tr>
<tr>
<td>3</td>
<td>• Full UQ integration with integrated models</td>
<td>Uncertainty quantification for inverse inference models</td>
<td>Test and evaluate transferability of learned models to partially unknown or limited data environments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>Test and evaluate transferability of learned models to new testbeds &amp; environments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Catalog and curate existing data sets**
- **Collaborative testbed experiment 2**
- **Indirect data planning and collection campaigns**
- **Evaluate testbeds and experiments for transfer experiments**
- **Validate predictive models against extended range of testbed experiments**
- **Apply & evaluate performance for partially unknown proliferation scenarios**
- **Detect & characterization performance enhancement in high-noise, sparse and missing data environments**
- **Check predictive model development – baseline scenarios**
- **Full UQ integration with integrated models**
- **Test and evaluate transferability of learned models to new data sets on the same testbed**
- **Detection and characterization development – inverse inference methods**
- **Uncertainty quantification for inverse inference models**
- **Detection and characterization development – baseline methods**
- **Catalog and curate new data sets**
- **Develop uniform cataloging, curation, and control across DNN**
- **Predictive model development – expanded observables – both direct and indirect**
- **Test and evaluate transferability of learned models to partially unknown or limited data environments**
- **Detection and characterization development – inverse inference methods**
- **Catalog and curate existing data sets**
ADAPD – Questions?