

NREL is a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, operated by the Alliance for Sustainable Energy, LLC.

# The NREL High Throughput Experiments for Materials (HTEM) Database **A Prototype Project-Specific Analytics Database Enabling the Application of Machine Learning to Experimental Data**

John Perkins, Andriy Zakutayev, Caleb Phillips, Jacob Hinkle, Robert White, Kristin Munch, Marcus Schwarting

#### john.perkins@nrel.gov

#### **Overview**

In the process of accelerating our analysis of large volumes of x-ray diffraction data, we have created a prototype data analysis ecosystem for creating project specific data analytics tools. We leverage existing NREL data harvesting capabilities and open source machine learning code libraries to create custom user-friendly analysis tools.

#### **Data Ecosystem for Data Analytics**

NREL Data Harvesting System Architecture



### **User Tool Development**

Leverage Open Source Machine Learning Resources



## The Data Analysis Need: **Rapid Analysis of XRD data**

880 Unanalyzed XRD Patterns (Data)



Structure Phase Map (Knowledge) Faster ?

### **Pieces of the Solution**

Peak Identification in Background Subtracted Data



	Accessible to every Built on NumPy, Sc Open source, comn	body, and reusable in various contexts iPy, and matplotlib nercially usable - BSD license	
Classification Identifying to which category an object belongs to. Applications: Spam detection, Image recogni- tion. Algorithms: SVM, nearest neighbors, random forest, — Examples	Regression Predicting a continuous-valued attribute associ- ated with an object. Applications: Drug response, Stock prices. Algorithms: SVR, ridge regression, Lasso, – Examples	Clustering Automatic grouping of similar objects into sets. Applications: Customer segmentation, Group- ing experiment outcomes Algorithms: k-Means, spectral clustering, mean-shift, – Examples	• Incl ma tas
Dimensionality reduction Reducing the number of random variables to consider. Applications: Visualization, Increased efficien- cy Algorithms: PCA, feature selection, non-nega- tive matrix factorization. — Examples	Model selection Comparing, validating and choosing parame- ters and models. Goal: Improved accuracy via parameter tuning Modules: grid search, cross validation, metrics. – Examples	Preprocessing Feature extraction and normalization. Application: Transforming input data such as tor use with machine learning algorithms. Modules: preprocessing, feature extraction. – Examples	• Alw • Fre

ncludes variety of machine leaning asks Always growing

Create Custom Python Package: *unmix xrd* 

#### • Custom python library & command-line program

- use routines in scikit learn
- can be run from Terminal or Python Notebook - single line commands run custom analysis

e.g. \$ unmix\_xrd cluster ---method spectral -o ./out/G\_ -n 4 ./data ----

- Includes peak detection, clustering, sparse unmixing, visualization • Pulls data from public & private databases resources - materialsproject.org, materials.nrel.gov, HTEM • Extensible
- Can be called from commercial analysis programs
  - integrate into existing analysis workflow
  - easily create custom user analysis tools
  - e.g. Menu Driven Cluster Analysis in Igor Pro

			Set Active Project in Igor		Contraction State Contraction	
0.0.0	Data Browser		New Ternary Cluster Graph		TRDAnalytics	
Current Data Folder: root		*	Run Spectral Cluster Analysis - Single N Bun Spectral Cluster Analysis - Multi N		#pragma TextEncoding = "UTF-8" #pragma rtGlobals=3 // Use modern global access method and strict wave access.	
Display	< 🔶 🍐 root	0	Load Cluster Data	Load Spectral Cluster Results	3 #pragma IgorVersion = 7.00	
✓ Waves	Name	A	Paul and Table		4 #pragma version = 0.1 5 #pragma ModuleName = XRDAnalytics_JDP	
✓ Variables	V Elec		Package loois	Load data:samples.comp.csv	6 7 Chatle Constant Mundanthation - 0.1	
✓ Strings	S fileName		3 8 ° 8 😽	Load data:samples.scalais.csv	8 Static StrConstant ksModuleName = "XRDAnalytics_JDP"	
✓ Info	S_path		6 6 28 8		9	
✓ Plot	S_waveNames		S And Carlo	S70	10 #include dirings as Lists menus-0	



"Heat" Map many sample, many ref





Samples in composition space







#### Implemented via simple system calls to *unmix\_xrd* package



#### **The Take Away**

 Custom Data Tools Can Accelerate Research Automated Processing, Visualizations, Easy Access, Machine Learing, ...



#### Results of spectral clustering on affinity matrix





