SAND2009-5097P

Ground-based Nuclear Explosion Monitoring Research at Sandia National Laboratories

Presented at the IDC, Vienna, Austria

August 18, 2009

Chris Young

GNDD Deputy Lab Leader, Sandia National Laboratories

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Database Access & Merging

DBTools and DBUtillib

Statement of Work

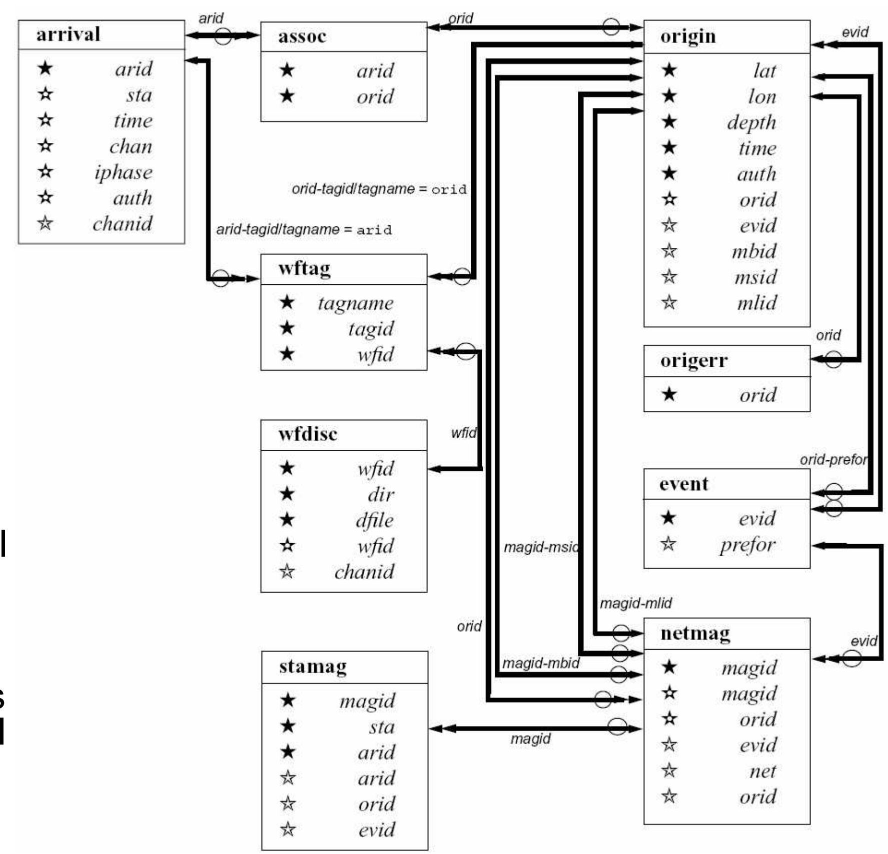
- SNL makes extensive use of oracle databases for storing information, but we have found the tools available for accessing information of interest for a given application to be cumbersome. Hence, we chose to design software to make it easier to interact with a database for various tasks

Work Accomplished

- SchemaSchema/Table Definition Tables (with Richard Stead): information about the schema is stored in database tables
- Rowgraphs: the compact representation of all related information within a schema to the level of detail specified by the user. Allows an application developer to easily get all the needed information without doing the full join
- EvLoader: tool to merge event catalogs with as extensive a schema as is needed to capture all related information
- DTX: general tool to intelligently merge or extract information from a database
- Wfmerge: merges waveform segments

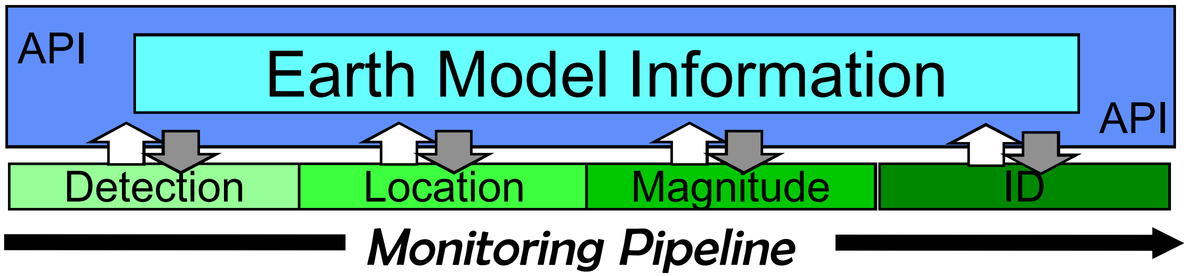
What's Next

- Improve tools as needed (no new tools anticipated)



Earth Model APIs

API Overview



Statement of Work

- SNL develops APIs to access Earth Model information for pipeline processing (libpgl, libslbm, libcorr)

Work Accomplished

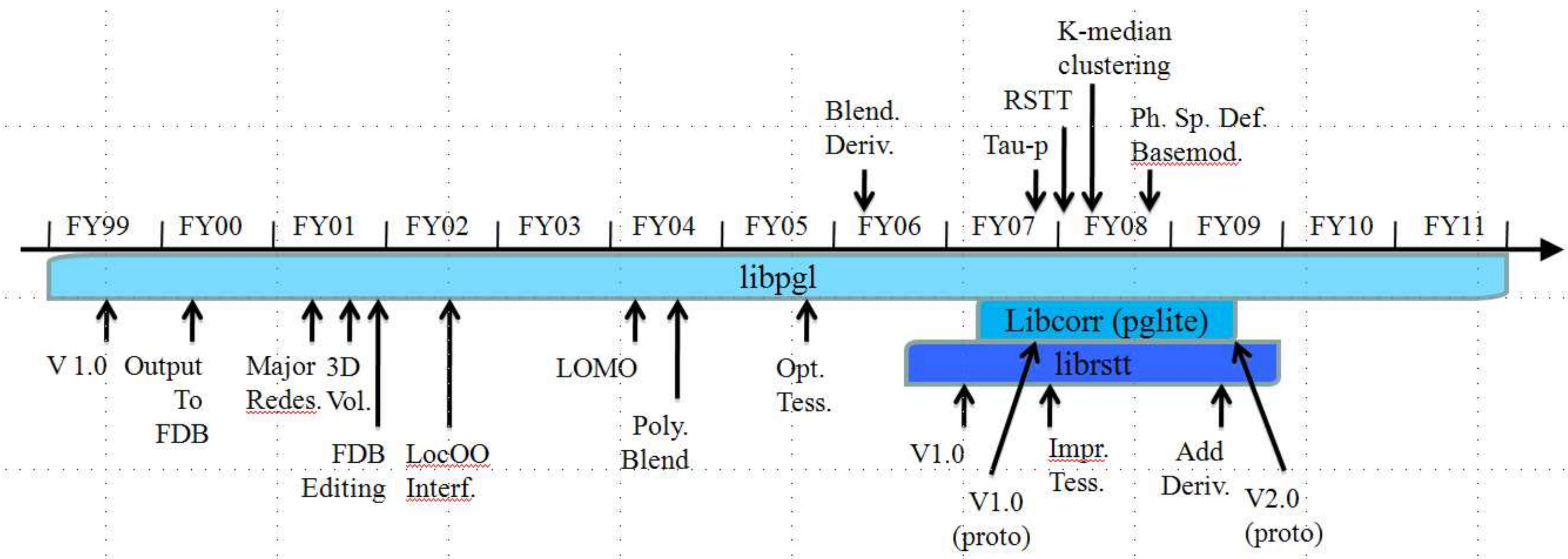
- Libpgl (C++)
  - Our largest, most comprehensive library (includes capabilities for other libraries)
  - Supports development and use of 1D, 2D, 3D models as well as interpolation of corrections surfaces
  - Developed to meet operational requirements for speed, accuracy, and robustness
- Librstt (C++ with C and Fortran interfaces)
  - Calculates travel times through "2.5D" models of crust and upper mantle
- libcorr
  - Stream-lined version of PGL that only does empirical corrections to a 1D or 2.5D basemodel. Corrections are done by linear interpolation off of a pre-calculated tessellation: no kriging-on-the-fly.

What's Next

- Librstt: implement moving least squares to give smooth derivatives across node boundaries

Sandia National Laboratories

API Development History



Event Location

LocOO

Statement of Work

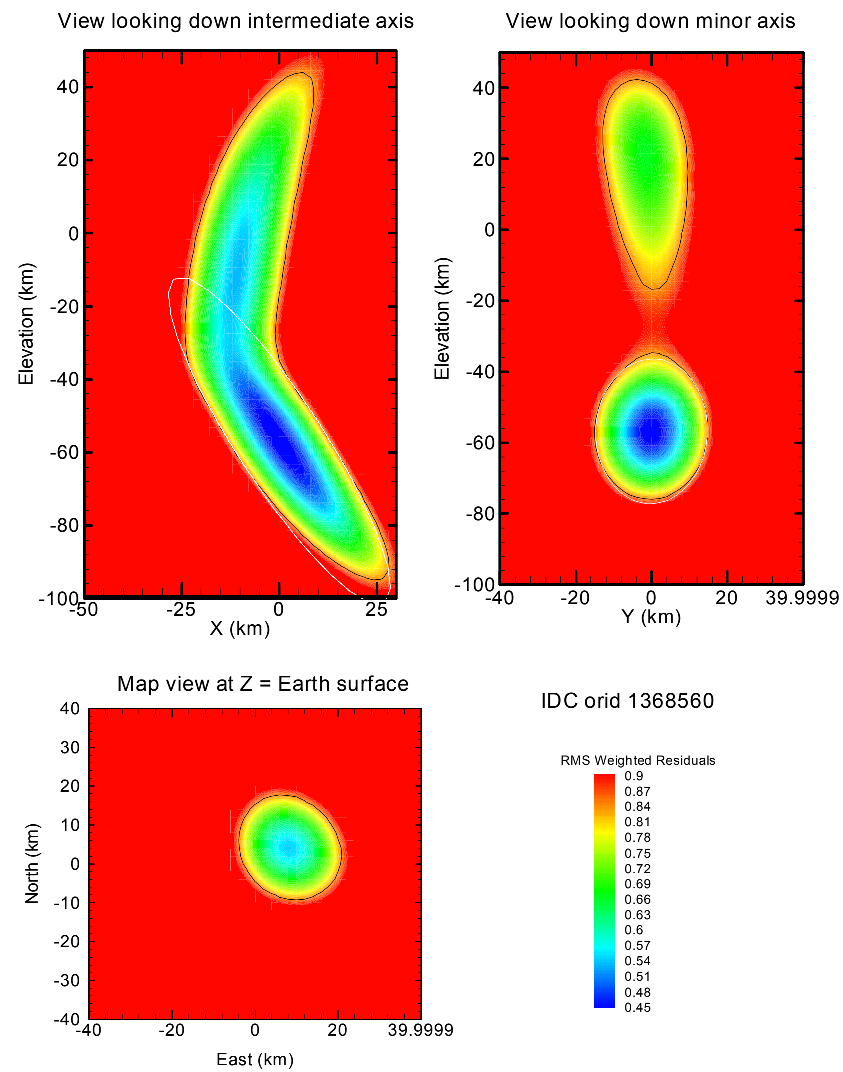
- SNL needs high-quality, robust, modern event location software that we can easily modify to support our monitoring research needs.

Work Accomplished

- SNL developed and continues to refine an object-oriented (C++) locator (LocOO) that meets operational performance requirements and has been shown to produce results equivalent to EvLoc/libloc
- LocOO uses libpgl for travel time, slowness, and azimuth prediction and so has access to all types of models supported by that library (1D, 2D, 3D, correction surfaces)
- Improvements vs. EvLoc/libloc
  - Uses levenberg-marquardt algorithm to deal with problematic non-linearity
  - Can calculate non-linear uncertainty estimates
  - Accounts for correlated observations (non-zero covariances)

What's Next

- LocOO port to Java is nearly complete
- LocOO will be used to quantify improvement in locations due to 3D tomographic models that we are developing



Sandia National Laboratories

Network Capability Assessment

Network Capability Assessment Program (NetCAP)

Statement of Work

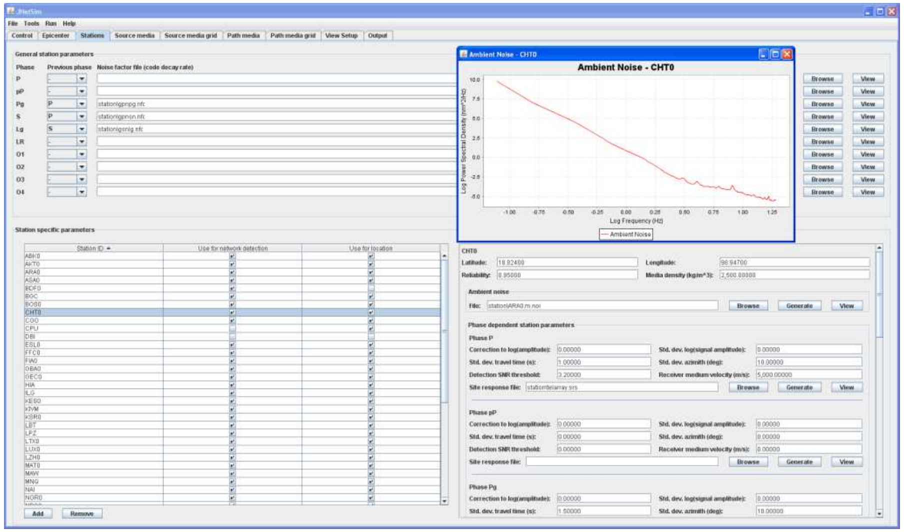
- SNL will develop a new tool based on the NetSim design, but coded in a modern, object-oriented language. Features and architecture will be developed in collaboration with potential users.

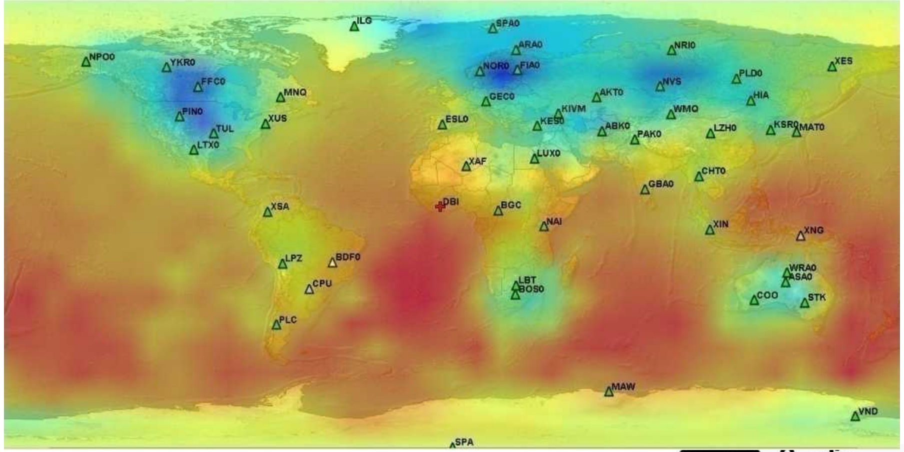
Work Accomplished

- Duplicated NetSim network detection capabilities using existing file formats for backwards compatibility
- Developed GUI for initializing, running, and displaying simulation output
- Created output file formats based on users' needs
- Interfaced with LocOO and libpgl for location uncertainty calculations
- Currently integrating MDAC parameters to improve fidelity of detection simulations
- Investigating extensions for event discrimination and other monitoring technologies

What's Next

- Further refinement of regional amplitude predictions (MDAC continues to evolve)
- Develop simulation model for event identification
- Add high-fidelity infrasound simulation capability (including effects of wind on signal propagation)
- Hydro via Hydrocam.



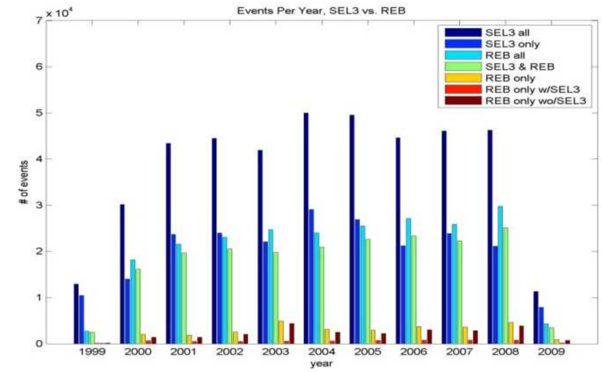


Sandia National Laboratories

Data Mining/Machine Learning

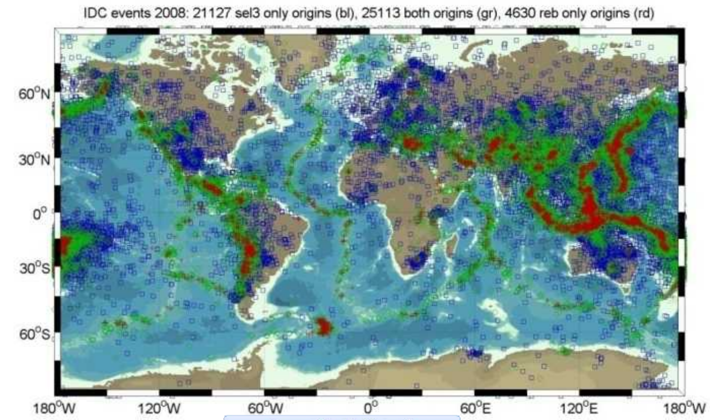
Sandia National Laboratories

Data Mining/Machine Learning Overview



Statement of Work

- Use Machine Learning/Data Mining methods to improve pipeline processing efficiency based on archived data
- Current approach relies heavily on analysts to fix problem events and find missed events

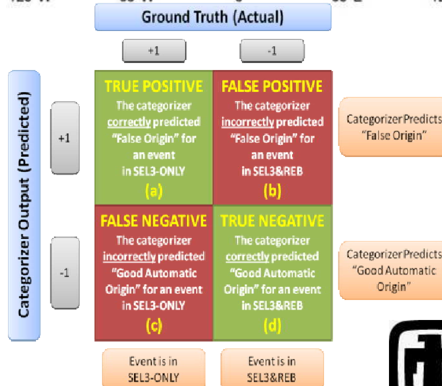


Work Accomplished

- We have acquired the full IDC database (10+) years to support advanced concepts research
- We are analyzing the IDC data using latest Machine Learning/Data Mining methodologies to see if improved algorithms and/or improved calibration can improve quality and efficiency of the automatic processing

What's Next

- Presented analysis of IDC data at ISS
- Will resent analysis of IDC data at MRR
- Submit paper based on IDC data analysis to a journal (early 2010?)
- Acquire waveform data set to extract additional measurements (attributes)



Algorithm	Feature(s) Used	False Negative Rate	False Positive Rate	Accuracy
Baseline (Single-Split)	avg(SNR)	59.5%	9.0%	65.7%
Baseline (Single-Split)	ndef	11.6%	34.5%	76.9%
Decision Tree	ndef & avg(SNR)	13.8%	29.8%	78.2%
Support Vector Machine	All features	14.4%	27.8%	78.9%
Random Forest	All features	15.8%	18.2%	83.1%

Sandia National Laboratories

3D Earth Models

mc1

Overview

Background

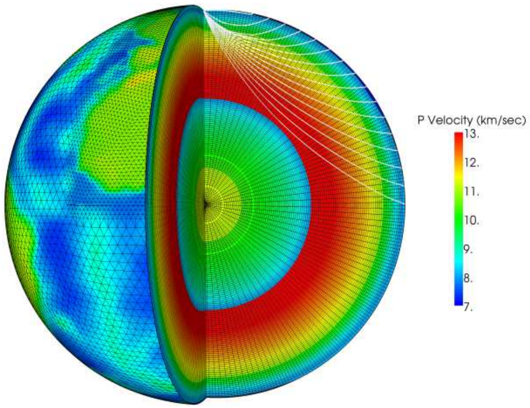
- Over the past decade, we have developed and refined the capability to compute highly accurate travel time, azimuth, and slowness estimates in seismically active areas by interpolating GT-based corrections to an underlying earth model. This empirical corrections lead to better event locations and better uncertainty estimates
  - Currently supported earth models are 1D (e.g. ak135) or 2.5D (e.g. RSTT).
- Complex 3D earth models (that more accurately represent the structure of the earth) would improve monitoring capabilities, especially in areas of low seismicity.

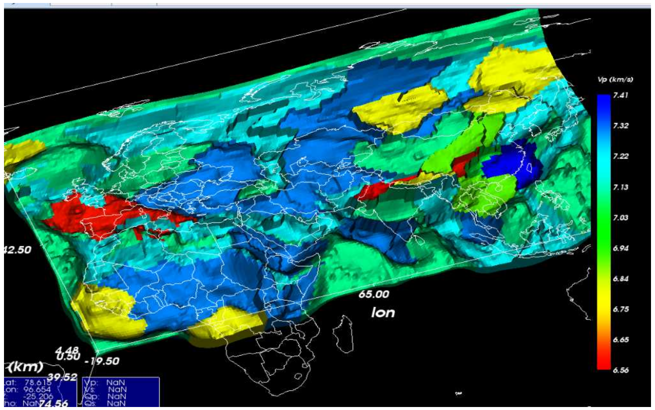
Objectives

- Develop advanced 3D earth models to improve ground-based nuclear detonation detection, location, and characterization.
  - Achieved through: Algorithm development, software development, research, testing, and validation
- Design a framework and implement software to represent 3D earth models, compute models/predictions, and locate events (accurately and quickly).
- Perform R&D to demonstrate how 3D earth models might be used operationally

Tasks

- Architecture Development
- Earth Model Representation
- Algorithm Development
- System Development





Sandia National Laboratories

Slide 14

mc1 The seismological research community has already shifted towards the use of 3D earth models.
Marcus Chang, 4/6/2009

3D Earth Models: Architecture Development

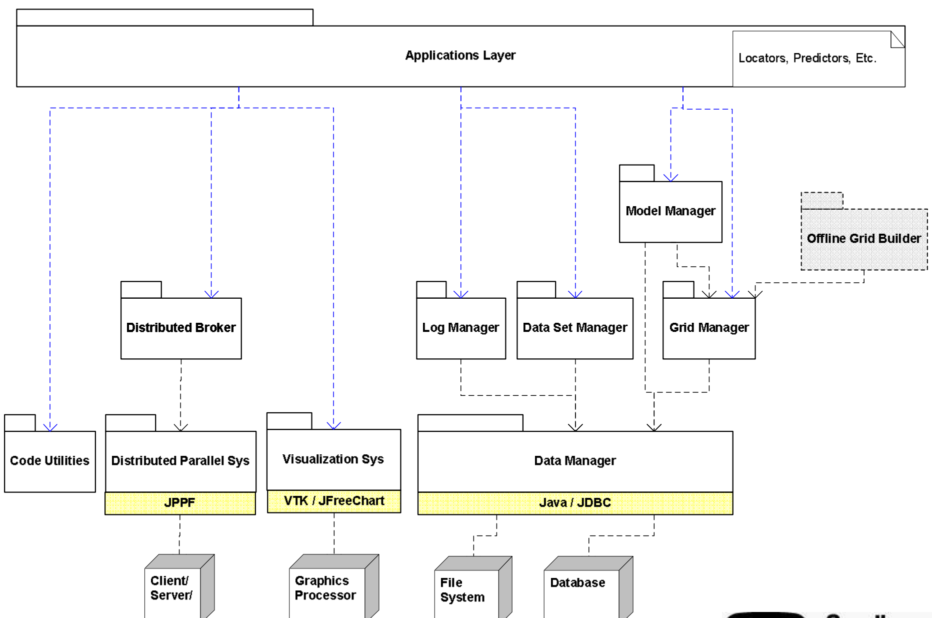
Statement of Work

- SNL will define a software architecture to support the 3D earth modeling research and development efforts.
- A well-designed architecture is critical to ensure the system is easily modifiable to support the latest research approaches.

Modifiability	Quality Attribute Scenario
Source	Developer
Stimulus	Incorporate new algorithm
Artifact	System
Environment	At design time
Response	Make modification without affecting other functionality
Response Measure	Changes are isolated; Time to incorporate existing algorithm is less than one week (if algorithm already supports information needed);

Work Accomplished

- Gathered use cases and requirements
- Developed quality attribute scenarios to identify desirable qualities of the system
  - Modifiability, scalability, performance, testability, usability
- Designed a conceptual architecture using various architectural tactics and patterns

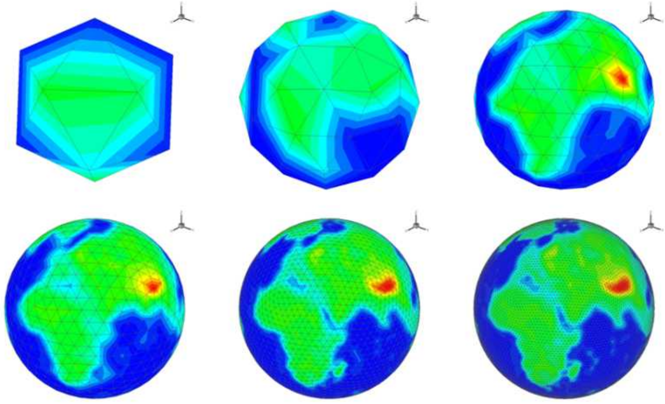


Sandia National Laboratories

3D Earth Models:
Earth Model Representation

Statement of Work

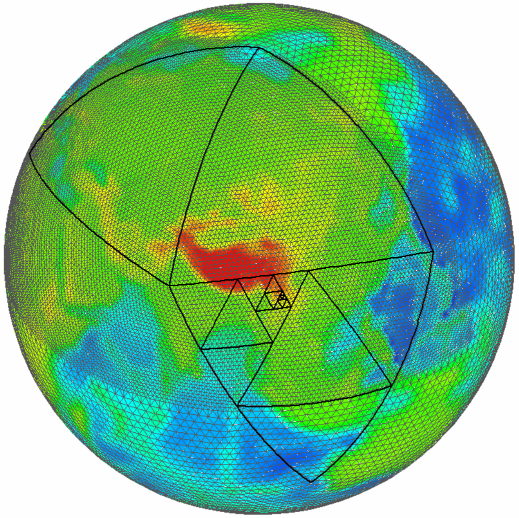
- SNL will develop a comprehensive Earth model representation data structure and storage format that will allow us to represent virtually any class of Earth model that has been developed: regular grid, layered models, unstructured, etc.
  - Internal representation structure will support true multi-resolution models



Variable resolution is achieved by only subdividing triangles where higher resolution is required

Work Accomplished

- Refined multi-resolution earth model representation software library
  - Added variable resolution in the radial direction, which is critical for accurately representing the earth



Sandia National Laboratories

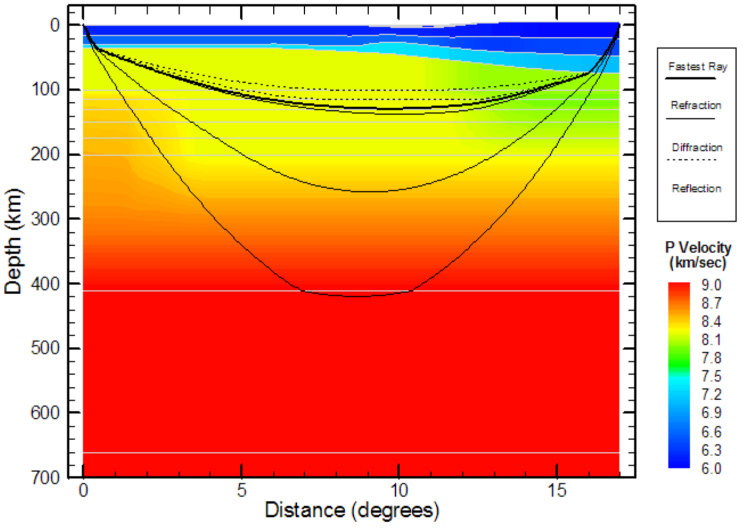
3D Earth Models: Algorithm Development

Statement of Work

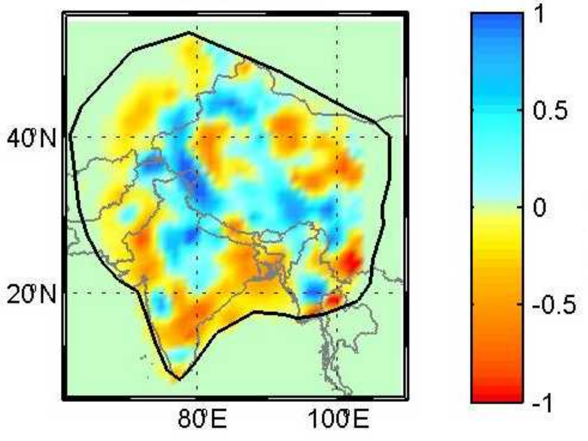
- SNL will research the latest algorithms for prediction, tomographic inversion, and location with 3D earth models. Work will include analysis, implementation, and testing of various algorithms to determine the top algorithms for potential monitoring use.

Work Accomplished

- Prediction algorithms:
  - Implemented and refined a pseudo ray bending algorithm which computes 3D ray paths and travel-times, azimuths, and slownesses for use in tomography
  - Performed comparison tests of the FMM and pseudo-Bender algorithms
    - Determined the pseudo-Bender is the best fit for performing accurate predictions in a timely manner on a parallel system
- Tomography software:
  - Developed tomographic inversion library with predictions performed by the Bender software running on our distributed parallel framework
    - Added variable damping, leave-one-out testing, and optimal regularization capabilities
- Continued to research ways to improve our prediction, tomography, and location algorithms
- Eventual goal is to use our system to develop a global 3D model



Vp at 300 km



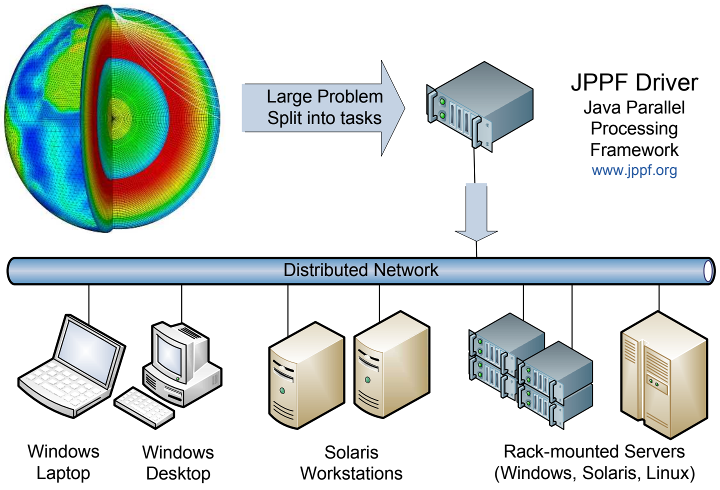
3D Earth Models:
System Development

Statement of Work

- SNL will build and maintain a distributed parallel computing system comprised of commodity multi-core desktops and servers. The system will be designed to process tasks in parallel using the Java Parallel Processing Framework (JPPF).

Work Accomplished

- Developed Python scripts to automatically build and deploy software libraries to every machine on the network
- Designed a new Node Resource Manager system for robustly managing the processing nodes across the system
  - New system enables multiple users to run different versions of the 3D research software simultaneously
- Maintained hardware and expanded the system as opportunities arose



Our current system includes:

*10 * 16-core **Windows** servers, up to 128 GB memory each*

*3 * Dual-core **Linux** servers*

*1 * 8-core **Solaris** workstation*

*7 * 8-core **Windows** desktops, 16 GB memory each*

N Staff desktops

We typically run with ~180 processes running in parallel with 92% efficiency.

Node

Sandia National Laboratories

3D Earth Models: What's Next

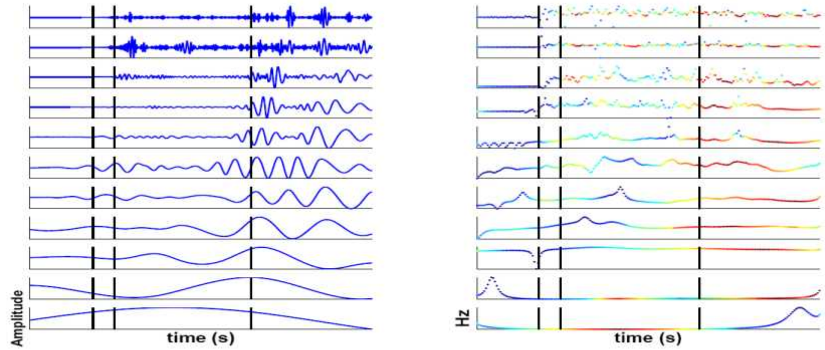
- Enhance system framework for managing processing nodes and algorithms
- Continue to investigate multi-core and GPU commodity hardware for high-performance computing
- Develop a Java-based location capability for use with the Java Parallel Processing Framework
- Create a global 3D tomography model with data from the EHB catalog
- Investigate ways to estimate model uncertainty: analytic vs. stochastic methods
- Add additional algorithms such as multiple dataset inversion, double-difference location, and pre-computed performance-optimized 3D travel-time tables.

Phase Detection and Identification

Phase Detection and Identification Using Advanced Signal Processing Techniques

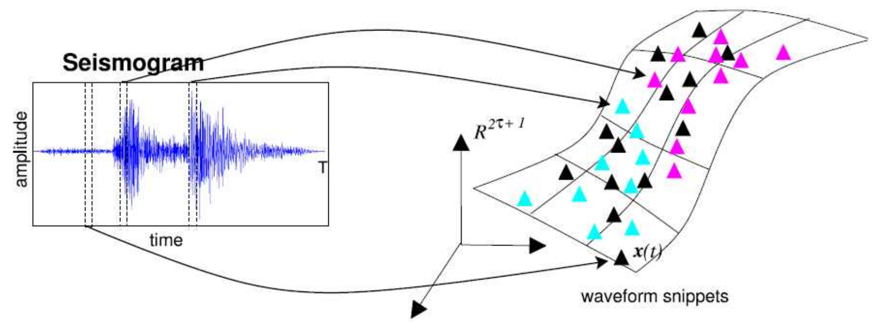
Statement of Work

- SNL seeks to develop improved means to detect and identify seismic signals



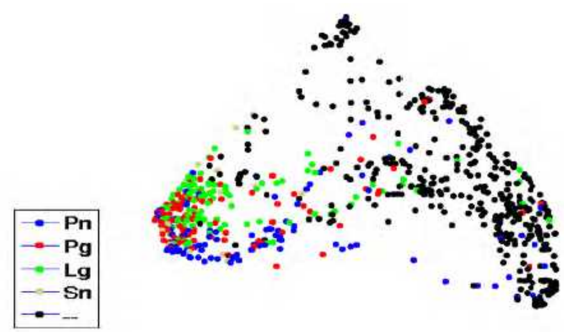
Work Accomplished

- SNL has a small contract with CU-Boulder to investigate applying cutting edge signal processing and Machine Learning algorithms to the signal detection and phase identification problems
- Presented a poster at Spring 2009 SSA



What's Next

- Write up CUB work for a journal article (submit in late August)
- Continue work with CUB
  - Evaluate new techniques against standard algorithms (e.g. STA/LTA) to quantify improvement
  - Test algorithms in several source regions to establish robustness



Sandia National Laboratories

Waveform Correlation

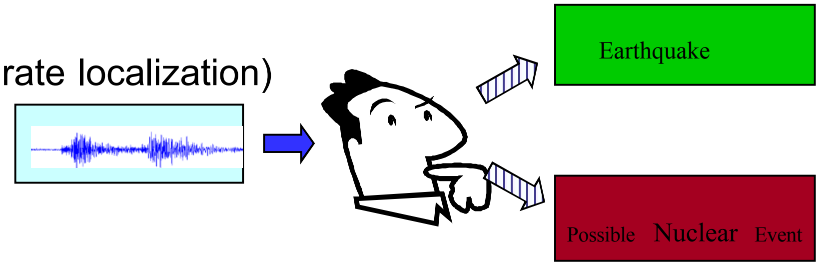
Waveform Correlation Overview

Background

- Monitoring data from a global seismic network for possible nuclear explosions implies detecting and screening hundreds of events per day.
- Depending on the area, a significant portion of the seismicity may be from similar events (same location, same source type)
- Our research seeks to make monitoring systems better able to deal efficiently with similar events, particularly earthquake and mining swarms/sequences, by exploiting the high degree of waveform similarity expected in such scenarios. ("If it looks the same, it is the same")
- Using Waveform Correlation to separate out similar earthquake and mining events allows the analyst to spend more time on possible nuclear explosions.

Objectives

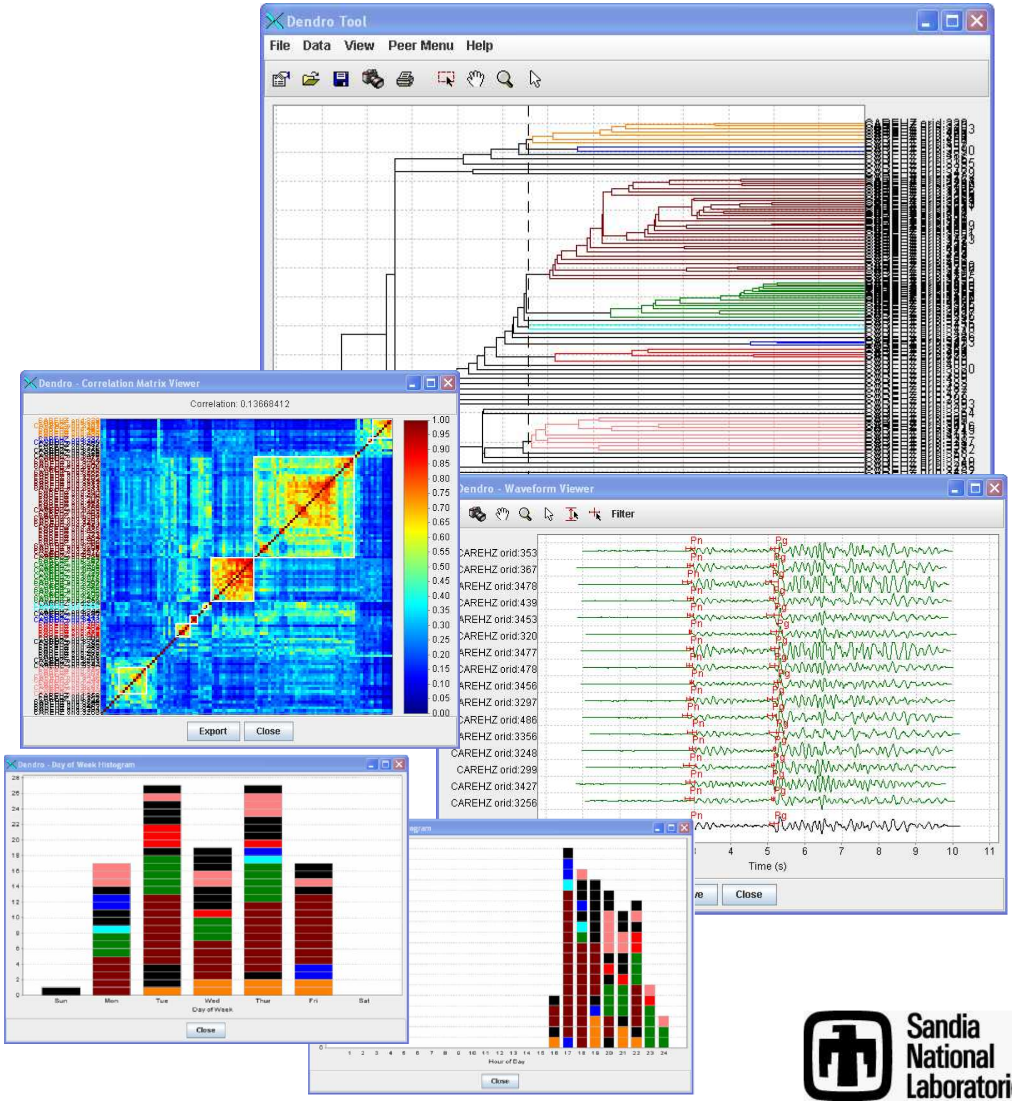
- Develop and implement methods to aid and improve analyst sorting routines:
  - Faster Processing
  - Reduction of detection threshold
  - More accurate phase picks (leads to more accurate localization)
  - Better use of analyst resources



Event Clustering/Identification (Dendro Tool)

Work Accomplished

- Built and extensively tested a Matlab-based tool for identifying similar events using agglomerative hierarchical clustering (dendrograms)
  - Takes as input a set of events
  - Identify clusters based on waveform similarity at one or more stations
  - User decides what level of similarity to use to define families (clusters)
  - Various analysis tools to help assess clusters (maps, histograms, etc.)
- Ported the tool to Java for improved performance and robustness

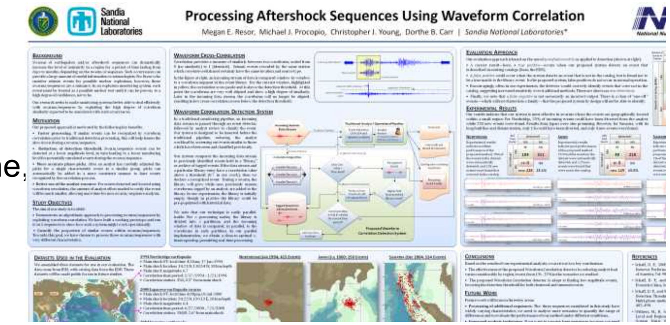


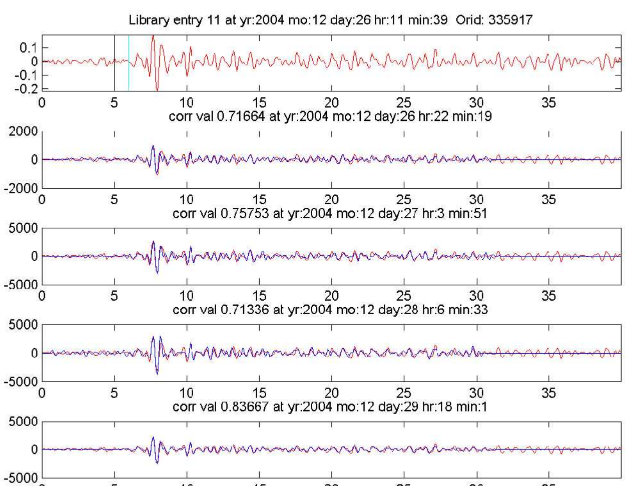
Sandia National Laboratories

Waveform Correlation Event Detection

Work Accomplished

- Implemented a Waveform Correlation system
  - Designed to be inserted before the analyst in the monitoring pipeline, and screen out similar events.
- Processed a week of data from 3 earthquake datasets, and evaluated performance.
  - Reduced analyst load up to 25%
  - Found hundreds of new events not in the catalog (lowering detection threshold).
  - Faster than Real Time speed





Master event

4 events in the next 4 days matched the library entry.

All were later localized to ~5.3 lat, ~94.4 lon

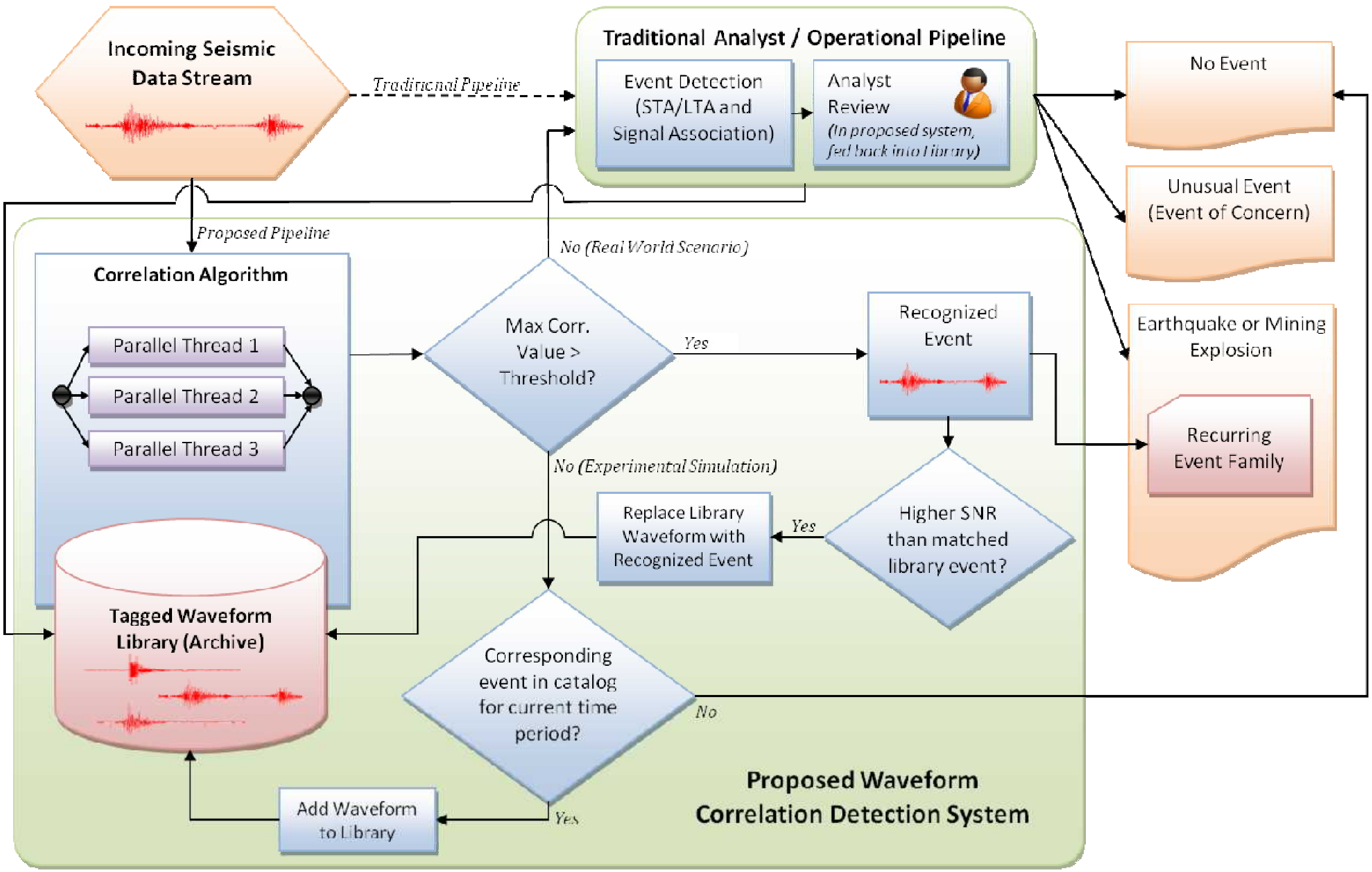
NORTHRIDGE

Reduced analyst load 25%
Found 228 new events

		Detected?	
		YES	NO
In Catalog?	YES	TP 104	FN 311
	NO	FP: 0	
		NEW: 228	

Sandia National Laboratories

Waveform Correlation Event Detection Pipeline



Sandia National Laboratories

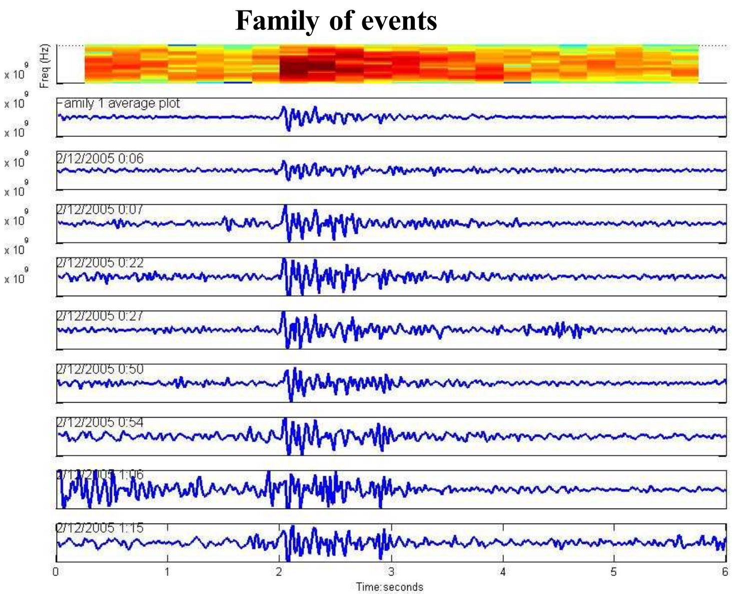
Self Scanner Algorithm

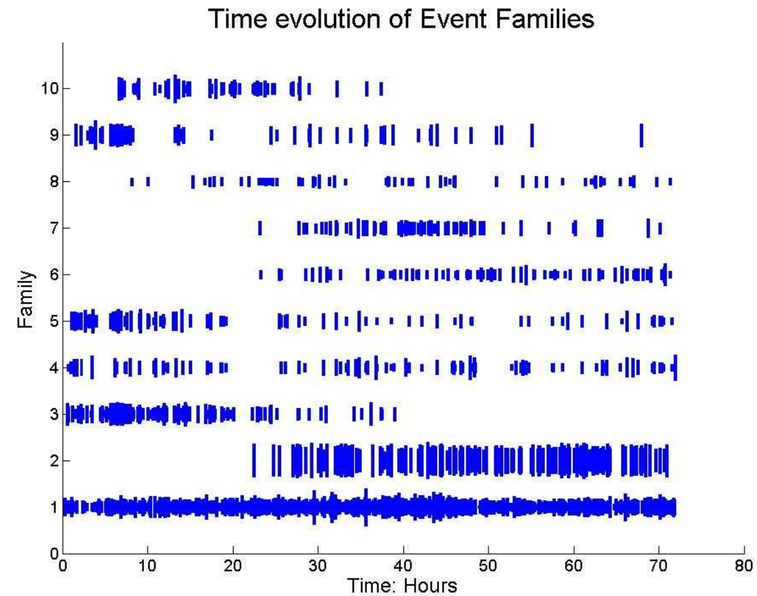
```
1. for t = 0 to t = end
2.       take an N-second window of data D1 (Red window)
3.       for D1 time to end of dataset
4.         take another N-second window of data D2 (Green window)
5.         calculate the correlation coefficient of D1 and D2
6.         if the correlation coefficient >  THRESHOLD
7.             save the correlation value and the location of the matching data segments
8.         slide Green window over 1 sample
9.       shift the Red window N/2 seconds
```

Self Scanner

Work Accomplished

- Implemented a Self Scanning System
  - Designed to find ALL repeated events in a data sequence, without requiring a master waveform
- Processed 4 days of volcanic swarm data and clustered results.
  - Ability to track how clusters change through time
  - Faster than Real Time speed





Waveform Correlation Future Work

•Develop metrics for Correlation Threshold selection

Help the analyst choose a correlation threshold which gives them a given tradeoff between false positive (FP) and true positive (TP) rates. Since the threshold is chosen before an operational system begins running, an educated guess is very important.

•Characterize correlation results vs station distance from event

Understanding how correlation results vary as the station distance changes gives us valuable intuition about the ultimate usefulness of this technique, and knowledge of when results are more/less accurate. Ultimately it will allow us to recommend on which stations waveform correlation should be used for a given event.

•Characterize effect of event magnitude on correlation results

There is some concern in the seismology community that events that differ in 2 or more levels of magnitude will not correlate well. This is likely due to the known low pass filter effect of the earth which increases with magnitude. We examine this hypothesis and determine the effect (if any) magnitude has on our waveform correlation results.

•Apply self-scanner to data sets of direct interest for nuclear explosion monitoring

The utility of the technique has been proven for a volcanic data set, but our interest is for nuclear explosion monitoring. We will analyze some data sequences of interest, such as a recording of a mine collapse and subsequent triggered events.

This work will help get us from a proof of concept to a system which can be used with a high degree of confidence.

Sandia National Laboratories