Talk Abstracts

Nonlinear analysis and forecasting of volcanic inflation and deflation processes

Marian Anghel
Los Alamos National Laboratory

Observations of active silicic volcanoes have shown recurrent inflation and deflation processes that are often cyclical. This type of behavior, which doesn't seem to be inherently stochastic, has been observed in other sciences, including biology, physiology, and economics. In these cases, it has been often discovered that low-dimensional, nonlinear, dynamical systems can explain this type of behavior. Nonlinear time series analysis methods have been devised to characterize the underlying dynamics of the system producing the observed time series just from scalar observations recorded in a single time series.

We use nonlinear time series analysis methods to characterize the recorded tilt data at Santiaguito volcano in Guatemala. An analysis of a week long tilt data using time delay embeddings and correlation dimension estimates, has shown that over the observed time scale the underlying dynamics producing the recorded tilt data has a deterministic, low-dimensional, non-chaotic attractor. This deterministic structure can be used for short-term predictions using a simple time delay embedding approach. Nevertheless, due to the long term memory of the dynamics, a machine learning approach using recurrent neural networks, with internal memory enabling the processing of long input sequences, produces better short-term predictors for the volcano's inflation and deflation process.
Fingerprint and Similarity Thresholding (FAST): A Data-Mining Approach for Earthquake Detection

Gregory Beroza
Stanford University

The Fingerprint and Similarity Thresholding (FAST) earthquake detection algorithm finds small earthquakes in continuous seismic data through uninformed similarity search (Yoon et al., 2015). FAST does not assume prior knowledge of templates nor does it use labeled examples of earthquake waveforms, rather it treats earthquake detection as a data-mining problem. FAST extracts a set of features, called fingerprints, which are compact binary representations of short-duration time windows that span long-duration data sets. We design fingerprints to be discriminative such that similar waveforms produce similar waveform fingerprints, and fingerprints corresponding to noise, which dominates most seismic data sets, have low similarity. FAST uses locality-sensitive hashing to index the data, and queries the index to identify similar fingerprints. FAST outperforms naive, brute force similarity search by carrying out approximate similarity search that identifies similar waveforms with high probability. The improved scalability that results allows us to search up to a decade of continuous data. Because we are detecting weak signals in very long duration data sets, we are susceptible to false detections due, for example, to sources of persistent noise. To address the false detection problem, we developed a method for extending single-station similarity-based detection over a network (Bergen and Beroza, 2017). We designed pair-wise pseudo-association to leverage the pair-wise structure of FAST output. Unlike the association typically carried out for earthquake detection, pseudo-association does not explicitly account for move-out. Instead, we exploit the fact that the relative arrival time of a pair of events at two different stations will be the same at all receivers. Pairwise pseudo-association and the supporting techniques, event-pair extraction and event resolution, complete a post-processing pipeline that combines single-station similarity measures from each station in a network into a list of candidate events. We have applied network-FAST to the Iquique, Chile foreshock sequence and found that it is sensitive and maintains a low false detection rate: we identify nearly five times as many events as are present in the local seismicity catalog (including 95% of the catalog events), and less than 1% of these candidate events are false detections.
Locally-sparse travel time tomography: sparse modeling of slowness patches and dictionary adaptation by unsupervised machine learning

Michael Bianco
University of California San Diego NoiseLab, Scripps Institution of Oceanography

We present a 2D travel time tomography method [1] which regularizes the inversion by modeling rectangular groups of slowness pixels from discrete slowness maps, called patches, as sparse linear combinations of atoms from a dictionary. These dictionary atoms can be considered ‘elemental’ slowness patches, which represent well the features in all the patches. We propose to learn the atoms in parallel with the inversion using dictionary learning, a form of unsupervised machine learning. Thus the dictionaries are adapted to features in specific slowness maps, which can be smooth and discontinuous (see Fig. 1).

This patch regularization, which we call the local model, is integrated into the overall slowness map, called the global model. Whereas the local model considers small-scale variations using a sparsity constraint, the global model considers larger-scale features which are constrained using L2-norm regularization. This local-global modeling strategy with dictionary learning has been successful for image restoration tasks such as denoising and inpainting [2], where diverse image content is recovered from noisy or incomplete measurements.

We use this local-global strategy in our locally-sparse travel time tomography (LST) approach to model simultaneously smooth and discontinuous slowness features. This is in contrast to conventional tomography methods [3,4], which constrain models to be exclusively smooth or discontinuous. We develop a maximum a posteriori formulation for LST, which is solved as an iterative inversion algorithm until convergence is achieved. We demonstrate the LST approach on densely, but irregularly sampled synthetic slowness maps and compare LST results with a conventional tomography method [4]. Synthetic slownesses and inversion results without travel time error are shown in Fig. 1.

References:
Characterizing precursory phenomena in laboratory stick-slip failure events using unsupervised machine learning

Chas Bolton
Pennsylvania State University

Earthquake forecasting is an important problem for mitigating seismic hazard and illuminating the physics of earthquake nucleation. However, progress in this area depends strongly on the ability to track and characterize seismic precursors, which is challenging even under laboratory conditions. One approach is to monitor frequency-magnitude statistics of acoustic emissions (AEs) throughout the laboratory seismic cycle. However, this technique has traditionally relied on the development of AE event catalogs, which requires thresholding and other methods to identify events. Catalog construction introduces biases and other unknown effects associated with filtering the time series. Therefore, it is not unreasonable to assume that valuable information is lost and along with that key insights into the earthquake nucleation process.

Here, we propose that by employing machine learning (ML) and using the entire time series of elastic radiation we can better characterize the evolution of AE precursors during the seismic cycle compared to event catalog-based approaches. We use data from a suite of friction experiments to study laboratory earthquakes using a servocontrolled biaxial deformation apparatus. Experiments were conducted in the double-direct shear configuration where two fault zones are sandwiched between three rigid forcing blocks. Normal stress on the faults was maintained constant at 2 MPa and a shearing velocity of 10 microns/s was imposed at the fault zone boundaries. Acoustic time series signals were recorded continuously at 4 MHz from two p-polarized piezoelectric transducers located adjacent to the fault. To investigate acoustic precursors, we first computed 42 statistical features of the continuous time series signal using a moving window. Each time window was 1.36 (s) and overlapped the previous window by 90%. The ML mean-shift clustering algorithm was then applied to the statistical features in order to find structure within the time series signal.

Our results show a systematic evolution in the acoustic properties as the fault approaches failure. Specifically, we identify five unique clusters in the stick-slip cycle and find that each cluster is associated with a specific period of the seismic cycle. In addition, our results suggest that the key information is contained in just a few statistical features, including the acoustic variance, kurtosis, and skewness. Using these features, we identify systematic changes in the clusters throughout a stick-slip cycle. This work suggests that basic statistical features of the time series signal evolve systematically as the fault approaches failure and provide important precursory information about upcoming failure events. Our results dovetails nicely with additional work conducted by our group showing that failure time of lab earthquakes can be predicted using ML. We show how an improved understanding of the microphysical mechanisms that contribute to changes in acoustic properties during the seismic cycle can be used to illuminate the processes responsible for precursors to failure and prediction of lab earthquakes.
Clustering P-Wave Receiver Functions To Constrain Subsurface Seismic Structure

Chengping Chai
University of Tennessee

The acquisition of high-quality data from permanent and temporary dense seismic networks provides the opportunity to apply statistical and machine learning techniques to a broad range of geophysical observations. Lekic and Romanowicz (2011) used clustering analysis on tomographic velocity models of the western United States to perform tectonic regionalization and the velocity-profile clusters agree well with known geomorphic provinces. A complementary and somewhat less restrictive approach is to apply cluster analysis directly to geophysical observations. In this presentation, we apply clustering analysis to teleseismic P-wave receiver functions (RFs) continuing efforts of Larmat et al. (2015) and Maceira et al. (2015). These earlier studies validated the approach with surface waves and stacked EARS RFs from the USArray stations. In this study, we experiment with both the K-means and hierarchical clustering algorithms. We also test different distance metrics defined in the vector space of RFs following Lekic and Romanowicz (2011).

We cluster data from two distinct data sets. The first, corresponding to the western US, was by smoothing/interpolation of receiver-function wavefield (Chai et al. 2015). Spatial coherence and agreement with geologic region increase with this simpler, spatially smoothed set of observations. The second data set is composed of RFs for more than 800 stations of the China Digital Seismic Network (CSN). Preliminary results show a first order agreement between clusters and tectonic region and each region cluster includes a distinct Ps arrival, which probably reflects differences in crustal thickness.

Regionalization remains an important step to characterize a model prior to application of full waveform and/or stochastic imaging techniques because of the computational expense of these types of studies. Machine learning techniques can provide valuable information that can be used to design and characterize formal geophysical inversion, providing information on spatial variability in the subsurface geology.
Deep Learning for Making Sense of Ambient Seismic Noise

Julien Clancy
Massachusetts Institute of Technology

We apply recent advances in deep neural networks to three classes of geophysical problems stemming from ambient noise imaging: wavespeed inversion in homogeneous media in the presence of anisotropic sources, local wavespeed inversion in inhomogeneous media, and source directionality estimation in homogeneous media. Our networks are inspired by those commonly in the signal processing literature, such as convolutional networks and LSTMs, but use a training procedure that appears unique to physical problems: data is generated on the fly and only used to compute a single gradient, then discarded and never seen again. These techniques prove to be highly performant and quite flexible—they easily accommodate for data gathered from different sensor geometries, or for different priors in the data generating procedure. We also find preliminary evidence that, in simplified analogues of these problems, the nodes in deeper layers of our networks are computing physically meaningful quantities. One caveat: all our numerical experiments use simulated data, not real data.
Mining application of machine learning for failure prediction

Tjaart de Wit
Institute of Mine Seismology

Machine Learning is an exciting new method in data analytics that is used to devise complex algorithms to predict future behaviour. A recent result from Los Alamos National Laboratory revealed that supervised Machine Learning can predict failure in laboratory earthquakes by examining raw continuous data[1]. Many underground mines have dense and modern seismic monitoring systems that provide a natural laboratory that can bridge the gap between laboratory stick-slip experiments and crustal seismology.

We apply this technique to continuous data recorded at an underground block-caving mine in Australia and examine if this can be used to predict large mining induced events. Continuous seismic data were recorded on more than 20 tri-axial short-period geophones for 1 year. During this time, more than 40,000 micro-seismic events (moment magnitude between -2.0 and 3.0) were processed and located. We expand upon the set of features used in the the random forest model, outlined by [1], to determine which of these features dominate at different scales of time-to-failure in this intermediate setting between a laboratory and crustal scale.

References:
A Convolutional Neural Network for intermediate-depth earthquake detection

Manuel Florez
Massachusetts Institute of Technology

Earthquake detection remains one of the fundamental operations in observational seismology. Despite the explosion in quality and quantity of seismic data our ability to build dense and complete seismicity catalogs remains limited. The classical techniques, routinely used at seismological observatories around the world, rely on high signal-to-noise ratios of the incoming waveforms and, thus, are only able to pick out the larger events. Detection thresholds of about Mw 3.0 are common. More recent approaches exploit the self-similarity of earthquakes and rely on using the waveforms of a few well known earthquakes as templates. Terabytes of continuous seismic signals at multiple stations are cross-correlated against these templates and thousands of events with high waveform similarity may be easily identified. This procedure, known as template matching, is computationally expensive and fails for events whose waveforms are significantly different from the templates. Lack of completeness in earthquake catalogs is a more severe issue for intermediate-depth (80-300 km) and deep focus earthquakes (300-700 km). Event-station distances are larger than for typical crustal earthquakes and signal-to-noise ratios for even moderate events (Mw 3.0 to Mw 4.0) are considerably smaller. We propose a Convolutional Neural Network (CNN) architecture for the detection of intermediate-depth earthquakes and the precise picking of p- and pP- wave arrival times. We test our implementation using a synthetically generated dataset and propose a training scheme that leverages both real and synthetic data for optimal results. As an example we apply our technique to an intermediate-depth cluster in northern Chile. We are able to detect 8 times more earthquakes than in the initial catalog and use the picked arrival times to relocate the events. We clearly resolve a double-planed structure at depth. We also compare our approach with the template matching technique; when events are similar to the templates both methods yield comparable results, however our CNN was able to detect a new set of 621 small magnitude events that correlated poorly with the templates.
Global Sensitivity Applied to Dynamic Combined Finite Discrete Element Methods for Fracture Simulations

Humberto Godinez
Los Alamos National Laboratory

Fracture propagation play a key role for a number of application of interest to the scientific community. From dynamic fracture processes like spall and fragmentation in metals and detection of gas ow in static fractures in rock and the subsurface, the dynamics of fracture propagation is important to various engineering and scientific disciplines. In this work we implement a global sensitivity analysis test to the Hybrid Optimization Software Suite (HOSS), a multi-physics software tool based on the combined finite-discrete element method, that is used to describe material deformation and failure (i.e., fracture and fragmentation) under a number of user-prescribed boundary conditions. We explore the sensitivity of HOSS for various model parameters that inuence how fracture are propagated through a material of interest. The parameters control the softening curve that the model relies to determine fractures within each element in the mesh, as well a other internal parameters which in fracture behavior. The sensitivity method we apply is the Fourier Amplitude Sensitivity Test (FAST), which is a global sensitivity method to explore how each parameter inuence the model fracture and to determine the key model parameters that have the most impact on the model. We present several sensitivity experiments for different combination of model parameters and compare against experimental data for verification.
Inverse Modeling in the Oil and Gas Industry as a Machine Learning Problem

Thomas Halsey
ExxonMobil Upstream Research Co.

Traditionally, model-based decision support for oil and gas production is achieved through history-matching reservoir models to historical production data. However, history matching is an ill-posed problem; furthermore, it is cumbersome to create detailed reservoir models through history matching when only the corresponding flow predictions and their uncertainty drive business decisions. We propose an alternative approach, combining data-driven and reservoir modeling methods to directly infer a “quantity of interest” relevant to a decision. The method relies on the availability of an ensemble of reservoir models corresponding to a plausible prior probability distribution of these models, combined with machine learning methods to infer the quantity of interest corresponding to an actual decision. We demonstrate this approach on a three-dimensional model based on a real oil field, where the decision to be made is to drill (or not) an infill well.
Audio-based unsupervised machine learning reveals cyclic changes in seismicity mechanisms in the Geysers geothermal field, California

Ben Holtzman
Lamont Doherty Earth Observatory, Columbia University

The earthquake process reflects complex interactions of stress, fracture and frictional properties. New unsupervised machine learning methods reveal patterns in time-dependent spectral properties of seismic signals and enable identification of changes in faulting processes. Our methods are based closely on those developed for music information retrieval and voice recognition, using the spectrogram instead of the seismogram. The spectrograms are first decomposed using non-negative matrix factorization (NMF) to characterize frequency patterns, followed by hidden Markov modeling (HMM) of the temporal patterns of frequency changes, yielding low dimension fingerprints for each event. These fingerprints are then clustered using the k-means algorithm. The NMF and HMM act to extract features that are common to all signals, enabling identification of subtle differences among each signal.

Clustering of 46,000 earthquakes of $0.3<M_L<1.5$ from the Geysers Geothermal Field (CA) yields groupings of events that have no reservoir-scale spatial patterns, but clear temporal patterns. Events with similar spectral properties repeat on annual cycles within each cluster and track closely the water injection rates into the Geysers reservoir, indicating changes in the reservoir acoustic properties and faulting processes with changes in thermo-mechanical state. Listening tests confirm that the clusters represent subtle but consistent spectral variations among seismic signals. Other studies using detailed analysis of waveforms have identified systematic changes in moment tensor and stress drop associated with the water injection rate history, on smaller sets of larger events in sub-regions of the Geysers. These studies motivate our interpretation of the temporal clustering as being due to changes in seismicity mechanisms, affected by fluid injection rates. These methods open many possibilities for identification of subtle changes in seismicity for earthquake forecasting, and also for real-time identification of changes in geothermal reservoir seismicity mechanisms, for improving the efficiency of energy production.
Estimating the Physical State of a Slow Slipping Fault from Seismic Signals

Claudia Hulbert
Los Alamos National Laboratory

Over the last two decades it has become apparent from strain and GPS measurements, that slow slip on earthquake faults is a widespread phenomenon. Slow slip is also inferred from small amplitude seismic signals known as non-volcanic tremor from low frequency earthquakes (LFEs) and has been reproduced in laboratory and simulation studies, providing useful physical insight into the frictional properties associated with the slip. By applying machine learning to the acoustic signal originating in the fault, we show that in the laboratory these acoustic emissions are a fingerprint of the fault frictional state.

We find that certain characteristics of the acoustic signal are tightly linked to fault friction, and that these same characteristics allow us to determine the fault displacement, as well as upcoming slow slip failure initiation and slip termination, directly from the acoustic signal. We can also estimate the upcoming event magnitudes by a second machine learning procedure based on predicted inter-event times. By applying machine learning approaches to continuous seismic data, new insight into the physics of faulting could be obtained in Earth, with the potential to improve earthquake hazards forecasting and earthquake real time warning.
Automated ambient noise processing applied to fiber optic seismic acquisition (DAS)

Fantine Huot
Stanford University

Distributed acoustic sensing (DAS) is an emerging technology used to record seismic data that employs fiber optic cables as a probing system. By measuring the backscattered energy of a pulsing laser transmitted down a fiber optic cable, it is possible to measure the strain rate occurring within different sections of the cable. DAS recording systems have been shown to measure data comparable with conventional geophones and have been successfully used in exploration and earthquake seismology settings.

Recently, a DAS array has been deployed beneath Stanford campus in the existing fiber optic telecommunication conduits. Because we can so easily use our telecomm infrastructure for continuous, dense, urban seismic acquisition, data collected in such a manner will go to waste unless we significantly automate ambient noise processing. Herein we introduce relevant data features for exploratory data analysis and use clustering algorithms to quickly identify coherent, repeating noises which inhibit reliable extraction of useful signals. We then train a convolutional neural network for detecting traffic noise and selectively filter it out of the data in order to generate ambient seismic noise fields that are suitable for interferometry purposes. We use Markov decision processes to reconstruct the array geometry from the data, which gives us the potential to extend this type of acquisition to other existing fiber optic networks in urban areas. This opens up the possibility of easily plugging into unused fibers in telecom bundles in areas where traditional geophones can not be deployed.
Deep Learning Seismic Tomography

Joseph Jennings
Stanford University

Velocity model building (VMB) is a key step in hydrocarbon exploration; The VMB main product is an initial model of the subsurface that is subsequently used in seismic imaging and interpretation workflows. Reflection or refraction Tomography and full waveform inversion (FWI) are the most commonly used techniques in VMB. On one hand, Tomography is a time-consuming activity that relies on successive updates of human-based analysis of seismic gathers. On the other hand, FWI is highly computationally demanding, without global convergence guarantees. We propose and implement [1] a novel concept that bypasses these demanding steps, directly producing an accurate gridding or layered velocity model from shot gathers. Our approach relies on training a deep neural network (DNN); and the resulting predictive model maps relationships between the data space and the final output (particularly, the presence of high velocity segments that might indicate salt formations). In terms of computation time, the training task takes several hours for 2D data, however the inference step (predicting a model from previously unseen gather) takes only seconds. Feature extraction is a key step in our workflow as it can greatly improve the training of the DNN by providing it with the most relevant data for learning. Specifically, we perform velocity analysis and provide semblance panels for different common midpoint (CMP) locations as the input feature. To calculate the semblance panel for a given midpoint, we first apply a normal moveout correction to a CMP gather. The set of semblance panels composes a semblance cube, which is the input feature to the DNN, as depicted in the figure (left). The DNN is composed of convolutional and fully-connected layers, and its output is a predicted velocity model. Predicted images quality evaluation with 10,000 models (80% train and 20% test) provide an average SSIM [2] of 0.8717 , relative to the ground truth models, which clearly indicates very high visual quality. Given the flexibility and modularity of our approach, we foresee new Geoscience applications, beyond hydrocarbon exploration.
Waveform classification using statistical learning algorithms to characterize data features from a dense deployment of geophones atop an active fault zone

Christopher Johnson
University of California San Diego

The objective of this study is to develop data based classifiers for earthquakes, microseisms, and ambient noise with a compressed parameterization of the waveforms and utilize statistical learning techniques to train a model to identify previously undetected signals in upper portions of the crust. The detection of shallow microseismic events requires the careful distinction between shallow seismic energy and local high-frequency ambient noise or atmospheric processes. The classification procedure involves manually identifying and labeling signals present in subsets of the data and use them for statistical model training. The process requires the calculation of features as scaler metrics, i.e. hundreds of temporal and spectral properties of the time series, that represent a compact expression of the data. The scaler metrics are derived using seismological knowledge of waveforms to maintain a physics based representation of the data. Each waveform is labeled to represent a specific source type, e.g. a local or regional earthquake, anthropogenic machines, atmosphere coupling, etc., and this requires exploratory data analysis to properly define the required features for each classification type and properly weigh them for increased model performance. The labeled features are used with supervised learning techniques to train an advanced statistical model to classify data using only these compressed features.

In May and June 2016 a dense array deployment of >1,100 vertical geophones recorded ground motions at the San Jacinto Fault Zone for 5 weeks at the same location as a borehole seismometer and strainmeter. The data set provides a unique opportunity to characterize the shallow subsurface from 0-1 km depth and fully describe the fault zone environment from the surface to the brittle-ductile transition zone at ~18 km depth. Here, we are using the >1,000 hand-picked earthquakes from the dense deployment data as waveform templates to develop the compact features. By creating waveforms using a subset of the dense array we obtain >10,000 samples for training the models. We then use the Support Vector Machine and Random Forest classification algorithms with the labeled data features. The features are evaluated for significance in the model performance and those not contributing are removed. The beginning stages of the 2-year project are underway and the waveform classification is at 100% recall for the simple case of earthquake and noise in a binary classification framework. When labeling near-field and regional waveforms recorded by the dense array the recall drops to ~80%. The inclusion of high-frequency atmosphere coupled signals also produces a lower recall of ~75%. These preliminary results are providing new information to redesign the data features to improve model recall. The next step following successful training is to apply the model to the continuous data in the dense deployment and search for undetected events. The project is ambitious in its nature and the expected outcomes will provide new scientific insights to processes in shallow crustal layers. In addition, we aim to develop a robust set of algorithms that are appropriate for use with dense temporary geophone deployments as well as regional broadband networks.
The role of machine learning in building a global smartphone seismic network

Qingkai Kong
UC Berkeley

MyShake is a global crowdsourcing smartphone seismic network to monitor and detect earthquakes. After it released to the public in 2016, we arrived at more than 270,000 downloads with more than 700 detected earthquakes globally within 1.5 years. Machine learning plays a critical role in MyShake that makes everything happen. In this talk, I will present the details of how we use the artificial neural network to distinguish earthquakes from the human activity movements recorded on a single phone in real-time. This includes how we do the data acquisition, pre-processing the data, addressing imbalanced datasets, feature engineering/selection, and evaluating the model. I will also talk the convolutional neural network (CNN) we built that run on the server to further classify the whole waveforms to find that caused by earthquakes. This includes how we cast the earthquake waveforms into images that suitable for the CNN to learn the difference, and some ideas of apply transfer learning on to it. In the end, I'd like to share some of my thoughts on applying machine learning in geosciences.
Generating Seismograms with Deep Neural Networks

Lion Krischer
ETH Zurich

Synthetic seismograms are paramount for many applications in seismology and various approximate techniques for their calculation have been developed over the decades; each with their own shortcomings and trade-offs. Realistic numeric calculations through three dimensional media at high frequencies on the other hand are computationally too expensive to run routinely and will remain so for the foreseeable future.

Seismic velocity models at global and continental scales are comparatively smooth and seismograms at adjacent stations, or from similar sources, vary little. Given enough training data, deep neural networks might be able to learn some underlying function or representation to quickly generate synthetic seismograms for unseen sources and new receivers. Possible applications for this are numerous and range from probabilistic source inversions, over optimal experiment design, to seismic hazard estimations.

We report the design and rational of a deep neural network architecture made up of a series of fully connected layers converting the inputs to an intermediate representation. A number of transposed convolutional layers follows to produce the output. Inputs to the network are source and receiver parameters, its outputs are the numeric values of the synthetic seismograms.

Successful applications to three dimensional complete seismograms including body and surface waves as well as physics like anisotropy and attenuation, but with a radially symmetric velocity model, are shown. Unlimited training as well as validation data is available for that case. Furthermore we present first experiments of training a neural network on data calculated through a three dimensional velocity model with severe restrictions on the available data.
Robust Learning with Noisy Labels in Multispectral Source Rock Characterization

Weichang Li
Aramco Services Company, Aramco Research Center - Houston

Performance of machine learning algorithms can be severely limited by the quality of the data and labels used in training. This is especially the case with scientific and engineering applications where collecting data and generating labels can be costly and error prone, especially if there is a lack of adequate domain knowledge. In geoscience applications, examples like this are abundant, including micro-scale geological sample classification and field-scale seismic structure identification. As the interest in applying machine learning to these problems grows rapidly, a common practice is to rely on human inputs or existing rule-based workflows to generate labels, both can be subjective and inaccurate. Consequently, blindly applying machine learning techniques without addressing label quality issues can produce misleading results that are subject to the same limitations as the label generating process. In this paper we propose a robust machine learning framework that combines supervised with unsupervised learning methods. Specifically we formulate the learning objective as a combination of the label prediction error penalty associated with supervised learning, a regularization term defined over a certain metric distance space for which the selected unsupervised learning method is defined, and a spatial regularization in the form of a Markov Random Field (MRF). We apply this technique to the problem of source rock characterization. As shown in Figure 1, the advantages of this robust approach include i) improving the prediction confidence level where it was low; ii) identifying mislabeled cases and using the corrected labels to obtain more accurate predictions.
Machine learning for laboratory earthquakes using catalogues of varying fidelity

Nicholas Lubbers
Los Alamos National Laboratory

Machine learning using continuous-waveform data has led to remarkable prediction of the mechanical state of laboratory earthquake machines. Here we show that we can achieve comparable accuracy using state-of-the-art instrumentation to generate a high-quality catalog. Furthermore, we measure how the prediction accuracy degrades as a function of the magnitude of completeness of the catalog. This points towards portable machine learning methodology that could apply to catalogs from natural faults.
Earthquake Mechanics and the Role of Machine Learning in Illuminating Fault Zone Processes During the Seismic Cycle

Chris Marone
Pennsylvania State University

Earthquake source theory is grounded in the mechanics of frictional instability and the dynamics of expanding rupture patches. Earthquake source spectra and scaling laws can be described by propagating dislocations for which seismic slip scales with earthquake source dimension. Slip patch dislocation mechanics also provides a framework for investigating earthquake nucleation by relating slip and frictional weakening to local elastic stiffness. Such nucleation models together with laboratory-derived friction constitutive laws indicate that failure should be preceded by observable changes in physical properties, associated with slip and deformation within the nucleation zone. Moreover, a wide range of lab data show that failure is always preceded by fault slip, the magnitude of which scales with the active fault zone width. Thus, from a wide range of viewpoints, earthquake faults should exhibit a systematic, and perhaps predictable, evolution during the seismic cycle. Laboratory observations of repetitive stick slip failure support this view, and recent works document the role of machine learning (ML) in predicting the timing of failure events. ML techniques have been used successfully on continuous records of acoustic emission to identify the factors that allow prediction of laboratory earthquakes. Parallel efforts are underway to use ML on active source measurements of fault zone elastic properties. The potential impact of these efforts can benefit from a focus on fault zone structure and the evolution of physical processes during the seismic cycle.

Recent works on granular and clay rich laboratory fault zones show systematic changes in elastic properties during the seismic cycle of stick-slip failure. Lab measurements of fault zone acoustic properties show that elastic wave speed begins to decrease well before failure. The precursors to failure extend across the complete transition from slow to fast laboratory earthquakes. These works show a systematic evolution of elastic wave speed during the seismic cycle indicated by: 1) an interseismic period of elastic loading and fault zone dilation with increasing wave speed, 2) inelastic load increase due to initial fault slip, which coincides with a period of roughly constant wave speed, and 3) the initiation of frictional weakening, macroscopic fault slip and reduction of elastic wave speed. These data are interpreted in terms of a fault zone model of localized slip surrounded by a damage zone that contains inactive ‘spectator’ regions. Time and slip dependent healing processes cause contact junctions to grow and stiffen, leading to increased elastic wave speed, whereas contact rejuvenation causes reduced wave speed. During the period of low resolved shear stress immediately after failure the whole fault zone undergoes time dependent healing and contact stiffening, which leads to an increase in elastic wave speed. As shear stress increases, interparticle slip begins within thin zones and along folia within a localized zone of shear. During the initial stages of slip, spectator regions continue to experience contact aging and stiffening, which together with localized slip leads to a reduced rate of wave speed increase and eventually quasi-constant wave speed within the fault zone when parity occurs between contact rejuvenation, in the active shear zone, and aging within spectator regions.

Machine learning can be applied to test models for the physical processes that occur during the seismic cycle. In particular, continuous measurements of elastic radiation can be used to illuminate fault zone structure and document the existence or absence of shear zones and spectator regions. Comparison of recent lab data with field measurements is promising. In both cases, elastic wave speed increases during the interseismic period and fault slip during earthquakes and slow slip events results in reduced wave speed. However, the magnitudes are quite different. In the lab, the total change in p-wave speed during the seismic cycle is of order 1% with one tenth to one quarter of that occurring during the preseismic phase, compared to field based measurements that show a total change in p-wave speed of 0.05 to 0.1% and a similar fraction of that occurring preseismically. These numbers are highly uncertain, as only a few well documented lab datasets exist and fewer field observations exist. Additional lab work is needed to document whether the magnitude of this effect scales with shear localization dimension or other factors. In addition, we need to locate acoustic precursors and identify their role in effecting changes in fault zone wave speed. These studies should be connected with ongoing ML work which shows that the variance of the acoustic signal scales directly with shear strength during the complete seismic cycle. Successful earthquake forecasts must be rooted in seismic source theory and rupture models, which means locating the acoustic precursors to laboratory earthquakes. However, even in laboratory settings, the suite of possible signals for earthquake forecasting, ranging from elastic moduli to frictional strength, and electromagnetic properties generally exceeds the number and spatiotemporal resolution of available
measurements. Generally, we either do not or cannot make the right measurements to reliably observe precursors to failure. Machine learning has proven that failure time can be predicted in advance of laboratory earthquakes but we do not know if acoustic precursors have distinct spectral signatures and how those evolve during the seismic cycle. Existing works portend a bright future of machine learning for earthquake forecasting, and perhaps ML can be used to predict how bright that really is.
Determining earthquake locations with a pattern recognition, delay-and-fire spiking NN approach

Ian McBrearty
Los Alamos National Laboratory

Determining an earthquakes source location is a ubiquitous and longstanding problem in observational seismology. The problem is most often solved by minimizing the residual between observed and theoretical travel times to all stations of interest using some pre-specified velocity model discretized onto a set of representative spatial grid nodes. Implicit to this formulation of the problem is the assumption that arrival times between all stations can be associated to a common event a priori, prior to solving the optimization problem. However, as the magnitude of completion of earthquake detections continues to decrease in response to improved capabilities and processing algorithms, the uncertainty regarding which arrival times truly associate to a common event easily becomes the principle challenge in earthquake location routines. False detections, large networks, and high rates of seismicity all compound this difficulty. Naively, the issue can be solved either through some iterative permutation scheme of “trying many combinations of arrivals”, or through back-projection which can separate arrivals into realistic combinations utilizing the combination of the spatial and temporal domains, and the theoretical travel time move-outs. However these approaches can lead to challenging nuances of their own, and can also add significant computational overhead.

In my work I present an alternative scheme which approaches the problem more from the perspective of classical pattern recognition. To a first order, it is clear that with nearly any velocity model the set of all possible theoretical travel time vectors to a set of n stations share a great deal of similarity to one another, and form a very sparse subset compared with the full $R^n$ space of any possible n dimensional vector of real values. I present a two step scheme in which first, (1) a representative sparse set of “template” travel time move-outs are created with the Self Organizing Map algorithm, which optimally “cover” all of the data space of possible travel time move-outs, and then (2) create a set of corresponding “neural nodes” for each template, which act in a “delay-and-fire” scheme of spiking neural networks to map continuous streams of arrival time “spikes” onto all of the nodes, and which constructively sum together and cause a “trigger” whenever one of the template patterns of move-out times is observed in the continuous arrival time data. Since the “spikes” can pass a Gaussian PDF of variable width to each node, and the triggering threshold can be arbitrarily set, the method is robust to noise in the arrival time data, and to false or missing detections. Triggered nodes effectively highlight when an event has occurred, and which set of arrival times are associated with the event. Absolute locations can be determined with standard algorithms, using the now associated arrival times.
Earth Science in the age of AI

Brendan Meade
Harvard University

Over the past 5 years, deep learning has rebuilt most every branch of applied computer science from machine vision to natural language processing. These fields have seen decades’ worth of theory, heuristics, and best practices replaced by astonishingly powerful neural networks. Earth science stands on the brink of a similar revolution and one that that is extraordinarily accessible due to the availability of easy-to-use APIs, pre-trained networks, and expertise from outside disciplines. Here we survey a class of solid-earth problems (wave propagation, earthquake prediction, earthquake triggering, precursor identification) and identify existing networks that may enable us to solve these problems. As an example, we demonstrate that a deep neural network can be used to quantify the relative effectiveness of different earthquake triggering mechanisms from a data set of 200+ mainshocks and 100,000+ aftershocks.
Automatic discrimination of deep and shallow induced-microearthquakes using neural network

Mostafa Mousavi
Stanford University

We develop an automated strategy for discriminating deep microseismic events from shallow ones on the basis of the waveforms recorded on a limited number of surface receivers. Machine learning techniques are employed to explore the relationship between event hypocentres and seismic features of the recorded signals in time, frequency, and time-frequency domains. We applied the technique to 440 microearthquakes $1.7 < M_w < 1.29$, induced by an underground cavern collapse in the Napoleonville Salt Dome in Bayou Corne, Louisiana. Forty different seismic attributes of whole seismograms including the degree of polarization and spectral attributes were measured. A selected set of features was then used to train the system to discriminate between deep and shallow events based on the knowledge gained from existing patterns. The cross-validation test showed that events with depth shallower than 250 m can be discriminated from events with hypocentral depth between 1000 and 2000 m with 88 percent and 90.7 percent accuracy using logistic regression and artificial neural network models, respectively. Similar results were obtained using single station seismograms. The results show that the spectral features have the highest correlation to source depth. Spectral centroids and 2-D crosscorrelations in the time-frequency domain are two new seismic features used in this study that showed to be promising measures for seismic event classification. The used machine-learning techniques have application for efficient automatic classification of low-energy signals recorded at one or more seismic stations.
Precursors that characterize the eruption dynamics of CO2-driven cold-water geysers using machine learning

Maruti Mudunuru
Los Alamos National Laboratory

Thermally driven geysers (such as Yellowstone) are characterized by frequent eruptions of liquid water and steam. Another subsurface system capable of producing periodic eruptions (similar to thermal geysers) is CO2-driven cold-water geysers. They erupt for over 24h at a time with relatively high velocity CO2-driven discharge from wellbores. Growing interest in geologic carbon storage has brought attention to CO2-driven cold-water geysers because of its similarity to high velocity wellbore leakage process. In the CO2-driven cold-water geysers, CO2 (gas) evolves by the pressure reduction (flashing) of CO2-rich fluids. Once the internal pressure of CO2 (aqueous) becomes greater than that of the surrounding fluid, CO2 separates from the fluid causing bubbles to nucleate, grow, and coalesce. Hydrostatic pressure reduction resulting from increasing CO2 gas volume fraction enhances expansion of CO2 bubbles leading to the eruption. In this talk, we present a feature extraction framework to identify a set of precursors to understand the eruption dynamics from timeseries signals (seismic and/or acoustic) using machine learning. To be specific, we extract/decompose signals that characterize the periodic eruption events from noisy data sets through time series feature engineering and source separation methods. This decomposition of signals (from sensors that are close and as well as from sensors that are far away from the geyser) into independent components (pre-eruption signatures and anthropogenic activities) can help in better understanding of the behavior of eruption times of CO2-driven cold-water geysers.
Bayesloc: Machine Learning Methods Applied to Seismic Multiple-Event Location

Stephen Myers
Lawrence Livermore National Laboratory

Seismic tomography relies on a high-quality data set of event locations and observed arrival times. Voluminous bulletin data sets are publicly available, but uneven data quality results in poor event locations that blur tomographic images even when outlier arrival-times are removed. We highlight the use of the Bayesloc method to simultaneously relocate events, identify outlier data, and assess the uncertainty of data and event location parameters. Bayesloc utilizes machine-learning methods to formulate the joint probability function for parameters comprising the multiple-event seismic location problem. The formulation emphasizes optimization and uncertainty quantification of physical parameters, and it includes stochastic parameters to account for error processes that are outside of the physical model. Markov-Chain Monte Carlo (MCMC) methods are used to draw samples from the joint probability function. Recent efforts to include precise differential time measurements based on waveform cross correlation result in inefficient MCMC sampling because event locations become strongly correlated (i.e., relative locations precisely constrained). We implement Langevin-Hastings sampling, based on probability gradients, to more efficiently sample event locations. We conclude by showing differences between data sets and images produced using traditional methods and those produced using Bayesloc. Prepared by LLNL under Contract DE-AC52-07NA27344.
We aim to move from the largely empirical, linear models of ore formation that have dominated the discipline over the past century to nonlinear models based on the remarkable developments in nonlinear dynamics in other physical-chemical systems over the past 30 years.

Hydrothermal systems are chemical reactors held far from equilibrium by the influx of heat, of chemical components in solution and of momentum. They operate far from equilibrium for as long as these inputs are maintained and suitable reactants exist within the system. The processes involved are intrinsically nonlinear so that the evolution of the system is unsteady and possibly chaotic. Such systems evolve to a nonequilibrium state characterised by the formation of spatio-temporal patterns where the supply of energy and mass balances the dissipation and rates of consumption of mass and energy.

Suitable methods for analysis should be those developed recently in nonlinear analysis rather than standard statistical methods applicable to linear systems. The signals we have from hydrothermal systems are the distributions of alteration and of mineralisation, well portrayed in drill core. There are many methods, including multifractals, for quantifying nonlinear dynamical systems. Another method of analysis involves the construction of recurrence plots which quantify the spatial patterns in the alteration and mineralisation and ideally enable the dynamics of the system to be established. The analysis is quantitative and enables a large variety of measures to be established for the spatial patterns. Good analysis requires very large numbers of data points measured in the 1000?s or larger.

The recurrence plot provides a fingerprint for the alteration and our aim is to build a library of such fingerprints to distinguish different styles of alteration and mineralisation. The ideal outcome is to be able to distinguish well-endowed systems from less-endowed systems on the basis of measures such as these recurrence plots. We also consider progress in using recurrence quantification instead of kriging in the construction of block models.
Learning high-flow and high-transport subnetworks in three-dimensional discrete fracture networks

Allon Percus
Claremont Graduate University

Structural and topological information play a key role in modeling flow and transport through fractured subsurface rock. Discrete fracture network (DFN) simulations reveal that that a small backbone of fractures accounts for most flow and transport in such porous media. Restricting the flowing fracture network to this backbone provides a significant reduction in the network’s effective size. However, the particle tracking simulations needed to determine the reduction are computationally intensive. Such methods may be impractical for large systems or for robust uncertainty quantification of fracture networks, where thousands of forward simulations are needed to bound system behavior.

We develop an alternative network reduction approach to characterizing transport in DFNs, by combining graph theoretical and machine learning methods. We consider a graph representation where nodes signify fractures and edges denote their intersections. Using random forest and support vector machines, we rapidly identify a subnetwork that captures the flow patterns of the full DFN, based primarily on node centrality features in the graph. Our supervised learning techniques train on particle-tracking backbone paths found by the dfnWorks computational suite, but run in negligible time compared to those simulations. We find that our predictions can reduce the network to approximately 20% of its original size, while still generating breakthrough curves consistent with those of the original network.
Optimising Insights from Machine Learning in Geoscience

Anya Reading
University of Tasmania

This presentation provides an overview of a significant and evolving body of research that has demonstrated success in drawing insights from spatial data using Machine Learning, and communicating those insights to scientific peers and end-users alike.

While the majority of data inference in solid Earth geoscience is carried out through deterministic physics-based modelling, we recognise that there is information contained in geoscience datasets that may be usefully inferred through statistics-based learning. We have carried out a number of demonstration studies that aim to form predictive maps through supervised learning, or aim to detect patterns through unsupervised learning. The multiple input data layers may include airborne geophysics, remote sensing data and geochemical data. Applications include lithology mapping in challenging locations and investigating multiple influences on regolith including bedrock and environmental history. We have developed a framework for machine learning that emphasises data preparation, the machine learning process itself and a rigorous process of prediction evaluation.

We find that prediction evaluation is especially important in geoscience and that the evaluation metrics can be as revelatory as the output of the machine learning process itself. While any Machine Learning Algorithm could result in informative outputs, we present evidence that Random Forests is a good first choice for prediction in spatial geoscience applications, especially where the practitioner is a non-computer specialist. We also discuss the success of using a Self Organising Map for pattern detection. In both cases, we provide a comparison to other algorithms. Furthermore, it is often the case that Machine Learning can be thought of as part of the computational geoscience toolbox for use as part of a workflow with more standard, deterministic physics-based approaches, and we provide a demonstration of this usage.

In this contribution, we summarise the interplay between the MLA and input data. We provide strategies for optimising insights, especially in situations where spatial sampling is not ideal. We also provide some suggestions with regard to the effective communication of ML results, and how to provide a credible explanation of how a given MLA is working. These are important considerations as ML techniques become more widely used, and perhaps sometimes misunderstood or misused, in geosciences.
Estimating the state of faults from continuous seismic data

Bertrand Rouet-Leduc
Los Alamos National Laboratory

Nearly all aspects of earthquake rupture are controlled by the friction along the fault that progressively increases with tectonic forcing, but in general cannot be directly measured. We show that fault friction can be determined at any time, from the continuous seismic signal.

In a classic laboratory experiment of repeating earthquakes, we find that the seismic signal follows a specific pattern with respect to fault friction, allowing us to determine the fault's position within its failure cycle. Using machine learning, we show that instantaneous statistical characteristics of the seismic signal are a fingerprint of the fault zone shear stress and frictional state. Further analysis of this fingerprint leads to a simple constitutive law quantitatively relating the seismic signal power and fault friction. These results suggest that fault zone frictional characteristics, and the state of stress in the surrounding rock, can be inferred from seismic waves, which could provide a powerful technique for seismic hazard assessment and earthquake warning systems.
Constraints on the Mantle Parameters Using Machine Learning Algorithms

M. Hosein Shahnas
University of Toronto, University of Minnesota

The convection in the interior of the Earth-like planets is highly nonlinear and the thermoeelastic parameters of the mantle are highly pressure and temperature dependent. Numerical and laboratory convection models have been widely used to constrain the mantle parameters. The significance of the accurate assessment of the mantle parameters rely on the fact that many Earth processes including the plate tectonics and continental drift, seafloor spreading, oceanic crust production, magnetic field of the Earth and geodynamo, as well as the composition of the Earth’s atmosphere are influenced by the vigour of convection in mantle. Although the thermochemical structure of the upper mantle processes can be inferred from seismic and mineral physics models, since the mantle material at higher mantle pressures may undergo phase-, structural-, spin- and dissociation-transitions, deeper processes are less well understood. All these transition parameters are both temperature and pressure dependent and the material properties of the main constituents of the lower mantle, perovskite (Pv) and ferropericlase (Fp) may significantly vary by the content of their basic elements.

The purpose of this study is to show that how machine learning (ML) algorithms can be employed in estimation of the lower mantle parameters. As an illustrative problem, we will focus on the flow stagnation at the mid-mantle depths by spin transition-induced density anomalies. In previous forward modelling, we have shown that the spin transition in iron in the mantle minerals may cause flow stagnation in the lower mantle and the degree of this stagnation depends on the magnitude of the spin transition-induced density anomaly. The training and testing samples for the machine learning models are produced by the forward numerical convection models with known magnitudes of density anomaly (as the class labels of the samples). We run a sufficient number of mantle convection models with different density anomalies for 200 Myr when the convection models have become sufficiently mature and have captured the impact of the spin transition effects in the convection models. The temperature fields carry the signature of the stagnation and hence can be used as learning data. Using the temperature fields as the training samples, we employ supervised machine learning (SML) to train a machine learned model for estimating the spin transition density anomaly. We show that SML techniques can successfully predict the magnitude of the mantle parameters. The method can be extended to more realistic 3D-convection models with many physical properties of the mantle, and the capability of the model in manipulation of elastic fracture processes which increases the complexity of the problem, but provides insights on constraints for an enhanced group of mantle parameters.
Building a Synthetic Training Dataset for Distributed Acoustic Sensor Measurements through Geomechanical Modeling

Christopher Sherman
Lawrence Livermore National Laboratory

Recent advances in distributed acoustic sensor (DAS) technology have produced a source of abundant new data for monitoring processes in the subsurface, such as geothermal energy production or hydraulic stimulation for hydrocarbon production. Because of the massive dataset (Tb per day) size, developing a machine learning approach for interpreting DAS data is essential for effective use, such as in operational situations, which require near-realtime results. In our work, we use the massively parallel multiphysics code GEOS to generate a catalog of synthetic DAS measurements that are typical of those recorded during the stimulation of a hydraulic fracture. We then relate physical observables in the model such as the extents of the generated fractures, fluid flow, and interactions with pre-existing rock fractures, to lay the foundation for a synthetic training dataset. A subset of this training dataset, which includes DAS measurements for three different wells and a variety of numerical model parameters, is given in Figure 1. These examples demonstrate the potential richness of DAS measurements and their use for subsurface monitoring. We are exploring the use of various algorithms to fully exploit the rich potential of this dataset.
Predictions of Flow and Transport in Fractured Media in the Subsurface using Graph-based Machine Learning

Gowri Srinivasan
Los Alamos National Laboratory
LA-UR-18-20895

Flow through fractured media in the sub-surface is heavily dependent on microstructural information. Continuum models often eliminate features critical to accurately predicting macroscale behavior but are commonly used since resolving thousands of fractures individually is computationally intractable. We overcome this hurdle by developing compact graph representations of the fracture networks, and using machine learning algorithms to mimic the detailed physics at the microscale. The resulting workflow has been shown to achieve up to 4 orders of magnitude computational speedup while maintaining accuracy.
Prediction of flux through fractures in a DFN through machine learning from topology attributes and flow physics

Shriram Srinivasan
Los Alamos National Laboratory

In geological sites composed of low permeability media like granite, shale or limestone, the interconnected fractures in the rocks provide the primary pathways for fluid flow. These sites are modelled using discrete fracture networks (DFN) which are expensive from the point of view of computation of flow and transport over the domain. However, studies have shown the prevalence of preferential flow-pathways and channeling effects which imply that sub-networks within the domain dictate the behaviour, and it provides a natural motivation to construct reduced-order models of these fracture networks by identifying such sub-networks. Prior efforts towards this goal used graph-theoretic algorithms and ideas to extract the backbone of the flow network. Recently, a supervised machine learning approach was used to classify fractures as members of the backbone based on various topological measures learned from backbones obtained through particle tracking data from high fidelity simulations. However, performing particle tracking to obtain backbones for training data is itself computationally intensive, hence we propose 2 different approaches.

Using only the results of the flow solver, it is possible to get data for flux through each fracture of the DFN network. Alternatively, since it is inexpensive to simulate flow and transport on graph representations of a DFN, we use them to get data on flux through graph edges and infer the flux through each fracture. Given the flow data on fractures in the DFN network, it is known that the flow rate through it must depend on both topological and hydrological properties such as permeability, aperture, length, topological degree etc. Our aim is to use a supervised machine learning algorithm to train on the existing data so that given a DFN network, fractures with flow rates greater than a critical value can be identified \(\text{a priori}\). Such identification will readily yield the flow backbone of the given DFN network.
New approaches to inverse problems in solid Earth geophysics: Insights from machine learning

Andrew Valentine
Australian National University

Can machine learning help us build a better understanding of the Earth’s internal structure and processes? Much of our current knowledge comes from solving inverse problems—by combining physical theory with observational data, it is possible to constrain unseen properties of the Earth system. Broadly speaking, two classes of approach are in common use in geophysical inverse theory: linearised methods, where a single solution is sought using an optimisation-based approach, and Monte Carlo-style techniques which aim to build an ensemble of possible solutions.

The former strategy is computationally-efficient, but often fails to properly capture the non-uniqueness and uncertainty inherent to most geophysical images; the latter is too computationally-expensive to be applied to many problems. In order to make full use of recent advances in computational modelling and the wealth of high-quality data now available, new imaging techniques are required. Underpinning the recent explosion in machine learning is a wealth of mathematical and statistical research exploring questions of inference, regression, and prediction. We hope that these ideas can be leveraged to develop novel strategies for tackling geophysical inverse problems, and this has been the focus of our research effort over several years. In this contribution, we will describe some of the key outcomes of this work—including the ‘prior sampling’ methodology and Gaussian process-driven inference—and demonstrate how they are being used to address previously-unsolvable problems in Solid Earth geophysics.
Combining Physically-Based and Data-Driven Models to Improve Forecasts of Groundwater Flow

AI Valocchi
University of Illinois

Physically-based numerical groundwater models (PBMs) are powerful quantitative tools to manage scarce groundwater resources and assess risk of subsurface contamination. As these models are being used to inform decisions and policies with major social, political and economic impact, it is important to improve the accuracy of these model and quantify their intrinsic uncertainties. Groundwater models suffer from uncertainty associated with errors in model structure, parameter values, input data and measurements. However, current state-of-the-art does not properly treat model structural error, which is ubiquitous in groundwater models, for example due to improper interpretation of geological structure and simplified conceptualizations of flow and contaminant transport processes. Current practice is to develop one or multiple groundwater models, and use field data as targets for calibrating parameter values. However, when calibrating an imperfect model, parameters may be over-adjusted to compensate for model structural error. This can lead to biased model forecasts for scenarios different from historical conditions reflected by calibration data. We have worked on two data-driven frameworks to reduce the detrimental effects of model structural error on model forecasts. The post-processor (also called the “complementary modeling”) framework constructs error models using machine learning techniques to correct for bias of an existing calibrated model. The error-explicit Bayesian framework jointly estimates model structural error (described using nonparametric kernel methods) with the physically-based model parameters. We demonstrate the performance of these frameworks for several real-world large-scale regional flow models. The data-driven framework brings together the strength of physically-based groundwater models and inductive data-driven statistical learning techniques, and is in harmony with new trends toward increased data availability and promotion of hydrologic observatories.
Potential Solution for Big Data Analysis from V-Cluster a system based on Graph Computing

David Yuen
University of Minnesota

Today it is clear that Big Data Revolution is taking place and is gradually but surely replacing high-performance computing and numerical simulation vis a vis partial differential equations as the main technical driver in society because of its pervasive nature. Outside the immediate scientific arena, the Big Data market encompass much more than any single professional society, e.g. AGU or GSA. There are many sectors in society Big Data can ably serve, such as city or provincial governments, financial sectors, hospitals, tourism, and last by not least, scientific and engineering problems, such as computational chemistry, jet engine design, structural engineering. In many countries, education has not kept apace with the demands from students outside of computer science, who want desperately to get into Big Data swing of things. Both Ultimate Vision and Mac Teach are start-up companies based at Beijing planto address this die need in China and also elsewhere by focusing our energy in education and training outside the immediate university environments. Ultimate Vision (UV) plans a strategy to maximize profits in our beginning. Therefore, we will focus on growing markets such as provincial governments, medical sectors, mass media and also education in the way of training. In terms of Big Data networking and message passing, we will not address issues such as network performance for scientific collaboration, such as seismic networks, where the market share and profits are small by comparison to hospitals and banks. We have developed a software and hardware system called V-Cluster, built with latest NVIDIA GPU and Intel CPU with ample amounts of RAM and local storage, over a couple of Tbytes. This hardware is generic. We have put in internal network with fast bandwidth over 100 Gbits/sec. Each core of V-Cluster can run at around 70 Tflops. The hardware system scales linearly with the number of cores. Our main strength in data analytics is the use of graph-computing paradigm for optimizing the transfer rate in collaborative efforts. We focus in training and education with our clients in order to gain experience in learning about new applications. We will lay out the philosophy of this second generation of of this Data Analytic system, whose total costs fall far below those offered elsewhere.

Our objectives are:
1. Solve problems in data analytics (finance, precision medicine, logistics), using our V-Cluster systems and graph computing software.
2. Training and Educating our clients and students ranging from undergraduates to accomplished professors
3. Consulting our clients at different levels in their path of adapting and gaining expertise at our technological approach.
4. Carry out research in industrial applications relevant to our incumbent and potential clients.

Our mantra lies in graph computing the common feature of big data and big models is the sparse, long-ranged relations among diverse entities.
Automatic seismic event detection and phase picking using deep convolutional neural network (CNN)

Lijun Zhu
Georgia Institute of Technology

Although manual picking of P and S waves is time-consuming and prone to human error, it is still the most reliable technique for seismic event detection and location. However, large volumes of continuous recording from various seismic networks create a significant challenge to such manual approach. Thus, much effort has been devoted to algorithms for automatic phase picking. P-wave picking algorithms include short-term average/long-term average (STA/LTA), the envelope function, autoregression, and kurtosis, all of which require tweaking a number of parameters to manually adapt to the waveforms of different types of seismic events under various conditions. Picking of S-wave arrivals is even more challenging due to contamination by the P coda and converted phases. A phase picking contest was sponsored by Alibaba cloud and the China Earthquake Administration jointly in Summer 2017 to improve aftershock catalogs of the 2008 MW7.9 Wenchuan earthquake in Sichuan, China [Fang et al., 2017]. The dataset, which contains both the continuous waveform and manually labeled P and S phases, motivated us to develop a series of machine learning based methods that we reported at AGU 2017 [Zhu et al., 2017].

This abstract focuses on the convolutional neural network (CNN) that we trained after the contest to classify all 20-second time frames extracted from the continuous waveform as P-phase, S-phase or noise-only. CNN is a special artificial neural network (ANN) that imposes a specific structure on its neurons, which have learnable weights and biases, such that spatial/local variance is captured through convolutions between input data and kernel filters. An objective of this work is to verify that P and S waves are distinguishable from each other and the background noise directly using raw data with no prior information. Thus, the data frames taken from all stations (with different ray paths and source functions) are mixed and passed directly to CNN without any pre-processing steps, such as a band-pass filtering. Because seismic events recorded at different stations have different amplitudes, a soft clipping is applied to keep most frames in the same dynamic range. Training is performed on 40000 frames (20000 P/S phases and 20000 noise-only frames) with known labels (P, S, or noise).

Our CNN achieves a 97% accuracy on the 20000 frames (10000 P/S phases and 10000 noise-only frames) unseen during the training stage. We also tested the classifier on continuous data by breaking the waveform into 20-second frames with a 0.1-second step. Not only are the strong events (with known labels from manual picking) detected correctly, but we also observe many weak events that were missed previously. Training can be accelerated by using Nvidia GPUs; for example, evaluating new frames takes only 0.4 msec using a GTX 1080 Ti GPU. Hence, applying our CNN algorithm to the entire one-month of continuous waveforms (14 stations) takes only 40 hours on a single GPU.

To better understand the performance of this CNN method, we plan to benchmark it against a template-matching technique on the same Wenchuan dataset. We also plan to test it on an open dataset from a different region to verify its general applicability. The proposed method can also be easily extended to an array detection algorithm using all stations in a network. Updated results will be presented at the meeting.
Poster Abstracts

Machine learning approach to rupture dynamics

Sabber Ahamed
Center for Earthquake Research and Information

Simulating dynamic rupture propagation is challenging due to uncertainties involved in the underlying physics of fault slip stress conditions and frictional properties of the fault. The trial and error approach is often the only way of determining the unknown parameters but running many rupture simulations can be computationally expensive. To reduce the computational cost and improve our ability to determine reasonable stress and friction parameters, we take advantage of machine learning approach. We created two models using the artificial neural network (ANN) and the random forest (RF) algorithms. We then train the models using a database of 1000 dynamic rupture simulations. Fault geometry, stress conditions, and friction parameters vary in each simulation. We cross-validate and test the predictive power of the models using additional 600 and 400 simulated datasets respectively. Both RF and ANN models predict with more than 81% testing accuracy and 84% true positive rate (recall). Both of the models are computationally efficient such that the 400 testings took a fraction of a second, leading to potential applications of dynamic rupture that have previously not been possible due to the computational demands of simulations.
Extending FAST to Detect Earthquakes over a Seismic Network

Karianne Bergen
Stanford University

Extracting earthquake signals from continuous waveform data recorded by networks of seismic sensors is a critical and challenging task in seismology. The Fingerprint and Similarity Thresholding (FAST) earthquake detector (Yoon et al., 2015) enables waveform-similarity-based earthquake detection in long duration continuous seismic data. FAST leverages locality sensitive hashing (LSH), a data mining technique for identifying similar items in large data sets, to efficiently detect similar waveforms without templates. Blind detectors like FAST are capable of identifying events with previously unknown sources, but they can be susceptible to false detections due to local correlated noise sources.

To improve detection accuracy and reduce false detections, we have developed a new method to extend FAST over a seismic network. We introduce pairwise pseudo-association, a technique that leverages the pairwise structure of event detections to identify events observed at multiple stations in the network without modeling the expected move-out. Pairwise pseudo-association and supporting techniques, event-pair extraction and event resolution, create a network detection pipeline that combines the sparse similarity matrix outputs from single-station FAST, applied independently to each station in a network, to produce a list of candidate events recorded across the network. Although our method was designed with FAST in mind, it is general and can be applied to any detector that produces pairwise waveform similarity measures for each station in a seismic network. Our approach automatically handles the unknown move-out across the network, is robust to missing or low-quality data at one or more stations, and is flexible enough to be applied to a variety of detection tasks and network geometries. Using the 2014 Iquique foreshock sequence as a test case, we show that our method is sensitive, identifying nearly five times as many events as the local seismicity catalog (including 95% of the catalog events), and has a low false discovery rate of less than 1%.
Estimating the Maximum Magnitude of Induced Earthquakes

Elizabeth Gilmour
University of Memphis

Seismicity in Oklahoma has sharply increased as a result of wastewater injection. Induced earthquakes currently dominate central and eastern United States seismicity (Keranen et al. 2016). These earthquakes have been occurring for a relatively short period of time, and thus there are large uncertainties regarding the maximum possible earthquake magnitude that might occur as a result of injection. We take a physical approach to estimating the maximum magnitude of induced earthquakes in Oklahoma through the use of dynamic rupture models and machine learning techniques.

We train a neural network to find the magnitude of a rupture as calculated by a dynamic rupture model. Dynamic rupture simulations calculate fault slip based on the mechanics of faulting, but not all parameter values and initial conditions for rupture simulations are well constrained by observational data. Additionally, dynamic rupture models are computationally intensive, and cannot be directly used to invert for parameters. The neural network is trained using a suite of dynamic rupture models, and reduces the computation time from approximately one hour per rupture to a fraction of a second for a large ensemble of rupture models.

The neural network will be used to match the observed magnitude-frequency distribution of seismicity with one produced by the dynamic rupture simulations. The earthquake catalog data can be inverted to find parameters consistent with earthquake occurrence for the area. To perform this inversion, we use a Markov chain Monte Carlo to randomly move through parameter space and select sets of rupture simulation parameters that are consistent with the earthquake observations. For each random set of parameters, we will use the neural network to perform the forward simulation to predict the magnitude-frequency distribution and compare with the original catalog. Once parameters are constrained in this way, we will be able to use the constrained input parameters to run a final suite of dynamic rupture models and estimate the maximum magnitude.

Our approach shows that machine learning has the potential to incorporate more complex physics-based simulations into geophysical inversions and seismic hazard estimates.
Identifying primary subnetworks in sparse three-dimensional discrete fracture networks using weighted graphs

Jeffrey Hyman
Los Alamos National Laboratory

Discrete fracture network (DFN) models explicitly use fracture geometry and network topology to simulate flow and transport through fractured systems. Recent advances in high performance computing have opened the door for flow and transport simulations in large explicit three-dimensional DFN. However, this increase in model fidelity and network size comes at a huge computational cost because of the large number of mesh elements required to represent thousands of fractures (with sizes that can range several orders of magnitude, from mm to km). We will discuss coarse scale graph representations of DFN and how they can be used to exploit geometric and topological properties of DFNs to perform system reduction without loss of accuracy for key quantities of interest. In particular, different graph-representations of DFN models and how they can be used to reduce the computational burden associated with DFN models will be demonstrated.
An Overview of Computational Learning Theory for use in Ground Motion Simulation

Naeem Khoshnevis
Center for Earthquake Research and Information

We study the challenges that exist in training an artificial intelligence (AI) system to simulate earthquake ground motions based on a predefined set of parameters. We use an artificial neural network to develop a prediction model, and study two important aspects about it: the sample complexity and the computational complexity; that is, the size of the training sample that should be used, and the running time that is necessary for the algorithm to learn. We analyze to what level can different tasks be learnable (as measured by the ability to estimate different features in a simulation), and the versatility of the developed system (as measured by its flexibility with respect to the configuration parameters).

In recent years, high frequency physics-based (deterministic) ground motion simulation techniques have improved considerably. However, due to uncertainties in the source, path, and site parameters, there is a growing demand for simulations which only have marginal variations in their parameters. Considering such simulations carry significant computational cost (in resources and time), we are interested in investigating whether it is possible to solve the optimization problem associated with determining the parameters controlling an arbitrary ground motion simulation using an AI system as a cost function, and whether the cost of training such a system can help reduce the long term cost of simulations. To test this idea, we study three common conditions in deterministic ground motion modeling, namely those of an elastic, anelastic, and inelastic propagating media, with individual and combined variations in source, path, and site parameters, in order to address both the sample and computational complexities. We find that the system can be trained for wave propagation in elastic media with fairly lower number of training data. Our study suggests that developing a prediction model for inelastic wave propagation, on the other hand, requires additional training data and it may not be easy to learn.
Detection of Seismic Events in Oklahoma using Convolutional Neural Network

Chenyu Li
Georgia Institute of Technology

Oklahoma has experienced an abrupt increase of induced seismicity in the last decade. Detection of small-magnitude earthquakes is essential because a complete earthquake catalog is the basis to understand injection-induced seismicity in Oklahoma. Machine learning, especially deep learning, provides robust tools for image classification and feature extraction with complex structures. For example, convolutional neural networks (CNN) have been recently applied to continuous seismic waveform recordings to perform phase picking and event detection, and is proved to be efficient (Zhu et al. 2017). In this work, we would like to verify the idea of transfer learning, which refines an existing classifier trained on a large dataset to a small dataset with a limited number of labeled samples in a different geographic region. Specifically, we utilize our recently developed CNN trained on a dataset in China after the 2008 M7.9 Wenchuan earthquake (Zhu et al. 2017). Using the single station CNN learned from Wenchuan dataset, together with ~2000 local/regional catalog events recorded by station OK029 in central Oklahoma (Chen et al., 2018), we expect to extract useful waveform features of local seismicity in Oklahoma. Updated results will be presented at the meeting.
Sparse Super-Pixel Representation Based Vegetation Classification from High-Resolution UAS Imagery

Youzuo Lin
Los Alamos National Laboratory

Vegetation, present at many sites of interest, may mask significant features of interest on the ground surface. We employ Unmanned Aerial System (UAS) to obtain remotely sensed surface imagery, which can be used to identify different species of vegetation. Supervised dictionary learning methods show supreme performance in classifying remotely sensed imagery. However, it yields poor classification results by applying supervised dictionary learning techniques directly to our UAS imagery. The reason is that there is only a limited number of spectral information (Red, Green and Blue) available in consisting the dictionary. On the other hand, our UAV imagery is highly spatially resolved (0.01 m/pixel). In order to compensate for the lack of spectral information and enrich the dictionary, we develop a novel supervised dictionary learning using both spectral and textural information. Spatial information can be also critical in characterizing various types of vegetation species. To further incorporate the spatial information, we generate super pixels by grouping similar pixels together and use them for classification. We illustrate the performance of our new classification algorithm using UAS imagery, and demonstrate that the classification accuracy of our algorithm can be much higher than conventional classification methods. Therefore, our classification method has great potential in characterizing the surface vegetation species.
How to tell the difference between stress transfer and fluid pressure pulses from accurate relocation of many earthquakes

Gabriele Morra
University of Louisiana at Lafayette

Recent advances in localization of earthquakes by using clustering of events with similar waveform has allowed to identify with unprecedented precision the shape and position of faults. By using the information provided by over 100 thousands pre and aftershocks of earthquakes in central Italy associated to a series of events of Mw ~ 6.0 and more along a 60km long fault system between 2009 and the present. By using the Boundary Element Method, based on the integration of a large number of Okada solutions, we invert the displacement measured by the tens of GPS stations present near L'Aquila and the time of the 2009 event. Once the fault slip is identified, we use the minor events to identify the geometry of existing faults. Faults are reconstructed using surface triangulation and square minimization from 90% of the closest events. Finally the Coulomb Stress associated to the seismic events themselves is calculate on the faults and compared to the amount of slip of each event. The entire data processing and numerical models are done using the numerical libraries of Python, and are available as a single Jupyter interactive Notebook to whom aims at integrating this research.

Both the 1997 and the 2009 main shocks were preceded by a series of small pre-shocks occurring in proximity to the future largest events. It has been proposed that the seismicity pattern of the two foreshocks sequences was caused by active dilatancy phenomenon, due to fluid flow in the source area, which would have produced pore pressure pulses. By analyzing this very large set of data (over 100k seismic events) we aim at constraining which seismic events are due to Stress Transfer from a fault to near one, which ones to fluid propagation and porosity waves, and whether other mechanisms might enter into play.

Northern and central Italy transitioned from convergence to post-orogenic extension. This has produced a unique and very complex tectonic setting characterized by superimposed normal faults, crossing different geologic domains, that allows to investigate a variety of seismic manifestations. Extensional systems are also characterized by release of gravitational energy and are expected to create a greater porosity. Since global compilations show that b-values for extensional systems are substantially greater than other tectonic setups (e.g. displaying many small events and few large events), central Italy with its continuous activity and ample database, is the ideal laboratory to investigate continental extensional systems.
Identification of sweet spots for hydraulic fracture in Avalon Shale, Permian Basin

Zohreh Souri
University of Houston

Kerogen content and Brittleness are the most important factors for investigating unconventional resources. Sweet spots for Hydraulic fracture are marked by high kerogen content (high TOC) and high brittleness. A 2-D map of sweet spots is made using lithofacies classification in Avalon Shale, Delaware Basin of the Greater Permian Basin. The proposed methodology may be helpful in sweet spot identification. This work takes advantage of well log analysis, artificial neural network and sequential Gaussian simulation to make a 2-D depth-section of lithofacies between two wells in New Mexico.
Accelerating USArray data processing using ensemble learning

Jiajia Sun
University of Houston

The USArray Transportable Array (TA) has traversed the continental United States and collected voluminous amounts of broadband seismic data that contain extremely rich information for mapping the structures of the Earth’s interior underneath North America. P-wave receiver function method has proven effective for studying crustal structures. Receiver functions are obtained after applying filtering, rotation and deconvolution to the recorded raw seismic data. It is critical to keep only ‘good’ receiver function traces for subsequent quantitative analysis. The task of classifying receiver function traces into ‘good’ and ‘bad’ categories is currently predominantly performed by seismologists through visual inspection. The whole quality control (QC) process requires some judgment calls on the part of the interpreter and could take up to days or weeks depending upon the amount of data involved.

Machine learning, due to its ability of learning complicated patterns and relations among input data, and making predications on new data instances, has been transforming many industries. In this study, we have investigated the feasibility of applying ensemble machine learning to the automatic classification of seismic receiver functions, with the hope that a trained predictive model can perform equally well as a seismologist but with a far greater speed. The seismological data used in our study are from earthquakes with a distance range of 30° to 90° and a magnitude of Mb 5 and above recorded at 201 stations in Alaska from the TA and Alaska Regional Network. We have obtained 12,597 receiver function traces, and manually labeled them as ‘good’ or ‘bad’.

We have explored two ensemble learning methods: random forest and gradient boosting, both using decision trees as the base estimator. The ensemble learning combines many weaker learners into a stronger one, and has found great success in many real-world applications. Specifically, random forest aggregates predictions from many independent decision trees, while gradient boosting trains many trees sequentially with each tree trying to improve upon the prediction from its predecessor. Among many boosting methods, we chose XGBoost, a scalable and distributed gradient boosting method.

We split our data set into a training and testing set with the proportion of 4:1. The training was completed within 20 seconds on a single PC with an overall accuracy of 90% and 92% for random forest and XGBoost, respectively. We are currently manually examining the mislabeled seismic traces and developing the best strategies to further elevate the prediction accuracy to human level. We are also investigating if classifying the data into 3 categories (good, bad and uncertain) is a better strategy, as this might improve the prediction accuracy for each category, and more importantly, allow human experts to focus primarily on these traces that are classified as uncertain.

Our preliminary results have shown that (1) machine learning is able to greatly accelerate the classification of receiver functions while maintaining a good classification accuracy, and (2) machine learning holds great promise for transferring some domain knowledge (e.g., determining if a seismic trace is good or bad) from a well trained expert (such as a seismologist) to others who can now indirectly incorporate the domain knowledge into his or her workflow.
Seismic Waveform Classification and Recognition Using Deep Learning Convolutional Neural Network

David Yuen
University of Minnesota

Today many digital seismic networks are deployed around the world. These produce a large number of continuous waveform records generated each day and demand new techniques on data processing. An efficient algorithm for automatically analyzing the contents of waveforms becomes the quest in seismic big data analysis. Indeed the traditional waveform auto-pick algorithms are not well adapted to the classification of complex earthquake events and noises from different tectonic regions. We also need to recognize them from continuous recorded waveform. Our study is based on more than 18,000 aftershocks and more than 60,000 background noise data manually selected by earthquake experts from 14 permanent stations in the Sichuan province in China from the end of June to the beginning of September 2008 following the devastating Wenchuan earthquake in May 12, 2008, and with data augmentation operations such as sliding window, filtering (0.1-20Hz). We train this large dataset by focusing on the basic features needed and then test this dataset. Afterwards we employ deep learning convolutional neural network for training and testing the models. The training accuracy reaches 95% and the verification accuracy reaches more than 90%. Compared with the traditional STA / LTA and AR-AIC methods, the identification results of the same data show that the convolutional neural network has lower false recognition rate and less time spent. Furthermore, we also verify the model accuracy and generalization ability of different sliding window steps and waveform segmentation length. In our test results, the models have a time interval of 120 seconds of seismic waveform signals and 2.5 seconds of sliding window yields the best performance. Then we use the trained optimal model and STA / LTA method to identify the real-time seismic data streams from 441 fixed stations in China's seismic network from December 12 to December 19, 2017. This program detected 7049 events, after correlated with adjacent stations as well as artificial identification, 3562 were identified and compared with the seismic catalog of China Seismological Network. Our AI detection rate of earthquakes above Magnitude 1 was found to be 95% and above Magnitude 3 was 99%. This finding shows conclusively that our Convolutional Neural Network (CNN) model has good potential for analyzing real-time seismic waveforms from different regions over a large area such as Sichuan province and can cover different signal-to-noise ratios. Our work shows the wide-ranging application prospects in the rapid detection of massive real-time seismic data streams. We now plan to further classify automatically the P and S phases from events, arrival time picking up and, last by not least, the source location of the earthquake.