Machine Learning in Seismology: Turning Data into Insights

by Qingkai Kong, Daniel T. Trugman, Zachary E. Ross, Michael J. Bianco, Brendan J. Meade, and Peter Gerstoft

ABSTRACT

This article provides an overview of current applications of machine learning (ML) in seismology. ML techniques are becoming increasingly widespread in seismology, with applications ranging from identifying unseen signals and patterns to extracting features that might improve our physical understanding. The survey of the applications in seismology presented here serves as a catalyst for further use of ML. Five research areas in seismology are surveyed in which ML classification, regression, clustering algorithms show promise: earthquake detection and phase picking, earthquake early warning (EEW), ground-motion prediction, seismic tomography, and earthquake geodesy. We conclude by discussing the need for a hybrid approach combining data-driven ML with traditional physical modeling.

INTRODUCTION

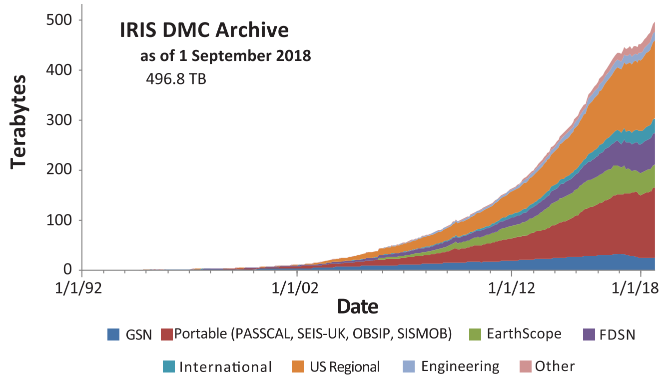
In a broad sense, machine learning (ML) is a set of related techniques that extract information directly from data using well-defined optimization rules. ML has recently drawn attention due to its wide ranging success in various fields (Murphy, 2012; Jordan and Mitchell, 2015; Witten *et al.*, 2016). Seismology has been a data intensive field since its very origin. As the years have progressed and the field has expanded, numerous methods and tools have been developed to detect and characterize earthquakes and to study earth structure. We as a community have already developed a rich set of techniques, but ML can bring a different and complementary set of useful tools.

In seismology, we are currently undergoing rapid changes in the "3 V's" often discussed by the big data community (Sagiroglu and Sinanc, 2013): volume, variety, and velocity. For example, the archive of seismic waveform publicly available from Incorporated Research Institutions for Seismology (IRIS) is increasing in size exponentially (Fig. 1). This dramatically increased volume of data (and the secondary products derived from the raw data) makes manual processing difficult. Many ML algorithms are designed with large datasets in mind: typically, more data gives better results. Dataset variety has increased too. Besides seismic data, other types of relevant geophysical datasets (e.g., Global Positioning System [GPS] time series and Interferometric Synthetic Aperture Radar [InSAR] images) are readily available from UNAVCO and other resource centers. The use of joint geophysical datasets might provide better resolution in certain problems, and carefully designed ML techniques can help analyze these datasets without introducing unnecessary complexity (Khaleghi *et al.*, 2013). Finally, velocity refers to the speed of data processing and distribution. This is important for real-time earthquake detection and earthquake early warning (EEW), which rely on rapid analyses of high-velocity data streams.

An exciting aspect of applying ML to seismology is the potential to find unseen patterns or new and significant features in our datasets. A recent example is using ML to predict the timing of the next slip event in laboratory slip experiments with features extracted from low-amplitude acoustic emissions that were previously considered to be noise (Rouet-Leduc *et al.*, 2017). As seismologists, we have the intuition and logic to analyze data, but ML could work beyond human intuition to facilitate the discovery of unconsidered patterns.

In essence, all ML algorithms learn from data using probability theory, which has been the mainstay of statistical methods for centuries. Most ML algorithms can be grouped into two main categories: supervised learning and unsupervised learning. Depending on whether the data have target labels or not (Fig. 2), one category may be preferable. Supervised learning, which comprises predictive modeling and operates on labeled datasets, can be further subdivided into classification and regression algorithms based on whether the target outputs are categorical (classification) or quantitative (regression). Unsupervised learning is subdivided into clustering and dimensionality reduction, depending on whether we are interested in grouping data into categories based on similarity, or simply reducing the input data dimensions. There are other more exotic types of ML algorithms such as semisupervised learning and reinforcement learning, for which we refer readers to more advanced texts (e.g., Murphy, 2012; Goodfellow *et al.*, 2016).

doi: 10.1785/0220180259



▲ **Figure 1.** The Incorporated Research Institutions for Seismology Data Management Center (IRIS-DMC) archive growth (modified from IRIS). The growth of the seismic waveform data at the IRIS DMC from the time it was established until 1 September 2018.

ML algorithms, although diverse in their implementation, tend to follow a basic workflow that includes the following steps (Fig. 3). In step 1: data collection, data are collected and partitioning into training and testing sets. A key aspect of ML is training the model on a random subset of the dataset, and then verifying the model on independent testing data. In step 2: preprocessing, data are cleaned and formatted, and missing data are removed or repaired. Feature extraction, which increases the performance of many ML algorithms by transforming the raw data into a more useful state for a given task, may also be performed. In step 3: model training, numerical optimization algorithms are used to iteratively tune the model parameters based on a cost function specific to the learning task of the problem. In step 4: model evaluation, model performance is evaluated on test data. Finally, in step 5: production, the finished ML model is applied in production mode to new data.

ML has received enormous recent interest across a wide range of disciplines (Jordan and Mitchell, 2015; LeCun *et al.*, 2015). We hope that this article will inspire both seismologists to further explore ML theory and techniques and data scientists to apply their latest ML algorithms in seismological fields. Although this article provides a high-level overview of potential ML applications to seismology, there are many wonderful textbooks and online courses that provide greater details about individual ML algorithms and their implementations (see Data and Resources). We next discuss recent applications of ML in seismology and their potential for obtaining new geophysical insights.

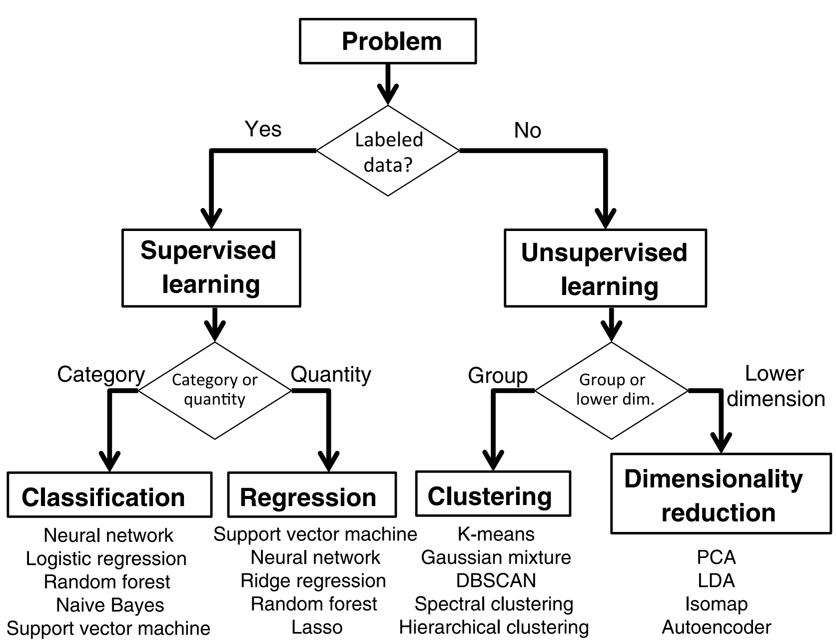
APPLICATIONS OF ML IN SEISMOLOGY

In the following, we present a detailed survey of five specific applications of ML to earthquake seismology, while acknowledging that there are many other worthy applications that merit discussion.

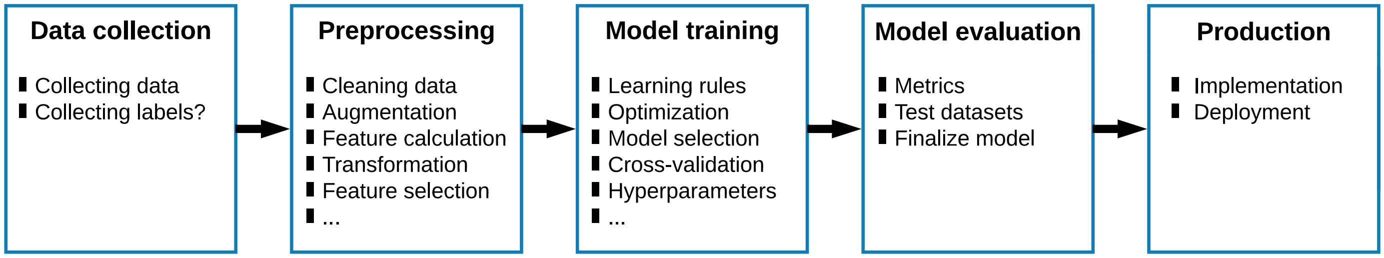
Earthquake Detection and Phase Picking

Automated detection and picking of earthquakes are longstanding problems in seismology, with the first algorithms developed in the late 1970s (e.g., Allen, 1978). Today, these subjects are still active research areas, and incorporate technology from computer science, electrical engineering, statistics, and many other fields to extract as much information from the data as possible. Some of the earliest applications of ML learning to seismology were to the problem of discrimination and classification of seismic events (e.g., Dysart and Pulli, 1990; Musil and Plešinger, 1996; Fedorenko *et al.*, 1999; Ursino *et al.*, 2001); and in recent years, this research has expanded to include utilizing ML to improve earthquake detection and phase picking capabilities (e.g., Dai and MacBeth, 1995; Tiira, 1999; Zhao and Takano, 1999; Wiszniowski *et al.*, 2014). These efforts have shown considerable promise to date, most notably in the area of deep learning, and suggest that many exciting new developments are coming in the near future. In particular, there is a distinct possibility that these algorithms will surpass the capabilities of human experts for the first time. Here, we outline some of the most promising examples of ML applied to the earthquake detection problem.

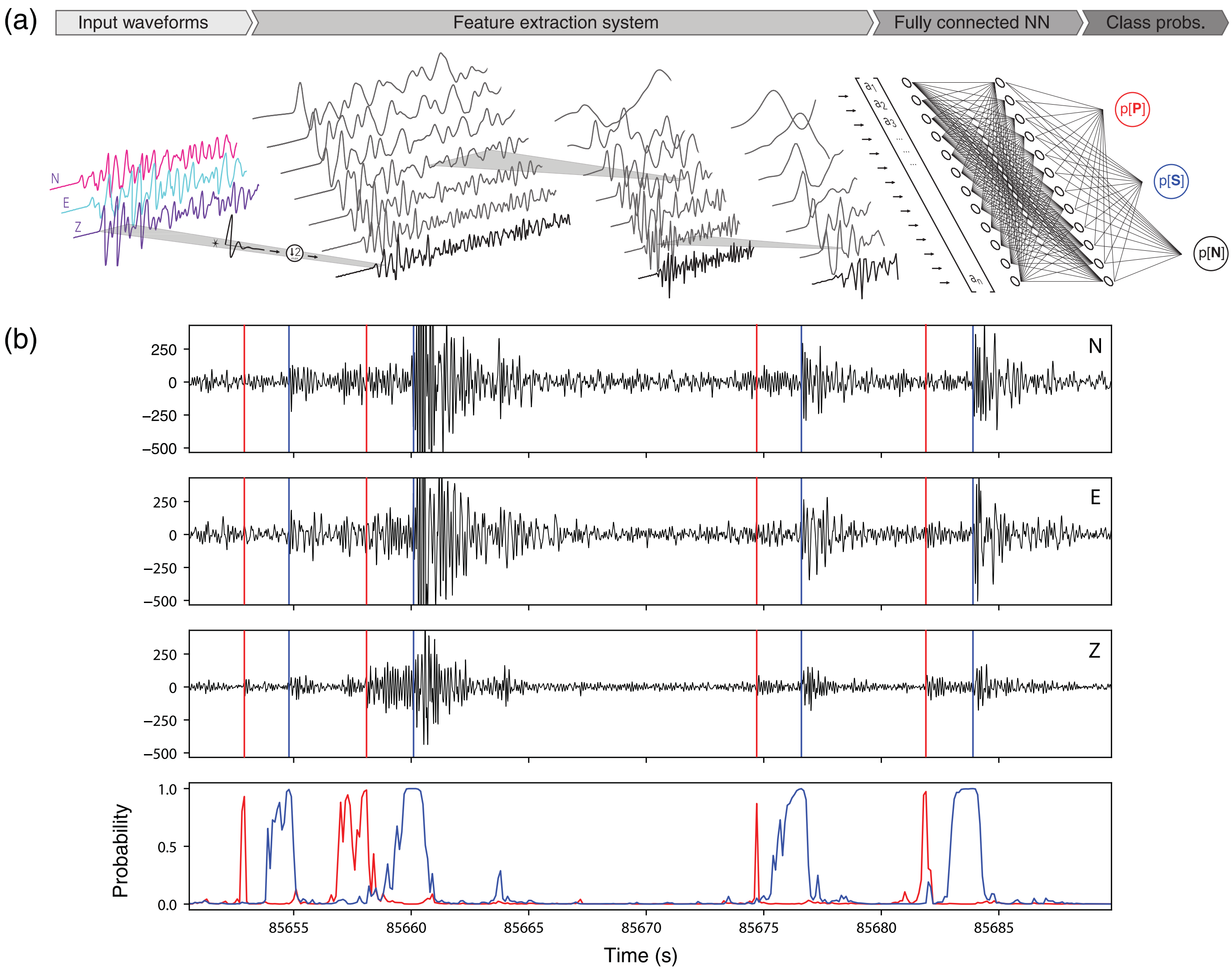
Over the last decade, there has been an explosion of interest in using the similarity of waveforms between nearby sources to detect previously unidentified earthquakes. This originally began with matched filtering (template



▲ **Figure 2.** Types of machine learning (ML) algorithms. Supervised ML operates on labeled datasets with the objective to develop models that predict either categorical or quantitative target variables. Unsupervised ML operates on unlabeled datasets with the objective to group data by similarity or reduce the dimensionality of the input datasets. Some common ML algorithms are listed at the bottom for each category.



▲ **Figure 3.** A generic ML workflow that guides many applications: (1) data collection, (2) preprocessing, (3) model training, (4) model evaluation, and (5) production.



▲ **Figure 4.** Example of generalized phase detection (GPD). (a) Cartoon schematic of a convolutional neural network (CNN) for GPD. A convolutional feature extraction system is combined with a fully connected neural network (NN) to produce class probabilities for *P* waves, *S* waves, and noise (Ross, Meier, Hauksson, *et al.*, 2018). (b) Application of GPD to the 2016 Borrego Springs, California, sequence. Red and blue colors indicate *P* and *S* waves, respectively. Vertical bars indicate automated picks.

matching), which uses waveforms of known events as templates to scan through continuous waveforms for new event detection (Gibbons and Ringdal, 2006; Shelly *et al.*, 2007; Peng and Zhao, 2009; Kato *et al.*, 2012; Ross *et al.*, 2017; Beaucé *et al.*, 2018; Chamberlain *et al.*, 2018). Recently, there has been an interest in applying ML and data mining algorithms for similarity-based event detection. In Perol *et al.* (2018), a convolutional neural network (CNN) was trained to simultaneously detect and locate earthquakes based on single-station waveform classification. For a given window of data, the goal is to predict which of several spatial regions the event occurred in, with the option for rejecting all of them. Alternatively, the Fingerprinting and Similarity Thresholding (FAST) algorithm (Yoon *et al.*, 2015; Bergen and Beroza, 2018) is a data mining approach that converts an entire continuous waveform dataset into a database of binary fingerprints. These fingerprints are compact representations of short segments of continuous waveform data and are organized in a special dictionary structure for efficient lookup. A key feature of FAST is that it is essentially unsupervised: earthquakes can be identified without prior knowledge of seismicity because, for highly similar waveforms, fingerprints are more similar to each other than those of random noise sources. In addition, FAST is computationally more efficient than template matching, which will help to facilitate automated processing of large waveform datasets.

A new category of earthquake detection algorithms that have recently emerged is generalized phase detection (GPD; Ross, Meier, Hauksson, *et al.*, 2018). Rather than search for near-identical waveforms, GPD instead trains convolutional networks to learn generalized representations of seismic waves from millions of example seismograms. This knowledge is then used to classify windows of data as *P*, *S*, or noise (Fig. 4). It has been shown to reliably identify *P* and *S* waves with excellent temporal sensitivity and performance in low signal-to-noise ratio conditions, resulting in typically 5–10 times as many events detected as conventional methods. GPD can simultaneously be used to pick arrival times with high precision. A key advantage of the method is that once trained, the model can be applied to datasets other than just those encompassed by the training set, such as data recorded in different tectonic regimes, large magnitude earthquakes, and active-source explosions. This is advantageous in situations in which a seismicity catalog is unavailable to use for template matching or in seismic monitoring.

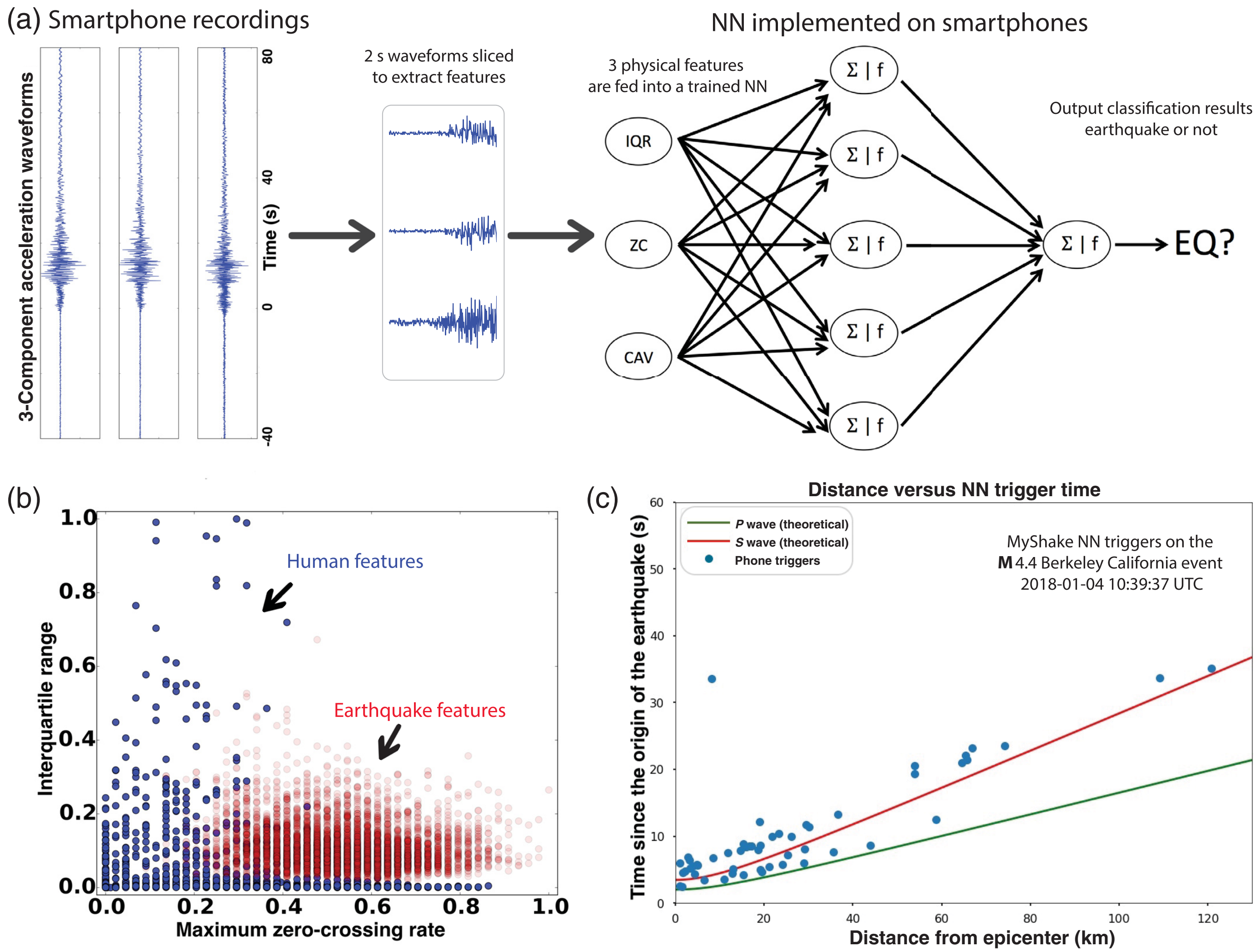
In addition to detecting earthquakes, there have been a number of noteworthy developments in algorithms for phase picking with ML. Chen (2018) developed an approach to pick seismic-wave arrival times using fuzzy clustering, which is based on the idea that the amplitudes of the seismic data before and after the arrival can be treated as separate, but possibly overlapping, clusters. This enables a decision boundary to be drawn that is taken as the arrival pick. Zhu and Beroza (2018) found great success in applying fully convolutional networks to pick *P*- and *S*-wave arrival times by training on millions of seismograms picked manually in northern California. Their method takes complete three-component seismograms as inputs and outputs probability time series corresponding to the likelihood of *P*- and *S*-wave onsets. They demonstrate state-of-the-art picking performance for both phase types, and their method further provides an important empirical mechanism for estimating the quality of the picks. This includes difficult cases such as clipped seismograms, in which even human analysts would have a difficult time. Ross, Meier, and Hauksson (2018) trained a CNN to pick *P*-wave onset times, but instead used the network as a regressor to predict the time index of the phase onset. They also trained a separate convolutional network to pick first-motion polarities of *P* waves, which are essential ingredients in calculating focal mechanisms. They demonstrated that the networks can often pick polarities more accurately than professional seismic analysts, as well as more frequently. This will lead to more detailed and expanded focal mechanism catalogs.

There have been several other exciting recent applications of ML to earthquake detection problems. Beyreuther *et al.* (2012) developed a hidden Markov model to classify and detect events for volcanic and geothermal areas. Treating event detections as an object detection problem, Wu *et al.* (2018) cascaded region-based CNN to capture laboratory slip events of different durations. Finally, Aguiar and Beroza (2014) and Zhang *et al.* (2014) combined insights from Google's PageRank and other image-based search engine methods to obtain waveform templates for low-frequency earthquakes.

EEW and Real-Time ML

EEW systems provide seconds to minutes of warning before the strongest shaking by taking advantage of the fact that electronic signals travel much faster than seismic waves, and that the *S* wave and surface-wave phases that produce the strongest shaking travel slower than the first *P*-wave arrivals (Allen *et al.*, 2009). There have been several recent efforts in EEW using ML algorithms, either based on hand-selected physical features extracted from seconds of waveforms (Kong, Allen, Schreier, *et al.*, 2016), or using deep learning algorithms to automatically extract features to identify the onset of the earthquakes at a single station (see more examples in the Earthquake Detection and Phase Picking section). Li, Meier, *et al.* (2018) trained a generative adversarial network (GAN) to learn the characteristics of both earthquake *P*-wave arrivals and background noise, resulting in a discriminator that mitigates false triggering. ML techniques such as support vector machine regression and neural networks (NNs) have also been used to estimate the magnitude, epicentral distance, and other relevant parameters using input features derived from a short-time window of waveform data following the *P*-wave arrival (Böse *et al.*, 2008; Cuéllar *et al.*, 2018; Ochoa *et al.*, 2018). Meier *et al.* (2015) proposed a method to estimate the magnitude and station–source distance by estimating the posterior probabilities from the observed frequency content to reduce the uncertainties. Böse *et al.* (2012) developed an image recognition-based algorithm to classify the observed ground-motion amplitudes into near-source/far-source regions and map a finite-fault rupture estimate automatically. Finally, Cua and Heaton (2007) proposed a unified framework for different components of the EEW, including real-time earthquake source estimation and alert decision making using a Bayesian approach. Using station locations, previously observed seismicity, and known fault traces as prior information could improve the system performance, especially at regions with low-station density (Yin *et al.*, 2018).

MyShake, a recent effort using smartphones to detect nearby earthquakes and provide EEW to the public, has many ML aspects and demonstrates promising results (Kong, Allen, and Schreier, 2016). MyShake has two levels of detections: a



▲ **Figure 5.** The NN used in the MyShake earthquake early warning (EEW) phone application. (a) The workflow of the NN algorithm on the phone, including extraction of features from recorded phone motion and implementation of an NN classifier to distinguish between motions from humans and earthquakes. (b) The interquartile range and maximum zero crossing rate are two important features for distinguishing between earthquake and nonearthquake motions (modified from Kong, Allen, Schreier, *et al.*, 2016). (c) Example application of MyShake at the network level to an **M** 4.4 earthquake that occurred in January 2018. NN triggers from individual users are compared against theoretical *P* and *S* arrivals.

single-phone supervised approach and an NN-based unsupervised approach. For individual phones, a trained NN is implemented on each phone to distinguish the earthquake signals from everyday human activities. Using a 2 s sliding window on a filtered three-component waveform, MyShake extracts three important features from the phone that represent the amplitude and frequency content of the movement. These features are then fed into an NN algorithm to classify whether the waveform is from an earthquake or from human activities (Fig. 5). When a phone detects an earthquake-like waveform, it sends a trigger message to the cloud server with a timestamp, location, and amplitude to make further confirmation of the earthquake by considering groups of phones within a region at a network level. The triggers from the phones are aggregated to a proper resolution grid to reduce the real-time computation burden, and the density-based spatial clustering of applications with noise (DBSCAN, Ester *et al.*, 1996) algorithm finds potential clusters of phones that have likely been triggered by an earthquake. Trigger messages in the clusters identified by DBSCAN are used to estimate the earthquake location and magnitude.

Ground-Motion Prediction Using Supervised Learning

Ground-motion prediction is a crucial aspect of earthquake hazard assessment, and although simple in concept it is challenging to perform in practice. At its core, ground-motion prediction answers the question: given a hypothetical earthquake source, how strong is the shaking likely to be? The basic physical factors controlling ground motion are well established: one can think of the ground motion observed at the surface as a convolution of source, path, and site effects (Boore, 1983). The classical approach to ground-motion prediction uses linear regression to model the first-order aspects of these effects (Campbell and Bozorgnia, 2008). In a linear ground-motion prediction equation (GMPE), the predicted ground-motion Y (in logarithmic units) is a normal distributed random variable that is a linear function of the input variables, which include the earthquake magnitude M and source–site distance R

$$\log Y = c_0 + c_1 M + c_2 \log R + \varepsilon; \varepsilon \sim \eta(0, \sigma^2). \quad (1)$$

Here, c_0, c_1, and c_2 are empirical coefficients and η denotes a normal distribution. The misfit term ϵ includes both epistemic uncertainty that can be reduced through better observational constraints and more sophisticated modeling approaches, as well as random variability that cannot be reduced (Douglas and Edwards, 2016). Probabilistic seismic hazard assessments are particularly sensitive to epistemic uncertainty in GMPEs (Anderson and Brune, 1999), and its reduction has been a primary focus in developing new GMPEs. For example, several of the most recent linear regression models developed for the Next Generation Attenuation relationships project (Bozorgnia *et al.*, 2014) include dozens of regression coefficients, which reduces data misfit but at the cost of increased model complexity. There has also been a significant recent effort to develop generic linear GMPEs that are regionally adjustable and hence exportable to different regions (Yenier and Atkinson, 2015).

Despite these advances, it is challenging to incorporate more complex source, site, and path effects within a linear GMPE. A viable alternative is to treat ground-motion prediction as a supervised learning problem, with well-defined input and target variables but considerably more flexibility on the model design. Moreover, the central focus on model validation inherent to the ML paradigm, including carefully partitioning of training and testing datasets, would help alleviate the traditional quixotic focus in GMPE model development on reducing the data misfit (Bindi, 2017), and instead allows us shift our attention to improvements in predictive validity.

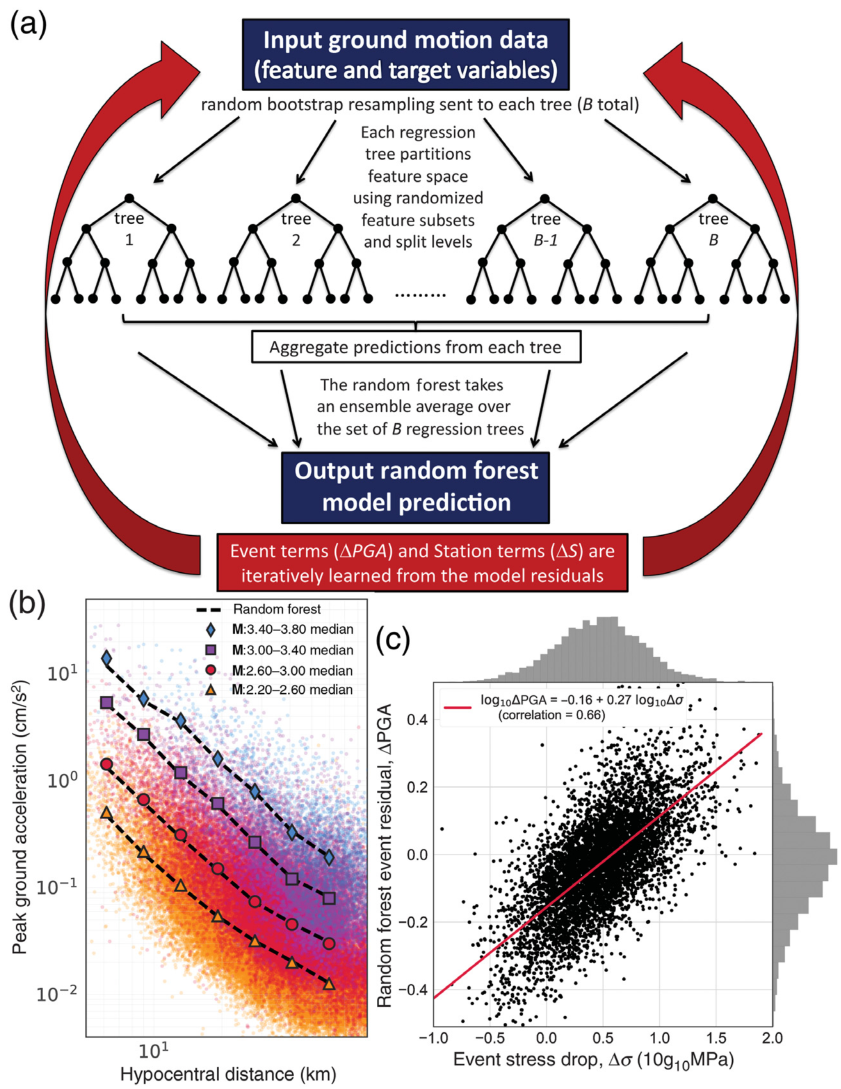
Some of the earliest ML GMPEs (Alavi and Gandomi, 2011) employed shallow NNs, and this approach is still the most commonly used. Derras *et al.* (2012) analyzed KiK-net records collected in Japan using an NN with a single hidden layer to predict peak ground acceleration (PGA) as a target variable using five input features: magnitude, epicentral distance, source depth, near-surface shear wavespeed, and site resonance frequency. Derras *et al.* (2014, 2016) generalized this approach to multiple target variables, including PGA but also peak ground velocity and pseudospectral accelerations at periods of interest for structural design. NNs are only one of many viable applications of supervised learning techniques to ground-motion prediction. Alimoradi and Beck (2014) developed a technique to synthesize realistic strong-motion records by applying Gaussian process regression to a sparse, orthonormal set of basis vectors called eigenquakes, which represent characteristic earthquake records. Thomas *et al.* (2016) developed a randomized adaptive neuro-fuzzy inference system to analyze records from the Pacific Earthquake Engineering Research database. Although these studies differ in technical details, viewed holistically they demonstrate the potential for improved predictive performance over linear GMPEs using similar input and target variables.

The modeling flexibility inherent to supervised learning also allows for the examination of input features that are not traditionally incorporated in linear GMPEs. To this end, Trugman and Shearer (2018) used a generalization of the random forest supervised learning algorithm (Breiman, 2001) to quantify the relation between earthquake stress drop and PGA for earthquakes in the San Francisco Bay Area. Although the basic correlation is intuitive—higher stress drop events are enriched in high frequencies and should have systematically larger ground-motion amplitudes (Baltay and Hanks, 2014)—there exist a few quantitative estimates for the importance of this effect and how it varies with magnitude and source–site distance. Trugman and Shearer (2018) demonstrate that the event residual terms learned by the random forest GMPE have a physical basis in the variability in earthquake stress drop (Fig. 6), thus highlighting the utility of ML techniques in ground-motion modeling.

ML tends to work best in scenarios where high-quality data are plentiful and easily available. This presents a significant challenge in ground-motion prediction, in which near-source records of large magnitude earthquakes—which pose the greatest hazard—are sparse. Future studies may focus on the best ways of integrating limited observational data in this regime with synthetic data from broadband rupture simulations, which will become increasingly prevalent in the coming years (Khoshnevis and Taborda, 2018). Similarly, ML approaches to nonergodic GMPEs, in which the predicted ground motions vary spatially (Anderson and Brune, 1999), have yet to be fully explored. Although ML itself is not a panacea for the outstanding questions in earthquake hazard analysis, an ML approach to ground-motion prediction may prove to be a powerful new tool in the next generation of seismic hazard assessments.

Tomography and Illuminating Geophysical Structure with ML

ML in seismic tomography has shown great promise for improving our understanding of subsurface geophysical structure. Seismic tomography methods obtain subsurface models or images from sensor array observations of seismic waves, which are generated by anthropogenic sources, earthquakes, or ambient noise processing. Seismic tomography is critical for deducing geophysical structure and characterizing seismic hazard (Rawlinson *et al.*, 2010). However, the demands placed on these methods are great, as tomography models are often estimated from limited and noise corrupted observations with nonlinear forward models. Such ill-posed inverse problems require regularization or assimilation of hypothesized geophysical structure to obtain physically plausible solutions. ML represents a modern paradigm for signal processing, with more sophisticated model priors and latent representations (Murphy, 2012) than classic inverse methods like Tikhonov or total variation regularization (Aster *et al.*, 2011). ML priors include sparsity constraints and latent dictionaries. The nonlinear general function approximation capability of NNs (Bishop, 2006) permits replacement of seismic data sim-



▲ **Figure 6.** Random forest ground-motion prediction equation (GMPE; Trugman and Shearer, 2018) and earthquake stress drop versus peak ground acceleration (PGA). (a) Schematic workflow for training the random forest GMPE. (b) PGA versus hypocentral distance for seismicity in the San Francisco Bay Area. Each point represents a site-corrected PGA measurement from an earthquake at a single station. Also shown is the median value in equally spaced magnitude–distance bins (large markers) and predicted values from the random forest GMPE (dashed lines). (c) Event PGA residuals learned from random forest GMPE versus earthquake stress drop. The least-squares linear fit and correlation coefficients are marked for reference.

ulation and inversion procedures with NNs. In the following, we introduce seismic tomography and show how it has benefitted from ML theory, including unsupervised and deep learning.

Seismic tomography can be categorized as either travel-time tomography or full waveform inversion (FWI; Virieux and Operto, 2009). Travel-time tomography calculates slowness (the inverse of seismic wavespeed) perturbations to reference models using source–receiver travel-time measurements. FWI methods calculate perturbations to a reference model which best-predict "full" recorded seismic waveforms. For both methods, the basic optimization problem is

$$\min_x \Phi(x) = \sum_{i=1}^{I} p(y_i - G(x)), \quad (2)$$

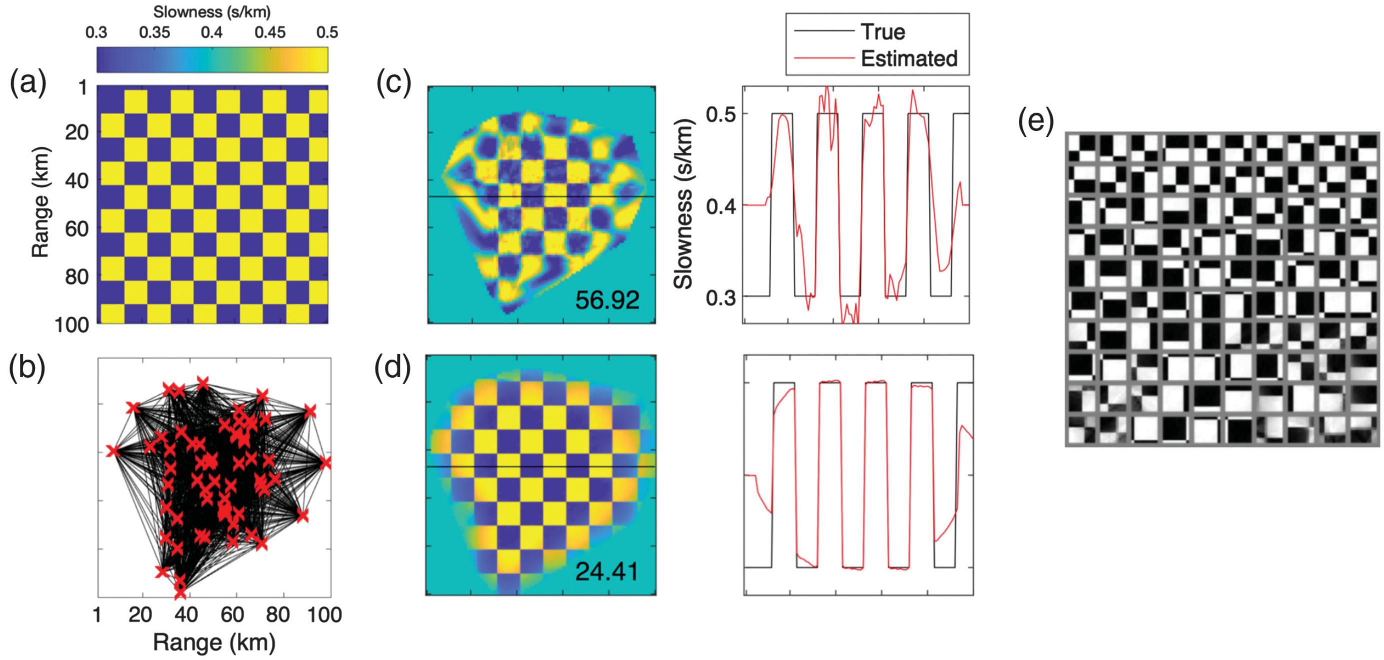
in which p is a penalty (e.g., least-squares), y_i is the data, G denotes the forward model, and x are model parameters. $\Phi(x)$ is minimized with respect to x, for I realizations of data. In travel-time tomography, y_i are travel-time observations, and G contains ray information relating the travel times to slownesses x. However, solving equation (2) directly will almost certainly give poor results, because it is ill-posed and ill-conditioned by the nonuniqueness, nonlinearity, and sensitivity to noise of the forward operator G. ML provides a means of constraining geophysical features in such models, but it is reliant on adequate training data to obtain reasonable performance.

The application of simple ML implementations to the seismic tomography problem is problematic due to a lack of training data, because in regional to global-scale seismic tomography, no geophysical ground-truth data exists. This issue has driven development of more advanced ML-based methods in seismic tomography that do not depend on large volumes of training data, or that are trained on simulations. Methods that do not require ground truths are based on adaptive, unsupervised learning frameworks (Elad, 2010; Mairal *et al.*, 2014). In these adaptive approaches, data observations themselves are used for training with unsupervised learning.

Adaptive ML-based seismic tomography methods, inspired by image denoising (Elad, 2010) and medical imaging (Ravishankar and Bresler, 2011; Greenspan *et al.*, 2016), have achieved compelling results. These methods combine sparse modeling with unsupervised learning. In sparse modeling, signals are represented using few (sparse) atoms from a dictionary of atoms D. Such atoms are solved using a least-squares objective function with a sparsity inducing prior. For example, a sparsity constraint is added to equation (2), with $x = D\alpha$, as

$$\min_\alpha \Psi(\alpha) = \sum_{i=1}^{I} p(y_i - G(D\alpha)) + \lambda||\alpha||_0, \quad (3)$$

in which α are the sparse coefficients, the l_0-norm enforces sparsity by counting the number of nonzero coefficients, and λ is a tuning parameter controlling sparsity that is analogous to a regularization constant in classical inverse methods. The l_0-norm is nonconvex and is typically solved using greedy methods. Under certain conditions, the l_0-norm can be replaced by the the l_1-norm, which is convex (Elad, 2010). The atoms in the dictionary D represent elemental geophysical features and can be represented by functions such as wavelets.



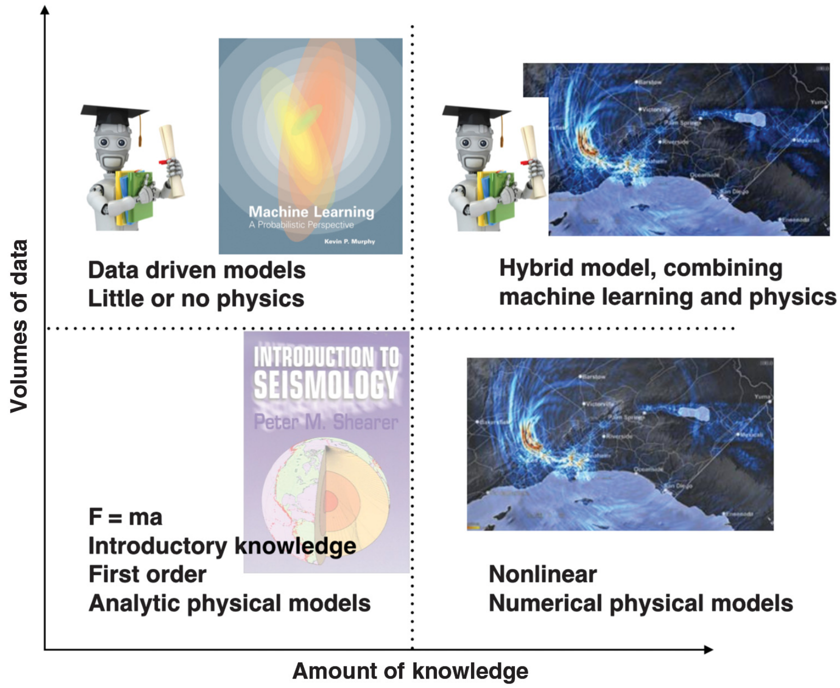
▲ **Figure 7.** Locally sparse travel-time tomography (LST; Bianco and Gerstoft, 2018) of checkerboard slowness. (a) Synthetic checkerboard slowness patterns with 100 × 100 pixel grid (km) are sampled by (b) 2016 straight rays from 64 seismic stations. (c) Conventional inversion using damping and smoothing regularization (Aster *et al.*, 2011) and (d) LST. Profiles from the 2D inversion are shown with true and estimated slownesses. The root mean square error (ms/km) estimated relative to the true slowness is printed on the 2D estimates. (e) Dictionary learned from LST contains checkerboard-like atom (100 atoms shown). Each atom (patch) is 10 × 10 pixels.

Sparse representations are appealing because they can model both discontinuous and smooth geophysical features (Loris *et al.*, 2007). In adaptive tomography, the dictionaries are learned directly from signal examples using dictionary learning, a form of unsupervised ML for which many algorithms exist (Mairal *et al.*, 2014). Such learned dictionaries can better represent specific signals than wavelets. Zhu *et al.* (2017) used this sparse and adaptive framework for FWI by iteratively learning the dictionary of seismic features, whereas Li and Harris (2018) incorporated a nonlocal similarity (Mairal *et al.*, 2014) in the dictionary learning procedure. Bianco and Gerstoft (2018) used a sparse and adaptive framework for 2D (surface wave) travel-time tomography, called locally sparse travel-time tomography (LST). Assuming dense, straight ray sampling, LST learns a dictionary of slowness features from patches of a least-squares inversion. The learned dictionary is then used to construct a sparse slowness model. Simulated travel-time inversions using the LST and conventional tomography (Aster *et al.*, 2011) methods are shown in Figure 7.

Seismic tomography approaches based on NNs, with early theory developed by Röth and Tarantola (1994), have also achieved compelling results. Moya and Irikura (2010) apply an NN approach to velocity model inversion. Gupta *et al.* (2018) address the challenges of limited measurements in travel-time tomography using subspace modeling and convolutional NNs. Moseley *et al.* (2018) present a fast approximate approach for seismic-wave propagation and inversion using deep learning, based on deep NN for speech synthesis. Araya-Polo *et al.* (2018) develop a formulation for FWI that replaces the iterative inversion scheme for velocity features with a deep NN. Lewis and Vigh (2017) use a deep NN FWI method to better detect salt domes. Such methods appear to be a future step for generative and inversion architectures.

Earthquake Geodesy and Noninertial Deformation

Although classical seismology has focused on high-frequency inertial deformation of the earth, the full spectrum of earthquake cycle behaviors also includes prolonged noninertial deformation (Ben-Zion, 2008). These motions include postseismic deformation (durations of years) and interseismic deformation (durations of decades), as well as slow or silent earthquakes (durations of weeks) (Peng and Gomberg, 2010; Beroza and Ide, 2011). Because these motions are noninertial, they are typically measured using geodetic techniques such as GPS and InSAR to estimate time-dependent displacements at Earth's surface. The precision of these measurements (GPS: ~0.1 mm/yr, InSAR: ~5 mm/yr) limits the resolving power of geodetic data to relatively large earthquakes that occur near Earth's surface. Earthquakes with $\mathbf{M} < 4$ are difficult to observe geodetically, because they are characterized by relatively small fault slip (< 1 cm) and static displacements in the elastic crust fall-off as the reciprocal of distance cubed for buried earthquakes. This suggests that there will be orders of magnitude fewer earthquakes observed geodetically than there are seismically. This bears on ML applications because it implies far fewer earthquakes available for creating labeled datasets from geodetic data. The case is similar for postseismic deformation as well as



▲ **Figure 8.** Geophysical insight will be maximized by leveraging the strengths of both physical and ML-based, data-driven models. Analytic physical models (lower left) give basic insights about physical systems. More sophisticated models, reliant on computational methods (lower right), can model more complex phenomena. Whereas physical models are reliant on rules, which are updated by physical evidence (data), ML is purely data-driven (upper left). By augmenting ML methods with physical models to obtain hybrid models (upper right), a synergy can be obtained that balances the complementary strengths of physical intuition with data-driven insights.

silent/slow-slip events. Currently, the total number of such geodetically observed events may be on the order of 1000.

Given this relative paucity of classically labeled data, ML applications to the noninertial part of the earthquake cycle may be somewhat different than those initially applied to seismic waveforms. In particular, seemingly obvious goals such as automating the search for slow/silent earthquakes may be challenging due to limited training data. Instead, other opportunities arise. For example, there are numerous nonexclusive mechanisms involved in postseismic deformation including both linear and nonlinear versions of afterslip, poroelasticity, and viscoelasticity. Here, ML approaches developed to infer the governing partial differential equations directly from observations (Rudy *et al.*, 2017; Long *et al.*, 2018; Raissi and Karniadakis, 2018) may play an essential role in resolving the nature and relative contributions of the mechanisms responsible for postseismic deformation. The core idea is that these ML approaches realize the mathematical structure of the governing physics (both linear and nonlinear) directly from observations of surface motions rather than relying on theory-driven concepts that have received traditional focus.

ML approaches also offer the possibility of radically accelerating generative models of earthquake cycle deformation. Numerical rupture and viscoelastic stress transfer models are widely used in earthquake science, but they are not ubiquitous. The primary reason for this is the computational cost of running these simulations and models. In some cases, it may be possible to train deep learning systems to emulate high-performance computing earthquake physics codes, so that they are represented in compact mathematical forms as NNs. The central concept here is that we tend to program calculations in terms of mathematical functions that are readily recognizable and comprehensible. However familiar these may be, there may exist far more compact nonlinear and nonorthogonal factorizations that enable the solution to be computed quickly, and NNs are free to construct over complete dictionary representation that may be vastly more computationally efficient (DeVries *et al.*, 2017; Moseley *et al.*, 2018).

OTHER APPLICATIONS AND FUTURE DIRECTIONS

There are many other exciting ML applications in our field of seismology. For example, the use of probabilistic graphical models and graph theory in seismology has become increasingly prevalent. The deployment of large-N arrays (Karplus and Schmandt, 2018) provides one such opportunity, in which weak event signals can be extracted using graph clustering (Riahi and Gerstoft, 2017) or similarity theory (Li, Peng, *et al.*, 2018). Separately, Trugman and Shearer (2017) use graph theory and hierarchical cluster analysis to obtain high-precision earthquake location estimates using differential travel times from pairs of earthquakes observed at a set of common stations. Telesca and Chelidze (2018) applied a visibility graph method to seismicity near a dam to find anomalous seismic activity.

Additional applications of ML to seismology extend well beyond the realm of graph theory. Araya-Polo *et al.* (2017) applied a deep NN trained on active seismic data for hydrocarbon exploration to detect subsurface fault structures. Krischer and Fichtner (2017) generate synthetic seismograms using GANs, training the networks using with synthetic seismic data. Using Bayesian networks, Hincks *et al.* (2018) modeled the joint conditional dependencies between parameters for the Oklahoma seismicity to understand the induced seismicity. Building on the preliminary analyses of Meade *et al.* (2017), DeVries *et al.* (2018) trained a deep NN to forecast aftershock locations using as input the static stress change tensor computed from finite-fault earthquake rupture models.

The ultimate realization of ML-based methods in seismology would leverage physical models to obtain synergy between the physical theory from domain scientists and the enhanced, data-driven constraints from ML and probability theory. Although the application of ML to seismology is becoming increasingly prevalent, ML is often currently applied without physical modeling (Fig. 8, upper left). Geophysical data sets tend to be poorly sampled, noisy, and incomplete and are often difficult to handle using standard ML techniques. Thus, often in seismology, we traditionally deferred to pure physics-based methods (Fig. 8, bottom). It would be transformative if we could develop a hybrid modeling framework that combines data-driven ML methods with explicit physical models (Fig. 8, upper right). This would provide a means of specifying a physical model as a component of the ML algorithm, or conversely, a means of using ML to train better physical parameterizations. Transparency of the learned algorithms would enable human learning and allow validation by testing for physical consistency.

In summary, seismology and ML benefit from each other. With its interesting problems and rich datasets, seismology supplies a real-world test bed for various ML algorithms, and even a driving force to compel the development of new algorithms. Although ML provides seismology with new tools to extract novel insights directly from the data, combining classical seismology techniques with ML in a hybrid approach might lead to radically new discoveries.

DATA AND RESOURCES

For further reading on machine learning (ML) fundamentals, we recommend the following textbooks and online course materials. Bishop (2006) is a more introductory text, whereas Murphy (2012) provides a more in-depth theoretical development. "Deep Learning" (Goodfellow *et al.*, 2016) provides a practical introduction to deep neural networks (NNs). There are also many excellent free online courses, such as Ng's "Machine Learning," Hinton's "Neural Networks for Machine Learning," Tibshirani and Hastie's "Statistical Learning," and Li *et al.*'s "Convolutional Neural Networks for Visual Recognition." This is not meant to be an exhaustive list of ML resources, but is a good place to get started.

ACKNOWLEDGMENTS

The authors thank Editor-in-Chief Zhigang Peng for the invitation to write this frontier article and three anonymous reviewers for their thoughtful comments which greatly improved the article. Q. Kong acknowledges support from the Gordon and Betty Moore Foundation through Grant Number GBMF5230 to UC Berkeley. D. Trugman acknowledges institutional support from the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory. Z. Ross acknowledges support from The Gordon and Betty Moore Foundation and the National Science Foundation. M. Bianco and P. Gerstoft acknowledge support from the Office of Naval Research (Grant Number N00014-18-1-2118).

REFERENCES

Aguiar, A. C., and G. C. Beroza (2014). PageRank for earthquakes, *Seismol. Res. Lett.* **85,** no. 2, 344–350.

Alavi, A. H., and A. H. Gandomi (2011). Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing, *Comput. Struct.* **89,** no. 23, 2176–2194.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68,** no. 5 1521–1532.

Allen, R. M., P. Gasparini, O. Kamigaichi, and M. Bose (2009). The status of earthquake early warning around the world: An introductory overview, *Seismol. Res. Lett.* **80,** no. 5, 682–693.

Alimoradi, A., and L. J. Beck (2014). Machine-learning methods for earthquake ground motion analysis and simulation, *J. Eng. Mech.* **141,** no. 4, 04014147, doi: 10.1061/(ASCE)EM.1943-7889.0000869.

Anderson, J. G., and J. N. Brune (1999). Probabilistic seismic hazard analysis without the Ergodic assumption, *Seismol. Res. Lett.* **70,** no. 1, 19–28.

Araya-Polo, M., T. Dahlke, C. Frogner, C. Zhang, T. Poggio, and D. Hohl (2017). Automated fault detection without seismic processing, *The Leading Edge* **36,** no. 3, 208–214.

Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke (2018). Deep-learning tomography, *The Leading Edge* **37,** no. 1, 58–66.

Aster, R. C., B. Borchers, and C. H. Thurber (2011). *Parameter Estimation and Inverse Problems*, Academic Press, Cambridge, Massachusetts.

Baltay, A. S., and T. C. Hanks (2014). Understanding the magnitude dependence of PGA and PGV in NGA-West 2 data, *Bull. Seismol. Soc. Am.* **104,** no. 6, 2851–2865.

Beaucé, E., W. B. Frank, and A. Romanenko (2018). Fast matched filter (FMF): An efficient seismic matched-filter search for both CPU and GPU architectures, *Seismol. Res. Lett.* **89,** no. 1, 165–172.

Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes, *Rev. Geophys.* **46,** no. 4, doi: 10.1029/2008RG000260.

Bergen, K. J., and G. C. Beroza (2018). Detecting earthquakes over a seismic network using single-station similarity measures, *Geophys. J. Int.* **213,** no. 3, 1984–1998.

Beroza, G. C., and S. Ide (2011). Slow earthquakes and nonvolcanic tremor, *Annu. Rev. Earth Planet. Sci.* **39,** no. 1, 271–296.

Beyreuther, M., C. Hammer, J. Wassermann, M. Ohrnberger, and T. Megies (2012). Constructing a hidden Markov model based earthquake detector: Application to induced seismicity, *Geophys. J. Int.* **189,** no. 1, 602–610.

Bianco, M. J., and P. Gerstoft (2018). Travel time tomography with adaptive dictionaries, *IEEE Trans. Comput. Imag.* doi: 10.1109/TCI.2018.2862644.

Bindi, D. (2017). The predictive power of ground-motion prediction equations, *Bull. Seismol. Soc. Am.* **107,** no. 2, 1005–1011.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer Verlag, New York, New York.

Boore, D. M. (1983). Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra, *Bull. Seismol. Soc. Am.* **73,** no. 6A, 1865–1894.

Böse, M., T. H. Heaton, and E. Hauksson (2012). Real-time finite fault rupture detector (FinDer) for large earthquakes, *Geophys. J. Int.* **191,** no. 2, 803–812.

Böse, M., F. Wenzel, and M. Erdik (2008). PreSEIS: A neural network-based approach to earthquake early warning for finite faults, *Bull. Seismol. Soc. Am.* **98,** no. 1, 366–382.

Bozorgnia, Y., N. A. Abrahamson, L. A. Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S.-J. Chiou, *et al.* (2014). NGA-West2 research project, *Earthq. Spectra* **30,** no. 3, 973–987.

Breiman, L. (2001). Random forests, *Mach. Learn.* **45,** no. 1, 5–32.

Campbell, K. W., and Y. Bozorgnia (2008). NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s, *Earthq. Spectra* **24,** no. 1, 139–171.

Chamberlain, C. J., C. J. Hopp, C. M. Boese, E. Warren-Smith, D. Chambers, S. X. Chu, K. Michailos, and J. Townend (2018). EQcorrscan: Repeating and near-repeating earthquake detection and analysis in Python, *Seismol. Res. Lett.* **89,** no. 1, 173–181.

Chen, Y. (2018). Automatic microseismic event picking via unsupervised machine learning, *Geophys. J. Int.* **212,** no. 1, 88–102.

Cua, G., and T. Heaton (2007). The virtual seismologist (VS) method: A Bayesian approach to earthquake early warning, in *Earthquake Early Warning Systems*, P. Gasparini, G. Manfredi, and J. Zschau (Editors), Springer Berlin Heidelberg, Berlin, Heidelberg, 97–132.

Cuéllar, A., G. Suárez, and J. M. Espinosa-Aranda (2018). A fast earthquake early warning algorithm based on the first 3 s of the *P*-wave coda, *Bull. Seismol. Soc. Am.* **108,** no. 4, 2068–2079.

Dai, H., and C. MacBeth (1995). Automatic picking of seismic arrivals in local earthquake data using an artificial neural network, *Geophys. J. Int.* **120,** no. 3, 758–774.

Derras, B., P. Y. Bard, and F. Cotton (2014). Towards fully data driven ground-motion prediction models for Europe, *Bull. Earthq. Eng.* **12,** no. 1, 495–516.

Derras, B., P.-Y. Bard, and F. Cotton (2016). Site-condition proxies, ground motion variability, and data-driven GMPEs: Insights from the NGA-West2 and RESORCE data sets, *Earthq. Spectra* **32,** no. 4, 2027–2056.

Derras, B., P.-Y. Bard, F. Cotton, and A. Bekkouche (2012). Adapting the neural network approach to PGA prediction: An example based on the KiK-net data, *Bull. Seismol. Soc. Am.* **102,** no. 4, 1446–1461.

DeVries, P. M. R., T. B. Thompson, and B. J. Meade (2017). Enabling large-scale viscoelastic calculations via neural network acceleration, *Geophys. Res. Lett.* **44,** no. 6, 2662–2669.

DeVries, P. M. R., F. Viégas, M. Wattenberg, and B. J. Meade (2018). Deep learning of aftershock patterns following large earthquakes, *Nature* **560,** no. 7720, 632–634.

Douglas, J., and B. Edwards (2016). Recent and future developments in earthquake ground motion estimation, *Earth Sci. Rev.* **160,** 203–219.

Dysart, P. S., and J. J. Pulli (1990). Regional seismic event classification at the NORESS array: Seismological measurements and the use of trained neural networks, *Bull. Seismol. Soc. Am.* **80,** no. 6B, 1910–1933.

Elad, M. (2010). From exact to approximate solutions, in *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, M. Elad (Editor), Springer New York, New York, New York, 79–109.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, *Kdd Proceedings*, Vol. 96, 226–231.

Fedorenko, Y. V., E. S. Husebye, and B. O. Ruud (1999). Explosion site recognition; neural net discriminator using single three-component stations, *Phys. Earth Planet. In.* **113,** no. 1, 131–142.

Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* **165,** no. 1, 149–166.

Goodfellow, I., Y. Bengio, A. Courville, and Y. Bengio (2016). *Deep Learning*, Vol. 1, MIT Press, Cambridge, Massachusetts.

Greenspan, H., B. van Ginneken, and R. M. Summers (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique, *IEEE Trans. Med. Imag.* **35,** no. 5, 1153–1159.

Gupta, S., K. Kothari, M. V. de Hoop, and I. Dokmanić (2018). *Deep Mesh Projectors for Inverse Problems*, available at http://arxiv.org/abs/1805.11718v1 (last accessed October 2018).

Hincks, T., W. Aspinall, R. Cooke, and T. Gernon (2018). Oklahoma's induced seismicity strongly linked to wastewater injection depth, *Science* **359,** no. 6381, 1251–1255.

Jordan, M. I., and T. M. Mitchell (2015). Machine learning: Trends, perspectives, and prospects, *Science* **349,** no. 6245, 255–260.

Karplus, M., and B. Schmandt (2018). Preface to the focus section on geophone array seismology, *Seismol. Res. Lett.* **89,** no. 5, doi: 10.1785/0220180212.

Kato, A., K. Obara, T. Igarashi, H. Tsuruoka, S. Nakagawa, and N. Hirata (2012). Propagation of slow slip leading up to the 2011 M_w 9.0 Tohoku-Oki earthquake, *Science* **335,** no. 6069, 705–708.

Khaleghi, B., A. Khamis, F. O. Karray, and S. N. Razavi (2013). Multisensor data fusion: A review of the state-of-the-art, *Int. J. Inf. Fusion* **14,** no. 1, 28–44.

Khoshnevis, N., and R. Taborda (2018). Prioritizing ground-motion validation metrics using semisupervised and supervised learning, *Bull. Seismol. Soc. Am.* **108,** no. 4, 2248–2264, doi: 10.1785/0120180056.

Kong, Q., R. M. Allen, and L. Schreier (2016). MyShake: Initial observations from a global smartphone seismic network, *Geophys. Res. Lett.* **43,** no. 18, 9588–9594.

Kong, Q., R. M. Allen, L. Schreier, and Y.-W. Kwon (2016). MyShake: A smartphone seismic network for earthquake early warning and beyond, *Sci. Adv.* **2,** no. 2, e1501055, doi: 10.1126/sciadv.1501055.

Krischer, L., and A. Fichtner (2017). Generating seismograms with deep neural networks, *AGU Fall Meeting Abstracts*, New Orleans, Louisiana, 11–15 December 2017.

LeCun, Y., Y. Bengio, and G. Hinton (2015). Deep learning, *Nature* **521,** no. 7553, 436–444.

Lewis, W., and D. Vigh (2017). Deep learning prior models from seismic images for full-waveform inversion, *SEG Technical Program Expanded Abstracts 2017*, Society of Exploration Geophysicists, 1512–1517.

Li, D., and J. M. Harris (2018). Full waveform inversion with nonlocal similarity and model-derivative domain adaptive sparsity-promoting regularization, *Geophys. J. Int.* doi: 10.1093/gji/ggy380.

Li, Z., M.-A. Meier, E. Hauksson, Z. Zhan, and J. Andrews (2018). Machine learning seismic wave discrimination: Application to earthquake early warning, *Geophys. Res. Lett.* **45,** no. 10, 4773–4779.

Li, Z., Z. Peng, D. Hollis, L. Zhu, and J. McClellan (2018). High-resolution seismic event detection using local similarity for Large-N arrays, *Sci. Rep.* **8,** no. 1, 1646.

Long, Z., Y. Lu, X. Ma, and B. Dong (2018). PDE-Net: Learning PDEs from Data, *Proc. of the 35th International Conference on Machine Learning*, 3208–3216.

Loris, I., G. Nolet, I. Daubechies, and F. A. Dahlen (2007). Tomographic inversion using ℓ1-norm regularization of wavelet coefficients, *Geophys. J. Int.* **170,** no. 1, 359–370.

Mairal, J., F. Bach, and J. Ponce (2014). Sparse modeling for image and vision processing, *Found. Trends Comput. Graphics and Vision* **8,** nos. 2/3, 85–283.

Meade, B. J., P. M. R. DeVries, J. Faller, F. Viegas, and M. Wattenberg (2017). What is better than Coulomb failure stress? A ranking of scalar static stress triggering mechanisms from 10^5 mainshock-aftershock pairs, *Geophys. Res. Lett.* **44,** no. 22, doi: 10.1002/2017GL075875.

Meier, M.-A., T. Heaton, and J. Clinton (2015). The Gutenberg algorithm: Evolutionary Bayesian magnitude estimates for earthquake early warning with a filter bank, *Bull. Seismol. Soc. Am.* **105,** no. 5, 2774–2786.

Moseley, B., A. Markham, and T. Nissen-Meyer (2018). *Fast Approximate Simulation of Seismic Waves with Deep Learning*, available at http://arxiv.org/abs/1807.06873v1 (last accessed October 2018).

Moya, A., and K. Irikura (2010). Inversion of a velocity model using artificial neural networks, *Comput. Geosci.* **36,** no. 12, 1474–1483.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*, MIT Press, Cambridge, Massachusetts.

Musil, M., and A. Plešinger (1996). Discrimination between local microearthquakes and quarry blasts by multi-layer perceptrons and Kohonen maps, *Bull. Seismol. Soc. Am.* **86,** no. 4, 1077–1090.

Ochoa, L. H., L. F. Niño, and C. A. Vargas (2018). Fast magnitude determination using a single seismological station record implementing machine learning techniques, *Geodes. Geodynam.* **9,** no. 1, 34–41.

Peng, Z., and J. Gomberg (2010). An integrated perspective of the continuum between earthquakes and slow-slip phenomena, *Nature Geosci.* **3,** 599.

Peng, Z., and P. Zhao (2009). Migration of early aftershocks following the 2004 Parkfield earthquake, *Nature Geosci.* **2,** 877.

Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Sci. Adv.* **4,** no. 2, e1700578.

Raissi, M., and G. E. Karniadakis (2018). Hidden physics models: Machine learning of nonlinear partial differential equations, *J. Comput. Phys.* **357,** 125–141.
Ravishankar, S., and Y. Bresler (2011). MR image reconstruction from highly undersampled k-space data by dictionary learning, *IEEE Trans. Med. Imag.* **30,** no. 5, 1028–1041.
Rawlinson, N., S. Pozgay, and S. Fishwick (2010). Seismic tomography: A window into deep Earth, *Phys. Earth Planet. In.* **178,** no. 3, 101–135.
Riahi, N., and P. Gerstoft (2017). Using graph clustering to locate sources within a dense sensor array, *Signal Process.* **132,** 110–120.
Ross, Z. E., M.-A. Meier, and E. Hauksson (2018). *P*-wave arrival picking and first-motion polarity determination with deep learning, *J. Geophys. Res.* **123,** 5120–5129.
Ross, Z. E., M.-A. Meier, E. Hauksson, and T. H. Heaton (2018). Generalized seismic phase detection with deep learning, *Bull. Seismol. Soc. Am.* doi: 10.1785/0120180080.
Ross, Z. E., C. Rollins, E. S. Cochran, E. Hauksson, J.-P. Avouac, and Y. Ben-Zion (2017). Aftershocks driven by afterslip and fluid pressure sweeping through a fault-fracture mesh, *Geophys. Res. Lett.* **44,** no. 16, 8260–8267.
Röth, G., and A. Tarantola (1994). Neural networks and inversion of seismic data, *J. Geophys. Res.* **99,** no. B4, 6753.
Rouet-Leduc, B., C. Hulbert, N. Lubbers, K. Barros, C. J. Humphreys, and P. A. Johnson (2017). Machine learning predicts laboratory earthquakes, *Geophys. Res. Lett.* **44,** no. 18, 9276–9282.
Rudy, S. H., S. L. Brunton, J. L. Proctor, and J. N. Kutz (2017). Data-driven discovery of partial differential equations, *Sci. Adv.* **3,** no. 4, e1602614, doi: 10.1126/sciadv.1602614.
Sagiroglu, S., and D. Sinanc (2013). Big data: A review, *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47.
Shelly, D. R., G. C. Beroza, and S. Ide (2007). Non-volcanic tremor and low-frequency earthquake swarms, *Nature* **446,** no. 7133, 305–307.
Telesca, L., and T. Chelidze (2018). Visibility graph analysis of seismicity around Enguri high arch dam, Caucasus, *Bull. Seismol. Soc. Am.* doi: 10.1785/0120170370.
Thomas, S., G. N. Pillai, K. Pal, and P. Jagtap (2016). Prediction of ground motion parameters using randomized ANFIS (RANFIS), *Appl. Soft Comput.* **40,** 624–634.
Tiira, T. (1999). Detecting teleseismic events using artificial neural networks, *Comput. Geosci.* **25,** no. 8, 929–938.
Trugman, D. T., and P. M. Shearer (2017). GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish springs and Sheldon, Nevada, earthquake sequences, *Seismol. Res. Lett.* **88,** no. 2A, 379–391.
Trugman, D. T., and P. M. Shearer (2018). Strong correlation between stress drop and peak ground acceleration for recent M 1-4 earthquakes in the San Francisco bay area, *Bull. Seismol. Soc. Am.* **108,** no. 2, 929–945.
Ursino, A., H. Langer, L. Scarfì, G. Di Grazia, and S. Gresta (2001). Discrimination of quarry blasts from tectonic microearthquakes in the Hyblean plateau (southeastern Sicily), *Ann. Geophys.* **44,** no. 4, doi: 10.4401/ag-3569.
Virieux, J., and S. Operto (2009). An overview of full-waveform inversion in exploration geophysics, *Geophysics* **74,** no. 6, WCC1–WCC26.
Wiszniowski, J., B. M. Plesiewicz, and J. Trojanowski (2014). Application of real time recurrent neural network for detection of small natural earthquakes in Poland, *Acta Geophys.* **62,** no. 3, 469–485.
Witten, I. H., E. Frank, M. A. Hall, and C. J. Pal (2016). *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Burlington, Massachusetts.
Wu, Y., Y. Lin, Z. Zhou, D. C. Bolton, J. Liu, and P. Johnson (2018). DeepDetect: A cascaded region-based densely connected network for seismic event detection, *IEEE Trans. Geosci. Remote Sens.* A Publication of the IEEE Geoscience and Remote Sensing Society, 1–14.
Yenier, E., and G. M. Atkinson (2015). Regionally adjustable generic ground-motion prediction equation based on equivalent point-source simulations: Application to central and eastern North America, *Bull. Seismol. Soc. Am.* **105,** no. 4, 1989–2009.
Yin, L., J. Andrews, and T. Heaton (2018). Rapid earthquake discrimination for earthquake early warning: A Bayesian probabilistic approach using three-component single-station waveforms and seismicity forecast, *Bull. Seismol. Soc. Am.* **108,** no. 4, 2054–2067.
Yoon, C. E., O. O'Reilly, K. J. Bergen, and G. C. Beroza (2015). Earthquake detection through computationally efficient similarity search, *Sci. Adv.* **1,** no. 11, e1501057, doi: 10.1126/sciadv.1501057.
Zhang, J., H. Zhang, E. Chen, Y. Zheng, W. Kuang, and X. Zhang (2014). Real-time earthquake monitoring using a search engine method, *Nature Commun.* **5,** 5664.
Zhao, Y., and K. Takano (1999). An artificial neural network approach for broadband seismic phase picking, *Bull. Seismol. Soc. Am.* **89,** no. 3, 670–680.
Zhu, L., E. Liu, and J. McClellan (2017). Sparse-promoting full-waveform inversion based on online orthonormal dictionary learning, *Geophysics* **82,** no. 2, R87–R107.
Zhu, W., and G. C. Beroza (2018). *PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method*, available at http://arxiv.org/abs/1803.03211v1 (last accessed October 2018).

Qingkai Kong
Berkeley Seismological Laboratory
University of California, Berkeley
209 McCone Hall
Berkeley, California 94720 U.S.A.
kongqk@berkeley.edu

Daniel T. Trugman
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, New Mexico 87545 U.S.A.

Zachary E. Ross
Seismological Laboratory
California Institute of Technology
Pasadena, California 91125 U.S.A.

Michael J. Bianco
Peter Gerstoft
Scripps Institution of Oceanography
University of California San Diego
9500 Gilman Drive
La Jolla, California 92093-0238 U.S.A.

Brendan J. Meade
Department of Earth and Planetary Sciences
Harvard University
Cambridge, Massachusetts 02138 U.S.A.

Published Online 14 November 2018