ADVANCING EARTH AND SPACE SCIENCE

Reviews of Geophysics

REVIEW ARTICLE

10.1029/2021RG000769

Big Data Seismology

S. J. Arrowsmith[1], D. T. Trugman[2], J. MacCarthy[3], K. J. Bergen[4,5], D. Lumley[6], and M. B. Magnani[1]

[1]Roy M. Huffington Department of Earth Sciences, Southern Methodist University, Dallas, TX, USA, [2]Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, USA, [3]Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA, [4]Data Science Initiative, Brown University, Providence, RI, USA, [5]Department of Earth, Environmental and Planetary Sciences, Brown University, Providence, RI, USA, [6]Department of Geosciences and Physics, University of Texas at Dallas, Richardson, TX, USA

Key Points:

- Big Data Seismology is an emergent subdiscipline that uses "big data" inquiries to explore fundamental science questions in seismology
- Three drivers of Big Data Seismology are the growth of large data volumes, the development of new algorithms, and advances in computing
- Big Data Seismology is being applied to study earthquakes, to better resolve Earth structure, and to open new frontiers in seismology

Correspondence to:

S. J. Arrowsmith,
sarrowsmith@smu.edu

Citation:

Arrowsmith, S. J., Trugman, D. T., MacCarthy, J., Bergen, K. J., Lumley, D., & Magnani, M. B. (2022). Big Data Seismology. *Reviews of Geophysics*, *60*, e2021RG000769. https://doi.org/10.1029/2021RG000769

Received 31 DEC 2021
Accepted 18 APR 2022

Author Contributions:

Conceptualization: S. J. Arrowsmith
Visualization: S. J. Arrowsmith, D. T. Trugman, J. MacCarthy, K. J. Bergen, D. Lumley
Writing – original draft: S. J. Arrowsmith, D. T. Trugman, J. MacCarthy, K. J. Bergen, D. Lumley, M. B. Magnani
Writing – review & editing: S. J. Arrowsmith, D. T. Trugman, J. MacCarthy, K. J. Bergen, D. Lumley, M. B. Magnani

Abstract The discipline of seismology is based on observations of ground motion that are inherently undersampled in space and time. Our basic understanding of earthquake processes and our ability to resolve 4D Earth structure are fundamentally limited by data volume. Today, Big Data Seismology is an emergent revolution involving the use of large, data-dense inquiries that is providing new opportunities to make fundamental advances in these areas. This article reviews recent scientific advances enabled by Big Data Seismology through the context of three major drivers: the development of new data-dense sensor systems, improvements in computing, and the development of new types of techniques and algorithms. Each driver is explored in the context of both global and exploration seismology, alongside collaborative opportunities that combine the features of long-duration data collections (common to global seismology) with dense networks of sensors (common to exploration seismology). The review explores some of the unique challenges and opportunities that Big Data Seismology presents, drawing on parallels from other fields facing similar issues. Finally, recent scientific findings enabled by dense seismic data sets are discussed, and we assess the opportunities for significant advances made possible with Big Data Seismology. This review is designed to be a primer for seismologists who are interested in getting up-to-speed with how the Big Data revolution is advancing the field of seismology.

Plain Language Summary Seismology is the scientific discipline by which vibrational waves that travel through the Earth are used to study a range of natural processes, from the structure and layering of the subsurface to the dynamics of earthquake rupture. Long propelled by advances in instrumentation and the availability of waveform observations, seismology has entered an era of Big Data where new opportunities come hand in hand with new challenges. This review article gives an overview of recent scientific advances powered by Big Data Seismology, along with emerging trends, potential challenges, and future opportunities. The review encompasses a range of subjects from seismic imaging and earthquake source physics to education and outreach, and is designed to be a jumping off point for readers interested in how big data is changing the face of seismology research.

1. Introduction

1.1. Historical Perspective

Advances in the science of seismology have been so dependent upon advances in instrument-based measurements of ground motion that the significance of the seismometer is analogous to the telescope in astronomy. A number of important advances in instrumentation, new types of network, and new experiment designs have been closely followed by significant discoveries.

1.1.1. Passive Seismology

Advances in seismic instrumentation in the late 1880's enabled the first teleseismic earthquake to be recorded in 1889 (Dewey & Byerly, 1969; Dziewonski, 2018), closely followed by the discovery of the major layers of Earth in the early 1900's, such as the core-mantle (Oldham, 1906) and crust-mantle discontinuities. The value of seismic networks in seismology was recognized early on, but the first networks were of limited geographical scope. In the early 1960's, the World Wide Standardized Seismographic Network (WWSSN) was created with the goal of improving the capability to detect and identify nuclear tests (Oliver & Murphy, 1971). The WWSSN was the first global scale network that comprised standardized equipment and, crucially, incorporated procedures for sharing waveform data into the design (via photographic paper). The WWSSN established procedures that are now common practice in seismology (i.e., data sharing, continuous waveform recordings) and enabled seismologists to increase the number of earthquakes detected and located and build the first accurate global seismicity maps at the same time that geomagnetic observations provided key evidence of seafloor spreading. The sharing of waveform data also enabled the first maps of focal mechanisms to be produced using the WWSSN (Sykes, 1967). These two data products—seismicity maps and earthquake focal mechanisms—played a key role in the early development and confirmation of plate tectonic theory (Isacks et al., 1968) and its relation to earthquake processes.

A seismometer can be characterized by its dynamic range (the range of amplitudes it can measure) and bandwidth (the range of frequencies of motion it can record). WWSSN sites comprised short-period and long-period instruments that were designed to prevent the microseism peak from contaminating analog recordings. Further developments in seismic instrumentation in the early 1970's and 80's—in particular force feedback and very-broadband (VBB) seismometers—resulted in vast improvements in instrumental recordings of earthquakes. The recognition of the value of digital broadband data inspired the development of the earliest digital networks, most notably the GEOSCOPE network (Romanowicz et al., 1984) which included 20 stations at the end of the 1980's and currently comprises 33 sites. In the late 1980's, inspired in part by the success of GEOSCOPE, the Global Seismographic Network (GSN), was established by the Incorporated Research Institutions for Seismology (IRIS). Unlike the WWSSN, which was designed and funded for an applied purpose, the motivation for the GSN network was scientific. Early GSN stations used VBB Streckeisen STS-1 instruments (Wielandt & Steim, 1986) and 24-bit Quanterra digitizers (Butler et al., 2004). By the mid 1990's, the GSN network had reached the present-day footprint. In place of sharing photographic paper by mail, GSN data were telemetered and made available to researchers in near-real time by the IRIS Data Management System (IRIS-DMC). Early digital seismic data enabled multiple revolutions in global seismology including the first catalogs of moment tensor solutions (Dziewonski & Woodhouse, 1983), the first finite-fault inversions of rupture processes (see Hayes [2017] for a review), and the first global tomographic models (Woodhouse & Dziewonski, 1984), which provide key constraints on the structure and dynamics of the mantle. Science advances enabled by GSN data continue today; the reader is referred to the recent review by Ringler et al. (2022) for more details on the history of global seismic networks and related scientific discoveries.

At local and regional scales, a variety of long-running networks that include several national networks have been installed and operated primarily for real-time earthquake monitoring. These networks have also provided invaluable data for scientific research, which has been facilitated by the establishment of the International Federation of Digital Seismograph Stations (FDSN) in 1986 (Suárez et al., 2008). The FDSN provides high-quality seismic data through standard data formats—in particular the Standard for the Exchange of Earthquake Data (SEED) data format—and through exchange protocols. Data from the GSN stations are also shared through the FDSN, forming the global backbone of stations upon which local, regional, and national data sets are superimposed. The first earthquake with an estimated $M_W > 9$ to be recorded by GSN and FDSN stations—the 2004 Sumatra earthquake—provided a key demonstration of the scientific and hazard applications enabled by these modern networks (Ammon et al., 2010; Lay et al., 2005; J. Park et al., 2005). The exceptional bandwidth and dynamic range of modern permanent seismic networks through GSN and FDSN led B. Romanowicz and D. Giardini to state: "It is no longer the quality of the data, but primarily the spatial resolution, the centralized archiving, and the continuity in time of the archives that will be critical for progress in understanding the dynamics of the solid earth and the generation of earthquakes" (Romanowicz & Giardini, 2001).

In addition to permanent networks, portable deployments of seismometers have played an important role in seismic research. Since 1988, the development of the IRIS Program for Array Seismic Studies of the Continental Lithosphere (IRIS-PASSCAL) program made it much easier for individual researchers in the US to conduct experiments through a shared equipment pool, with data shared via the IRIS-DMC (Aster et al., 2005). Similar programs were also developed in other countries such as Germany, the UK, Australia, and New Zealand (Hammond et al., 2019). Scientific results from a number of portable experiments identified variations in seismic properties of the continental lithosphere at a range of spatial scales (Humphreys & Dueker, 1994). These studies, as well as a project called "Skippy" that involved a rolling network of seismometers across Australia (van der Hilst et al., 1994) motivated the comprehensive Earthscope program, which included a rolling array of 400

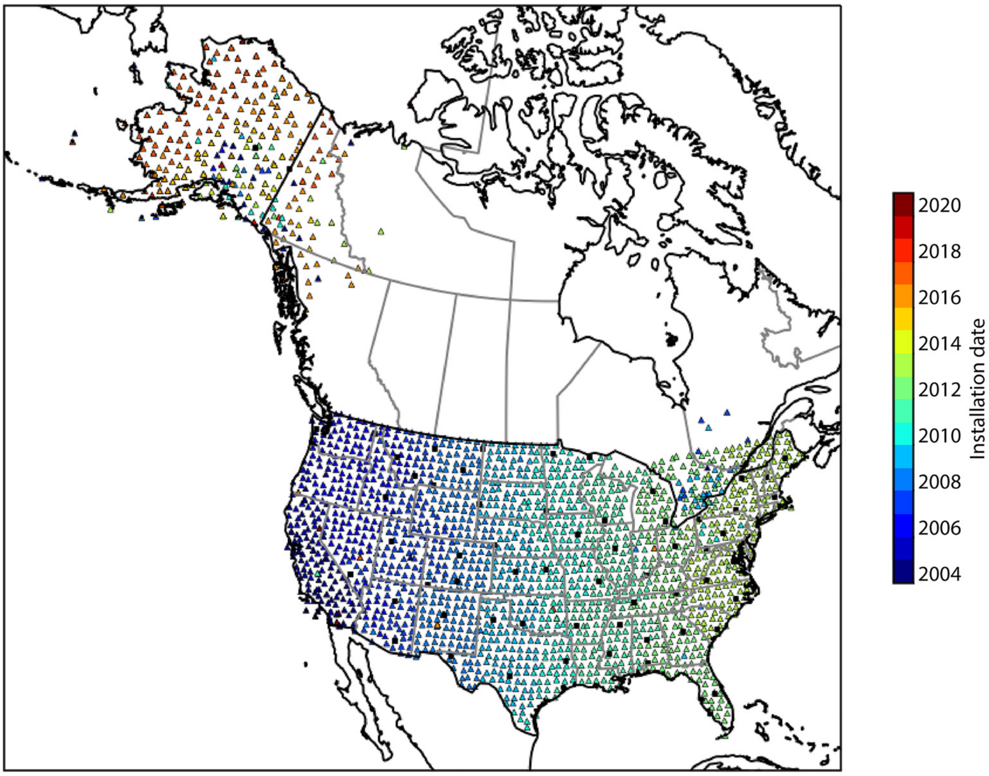

Figure 1. Map showing the locations of sensors in the USArray Transportable Array (triangles color-coded by the installation date) and permanent stations in the US National Seismic Network and Global Seismographic Network (black squares).

seismometers across the continental US and Alaska (USArray; Kerr, 2013) (see Figure 1). The USArray data has contributed to a wide variety of scientific studies, but in particular has allowed 3D spatial variations of the seismic wavespeed and anisotropy of the continental lithosphere beneath the North American continent to be resolved at unprecedented resolution (Burdick et al., 2014; Lin et al., 2011; Long et al., 2014; W. Shen & Ritzwoller, 2016). These observations have improved our understanding of the relationship between mantle processes and tectonic features at the surface (Moschetti et al., 2010).

Since seismometers only record the temporal characteristics of a spatiotemporal wavefield at a single spatial location, arrays of seismometers have been used to record both spatial and temporal properties of seismic waves since the 1960's (Rost & Thomas, 2002). Traditional seismic arrays, deployed such that signals of interest are coherent in both amplitude and phase at each array element, allow for the spatial filtering of the wavefield to extract signals at lower signal-to-noise ratios and can provide the direction-of-arrival (DOA) of an incoming wave. The use of traditional arrays in seismology has been limited by the cost and practicality (e.g., their geographical footprint requires more land access) of deploying and maintaining them. In the passive seismology community, seismic arrays have been more widely used for nuclear explosion monitoring (Koper, 2019) because they improve signal-to-noise ratio and provide backazimuth information for incoming events, which is important for monitoring at regional and teleseismic distances. Seismic arrays form the backbone of the seismic component of the International Monitoring System (IMS) network, a global network formed in 1996 that assimilates data from stations operated by different countries (Dahlman et al., 2011). As we explore below, techniques developed for processing data from seismic arrays are being leveraged or adapted for new Big Data Seismology data sets.

1.1.2. Exploration Seismology

The use of exploration seismology dates to the early 1900s, when the method was first used for the detection of a geological formation in the subsurface of Oklahoma near the Vines Branch area (Karcher, 1987). The experiment used small dynamite charges as seismic sources and three receivers positioned at increasing distances from the source along a line perpendicular to the shooting line. The shots were fired sequentially and the signal recorded

10.1029/2021RG000769

on photographic film. The timing of the recorded reflected signal, the layout of the source and receivers and knowledge of wave propagation speed in the subsurface allowed the pioneering team of geologists and physicists (William P. Haseman, J. Clarence Karcher, Irving Perrine, and Daniel W. Ohern) to calculate the distances traveled by the waves and the bouncing points in the subsurface, and therefore to reconstruct the depth of the reflecting horizon.

Today, seismic reflection profiles are typically generated by firing thousands of shots recorded by 300,000+ channels. Reflection and refraction data are acquired worldwide both on land and in marine environments, and exploration seismology is one of the most relied upon imaging methods for energy and resource exploration, geohazard assessments and scientific investigation of the crust and upper mantle. This expansion has been possible through numerous key innovations in instrumentation, survey design, and digital data analysis that took place over the years. For a comprehensive review of the history and development of the exploration seismology method, we refer the reader to Yilmaz (2001), Lawyer et al. (2001), and Bednar (2005).

1.1.2.1. Instrumentation

The invention and development of the vibroseis method in the early 1950s (Anstey, 1991; Lindseth, 1968) and the tuned airgun array concept (Avedik et al., 1996; Dragoset, 1984; Giles, 1968) revolutionized the seismic exploration field by providing control over the frequency characteristics of the seismic impulse produced by a seismic source. Around the same time, progress in seismometer construction dramatically reduced the size and weight of geophones, allowing flexibility in field operations and the employment of large numbers of receivers.

Improvement in marine acquisition came later, around 1980, when electronic digitizing modules were introduced within the seismic streamer, and signals from many sensors could be multiplexed and transmitted using just a few wires, thereby eliminating the limit on the number of receiver stations that could fit inside the streamer (Peardon & Cameron, 1986). Streamer lengths grew to thousands of meters and to more than 10,000 receiving channels. With longer cables came greater noise, as the constant motion of the streamer through the water generates pressure that is transmitted to the hydrophones. By the 1990s, solid streamers became available, and the light oil filling the streamer to provide neutral buoyancy was replaced by solid material, dramatically improving the signal-to-noise ratio (SNR; Brink & Spackman, 2004).

1.1.2.2. Survey Design

Two major innovations have improved survey design since the inception of the exploration seismology method: the development of the Common Mid Point (CMP) method, and 3D acquisition design. The concept of the CMP method, introduced in 1960 (Mayne, 1962), would eventually revolutionize the exploration seismology field and the way the seismic surveys were designed and processed. The CMP method derived from the use of source and receiver arrays to improve SNR, a practice that dates back to the late 1930s, when geophone arrays of 3–6 elements were used for seismic exploration surveys. Array elements grew to 100 by the late 1940s, and by the 1950s work on source and receiver array design boomed, and techniques were developed to preferentially attenuate noise traveling horizontally (e.g., ground roll) and to enhance vertically traveling wavefronts (Mosher & Simpkin, 1999). The CMP method hinged on the acquisition of redundant data in large arrays, and on using a time-shift design that would allow beamforming and thus the suppression not only of the incoherent noise (a goal already achieved by the existing receiver arrays) but also the coherent noise, such as multiple reverberations (Martin et al., 2000). The CMP method involved the acquisition and processing of unprecedented amounts of data, and its implementation forced a revolution in survey design and would not have been possible without innovations in data recording and storage. On land, roll-along methods were developed to expedite field operations. Data were recorded and stored on magnetic tapes, which later were replaced by digital tapes.

The expansion from 2D to 3D survey geometry began in the late 1960s to overcome the limitations and inadequacies of 2D imaging, particularly in complex regions. The transition was not without problems, particularly because 3D surveys were costly and lengthy to acquire and required new and more computationally intensive processing routines. The simultaneous recording of sources on multiple receiver lines (or streamers) demanded precise information of source-receiver positions at each time, which in turn required measuring the coordinates of all receivers (and the shape of the streamer) for each shot. In marine acquisitions the introduction of digital compasses along streamers (Goutorbe & Combier, 2010; later replaced by networks of acoustic receivers and transmitters distributed across the towed streamers), combined with the recorded ship position, produced absolute locations for receiver positions (Manin et al., 1988). As the number of streamers towed grew from one to eight in

10.1029/2021RG000769

the 1990s, the amount of metadata associated with a 3D survey grew comparable to the amount of recorded trace data. Today 3D marine acquisitions commonly tow more than 10 cables (up to 16), and streamers have reached lengths of more than 12 km and thousands of channels, allowing spatially dense recordings of sampled reflected and refracted arrivals. The improvement in streamer location, the introduction of steerable streamers, new and more accurate navigation technology, new noise reduction algorithms, and high-density 3D surveys paved the way for further expansion to 4D and time-lapse seismic surveys. Repeatability of seismic surveys was improved by these innovations, and today the use of the 4D method extends beyond detecting changes in hydrocarbon reservoirs to applications in monitoring carbon capture and storage operations, waste water injection, to mention a few.

1.1.2.3. Processing

The evolution of survey design and instrumentation in exploration seismology in the last hundred years has followed a path that points decisively toward a stunning increase in the amount of data recorded both in space (i.e., number of individual traces) and time (i.e., number of samples per trace; Ben-Zion et al., 2015; L. Han et al., 2016; Kiser et al., 2016). Digital technology innovations during this time period allowed the expansion in data volume recorded, stored and processed. Notably, the increase in computer processing power has been critical to the development and enhancement of seismic data processing functions. Of all the steps in seismic processing, imaging is perhaps the most complex and important, and one that has evolved through the years as more densely sampled data, faster computing, and then 3D data became available (Bednar, 2005). Today, as discussed in detail in Section 3, Full-Waveform Inversion (FWI) is the frontier of imaging and velocity model estimation, where all elements of the recorded wavefield—primaries and multiples, reflections and refractions—contribute to building the subsurface image. In addition, new recent advances in instrumentation are pushing the boundaries toward the overarching goal of constructing a truthful subsurface image. New, smaller micro-electro mechanical systems (MEMS) accelerometers have appeared on the market in the past few years (Mustafazade et al., 2020). These accelerometers can record in three components (3C), allowing multicomponent surveys, the recording of P- and S-waves, and of converted waves (Mougenot & Thorburn, 2004). Multicomponent surveys are now possible both on land, with 3C receivers, and in marine environments with the use of ocean bottom cables recording with three seafloor accelerometers plus a hydrophone (4C). Versatility in land sources, such as P- and S- wave vibroseis (i.e., 3C sources) shooting on 3C receivers results in recording of the full wavefield, providing further improvement in imaging and subsurface characterization (Gaiser & Strudley, 2005). Repeating the 3D seismic surveys and imaging procedure over calendar time, a technique developed in the 1990s, has become known as time-lapse (4D) seismology (R. Calvert, 2005; Lumley, 2001; Lumley et al., 2015; Nur, 1989; van Gestel, 2021).

One hundred years after that first experiment in Vines Branch, we are able to resolve structures of the Earth's crust down to meters, and to illuminate the lithosphere to depths of 150 km. Worldwide, knowledge of the structure and composition of the Earth's crust and uppermost mantle has come primarily from analyses of seismic reflection and refraction data. Such imaging power has defined the characteristics of lithospheric plates (A. Calvert et al., 1995; Christeson et al., 2007; Nelson et al., 1985; Oncken et al., 2003), their compositional elements and boundaries (Contreras-Reyes et al., 2008; Cook et al., 2004; Singh et al., 1999), and delineated the structures along which they deform (J. Li et al., 2018; Mackenzie et al., 2005; Makovsky et al., 1996). In combination with the imaging power of passive seismology, exploration seismology surveys have enabled discoveries that have laid the foundations of plate tectonic theory (Mutter, 1985; Vera et al., 1990; Zhao et al., 1993). Thanks to the resolving power of exploration seismology, methods such as seismic stratigraphy were developed to unravel how sedimentary successions are built, and, in the process, to understand the factors governing sedimentation in basins at particular times in earth history (Christie-Blick & Driscoll, 1995; Vail, 1987; Vail et al., 1977). Today seismic exploration methods are employed at all scales, from the ultra-shallow to lithospheric, to address questions and processes that range from the anthropogenic impact on soils to the detection of thermohaline staircase structures in the world's oceans (Holbrook et al., 2003). Exploration seismology is a thriving business, expanding its application to new fields while bridging disciplines. Section 2.1.5 further explores the current state-of-the-art and future opportunities in exploration seismology.

1.2. Big Data Seismology

This paper explores the emergence of a subdiscipline we refer to as Big Data Seismology, which is characterized by the use of "big data" inquiries to answer fundamental science questions in seismology that have remained elusive or have only emerged due to vast increases in data volume. While "big data" is a relative term, and the

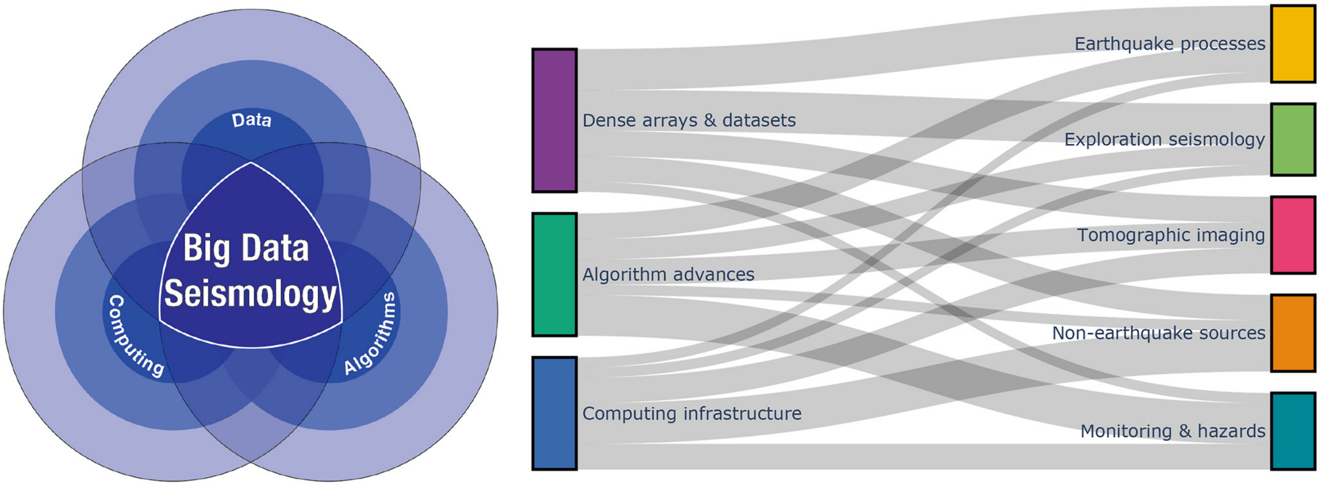

Figure 2. The growth of three drivers define Big Data Seismology: (a) dense arrays and data sets, (b) novel algorithms that can take advantage of these data, and (c) advances in computing infrastructure. Improvements in these three areas are leading to advances in fundamental and applied seismology, as outlined in the chart on the right.

definition of "big" is distinct for passive and exploration seismology communities, Big Data Seismology is defined in this paper as the set of seismological investigations made possible by advances in three areas: new types of data-dense sensor deployments, advances in computing, and the development of new types of data-hungry algorithm (Figure 2). Big Data Seismology practitioners will benefit from learning key lessons from both passive and exploration seismology communities. In particular, while the passive seismology community has decades of experience with sparse networks but continuous data recording, the exploration seismology community has extensive experience with dense but short-duration networks. This paper begins by exploring three drivers of Big Data Seismology. We next discuss the science that Big Data Seismology inquiries are enabling for studying both earthquake and non-earthquake processes, and for imaging. In the final section of the paper, we explore various opportunities and challenges of Big Data Seismology. Opportunities include practical applications like real-time monitoring and earthquake early warning. Challenges include retaining long-standing traditions of data accessibility and sharing, educating students on critical skills that cross-cut traditional degree programs, and ensuring that access to training and infrastructure is equitably distributed. Big Data Seismology is nascent, and the initial research reviewed in this paper highlights potential rather than reviewing fundamental discovery. However, as we have discussed in this section, advances in passive and exploration seismology have been driven by data. This paper is designed to offer a gateway for graduate students, early career scientists, or experienced seismologists into the emerging field of Big Data Seismology.

2. Drivers of Big Data Seismology

Major advances in a field of study are commonly summarized (however deficiently) by just a few transformative discoveries or experiments. As an analog to what the Sloan Digital Sky Survey (York et al., 2000) did for astronomy (Kremer et al., 2017; Y. Zhang & Zhao, 2015), the USArray (Figure 1) is an appropriate vehicle to introduce the three main drivers in the evolution of Big Data Seismology (Figure 2). While global, regional, and local seismic networks have always been deployed by the passive seismology community with some sense of survey style exploration, few achieved the level of uniformity, accessibility, and scale that USArray did in order to produce data-driven innovation into new areas of the field. As a distinct data set, it was relatively "well-behaved" in that it allowed researchers to make some simplifying assumptions that facilitated producing large-scale research products more easily. Notably, seismic instrumentation and sampling rates were largely uniform, station locations were gridded at a roughly consistent spacing of 70 km, and data gaps and quality problems were rare (Busby et al., 2018). USArray data were easy to access and download through a number of convenient request mechanisms (USArray Data at the DMC, 2015). Finally, the data set was large enough to challenge researchers to employ considerable computational resources to devise and perform novel large-scale data-hungry analyses, some of which have derived from the exploration seismology community (Burdick et al., 2014; Lin et al., 2011; Long et al., 2014). The success of well-behaved data sets like USArray to drive innovation is evident in like-spirited deployments that have followed it: IberArray in the Iberian Peninsula (Díaz et al., 2008), AlpArray in the Alps (Hetényi et al., 2018), and ChinArray in China (K. Wang et al., 2020; W. Wang et al., 2021; Zeng et al., 2020). Through these efforts and others, the continued growth of data volumes, computing availability, and the appetites of our algorithms is expanding the scope of seismic inquiry into emerging areas. Below, we highlight components of these drivers and how they are defining the boundaries of Big Data Seismology today and into the future.

2.1. New Data-Dense Sensor Deployments

Within the past 5–10 years, several technological innovations in seismic sensor systems have been driving the trend toward the use of spatially dense sensor networks in passive seismological investigations (Cochran, 2018). In particular, three that we discuss below include the development of nodal systems, community sensors, and distributed acoustic sensors. We also review the trend toward multi-sensor integration and the state-of-the-art in modern exploration seismology surveys.

2.1.1. Nodal Sensors

Seismic nodal sensors are an extension of geophones that, as described in the previous section, have been used in exploration seismology since the 1950s. Unlike broadband seismometers, geophones are passive analog instruments that comprise an inertial mass (with coil) that is free to move around a magnet that moves with the ground. The motion of the coil generates an electrical signal that is proportional to ground velocity. While they were conventionally installed as part of long cables, recent innovations developed self-contained nodal seismometers that combine sensor, data logger, GPS, and battery in a single unit (T. Dean et al., 2018). These nodes can be easily deployed and retrieved, and can typically record for up to 30 days. They are sold with harvesters, which are racks for storing, charging, and extracting data in volume from nodal instruments. The basic concept makes it easy for an academic research group to deploy and manage data from a relatively large number of sensors.

Deploying nodes at scale has led to the recent trend of dense networks of sensors that are referred to as "Large-N arrays" (the N simply refers to the number of instruments). One of the earliest Large-N deployments of nodal instruments was in Long Beach, CA (Lin et al., 2013). The Long Beach experiment was done by NodalSeismic on behalf of Signal Hill Petroleum, and contained two phases. In the first phase, 5,300 sensors were deployed over a 100 km^2 area for 6 months. In the second phase, 2,500 sensors were deployed over 3 months to extend the original survey to the east (Figure 3). The estimated total data volume for this experiment is 51 TB, which is approximately equal to the entire Earthscope experiment (Figure 4). While these data are not freely available, it subsequently motivated the deployment of a number of smaller Large-N arrays by the passive seismology community. Some important deployments are archived on the IRIS website (https://www.iris.edu/hq/initiatives/recording-the-full-seismic-wavefield) and include the IRIS Wavefields Community Experiment in Oklahoma (Sweet et al., 2018), the Imaging Magma Under St. Helens (iMUSH) experiment (Hansen & Schmandt, 2015), and the Sage Brush Flats array across the San Jacinto fault zone (Ben-Zion et al., 2015). A recent Focus Section in *Seismological Research Letters* explored research conducted on several of these initial Large-N nodal arrays (Karplus & Schmandt, 2018).

2.1.2. Community Sensors

Nodal arrays have provided extremely dense recordings, but are only practical for short-term deployments (weeks to months). Another recent innovation has been the development of low-cost personal seismometers that provide long-term measurements. These sensors typically exploit compact computers such as Raspberry Pi. The Raspberry Shake (Anthony et al., 2019) is a system that has been commercialized and sold to both hobbyists and seismic networks. It uses a geophone as the seismic sensor, and the Raspberry Pi as the digitizer and telemetry system, and also sells units that have an infrasound sensor. The data are openly available on IRIS, and can provide recordings in places where other seismometers are not available. In particular, their value in urban environments has been demonstrated in a recent paper that explored the reduction in anthropogenic noise following COVID-19 restrictions (Lecocq et al., 2020). Another network that relies on community involvement is the Community Seismic Network (CSN) in Los Angeles (Clayton et al., 2015). This network is designed to provide high-density urban measurements of strong earthquakes, in order to provide actionable information to first responders (i.e., where ground shaking is strongest). The CSN network uses low-cost MEMS accelerometers that are designed to record strong ground motions.

10.1029/2021RG000769

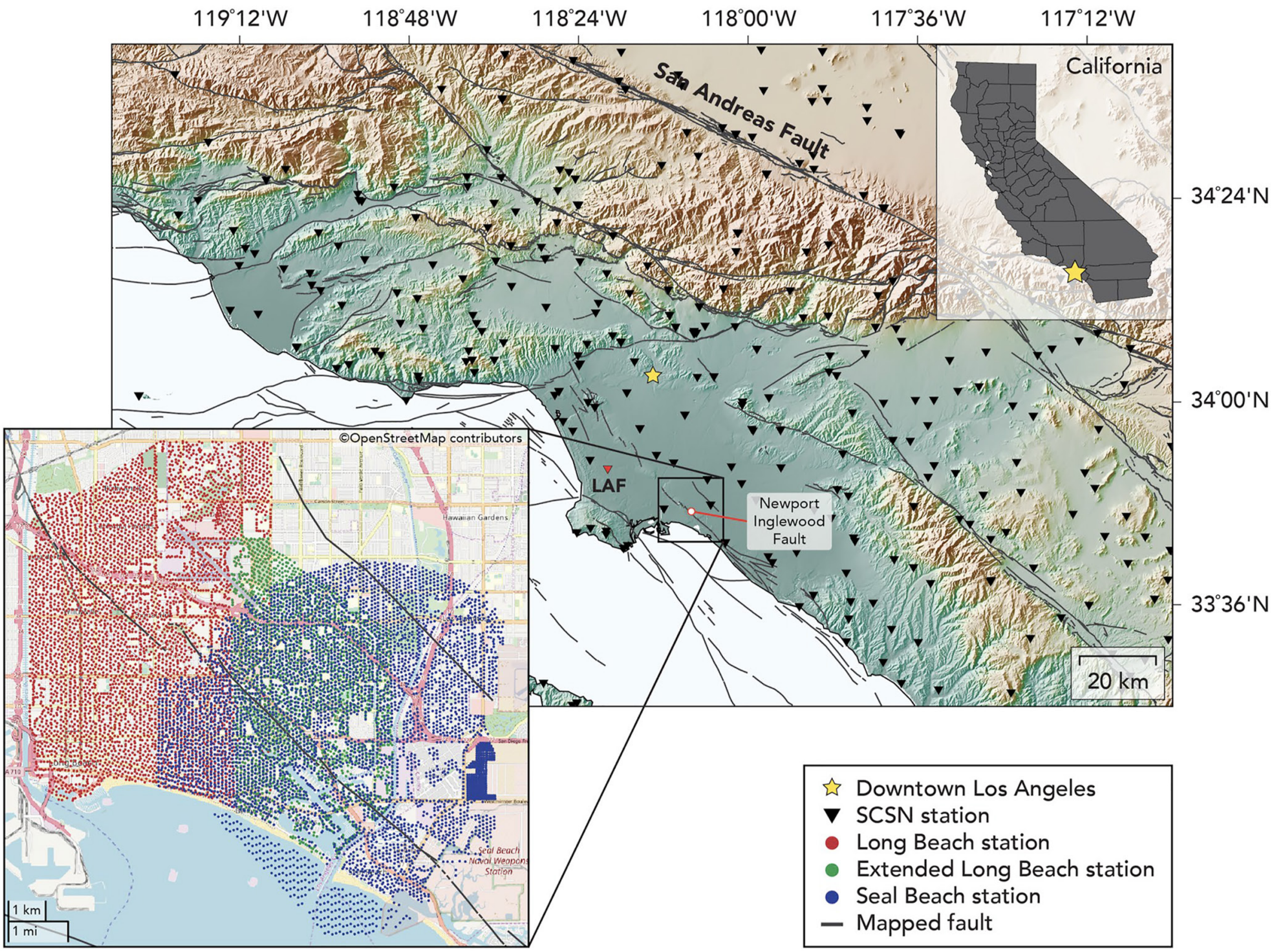

Figure 3. Map showing the Long Beach nodal array (inset) and the Southern California Seismic Network (SCSN) regional network (Castellanos & Clayton, 2021).

Modern smartphones can also provide useful seismic data through their built-in accelerometers, in particular by providing time-critical data in the event of earthquakes. Recent papers have explored the use of crowdsourcing, such as identifying relevant posts on Twitter, as triggers for detecting earthquakes rapidly (Earle et al., 2011). The rapid detection of earthquakes is useful in Earthquake Early Warning (EEW). The Berkeley Seismology Lab have developed a cellphone app (Kong et al., 2016) that exploits machine learning (ML)—a class of techniques discussed in detail in Section 2.3.4—to distinguish motions that are indicative of earthquakes from motions associated with the user moving around. Earthquake detections are transmitted to a datacenter where detections from multiple phones are compiled to trigger earthquake alerts. As smartphones have become ubiquitous, they have the potential to provide extremely dense (albeit noisy) measurements of earthquakes. For applications like EEW, having numerous low-quality sensors close to the earthquake source is often more useful than having more distant high-quality sensors, as algorithm performance hinges fundamentally on rapid earthquake detection and location.

2.1.3. Distributed Acoustic Sensing

In essence, nodal and community sensors are technological innovations that allow researchers to deploy or access classical translational sensors—geophones and accelerometers—at much greater spatial density. In contrast, Distributed Acoustic Sensing (DAS) is an emerging technology that exploits a fundamentally different type of sensor based on the use of Rayleigh backscattered light in fiber-optic cables to provide measurements of strain or strain rate (Lindsey & Martin, 2021; Zhan, 2020). In DAS systems, an interrogator unit (IU) sends pulses of laser light down a fiber and measures phase shifts in the backscattered light originating from discrete points (sites where the refractive index changes) along the fiber. The backscattered phase shifts, averaged from all points in a

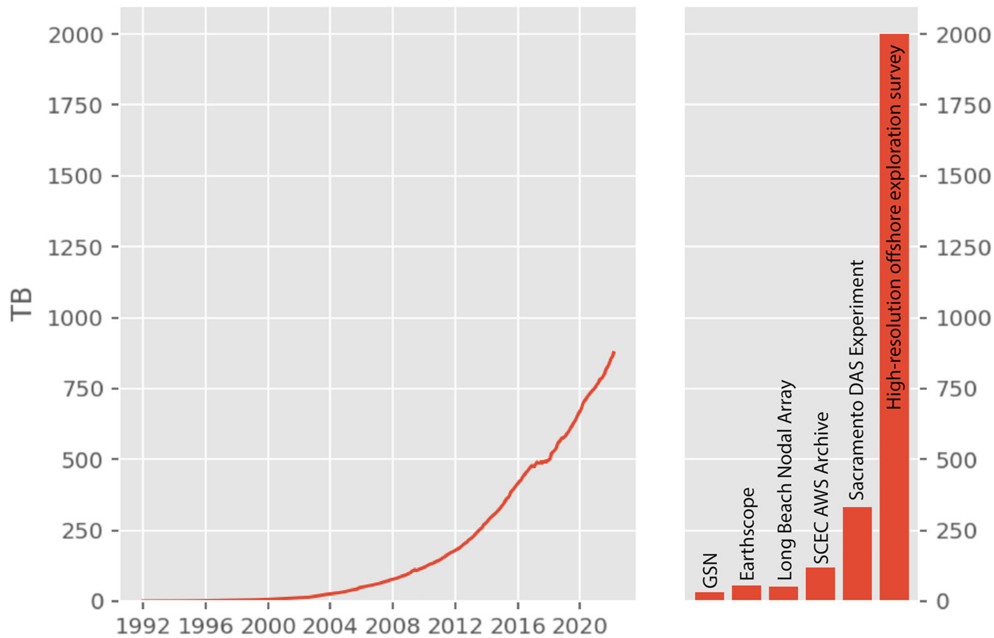

Figure 4. Data volume of the IRIS-DMC Archive as a function of time (left) and the corresponding data volumes of five benchmark data sets: GSN (1992–present), Earthscope (2005–present), the Long Beach Nodal array (Figure 3), the SCEC archive on Amazon Web Services (E. Yu et al., 2021), and the Sacramento Distributed Acoustic Sensing experiment (Ajo-Franklin et al., 2019). For reference, a "typical" modern high-resolution marine exploration survey is also shown (Jacobs, 2018).

segment of fiber (referred to as a gauge), is proportional to the change in path length over the gauge length. DAS systems can thus provide measurements of the strain or strain rate at very high spatial resolution (gauge lengths typically vary from 1 to 40 m long) and can operate continuously by using a single IU at the end of the fiber. Thus, power and data storage is only necessary at a single location, unlike nodal systems, which require power and data storage at each site. DAS technology was first employed for seismological purposes by the exploration seismology community in the early 2010's for short-duration active source experiments (Daley et al., 2013; Mestayer et al., 2011). However, the usage of DAS systems has subsequently expanded to include a range of local, regional, and teleseismic applications in the passive seismology community (Lindsey & Martin, 2021).

DAS systems have some important advantages over classical sensor systems, and some key limitations and challenges that are being actively explored. The main advantage is that DAS is an inherently spatially dense sensor array that can be operated for long time durations. DAS systems can also be operated in settings where conventional sensor systems are hard to deploy (e.g., offshore, on glaciers, or down wells), and can leverage pre-existing but unused fibers ("dark fiber"), which makes them particularly useful in urban environments (Fang et al., 2020). DAS systems also have an extremely broadband frequency response, which makes them useful for a wide-range of seismological investigations (Lindsey, Rademacher, & Ajo-Franklin, 2020). Analyses of DAS and colocated geophone or broadband observations have found that DAS measurements retrieve similar ground motions (H. F. Wang et al., 2018). The limitations include the fact that the instrument response and associated effect of coupling of the fiber-optic to the ground are often unknown (Paitz et al., 2020), IU units are extremely expensive and not widely available to the research community, and fiber installation details are sometimes poorly known (especially when using dark fiber). DAS deployments can also generate voluminous quantities of data, which create unique challenges with archival and data management (Quinteros et al., 2021). As an example, a recent 7-month recording of strain rate from a single IU attached to a dark fiber in Sacramento generated over 300 TB (Figure 4).

2.1.4. Multi-Sensor Integration

In addition to the development of specialized sensor systems that result in greater data volume, an increasing trend in seismological research involves the integration of different types of sensor systems (referred to as data "variety" in Section 2.3.4). Global Navigation Satellite System (GNSS) receivers are increasingly being used to

10.1029/2021RG000769

extend the traditional bandwidth of classical ground translation recordings for characterizing the magnitudes and rupture dynamics of large earthquakes (Larson, 2019; Melbourne et al., 2021). In common with the growth in data volumes of traditional seismic data, data volumes from GNSS observations are also increasing, however this is driven by land-surveying applications (Melbourne et al., 2021). The proliferation of GNSS receivers integrated into cellphones also provides an opportunity to provide time-critical information for Earthquake Early Warning (Minson et al., 2015). Measurements of barometric pressure and infrasound were added to the USArray TA and to many sites in the GSN and have been used to study geophysical sources that produce both seismic and atmospheric waves such as bolides (Arrowsmith et al., 2021), volcanoes (Sanderson et al., 2020), and earthquakes (Walker et al., 2013). These data streams have also opened up opportunities to study atmospheric phenomena such as atmospheric gravity waves (de Groot-Hedlin et al., 2014). The addition of meteorological sensors to many seismic sites (such as TA and GSN stations) has been motivated by the interest in characterizing the atmospheric effect on seismic ground motions but also provides potential collaborations with the meteorological community (Tytell et al., 2016).

2.1.5. Modern Exploration Seismology Surveys

2.1.5.1. 3D Land Surveys

Today, 3D seismic land surveys are acquired using Vibroseis source arrays, and either cabled geophone arrays or cable-free autonomous geophone nodes, deployed at nominal 10–50 m array spacing. A modern 3D seismic land crew can deploy 300,000 or more live cabled geophone channels, and uses a typical recording time of 10–15 s at 1–2 msec sampling for each vibrator source shot point. To increase areal coverage, cabled geophone array "patches" are often "rolled along" and source points are repeated to build up the full desired array aperture and density. Autonomous geophone nodal arrays can consist of 10,000 or more deployed nodes with sufficient battery life to record continuously up to 90 days at 2 msec sampling, and provide the additional capability to record passive seismic data during the quiet listen periods (e.g., overnight) between active source shooting (Lin et al., 2013). Additionally, buried fiber optic DAS areal arrays are increasingly being deployed to record active seismic source surveys, in addition to passive seismic data and ground deformation (Bakulin et al., 2020; Monk, 2020). Vibroseis source arrays typically consist of 4–8 or more vibrator trucks, each with up to 80,000 lbs peak ground force with a frequency sweep range of 2–250 Hz for up to 60 s. Vibroseis data acquisition times, density and volumes are being accelerated by use of simultaneous sources and compressive sensing. Most Vibroseis data are cross-correlated with the source chirp signal "on the fly" in real-time by the digital recording system, although the data can also be optionally stored in the larger uncorrelated form. As a result, a typical modern 3D seismic land survey acquired today will have a density of about 1 million seismic traces per square km, and a total data size of 100 Terabytes to 1 Petabyte per survey.

2.1.5.2. 3D Marine Surveys

Most modern 3D seismic marine surveys are acquired using compressed airgun source arrays, and either towed hydrophone streamer cable arrays, ocean bottom cable (OBC), or autonomous ocean bottom nodes (OBN). A modern 3D seismic marine streamer survey typically tows 4–18 hydrophone streamer cables of 6–12 km length each, at a depth of 5–15 m below the sea surface, representing over 10,000 live hydrophone channels in the water that record each airgun source point with a record length of 10–15 s at a 1–2 msec sample rate. To increase areal aperture coverage, this basic narrow-azimuth streamer geometry can be replicated with multiple boats (or repeated, or simultaneous-source shooting) to build up wide-azimuth coverage, and multiple azimuth 3D seismic surveys can be completely re-acquired at 30–60° azimuth compass headings to provide full 360-degree azimuthal aperture. Alternatively, buried OBC geophone or fiber optic cables, or autonomous OBN units, can be deployed on the deep seafloor by underwater ROVs (remotely operated vehicles), typically at up to 2,000 OBN's per deployed patch, and then the patches can be "rolled along" by the ROVs with repeated source points to build up 3D aperture and density (Beaudoin & Ross, 2007). Each OBN unit typically records 4C data (pressure and 3C particle velocity) continuously for up to 90 days at a 2 msec sample rate. As a result, a typical modern 3D seismic marine survey today will have a high-density areal data coverage (e.g., 10×10 m pixels) and a total data size of 1–10 Petabytes.

2.2. Improvements in Computing Infrastructure

The ability to access and analyze data has always been critical in seismology, but in the last decade or more, data volumes have been growing faster than our ability to easily access, store, and process them (Quinteros et al., 2021). The need for ways to perform analyses that have become intractable on desktops or small computing clusters has grown in parallel with data volumes. Such large-scale computing has traditionally been the domain of high-performance computing (HPC), which has long been available at institutional computing centers like in national laboratories, industry, and at large universities. These systems have a number of disadvantages, however, when used for data-intensive analyses (Ahrens et al., 2011). They commonly employ highly specialized hardware and highly coupled software that are optimized for running CPU- and memory-intensive simulations, where the goal is higher Floating Point Operations per Second (FLOPS) instead of data processing throughput. Further, they are tightly controlled and administered systems, commonly requiring employee status and proposals for computing time. Taken together, these properties can make HPC centers inhospitable to the diverse needs of many data-intensive analysis applications.

A significant shift in "big data" computing began in the mid-2000s, with the development and accessibility of large-scale data-intensive infrastructure and software ecosystems, on commodity hardware outside the domain of traditional HPC facilities. This shift is greatly influenced by the volume and commercial value of non-seismic data, and the success of new ML and data science techniques like deep learning, that are not a natural fit for tightly coupled HPC systems. The algorithms and models that learn from data (see Section 2.3) are commonly executed in parallel over a range of input parameters, a style of computing known as high-throughput computing (HTC; Arora, 2016). These commercial and non-commercial influences have produced an explosion of new open-source tools for data-intensive analyses, like TensorFlow (Abadi et al., 2016), PyTorch (Paszke et al., 2019), and Scikit-Learn (Pedregosa et al., 2011). They have also produced new scalable computing models, platforms, and technologies like MapReduce (J. Dean & Ghemawat, 2008), Hadoop (Apache Software Foundation, 2010), Spark (Zaharia et al., 2016), Docker (Merkel, 2014), and Kubernetes (Hightower et al., 2017). Finally, they have shaped (and been shaped by) the development of commercial cloud computing and storage services and platforms, like Amazon Web Services (AWS) and Google Cloud Platform (GCP). Below, we highlight a few examples of how the availability of data-intensive computing models, the growth of open-source analysis software and technologies, advances in computing hardware, and the maturation of commercial cloud computing services have set the stage for computing in Big Data Seismology.

2.2.1. Data-Intensive Computing Models and Technologies

With the growing volumes of data with business value associated with the internet and connected devices in the early and middle 2000s, the need arose for computing frameworks that could run on inexpensive commodity hardware and be programmed by nonspecialists. These needs were distinct from the physics-based simulation-driven requirements of existing HPC systems, in that they were driven by processing throughput, nonphysical models based on data, and HTC style computing. As a result there was a new commercial interest in efficiently storing, accessing, and analyzing large volumes of heterogeneous data, and ensuring that the data and analyses were robust in the face of hardware failures that are inevitable with commodity hardware.

Among the first approaches that emerged to address these needs was the Google File System (GFS; Ghemawat et al., 2003) distributed file system and the MapReduce (J. Dean & Ghemawat, 2008) programming model, published by researchers at Google, followed soon after by the Hadoop and the Hadoop Distributed File System (HDFS) implementations, originally supported by developers at Yahoo! and later open-sourced (Shvachko et al., 2010). MapReduce and Hadoop offered two significant contributions. The first was a file system that redundantly distributed data across a cluster of nodes, near to where computations would occur. The second was a relatively simple programming interface that expressed computations in terms of a *map* stage, in which an analysis function is applied across a set of inputs, and a *reduce* stage, in which the intermediate outputs of the previous stage are merged into a potentially smaller output and written to disk. The co-design of a distributed file system and a computing framework that managed parallelism for the user made it possible to run high-throughput fault-tolerant analyses on thousands of inexpensive computing nodes over terabytes to petabytes of data. These advances were explored further with the development of Spark, a computational framework that did not rely on writing intermediate results to disk, but instead distributed data across a computation cluster through a faster in-memory format, called Resilient Distributed Data sets (RDDs; Zaharia et al., 2010). The utility of accessible data-intensive computing software has led to an explosion of technologies and services that facilitate scalable and high-throughput analyses, including non-relational "no-SQL" databases like MongoDB and DynamoDB, parallelization frameworks in Python, like Dask and Ray, and companies that maintain and provide these technologies as a service, making their use in scientific research more attainable than ever before. We briefly describe applications of these technologies in seismology.

Addair et al. (2014) investigate the use of Hadoop to perform a global-scale cross-correlation analysis and compared its performance to a conventional local cluster implementation. Using a 1 TB waveform data set, they achieved an average data processing rate of 16.7 GB per minute on a 10-node cluster, with 120 cores and 960 GB of RAM. This represented a 19x improvement in speed compared to the performance of the traditional cluster implementation. In order to achieve this higher performance, the authors modified how they stored their data, and how they expressed their analysis code to use the data structures and interfaces in the Hadoop ecosystem. They transformed their seismic event and station metadata from being stored in highly normalized database tables into a de-normalized form, in which many data are duplicated and distributed across the computing nodes. They also employed a number of Java-based software frameworks that support Hadoop: Sqoop for transferring data from a relational database into the HDFS, and Apache Pig for expressing SQL-like queries. Finally, they profiled and optimized their analysis to identify and minimize bottlenecks, such as balancing the size of computing tasks that were distributed to the computing cluster. Magana-Zook et al. (2016) expanded on this study to also evaluate Apache Spark and used a much larger data set of over 40 TB. The authors found a similar speedup of approximately 15x compared to their traditional cluster implementation. Further, they noted that the data storage and transfer speeds of the traditional implementation would likely quickly become a bottleneck with the addition of more computing nodes to the cluster, compared to the Hadoop or Spark implementations. Both sets of authors noted the challenges of preparing large volumes of data to use these tools, and suggested that the one-time costs of migrating data and software could be amortized by improved computing speed over time.

Python is a popular language for research in seismology, and the Dask library is a tool for scalable parallelism that can also facilitate map-reduce style computations, as well as integrate with existing libraries in the scientific Python ecosystem (Rocklin, 2015). MacCarthy et al. (2020) used Dask on a commercial cloud service to request 5.6 TB of compressed data from the IRIS-DMC (IRIS Data Management Center) on-the-fly for a survey style detection application (Marcillo & MacCarthy, 2020). As one goal of the study was to avoid the complexities of data ingestion and storage noted by earlier authors, waveform data from FDSN Web Services interface requests to the IRIS-DMC were streamed directly into the computation without first being transformed into intermediate formats or even being saved at all. The authors acquired data and performed a spectral peak detection survey over windows of continuous data. They achieved an average total (computing + analysis) throughput of 1.7 GB per minute on a cluster of 50 nodes with 400 GB memory and 100 CPUs, improving to only 2 GB per minute for a cluster of twice the size. Similar to the earlier studies, data acquisition speed was the most significant bottleneck. In this work, however, the limitations were related to internet transfer speeds and data center service capacity instead of local storage and transfer speeds.

As machine learning and other highly parallelizable workflows have become widespread, there is an emerging trend of using Graphical Processing Units (GPUs) in complement to or instead of traditional CPU hardware systems. Originally designed for gaming and graphics applications, GPUs are appealing for scientific computations in which smaller tasks can be easily parallelized and distributed over thousands of memory-limited cores on a single computing unit. While not all problems are well-suited for this architecture, machine learning applications involving the training of deep neural networks (Section 2.3.4) are perhaps the canonical example (Goodfellow et al., 2016; LeCun et al., 2015). In seismology, GPUs have been used to accelerate seismic wave propagation codes (Komatitsch et al., 2010) that form the basis for tomographic inversions (Section 3.2) and as part of earthquake detection workflows (Ross, Trugman, et al., 2019) involving template matching or other related techniques (Section 3.1).

2.2.2. Cloud Computing Services and Platforms

While numerous data-intensive computing technologies, open-source software, and faster hardware have made Big Data Seismology more achievable on locally owned and administered computing systems, the time and specialized knowledge required to acquire and maintain local computing resources can make them out-of-reach for many researchers and teams. Recognizing this as a common limitation for businesses, commercial cloud (CC) providers like AWS, GCP, and Microsoft Azure have emerged since the middle 2000s to provide computing, storage, and software services for users interested in quick access to a diverse set of services, commonly with a pay-as-you go pricing model. The CC offers users who cannot access or prefer not to use institutionally maintained computing systems a wide array of services, often without the need to acquire and maintain hardware or install and update software. The three primary types of services offered by cloud providers are virtualized computing servers (*instances*), data storage, and numerous software services, such as managed Hadoop/Spark clusters and so-called *serverless* Functions-as-a-Service (FaaS). Together, CC services have lowered start-up costs and times, increased scalability, and widened the diversity of business applications, and have begun to do the same for scientific research applications. Here, we briefly review some applications in seismology that have taken advantage of the cloud.

A significant advantage of the commercial cloud over locally maintained resources is the diversity of services offered. MacCarthy et al. (2020) and Marcillo and MacCarthy (2020) performed a survey of USArray data for frequency domain signals indicative of industrial noise. They streamed waveform data from the IRIS-DMC into a computing cluster, through more than 12 million FDSN web services requests. This type of inundation of waveform requests would normally trigger service denials from the data center based upon a user's unique Internet Protocol (IP) address, but the authors were able obtain up to 100 ephemeral IP addresses from AWS, their cluster host. This allowed them to demonstrate a streaming workflow that bypassed the complexities of user-managed data storage and access, resulting in an order-of-magnitude increase in analysis throughput compared to a traditional download-and-store workflow. While this represented an improvement, the authors noted that fully in-cloud data and computation can also take advantage of the scalable storage bandwidth offered by cloud providers, and they cite throughputs of up to 43 GB per minute for the same computation performed on a 100-node cluster using data stored in cloud object storage, Amazon S3 (Amazon Simple Storage). To facilitate research using large volumes of seismic data, the Southern California Earthquake Data Center earthquake catalog and waveform archive were recently hosted on Amazon S3 (E. Yu et al., 2021). The data set includes continuous waveform data, event-triggered waveform segments, an earthquake catalog, phase picks, and station metadata. At the time of writing, the waveform holdings were ≈120 TB.

2.3. Techniques and Algorithms

Traditional algorithms used to process and analyze seismic data do not scale well to large data volumes and hence new methods are needed to optimally extract meaning from data. At the same time, the development of new techniques and algorithms are also enabling drivers for Big Data Seismology, as they allow the more complete analysis of large data sets. A critical workflow in passive seismology, which we use to illustrate how techniques can scale with data, involves processing data to find signals and events. Since the advent of digital recording in the 1970's, this process has largely been driven by algorithms, where the choice of algorithm depends on the following properties of the data being processed: the data volume, data density, and data knowledge (Figure 5). Data volume simply refers to the amount of data in a particular data set (as depicted for some benchmark data sets in Figure 4). In a spatial sense, data volume scales with the number of discrete locations where the wavefield is recorded. In a temporal sense, data volume scales with the duration and sampling rate of ground motion recording. In contrast, data density specifically characterizes the proximity of the measurements, either in space or time. Spatial density is particularly important for the choice of algorithm, as various techniques are based upon the coherence of waves between sensors (where coherence can sometimes include both amplitude and phase, while other times is based only on amplitude). Data knowledge refers to the amount of *a-priori* knowledge about the signals or noise that one has before processing the data, since this critically informs the type of algorithm used. In general, the more knowledge one has about the signals or events that may be present in the data set, the more effectively they can be detected, but at the cost of generality.

Different algorithms are designed for different levels of data volume, density, and knowledge, as shown in Figure 5. In effect, different algorithms have different data appetites. Algorithms with low data appetites have been used successfully for processing data from sparse networks for several decades, but do not necessarily scale well for the new types of data sets associated with Big Data Seismology.

2.3.1. Low-Volume, Low-Density Techniques

Passive seismology has traditionally been based on low-volume, low-density data sets such as the GSN and many regional and national networks (see Section 1). Processing data from these networks has thus required algorithms that have low data appetites. With low data knowledge, the ratio of the short-term-average (STA) to

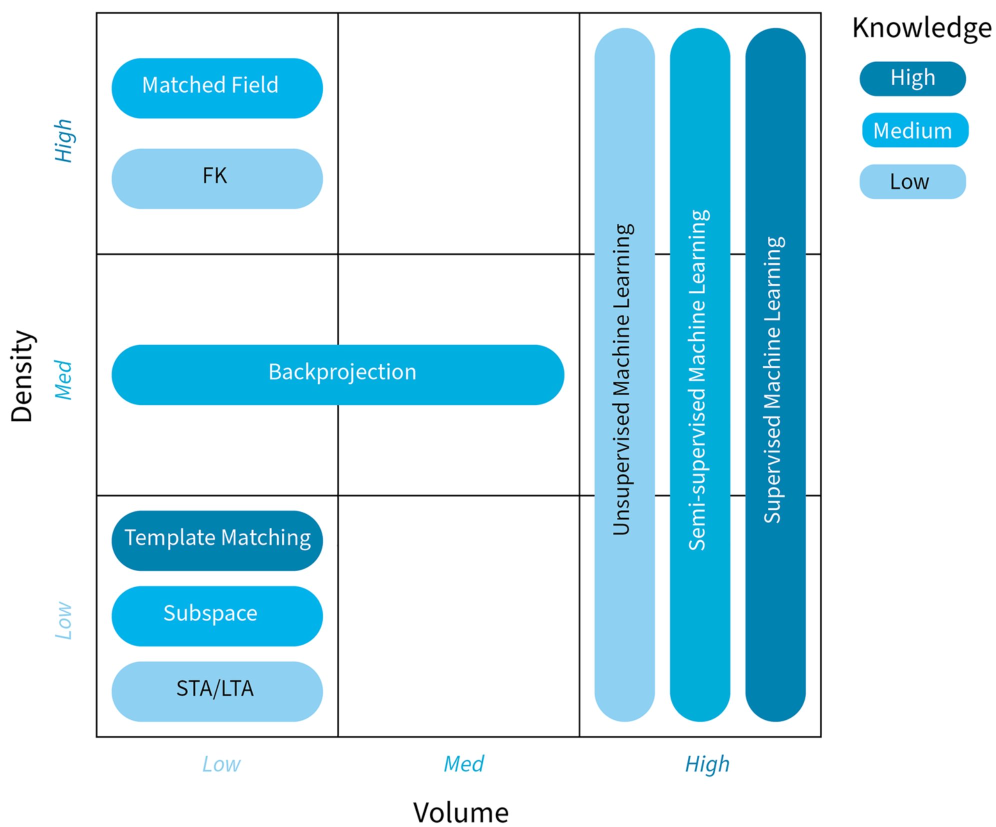

Figure 5. A critical workflow in passive seismology involves processing data to find signals and events. This figure explores how some widely used algorithms scale based on the data volume, density, and knowledge.

long-term-average (LTA)—referred to as STA/LTA—is widely used as the basis for detecting signals at single stations (R. V. Allen, 1978; Withers et al., 1998). The STA/LTA algorithm assumes that signals can be distinguished from noise based on power, and thus is general but limited to detecting signals with good SNR. Another type of low-volume, low-density algorithm exploits higher-order statistics of seismic data in moving windows (e.g., kurtosis or skewness) to detect differences in the statistical distribution that distinguish signals from noise (Baillard et al., 2014).

At the other extreme of data knowledge, when one has an exact template of the signal to search for, template matching is the optimal detection technique (Kay, 1993) and can detect matches at very low SNR (Gibbons & Ringdal, 2006; Ross, Idini, et al., 2019; Shelly et al., 2007; Skoumal et al., 2014). Template matching techniques are based on calculating cross-correlation coefficients between template waveforms and a continuous data stream, and thus assume that signals of interest are amplitude-scaled versions of the templates. Unlike STA/LTA, template matching is not general as it makes very specific assumptions about the signals to be found.

It is common to have some knowledge about the signals of interest, beyond the simple assumption that they have larger power than the noise, without knowing their exact form. For these medium levels of data knowledge, subspace detectors are a generalization of template matching that exploits features extracted from a set of prior signals (Harris, 2006). Subspace detectors thus provide a compromise between generality and performance that is useful in many practical applications.

2.3.2. Low-Volume, High-Density Techniques

As discussed in Section 1, seismic arrays have been used since the 1930s in exploration seismology, and since the 1960s in passive seismology, as a means to enhance SNR and provide estimates of the DOA of incoming seismic waves (Rost & Thomas, 2002). Classical seismic arrays were designed such that signals would be coherent across the array, but time-shifted at each array element based on the DOA. In passive seismology, classical arrays are typically composed of a relatively small number of sensors (a few to a few tens of elements) but capture the wavefield closely enough that the signals are coherent. Thus, we consider them low volume, high-density data sets.

With low data knowledge, detection using seismic arrays is based on the ratio of power in the beam to the residual power, which can be formulated as an F-statistic (Blandford, 1974). This approach can be thought of as analogous to STA/LTA, as it distinguishes signal from noise based on a comparison of powers. The difference is that arrays also provide spatial filtering and can therefore detect signals at lower SNR. For most applications, the signal DOA is unknown and therefore this is also estimated, typically using frequency-wavenumber (FK) analysis (Rost & Thomas, 2002).

With more data knowledge, matched field processing is an array-based technique that is analogous to matched filtering, but that compares the observed wavefield data against templates that capture the phase and amplitude differences of prior signals observed across an array (Harris & Kvaerna, 2010). In this sense, matched field processing is not restricted to the plane wave assumptions of conventional beamforming. Typically, matched field processing focuses on matching the moveout across an array and not necessarily the waveforms wiggle-for-wiggle, but it can be implemented to match both. In this sense, the technique can be less susceptible to source effects. Matched field processing has been used to discover small events below the nominal detection threshold (J. Wang et al., 2015).

2.3.3. Medium-Volume, Medium-Density Techniques

The techniques developed for processing waveform data from low-volume data sets, which are described above, treat each station or array separately. In general, these techniques are designed to detect signals at a given station or array; signals from separate stations and/or arrays are subsequently grouped using association techniques to build events (Draelos et al., 2015; Yeck et al., 2019). For medium volume/medium density data sets, a common approach is to process data across a network by migrating or backprojecting the waveform data back to a series of event hypotheses (Drew et al., 2013; Grigoli et al., 2014; Kao & Shan, 2004; Langet et al., 2014). While this approach fails for low density networks, it has been shown to work quite successfully for medium density networks such as regional networks in seismically active areas (Arrowsmith et al., 2018). Such techniques are typically referred to as "backprojection" techniques (Figure 5), and are similar in principle to migration techniques used in exploration seismology.

2.3.4. High-Volume Techniques: Machine Learning

Machine learning (ML) has emerged as a key set of tools for exploring, analyzing, and extracting patterns from large seismic data sets. Machine learning can be divided into several broad classes of learning problems, of which supervised learning and unsupervised learning are the most common in seismology. In *supervised learning* problems, the ML algorithm optimizes a prediction model using a collection of labeled examples referred to as *training data*. A well-trained ML model should learn to reproduce the key relationships in the data, and thus be able to generalize and make accurate predictions for new observations similar to, but distinct from, those used to build the model. In contrast, *unsupervised* ML algorithms learn to extract patterns or structure in a collection of observations without access to labeled examples corresponding to a pattern of interest. The wide range of ML applications in the geosciences have been discussed in recent review papers on ML in seismology (Kong, Trugman, et al., 2019) and solid Earth geoscience (Bergen et al., 2019; Dramsch, 2020) and on deep learning in geophysics (S. Yu & Ma, 2021). This section presents the role of ML in the analysis of large scientific data sets and highlights classes of machine learning tools that are particularly relevant to Big Data Seismology. A comprehensive review of ML for seismology is beyond the scope of this paper, and we refer the interested reader to the review papers listed above for further details on this topic.

Machine learning is well-suited for big data analysis in seismology, and can be applied to high-volume data sets with a range of data densities and levels of data knowledge (Figure 5). Machine learning, especially deep neural networks (DNNs), is particularly powerful at transforming large quantities of data into models (Goodfellow et al., 2016; Raghu & Schmidt, 2020). Neural networks are neither new inventions nor new to the field of seismology (McCormack, 1991); research over the last decade has made it possible to efficiently train these models on larger data sets (Krizhevsky et al., 2012; Raina et al., 2009). Training ML models is generally more effective when more training observations are available (Sun et al., 2017), especially when the data are high-quality (Figure 6). Larger data sets allow ML learning algorithms to learn the complex representations and mappings required to perform challenging analysis tasks with high accuracy. And large data sets containing diverse examples improve

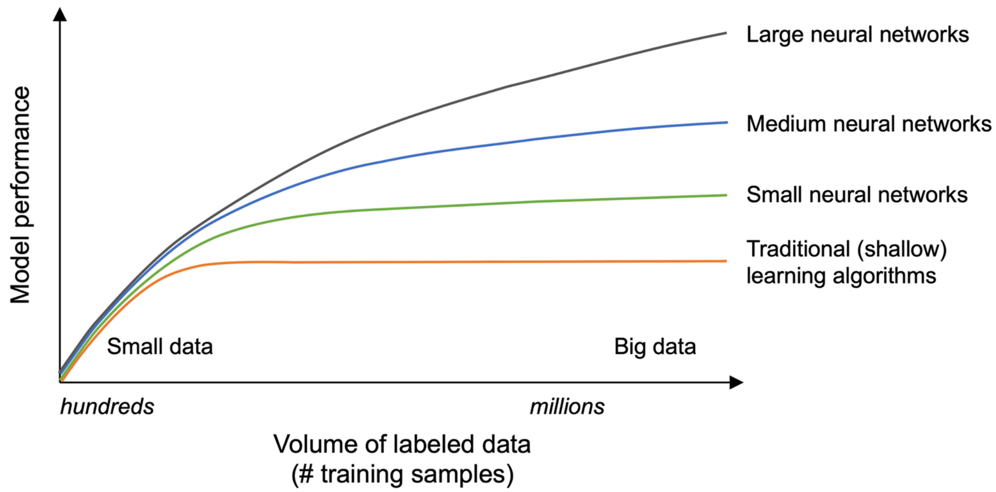

Figure 6. Large machine learning (ML) models, especially deep neural networks, can achieve improved performance by training on massive data volumes. Traditional, non-neural network ML algorithms, often called "shallow" ML algorithms, learn from data, but do not see the same benefit from training on massive labeled data sets (Ng, 2011).

the generalization capability of ML models (Shorten & Khoshgoftaar, 2019). For many ML models, including DNNs, the application of the trained model to new data for prediction is computationally efficient compared to training, requiring only a single pass through the model. Thus seismologists are increasingly adopting ML to automate data analysis and keep up with growing volumes of seismic data.

2.3.4.1. Machine Learning and Big Data Challenges

Machine learning offers tools to tackle a range of big data challenges. While the term "big" is field-, task- and context-dependent, so "big" data are broadly distinguished from small data by a set of characteristics referred to as the *five Vs*: volume, velocity, variety, veracity and value (L'heureux et al., 2017). Analyzing seismic data sets is often challenging due to the sheer *volume* of data that has been collected. Analysis of high-volume data can be automated using ML. For example, using a DNN and parallel computing a recent study (Dumont et al., 2020) identified regions of coherent energy in 170 GB (10 days) of DAS data with an inference time of 30 min. Deep learning frameworks enable large-scale training and fast inference with GPUs. For high-dimensional data, dimensionality reduction techniques from unsupervised learning can reduce the size of data sets and make data analysis more manageable. For large volumes of unlabeled data, unsupervised approaches including clustering and anomaly detection or semi-supervised learning techniques, which learn simultaneously from both labeled and unlabeled observations, can be used.

The *velocity* of data refers to the rate at which new data are accumulated or the rate at which incoming data must be analyzed. Examples of high-velocity data in seismology include real-time data collection from seismic networks or DAS systems for tasks that require rapid processing like near real-time monitoring or EEW (see Section 4.1). Machine learning models can be trained on previously collected batch data and applied to incoming data streams to analyze the data in near real-time (J. Chen & Ran, 2019). For such models it may be necessary to monitor the data stream for drift and periodically retrain models with recent examples (Gama et al., 2014). Other ML models can be trained in an online manner, with model parameters updated in near real-time (Gomes et al., 2019).

Data do not have to be high-velocity or volume to pose an analysis challenge. Scientific data sets are often unstructured, high-dimensional, or heterogeneous, and many tasks require jointly analyzing data from multiple sources or modalities. These high-*variety* data can prove challenging to analyze even if they are not extreme in volume or velocity. Machine learning has been successfully applied to data fusion tasks, which require modeling complex relationships between multiple data sources (T. Meng et al., 2020). Deep neural network architectures are particularly appropriate due to their flexibility, allowing DNNs to accept a range of data types as inputs and learn joint representations that integrate multisource or multimodal data (Ramachandram & Taylor, 2017).

High-variety data are poised to become increasingly central to Big Data Seismology as researchers seek to gain insights through joint analysis of seismic, geodetic, infrasound, meteorological, and other data sources (see Section 2.1.4).

Data *veracity* refers to data quality, including noisy data, missing entries, uncertain measurements or other inconsistencies within the data set. Machine learning can automatically assess data quality (Jiang et al., 2019), perform data pre-processing or cleaning steps such as denoising (W. Zhu et al., 2019) or identify clipped waveforms (S. Wang & Zhang, 2020).

Extracting *value* from low information-density data sets, including low-SNR data or crowdsourced data, is also a critical big data challenge in seismology (see Section 2.1.2). Machine learning is well-suited for these data sets, as the ability to identify patterns in large data sets that are too subtle to be recognized by traditional analysis techniques is a key strength of ML algorithms.

2.3.4.2. Machine Learning Techniques for Big Data Analysis in Geoscience

Representation Learning and Deep Learning

The performance of a ML algorithm depends on the representation of the data fed into the model. Availability of a good set of features is a prerequisite for successful application of these simple models; specifically, traditional ("shallow") ML models require a representation of the data for which the key patterns or structure in the data set can be easily revealed and modeled by the simple transformations performed by the algorithm. Thus hand engineering of task- and domain-specific features is a critical step in training these ML models (Zheng & Casari, 2018). In contrast, *representation learning* or *feature learning* (Bengio et al., 2013) learns a mapping from the original, unprocessed or lightly processed, high-dimensional measurements to a new data representation that captures key features that are useful for modeling the data or for prediction tasks. Representation learning can be used in unsupervised, semi-supervised or supervised learning tasks; data representations can be learned from large quantities of unlabeled data and these learned features can be passed to a supervised algorithm that leverages available labels for a specific task. In seismology, unsupervised feature learning has been used to extract features from waveform data that allow for better separation of signals for clustering analysis (Holtzman et al., 2018; Seydoux et al., 2020).

Deep neural networks learn a hierarchical data representation through a composition of transformations performed at each layer (LeCun et al., 2015). Using a large number of layers, these models can simultaneously learn a predictive model along with a feature representation for the data tailored to the prediction task. This is a major reason for the success of DNNs in seismology and other fields: representation learning performed by DNNs allows researchers to bypass careful feature engineering and learn features from the data itself as part of the learning task. In order to model hierarchical data representations and solve complex prediction tasks, DNN models contain a large number (often hundreds of thousands to millions) of model parameters. These parameters are learned using training data, so DNNs typically require very large data sets for training, on the order of tens of thousands to millions of observations for the neural networks of the scale currently used in geophysical applications. For example, 1.3 million labeled waveforms from the Stanford Earthquake Data set (Mousavi, Sheng, et al., 2019) were used to train the 372,000 parameters in the EQtransformer model for earthquake detection and phase picking (Mousavi et al., 2020).

An advantage of DNNs for big data analysis is their flexibility in the choice of model architecture and learning task. Deep neural networks can be used in supervised learning or unsupervised and semi-supervised learning tasks with an appropriate loss function. Deep neural network architectures may include many types of layers, which perform different types of transformations of the data, and the connections between layers can be complex, with a single layer taking inputs from or passing outputs to multiple layers. There are a number of widely used classes of DNN architectures, including convolutional neural networks (CNNs; LeCun & Bengio, 1995), long short-term memory (LSTM) and recurrent neural networks (RNNs; Hochreiter & Schmidhuber, 1997), transformer networks (Vaswani et al., 2017), generative adversarial networks (GANs; Goodfellow et al., 2014), and autoencoder networks (Kingma & Welling, 2013). Examples of deep learning in seismology are discussed in Sections 3.1.1, 3.3.2, 3.3.3, and 4.2.

Transfer Learning

A major obstacle preventing the adoption of deep neural networks for applications in Big Data Seismology is the "big data, small labels" problem: the availability of a large data set containing few high-quality labels relevant to the task of interest. In these applications *transfer learning* offers a way to build ML models from small data when there is not enough labeled data to train a DNN or other large-scale ML model from scratch. Transfer learning (Pan & Yang, 2009; C. Tan et al., 2018) is a set of ML techniques for improving prediction performance on a "target task," for which data may be scarce, by transferring information learned on a related "source task." The source task is typically a task for which labeled data are easier to acquire, and information is transferred by sharing the feature representation or model parameters learned for the source task. For example, representations learned from 1.2 million natural images have been successfully used to initialize DNN models for medical imaging tasks that can then be *fine-tuned* using a few hundred task-specific labeled medical images (Shin et al., 2016). In seismology, a transfer learning approach was used to train a phase picker for mesoscale hydraulic fracturing experiments with only 3,500 seismograms by fine-tuning model parameters initially trained on a larger catalog of local seismic data (Chai et al., 2020).

Transfer learning enables learning from heterogeneous data, including applications in which training and target data do not belong to the same distributions or in which the domains or target prediction task may differ from available labeled data. Transfer learning encompasses a broad set of learning problems and methods including: inductive transfer (different source and target tasks, same domain), domain adaptation (similar task, different domains), zero- or one-shot learning (predicting target class not included in training data; Lampert et al., 2009), and multitask learning (co-learning multiple tasks; Y. Zhang & Yang, 2021). Transfer learning has been effective in a number of scientific applications in which data for the target task are scarce or low-quality, such as combining large computational and small experimental data sets (Jha et al., 2019) or adapting general models to specific tasks with relatively few, low-quality labeled examples (Pesciullesi et al., 2020). In seismology, the potential for transfer learning was also explored in the context of generative models by T. Wang et al. (2021), where detection algorithms trained on generated synthetic waveforms were used to improve detection performance in observational data sets.

Other ML Tools and Techniques

Learning tasks such as semi-supervised learning and self-supervised learning offer alternatives to handling the big data, small labels problem in Big Data Seismology. *Semi-supervised* learning allows researchers to leverage both labeled and unlabeled observations during model training (Chapelle et al., 2010). One class of semi-supervised learning methods, called *active learning*, sequentially queries labels for small subsets from a larger unlabeled pool in order to improve prediction performance and reduce model uncertainty. Some ML methods, like GANs, can be trained in either an unsupervised or semi-supervised manner depending on data availability. *Self-supervised* learning is a learning task in which labels for supervised training are generated automatically from an unlabeled data set. For example, a self-supervised approach is used to train a CNN model with a U-net (Ronneberger et al., 2015) architecture for blind denoising of DAS data (M. van den Ende et al., 2021); no denoised data are available for training, so instead the model is trained to reconstruct a zeroed waveform from the waveforms at 10 neighboring channels. If large volumes of labeled data are not available but labels can be generated on the fly, *online learning*, which learns on a single pass through data, or *reinforcement learning* (Arulkumaran et al., 2017), which learns sequentially through a trial-and-error process, offer solutions for big data analysis. For data sets that cannot be easily shared, *federated learning* is a class of ML methods for decentralized training of ML models (Konečný et al., 2016). This enables learning from data sets that cannot be stored on a single centralized server due to data privacy concerns or technical barriers to data access and transfer.

3. Science Enabled by Big Data Seismology

This section explores the scientific opportunities that are being made possible by advances in data, algorithms, and computing. In Figure 2, the mapping between drivers and opportunities is denoted schematically. In this section, we expand on these opportunities in three major sections: earthquake processes, imaging (which includes tomographic imaging and exploration seismology), and non-earthquake sources. In Section 4 we address the practical opportunities that relate to monitoring and hazards.

3.1. Earthquake Source Processes

Some of the most profound scientific insights of the era of Big Data Seismology have come from an improved ability to characterize earthquakes and their rupture processes across multiple scales. Here we describe recent advancements in earthquake detection and source analyses enabled by new, high-resolution seismic data sets. The drivers of Big Data Seismology described in Section 2, from new array technologies to novel detection algorithms, have played a fundamental role in enabling these new insights into the physics of earthquakes.

3.1.1. High-Resolution Earthquake Catalogs: Insights Into Earthquake Nucleation and Triggering

Earthquake monitoring is one of the oldest and most important tasks in seismology. Across the globe, seismic network analysts monitor real-time waveform streams for phase arrivals generated by earthquakes and other natural or anthropogenic events. If phase arrivals at multiple stations are detected within a short timespan, they may be associated with a causative event for which the location and origin time can then be inferred by examining the timing of the arrivals across the network.

This basic workflow for seismic network monitoring, while well-established and broadly successful, has several important limitations. Traditional monitoring of this type relies extensively on manual analysis and expertise from human analysts to detect valid phase arrivals, associate them with real events, and provide timely and accurate hypocentral estimates. This poses a significant challenge as network densities increase, as the rapidly growing volume of network data will far outpace any growth in the expert workforce. Even with the help of semi-automated tools like STA/LTA detection algorithms (R. V. Allen, 1978), analysts will inevitably need to focus only on the larger events, leaving many of the smaller (though potentially detectable) events uncatalogued. Because of this, earthquake catalogs are inherently incomplete (Mignan & Woessner, 2012; Woessner & Wiemer, 2005), preferentially documenting the largest events while leaving out the smallest.

These "missing" small earthquakes, while not hazardous in and of themselves, can provide essential observational constraints on earthquake and stress transfer processes. With this in mind, in recent years there has been a substantial push to use advanced techniques from data science, signal processing, and ML to analyze large volumes of seismic data with an aim to significantly improve catalog completeness. The strategies that have been employed to do so are numerous. Perhaps the largest body of work has come on the earthquake detection problem through the use of similarity search algorithms (Rong et al., 2018; Skoumal et al., 2016; Yoon et al., 2015), including template matching (Gibbons & Ringdal, 2006; Ross, Idini, et al., 2019; Shelly et al., 2007; Skoumal et al., 2014) and ML algorithms trained on large volumes of human-labeled phase arrivals (Perol et al., 2018; Ross et al., 2018; Mousavi et al., 2020; Mousavi, Zhu, et al., 2019; W. Zhu & Beroza, 2019). Machine learning algorithms have also been used in phase association and earthquake location problems (McBrearty et al., 2019; Ross et al., 2020; Ross, Yue, et al., 2019; H. Shen & Shen, 2021). More commonly, large-scale waveform cross-correlation calculations have been applied extensively to refine the relative locations of earthquake hypocenters (Matoza et al., 2013; P. Shearer et al., 2005; Trugman & Shearer, 2017; Waldhauser, 2009; Waldhauser & Ellsworth, 2000), dramatically improving our understanding of the in situ fault structure on which earthquake sequences occur (Chamberlain et al., 2021; Ross et al., 2020; Ross, Trugman, et al., 2019; Schoenball & Ellsworth, 2017; Shelly, 2020; Shelly et al., 2016; Skoumal et al., 2019; Y. J. Tan et al., 2021; Trugman, Ross, & Johnson, 2020), see Figure 7. While such techniques can be applied at a small scale, regional-scale analyses are the true realm of Big Data Seismology and can require considerable computing resources even with highly optimized algorithm design. The Quake Template Matching catalog (Ross, Trugman, et al., 2019) of southern California involved several hundreds of thousands of GPU-hours to exhaustively search 10 years of continuous waveform data from hundreds of Southern California Seismic Network stations for hidden earthquakes buried in the noise generated by larger earthquake signals or other ambient processes. The application of template matching in this case improved catalog completeness by an order of magnitude (1.81 million detected events compared to 180,000 in the network-reviewed catalog), thus substantially expanding the sample size of observations to study fundamental earthquake processes. Because earthquake magnitudes obey a power law distribution where small earthquakes occur much more frequently than large ones, the improvements to detection capability enabled by Big Data Seismology can have considerable scientific leverage.

One important application of these new, high-resolution catalogs is in the study of the earthquake nucleation process—how earthquakes get started. It has long been recognized that some, though not all, earthquakes are preceded by a sequence of smaller events called foreshocks. Foreshocks are ubiquitous in well-monitored

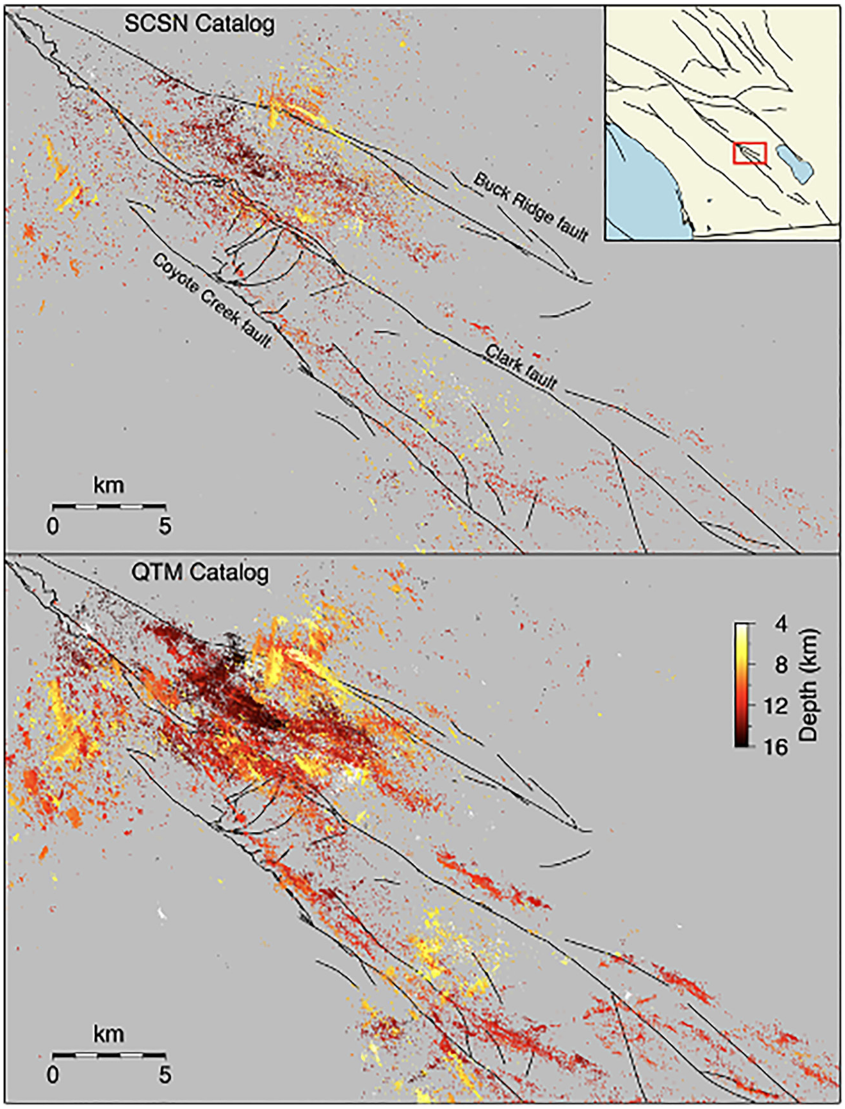

Figure 7. Improved resolution in fault structure near San Jacinto, California from large-scale template matching calculations. Each panel plots cataloged earthquake location, with each event displayed as a small circle color-coded by its hypocentral depth. The template matching catalog (bottom panel) shows a dramatic (10x) increase in detected seismicity and resolution of fault structure compared to the regional network catalog shown in the top panel. Adapted from Ross, Trugman, et al. (2019).

laboratory experiments (Bolton et al., 2019; Corbi et al., 2019; Hulbert et al., 2019; P. A. Johnson et al., 2013; Lubbers et al., 2018; McLaskey & Lockner, 2014; Rouet-Leduc et al., 2017; Thompson et al., 2009; Trugman, McBrearty, et al., 2020), and are commonly observed in natural fault systems (Abercrombie & Mori, 1996; X. Chen & Shearer, 2016; Dodge et al., 1996; Jones & Molnar, 1976; Seif et al., 2019; P. M. Shearer, 2012). This suggests that they play an important role in the nucleation process, but the underlying physical mechanisms are not yet well-understood. For example, foreshocks may play an active role in the nucleation process through a cascading transfer of stress culminating in the eventual mainshock (Dodge et al., 1995; Ellsworth & Beroza, 1995; Ellsworth & Bulut, 2018; Yoon et al., 2019). Alternatively, foreshocks may be more of a passive indicator of a broader, largely aseismic preslip process and play no direct role in nucleation (Inbal et al., 2017; Kato et al., 2012; McLaskey & Lockner, 2014; Mignan, 2014). In real fault systems, the cascade and preslip models of nucleation are not mutually exclusive and indeed may feedback on one another (Cattania & Segall, 2021; McLaskey, 2019; Noda et al., 2013). Recent observations from high-resolution earthquake catalogs around the world (Cabrera et al., 2022; Durand et al., 2020; Ellsworth & Bulut, 2018; Feng et al., 2021; Gardonio et al., 2020; Malin et al., 2018; H. Meng & Fan, 2021; Moutote et al., 2021; Sánchez-Reyes et al., 2021; Shelly, 2020; Trugman &

Ross, 2019; M. P. A. van den Ende & Ampuero, 2020; Yao et al., 2020; Yoon et al., 2019) are beginning to bridge the gap between laboratory and field scales by providing more complete and holistic observations of the nucleation process. The diverse range of physical processes highlighted by these recent studies suggest that earthquake nucleation does not follow a simple, uniform trajectory common to all earthquakes, but instead may be complex and highly dependent on the details of the faulting environment and stress regime. Earthquake nucleation is likely to remain a pivotal focus of scientific debate, and future studies enabled by enhanced earthquake catalogs may help elucidate or reconcile the disparate set of present observations.

Another area where high-resolution earthquake catalogs have shown significant value is in studies of earthquake triggering. Earthquakes redistribute stresses within the Earth, releasing stress along the faults that slip while concentrating stress at fault edges and in the surrounding region. Earthquakes also radiate seismic waves and initiate time-dependent postseismic processes that further perturb the stress field. The relative importance of static, dynamic, and postseismic stress transfer in the triggering of aftershocks is an important problem that has been widely studied in the literature (Freed, 2005; Gomberg & Davis, 1996; Hardebeck, 2014; Hardebeck et al., 1998; Kilb et al., 2000; King et al., 1994; Richards-Dinger et al., 2010; Stein, 1999; van der Elst & Brodsky, 2010). Having access to precise and highly complete earthquake catalogs can dramatically expand the sample size of events to measure changes in earthquake rate, thus providing essential observational constraints on these processes (Aiken & Peng, 2014; Delorey et al., 2015; Fan et al., 2021; Frank & Abercrombie, 2018; Miyazawa et al., 2021; Peña Castro et al., 2019; B. Wang et al., 2018). This is particularly important during dense aftershock sequences, where traditional network analysis is overwhelmed by larger earthquakes and high levels of background noise, leading to transient jumps in the completeness level of regional network catalogs (Hainzl, 2016). These same tactics can be used to study other triggering mechanisms in greater detail, such as that from tidal forcing (Cochran et al., 2004; Delorey et al., 2017; Heaton, 1975). In essence, better earthquake catalogs can enhance small signals and provide a sort of magnifying glass with which to view important physical processes (Brodsky, 2019).

3.1.2. Dense Arrays: New Constraints on Source Processes

It has long been recognized that dense arrays of seismometers have the potential to provide important and complementary information about geophysical processes beyond that available from analysis of traditional seismic network data. At the regional scale, the most impactful example of this may be the Earthscope USArray initiative, discussed in Section 2. USArray data have become a cornerstone of tomography and imaging studies (Buehler & Shearer, 2017; Burdick et al., 2017; Ma & Lowry, 2017; P. M. Shearer & Buehler, 2019).

While the emphasis on tomography is understandable given its direct analog with industrial arrays used in exploration seismology, it has also become increasingly clear that dense arrays have tremendous utility in characterizing the earthquake source. In particular, back-projection methods that take advantage of the temporal symmetry of the wave equation to form an image of the earthquake rupture (Ishii et al., 2007; Walker et al., 2005), have become widely used in global and regional studies of large earthquakes (Fan & Shearer, 2015; Kiser & Ishii, 2012, 2017; Koper et al., 2011; Mesimeri et al., 2020; Xu et al., 2009; Zhan et al., 2014). Array-based waveform analyses provide a complementary characterization of rupture kinematics to that which can be obtained through more traditional, finite fault inversions (Fan & Shearer, 2017; Kiser & Ishii, 2017; S. Park & Ishii, 2015). For example, while back-projection images cannot be used directly to constrain moment release, they give the most compelling observational constraints on rupture velocity and depth-dependent changes in the frequency content of radiated energy (Fan & Shearer, 2015; Kiser & Ishii, 2012; Koper et al., 2011; Walker & Shearer, 2009; Zhan et al., 2014).

In recent years, array seismology is becoming redefined through increasingly dense, local deployments of seismometers. These nodal arrays, with inter-station spacing on the order of hundreds of meters, allow for a more complete measurement of the seismic wavefield with minimal spatial aliasing (Anthony et al., 2019; Karplus & Schmandt, 2018). For example, earthquake sequences that are recorded by regional seismic networks alone will commonly have gaps in station coverage of tens of degrees or more in azimuth and takeoff angle even in the most optimistic cases, which can hinder the characterization of parameters like radiated energy or radiation patterns that require averaging across the focal sphere. Nodal array deployments, in contrast, largely eliminate such issues in station coverage (at least at a local scale), thus opening up new opportunities for targeted scientific studies.

A noteworthy early example of the use of nodal arrays in passive seismology was the deployment Long Beach array in southern California, described in Section 2. While the Long Beach array has been used primarily for

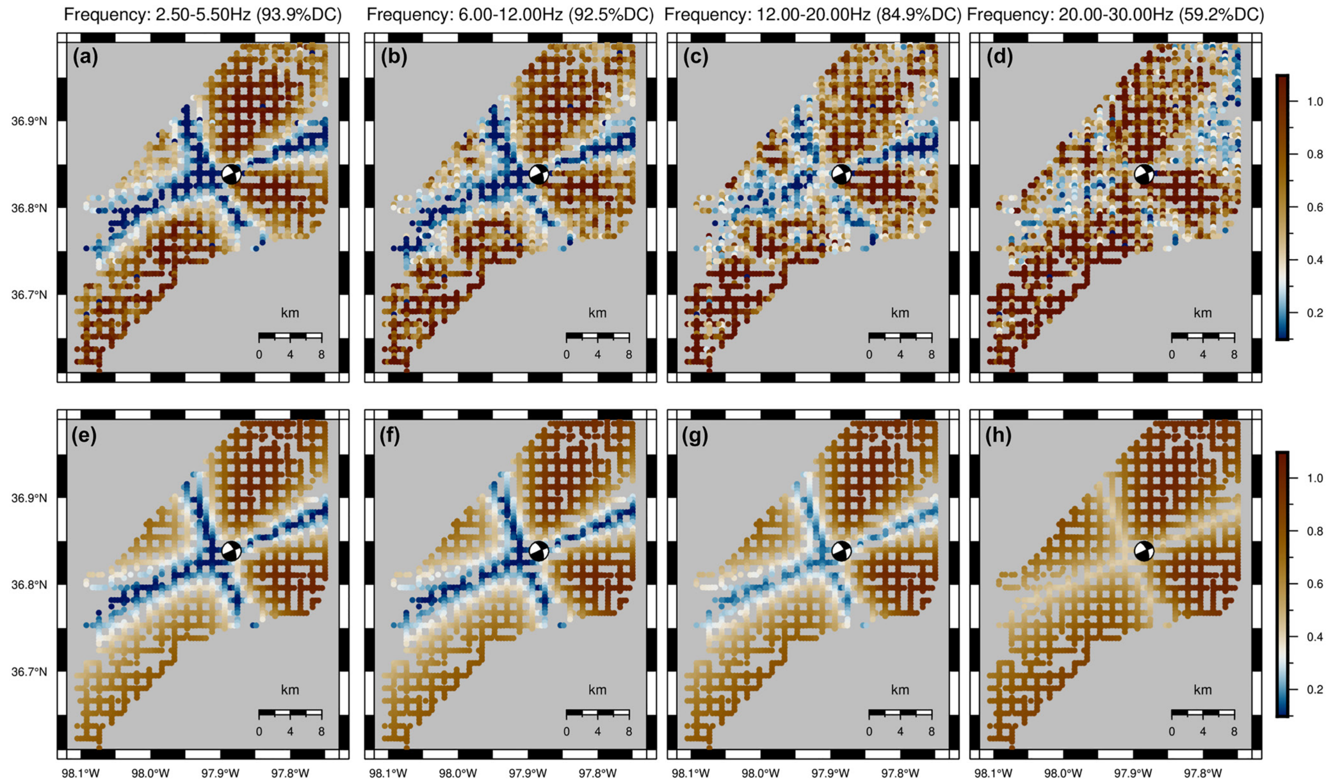

Figure 8. Using dense arrays to study frequency-dependence of earthquake radiation patterns. The top panels show the spatial pattern of P-wave amplitudes for a M_L 2.0 earthquake recorded by a dense nodal array in Oklahoma. Each top panel corresponds to a different frequency band, increasing from left to right. The bottom panels display predicted amplitudes from a radiation pattern model that gradually transitions from fully double couple to progressively more isotropic with increasing frequency. Adapted from Trugman et al. (2021).

tomography, other dense deployments have focused more directly on the earthquake source. In California, temporary deployments near the San Jacinto Fault (Ben-Zion et al., 2015; Y. Cheng et al., 2020; C. W. Johnson et al., 2019, 2020; H. Meng & Ben-Zion, 2018b; Roux et al., 2015; Share et al., 2017; Sheng et al., 2021; Y. Wang et al., 2019) and among the Ridgecrest aftershocks (Catchings et al., 2020) captured high rates of local seismicity, as did analogous deployments in areas of induced seismicity in Oklahoma (Dougherty et al., 2019; Sweet et al., 2018) and Alberta, Canada (Eaton et al., 2018). Array analysis has also played a foundational role in the detection and characterization of non-tectonic events such as low frequency earthquakes (Asano et al., 2008; Hutchison & Ghosh, 2017; Sweet et al., 2014, 2019; Thomas et al., 2013; Ueno et al., 2010) and volcano-tectonic seismicity (Glasgow et al., 2018; J. Han et al., 2018; Hansen & Schmandt, 2015; H. Shen & Shen, 2021).

Analyzing waveform data from these dense local arrays often requires the development of new detection and characterization methodologies (Y. Cheng et al., 2020; Z. Li et al., 2018; H. Meng & Ben-Zion, 2018b; Poulin et al., 2019), but the effort is very often worthwhile. For example, the San Jacinto array has been used to quantify spatial variations in ground motion amplitudes (C. W. Johnson et al., 2020) and identify different classes of seismic noise (C. W. Johnson et al., 2019). In Japan, the timely deployment of the Metropolitan Seismic Observation Network (MeSO-net) in Tokyo (Hirata, 2009; Sakai & Hirata, 2009) provided a unique imaging vantage point of the 2011 M9 Tohoku rupture process (Honda et al., 2011), while rapid-response aftershock deployments like those for the 2003 M7.3 Western Tottori sequence have enabled unprecedented detail in the measurement of earthquake source mechanisms (Hayashida et al., 2020). The Tony Creek Dual Microseismic Experiment (ToC2ME) has yielded fundamental insight into the nature of earthquakes induced by hydraulic fracturing (Eaton et al., 2018; Igonin et al., 2018, 2021; Ojo et al., 2021), while the LArge-n Seismic Survey of Oklahoma (LASSO) has done the same for earthquakes induced primarily by wastewater injection (Cochran et al., 2020; Kemna et al., 2020; Trugman et al., 2021) see Figure 8. While studies using aftershock monitoring arrays near Ridgecrest (Catchings et al., 2020) are still ongoing at the time of writing, data sets like these are likely to give a comparable or even more extensive set of new insights into the physics of earthquakes.

In the coming years, DAS is likely to play an increasingly prominent role in earthquake source characterization. Applications of DAS naturally extend well beyond earthquake seismology (Lindsey & Martin, 2021), but its usage in this realm has already begun to have a notable impact (Daley et al., 2013; Z. Li & Zhan, 2018; Lindsey et al., 2017; Lellouch et al., 2020, 2021; Nayak et al., 2021; C. Yu et al., 2019; Zhan, 2020). Since optical fibers are pervasive due to their usage in telecommunications, the potential for so-called "dark fiber" analyses are immense, whether as temporary deployments during aftershock sequences (Z. Li et al., 2021) or within urban areas (Fang et al., 2020; Lindsey, Yuan, et al., 2020; Spica, Perton, et al., 2020). Analysis of submarine DAS data sets (Lindsey et al., 2019; Spica, Nishida, et al., 2020) may be particularly important, as traditional seismic networks are necessarily data poor in the offshore region, despite the obvious hazard potential in subduction zones.

3.2. Imaging

Some of the most significant discoveries in seismology have exploited seismic waves to image inaccessible parts of Earth's interior. The three drivers of Big Data Seismology, described above, are enabling applications of seismic imaging to address new science questions. Since the resolution of Earth models is fundamentally limited by the spatial density of sensors, new data-dense sensor deployments are allowing for the inversion of 3D Earth structure at unprecedented resolution, particularly of crustal and near-surface structure. Recent innovations in techniques and algorithms, in particular the development of ambient seismic noise (ASN) interferometry, are critical for imaging using short-duration nodal arrays in aseismic regions, and for recovering temporal variations in shallow Earth properties. Finally, improvements in computing play a key enabling role for utilizing the large data volumes and performing seismic inversion, particularly using full waveforms.

In this section, we briefly review the most important methodological advances in imaging applicable to Big Data Seismology, and explore the scientific questions that these are beginning to answer.

3.2.1. Methodological Advances

3.2.1.1. Ambient Noise Interferometry

Traditional seismic imaging techniques in passive seismology have relied upon earthquakes, while exploration seismology has relied on controlled sources. A major advance in the early 2000's was the development of techniques that exploit seismic ambient noise for imaging (Campillo & Paul, 2003; Shapiro et al., 2005). Ambient noise interferometry exploits the fact that seismic noise, despite containing no easily identifiable signals, is composed of propagating wavefields that obey the wave equation. Under certain conditions, the cross-correlation function of the ambient noise wavefield between two stations can be used to retrieve the Green's function—the elastic impulse response of the ground between the two locations—by effectively turning one sensor into a virtual source (Nakata et al., 2019). The theoretical assumptions on which ambient noise interferometry is based are that the noise field comes from sources located isotropically around the receivers, or that the noise field is diffuse, and that the sources are uncorrelated. Despite the fact that these assumptions are not strictly valid with real seismic noise, the technique works effectively for many sources of noise such as the global microseisms in real-world conditions. The application of ambient noise interferometry to many cultural noise sources requires developments to the workflow (Ayala-Garcia et al., 2021; Sager et al., 2021). Since the composition of seismic noise is dominated by surface waves, the use of interferometry to retrieve body waves is more challenging and relies upon dense networks of stations (Nakata et al., 2015). However, body waves can provide more accurate depth resolution of 3D Earth structure, complimenting the lateral resolution possible with surface waves. Ambient noise interferometry has been critical for many applications of Big Data Seismology discussed below. Nodal arrays are impractical for long-term deployments, and thus do not typically capture sufficient events for earthquake-based imaging, so most applications of imaging have relied upon ambient noise interferometry. Ambient noise interferometry can also provide continuous time-lapse measurements of Earth properties.

3.2.1.2. Full-Waveform Inversion

The state-of-the-art technique for estimating high-resolution 3D models of physical properties (e.g., velocity, anisotropy, density, Q) in the earth is Full Waveform Inversion (FWI; Tarantola, 1984, 1988; Virieux & Operto, 2009). Unlike classical methods based on body wave travel time or surface wave dispersion, the goal of FWI is to use all of the information in the seismogram. FWI is posed as an inverse problem to find the optimal earth model $\vec{m}$ that matches the observed seismic data $\vec{d}$ with the computationally modeled seismic data $\vec{d}_m = F(\vec{m})$ such that the misfit error of the data residuals between d and d_m is minimized. In practice, to maximize efficiency with limited HPC computational resources, the FWI inverse problem is posed as an L_2 least squares optimization problem, and solved via conjugate gradient adjoint method techniques. The FWI L_2 gradient step is calculated by correlating the reverse-time ("adjoint") propagated wavefield of the surface data residuals $F^*\left(\Delta\vec{d}\right)$ with the forward modeled source wavefield $F(\vec{m})$ for all observations; locations in the subsurface that are "imaged" by the data residuals guide the algorithm where the velocity model is incorrect, and how it needs to be updated. To ensure stability and convergence of this strongly nonlinear optimization problem, many small gradient steps are iteratively calculated and carefully conditioned. Due to computational cost, most 3D full waveform inversions start with a smooth tomographic (traveltime-based) velocity model estimate, which is then updated through a few iterations of reverse-time migration (RTM). The high resolution (wavenumber) velocity model details are added last by running a few (<10) iterations of FWI to better match the data.

FWI was originally developed and applied in exploration seismology but has recently been applied to passive seismology. It relies heavily upon two of the drivers of Big Data Seismology, including advances in algorithms and computing (Tromp, 2020). Full 3D FWI has only barely become possible in the past 10 years or so due to recent phenomenal advances in HPC hardware and sophisticated algorithms (Benfield et al., 2017; Prieux et al., 2013; Sirgue et al., 2010; Solano & Plessix, 2019; Vigh & Starr, 2008; H. Wang et al., 2021; Warner et al., 2013). The growth in both volume and density of data sets used by the passive seismology community has only recently resulted in data sets suitable for full waveform inversion (Fichtner et al., 2009; Marone et al., 2007; Schaeffer & Lebedev, 2014; Tape et al., 2009; Yuan et al., 2014; H. Zhu & Tromp, 2013; H. Zhu et al., 2012). A good example of this is the TA, which has provided the unprecedented ability to make 3D full waveform inversion estimates of the elastic properties of the US continental crust including the Moho and upper mantle (Krischer, Fichtner, et al., 2018; H. Zhu et al., 2017, 2020). Since FWI is fundamentally limited by background noise levels and data quality, the advances in automated denoising and quality control techniques described in this article should only improve future applications of FWI.

3.2.2. Scientific Opportunities

Much of the large body of literature in seismic imaging has been focused on imaging deep crustal and mantle properties that change at the rate of plate tectonics. Advances in data, algorithms and computing are continuing to enable seismologists to resolve new features of the crust, upper mantle, and deep earth that relate to tectonics and mantle circulation. However, there are also many exciting scientific opportunities afforded by Big Data Seismology that lie in imaging shallow crustal properties relating to processes that occur over a wide range of timescales (from hours to millennia). To date, most of the literature on imaging using new data-dense deployments has focused on shallow Earth properties.

3.2.2.1. Imaging the Deep Earth

Our understanding of the deep Earth will continue to be improved by advances in data, algorithms, and computing. As described in the Introduction, the USArray has allowed 3D variations in seismic wavespeed and anisotropy to be mapped at high-resolution in the crust and upper mantle at continental scale, leading to an improved understanding of the relationship between processes in the mantle and tectonic features at the surface. Much of the literature is cited in the most recent continental-scale imaging studies (Bedle et al., 2021; Porritt et al., 2021). The deployment of similar high-density arrays in Asia have enabled detailed mapping of the seismic structure of the upper mantle in Eastern Asia (Tao et al., 2018), Tibet (Xiao et al., 2020), the Alps (Paffrath et al., 2021), and the Iberian peninsula (Chevrot et al., 2014).

Deep Earth studies have often relied on sparse global networks. From a big data perspective, the value of stacking large numbers of observations to enhance the signal-to-noise and make new discoveries has been importance since the digital revolution (Ritsema & Lekić, 2020). For example, stacking observations over long time histories have proven invaluable for imaging mantle discontinuities (P. M. Shearer, 1990; P. M. Shearer & Masters, 1992). In addition to studying the crust and upper mantle, the relatively dense station spacing in the USArray has been exploited to stack core-phases to unique insights into the structure of the core-mantle boundary (Sun et al., 2017), and to map the dimensions of scatterers in the lower mantle (Ritsema & Lekić, 2020).

Many of the algorithm drivers discussed in this paper have also played an important role in deep earth research. In addition to being widely used for studying the crust and upper mantle, the development of seismic interferometry has also found application for extracting signals of new seismic phases that penetrate the lowermost mantle and core (Pham et al., 2018). The use of unsupervised machine learning to separate the signature of seismic wave scattering from effects due to radial variations in Earth properties has been used to map the presence of lateral heterogeneity near the core-mantle boundary (Kim et al., 2020).

The most significant limitation for imaging the deep earth is the uneven volumetric sampling that is driven by the constraints imposed by the distribution of sources and receivers (Tkalčić, 2017). To improve the volumetric sampling, more station deployments over the ocean basins are needed. While the cost of ocean-bottom seismometers remains prohibitive, new ocean bottom seismometers including the Japanese S-net (Takagi et al., 2019)—a dense cabled ocean-bottom network—are providing opportunities to improve images of regional tectonic processes. At global scales, the recent engineering development of free-floating seismometers in the oceans has led to demonstrable improvements in resolution of the oceanic upper mantle (Nolet et al., 2019). Distributed acoustic sensing using submarine fiber-optic cables also has considerable potential to make much-needed capture of seismic ground motions over ocean basins (Marra et al., 2018).

3.2.2.2. Imaging in New Environments

Environments where it is challenging to image the subsurface using classical seismic sensors can be more easily explored using the new data-dense sensor networks described in Section 2. DAS systems and nodal arrays allow for the investigation of Earth properties in environments where it is difficult to install traditional sensor systems. These include in glacial environments, urban environments, and (for DAS) offshore environments. New environments also open up the opportunity to use new types of sources in ASN interferometry. By being able to image these new environments, new types of science question can be explored.

The increasing use of seismology for retrieving the elastic structure of glaciers and understanding their related dynamics is being driven by the need to understand the impacts of climate change on one of the most sensitive parts of the Earth System (Aster & Winberry, 2017; Podolskiy & Walter, 2016). Recent deployments of nodal arrays (Gimbert et al., 2021) and DAS sensors on glaciers (Booth et al., 2020) have highlighted the use of these new types of high-density deployment for making unique measurements. Seismic studies of ice have traditionally been active source, or exploited ice quakes for classical seismic inversion (Aster & Winberry, 2017; Podolskiy & Walter, 2016). A challenge of ambient noise interferometry in ice is the fact that the wavelengths of ambient ocean noise cannot resolve the scale lengths of important glacial structures. Further, the sources of higher frequency noise in glacial environments do not satisfy the classical assumptions of diffuse homogeneous noise required for interferometry (Section 3.2.2). However, recent work has demonstrated that spatial averaging on dense networks of nodes enables retrieval of Green's functions (Sergeant et al., 2020). An advantage of DAS sensors over classical translational sensors is the ease with which they can be deployed in boreholes, where they can sample strain rate continuously with high vertical resolution. The deployment of a DAS in a glaciological borehole on the Greenland Ice Sheet enabled the vertical mapping, at 10 m resolution, of changes in ice crystal fabric and temperature regimes associated with glacial transitions, and a subglacial layer of consolidated sediment (Booth et al., 2020).

The oceans, which cover two-thirds of Earth's surface, have always been a challenging environment to record ground motions. The gap in classical long-duration seismic observations in the oceans represents a major limitation in resolving Earth's deep interior, and of understanding tectonic processes near ocean-ocean and ocean-continent plate boundaries (Marra et al., 2018; Romanowicz et al., 2009). Seismic imaging of Earth structures beneath the seafloor has relied upon short-duration but high-density active source surveys using large research vessels, which have provided important insight on divergent and convergent plate boundaries in the ocean (e.g., Arnulf et al., 2018, 2021; Christeson et al., 2007), and a small number of longer-duration but ultra-low-density permanent ocean bottom seismometers (Marra et al., 2018). The use of DAS interrogators that can leverage existing dark fiber under the seafloor has great potential to help address this data gap (Marra et al., 2018). Submarine optical fiber cables cover over 1 million kilometers in length, and already cross many of the ocean basins, yet developments in the range of DAS systems, and additional supporting infrastructure would be still required before this could be fully exploited for seismology (Marra et al., 2018). In the meantime, a handful of recent studies have utilized DAS cables to image ocean sediments and faults in the Continental Shelf offshore California (F. Cheng et al., 2021; Lindsey et al., 2019), Belgium (Williams et al., 2021), and Greece (Lior et al., 2021). These studies have shown that ambient noise interferometry can be applied to Scholte waves to obtain shear-wave velocity profiles, and that the locations of faults can be identified with scattering.

Volcanoes are complex environments that comprise small-scale variations in three-dimensional seismic properties that are challenging to resolve with traditional seismic deployments (Nagaoka et al., 2012). Volcanic plumbing systems are also dynamic processes, with recent work demonstrating that temporal variations in seismic properties over timescales ranging from days to years can provide important insights on magmatic activity within volcanic edifices (Brenguier, Rivet, et al., 2016). One of the earliest nodal seismic arrays was deployed on Mount St. Helens in 2014 (Hansen & Schmandt, 2015), and has been used to image the 3D P and S wave structure using local tomography, resolving small-scale features with dimensions of 5–20 km interpreted as magma plumbing systems (Kiser et al., 2016; Ulberg et al., 2020). While ambient noise interferometry has been widely used for surface waves, a study using data from the Long Beach nodal array (Nakata et al., 2015) found that dense networks of arrays could be used to extract body waves using interferometry, which is important for resolving the depths of structures. A set of three dense seismic nodal arrays was deployed around Piton de la Fournaise volcano with the goal to use ambient noise interferometry to extract body waves traveling through the active magma chamber (Brenguier, Kowalski, et al., 2016). The application of ambient noise interferometry to data from these nodal arrays found subtle changes in seismic wave velocity associated with transient crustal deformation (Mao et al., 2019). Interferometry using coda waves is another new technique that is particularly useful in volcanoes, since many signals are emergent but have strong coda. Coda wave interferometry has been applied to a dense network of temporary stations on Erebus volcano (Chaput et al., 2015).

One of the original large-N nodal arrays was deployed in the urban environment of Long Beach, CA described in Section 2. While the deployment included an active source vibroseis survey, it also motivated the use of ambient noise methods for imaging in urban environments on dense networks (Castellanos & Clayton, 2021; Jia & Clayton, 2021; Lin et al., 2013; Nakata et al., 2015). The use of active source surveys in urban settings is costly, intrusive, and logistically challenging. The analysis of data from the Long Beach array, and additional Large-N nodal arrays deployed in the Los Angeles area, have utilized passive techniques including ambient noise interferometry of surface waves (Jia & Clayton, 2021; Lin et al., 2013), body waves (Nakata et al., 2015), and receiver functions (X. Wang et al., 2021) to reconstruct high-resolution 3D velocity models from the surface to depths of a few kilometers. High-resolution velocity models can provide important constraints for earthquake hazard assessment, which relies on models of peak ground acceleration (Castellanos & Clayton, 2021), a property that is very sensitive to crustal properties and can vary spatially across multiple scales. Nodal arrays have subsequently been deployed in other urban areas including Albuquerque and Singapore (Finlay et al., 2020; Lythgoe et al., 2020). As discussed in the section on non-earthquake sources, sensors in urban environments record a rich variety of signals from anthropogenic sources. From the perspective of seismic imaging, many of these sources can also be used as sources for ambient-noise imaging at high frequencies (Brenguier et al., 2019; Pinzon-Rincon et al., 2021). As described above, the challenge of using these sources, which violate some of the assumptions of ambient noise interferometry, is an active topic of research. There is also significant opportunity to exploit existing dark fiber in urban environments with DAS interrogators as a means to probe the near-surface structure (Fang et al., 2020; Spica, Perton, et al., 2020), with applications to hazard monitoring as well as to time-lapse imaging.

3.2.2.3. Imaging at High Resolution

The new types of data-dense observation and algorithms discussed in this paper are allowing academic seismologists to make industry style measurements at much lower cost, and are filling an observation gap between classical passive and exploration seismology. Passive observations of ambient noise are being explored to image small-scale structures like fault zones, which contain strong heterogeneities in seismic properties (Hillers et al., 2014) and extend to depths beyond those typically resolved with active seismic surveys (Mordret et al., 2019). Recent nodal array deployments in Long Beach (Lin et al., 2013) and across the San Jacinto fault zone (Ben-Zion et al., 2015) have resolved 3D variations in seismic wavespeed and anisotropy at high resolution. The application of ambient-noise interferometry to the Long Beach nodal array (Castellanos & Clayton, 2021) has allowed for the 3D resolution of fault zone structures, as well as small-scale structures associated with an aquifer and river (Figure 9). A recent DAS experiment on the Reykjanes Peninsula in Iceland was able to clearly resolve structural

10.1029/2021RG000769

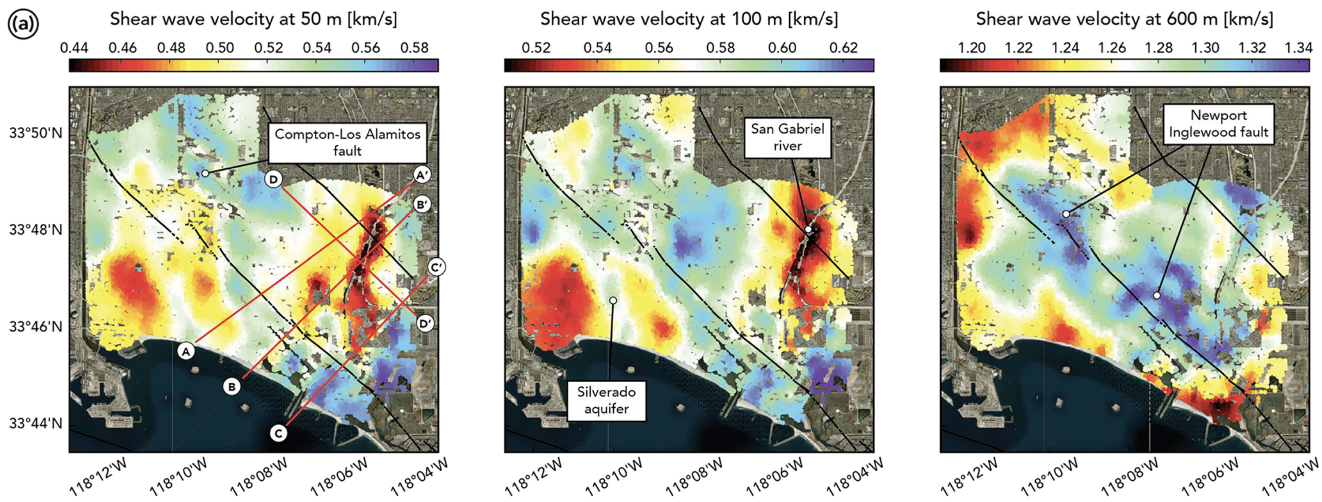

Figure 9. Horizontal slices through a shear wave velocity model obtained using ambient-noise interferometry across the Long Beach nodal array (Castellanos & Clayton, 2021). Structures associated with fault zones, aquifers, and rivers are observed in the image.

features such as normal faults and volcanic dykes (Jousset et al., 2018). The analysis of trapped waves in the fault zone allowed the detailed mapping of phase propagation within a fault-damaged zone using both earthquakes and microseismic noise (Jousset et al., 2018). While the era of Big Data Seismology allows academic researchers to image structures at scales that have traditionally been the domain of active source industry experiments, it is also opening up practical applications for seismic hazard assessment. The use of computational modeling to assess ground-motion accelerations from large earthquakes is greatly enhanced by the ability to resolve structures at the scales made possible by nodal arrays such as Long Beach (Castellanos & Clayton, 2021) and San Jacinto (C. W. Johnson et al., 2020). Further, passive methods with nodal arrays can resolve the structure of sedimentary basins in urban environments, where conventional active source seismic techniques are too obtrusive. Nodal arrays and passive imaging techniques such as interferometry (Castellanos & Clayton, 2021) and receiver functions (X. Wang et al., 2021) provide an alternative path to characterizing seismic hazard at high resolution.

In exploration seismology, a typical 3D digital earth model today may be 1,000 sq.km in area and up to 6–10 km in total depth (Figure 10), requiring over 1 billion computational voxels (Shragge, Bourget, et al., 2019; Shragge, Lumley, et al., 2019). Since each voxel must store 1–10 physical properties (velocity, density, anisotropy, Q, etc.), the typical size of the 3D digital earth model is on the order of 10–100 Gigabytes, often requiring virtual shared memory HPC systems, or computational domain decomposition across distributed memory systems. Most 3D seismic wavefield computations are performed using a finite difference (FD) or finite element (FE) method to solve an acoustic or elastic wave equation in the space-time or space-frequency domain. Due to the large size of the 3D computational domain, and the large computational effort to solve for the 3D seismic wavefield propagation, it can typically take on the order of 1 computational hour to model a single 3D seismic shot gather on a Top50-ranked parallel computing CPU or GPU cluster.

3.2.2.4. Imaging Time-Varying Processes

The development of time-lapse (4D) seismology (R. Calvert, 2005; Lumley, 2001, 2010; Lumley et al., 2015; Nur, 1989; van Gestel, 2021) was pioneered in exploration seismology based on repeating 3D seismic surveys and imaging procedure. The development of ambient noise interferometry, which allows the continuous measurements of Earth properties, has enabled seismologists to study temporal variations in Earth properties passively. For example, the response of crustal properties to dynamic stress perturbations, measured through changes in seismic velocity ($\delta v/v$) have been used to image coseismic and postseismic changes at fault zones (Brenguier, Campillo, et al., 2008), to delineate the structure of volcanic plumbing systems (Brenguier et al., 2014), and to measure precursors to volcanic eruptions (Brenguier, Shapiro, et al., 2008). This technique, in combination with the development of data-dense sensor networks, is enabling the use of seismology to study subtle temporal variations in Earth properties at shallow depths that are associated with environmental drivers. As an example, the measurement of $\delta v/v$ on the dense Hi-net seismic network in Japan has found variations in climate-induced

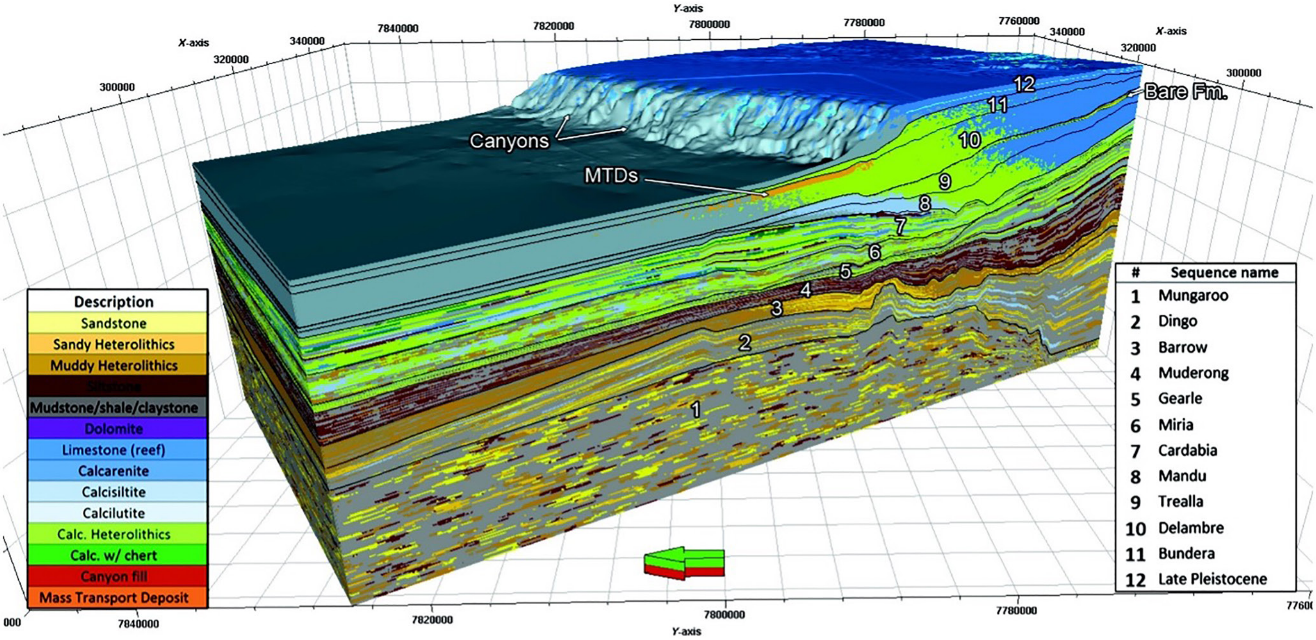

Figure 10. 3D digital earth model offshore NW Australia, 100 × 50 × 6 km sampled at 25 m lateral and 5 m vertical, representing more than 10 billion voxels and 12 computational hours per 3D seismic FD elastic shot gather on a Top50 HPC cluster (Shragge, Bourget, et al., 2019; Shragge, Lumley, et al., 2019).

pore pressure changes associated with spatiotemporal variations in precipitation and snow depth (Figure 11), which increase the pore-fluid pressure (Q.-Y. Wang et al., 2017). Increasing the data-density, and frequency of ambient noise sources, allows seismologists to study processes at the spatiotemporal resolutions necessary to study the near-surface response to hydrological and meteorological processes in the critical zone (Parsekian et al., 2015). For example, data from a nodal array at the Shale Hills Critical Zone Observatory was used to detect velocity changes in the critical zone with spatial resolutions of 1–100 m and temporal resolutions of 1 hr (Oakley et al., 2021). These changes were explained by variations in temperature and water infiltration. DAS sensors are

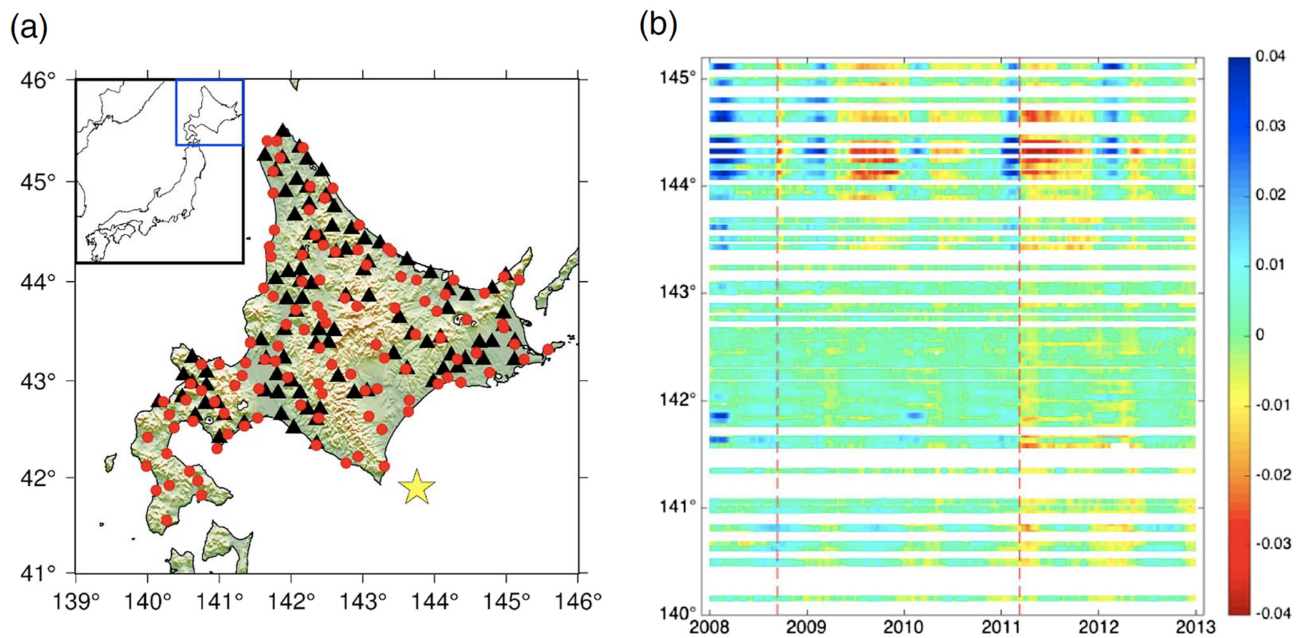

Figure 11. Map of Japanese Hi-net stations in Hokkaido (black triangles) and climate stations (red triangles) and estimates of $\delta v/v$ as a function of longitude showing changes due to earthquakes (red dashed lines) and seasonal hydrological effects, which are particularly pronounced on the east of Hokkaido (Q.-Y. Wang et al., 2017).

particularly well-suited to monitoring temporal variations in near-surface Earth properties as they provide high data-density and can be operated continuously for long durations. In an urban environment, ambient noise from traffic was used to measure shear-wave velocities in the top 20 m as a function of time, finding that the estimates did not change significantly but that repeated measurements provided estimates of uncertainty (Dou et al., 2017). DAS observations using a dark fiber in the Sacramento Valley of California have estimated variations in $\delta v/v$ that correlated with groundwater levels in an aquifer, suggesting a role for seismology in the dynamics of aquifers at unprecedented spatiotemporal resolutions (Rodríguez Tribaldos & Ajo-Franklin, 2021).

3.3. Non-Earthquake Sources

The discipline of seismology classically includes two main prongs—improving our understanding of earthquakes and imaging Earth's interior—for which we have shown that Big Data Seismology is opening up new opportunities. This section explores similar opportunities in other traditional subfields of seismology, including volcano seismology and forensic seismology. In the past few years, the drivers of Big Data Seismology have also enabled the growth of entirely new subfields of seismology, including environmental seismology and urban seismology.

3.3.1. Environmental Seismology

Environmental Seismology has been defined as the study of ground motions that originate from, or are affected by, causes outside the solid Earth (Larose et al., 2015). In this paper, we will restrict the definition of environmental seismology to the study of natural sources of ground motion, which distinguishes it from urban seismology. In this context, environmental seismology includes the study of ground motions that are caused by processes in the atmosphere, hydrosphere, and cryosphere. The same drivers of Big Data Seismology, discussed above, have played a key role in the recent surge of papers in environmental seismology. Much of environmental seismology involves the study of low-amplitude signals or continuous ground motions that can only be detected and studied using dense networks and new algorithms. These ground-motions can dominate the composition of seismic data from nodal and DAS sensors, and are a significant source of noise for studying earthquake processes. As described in Section 3.2, the discipline has widely exploited ambient noise interferometry to use these sources of ground motion to study near-surface properties. As explored in this section, Big Data Seismology is also beginning to use these sources of ground motion to study the associated Earth processes themselves.

Seismology is playing an increasingly pivotal role in understanding glacier dynamics (Aster & Winberry, 2017; Podolskiy & Walter, 2016). Seismic observations provide unique insight into physical processes that control basal motion, including stick-slip glacial earthquakes and subglacial flow, both of which affect the rate of flow of glaciers and are interconnected (Nanni et al., 2021). Seismic observations also provide information on englacial fracturing, and the role of fracturing on iceberg calving. Since these processes occur beneath ice and over a range of timescales, they are challenging to monitor with other techniques. The use of dense networks on glaciers, including nodes (Gimbert et al., 2021) and DAS (F. Walter et al., 2020), provide the opportunity to lower the event detection threshold and improve the subsequent location and characterization of events. Observations from a dense nodal array on a valley glacier in the French Alps have been used to locate impulsive events from basal stick slip and englacial fracturing sources as well as tremor-type signals from subglacial water flow. Data from the same nodal array have provided high-resolution spatiotemporal maps of seismic source locations, mapping the subglacial drainage system (Nanni et al., 2021). DAS measurements on glaciers have been used to perform full waveform source inversions of icequakes (Hudson et al., 2021), and to locate stick-slip icequakes with extremely high precision. Ultimately, improved observations of these glacial processes are needed in order to improve models of glacial dynamics and monitor how glaciers are responding to climate change.

The dominant cause of continuous ground motions between 5 and 25 s period is related to wind-driven oceanic waves and their interaction with coastlines or with opposing wind-driven waves (Aster et al., 2008). These ground motions, referred to as microseisms, have long been considered a source of noise by seismologists but have found widespread utility as a source in ambient noise interferometry, as discussed in Section 3.2. Long-running seismic stations, such as the stations in the GSN network, have also provided the opportunity to explore multidecadal climate-induced signatures in microseisms (Aster et al., 2008). While not necessarily "big" in terms of the number of spatial measurements, climate studies using seismic data are "big" in the sense they require long

10.1029/2021RG000769

time-histories of recordings and we consider them an application of Big Data Seismology. Seismic observations of microseisms have been used to extract information about ocean waves (Bromirski & Duennebier, 2002) and to identify effects due to climate drivers (Anthony et al., 2017; Aster et al., 2010). Subtle differences in hydroacoustic waves generated by earthquakes observed over multiple years have also been used to measure changes in the temperature of the ocean associated with climate change, in a technique referred to as ocean thermometry (W. Wu et al., 2020). Large spatial density observations of microseisms have recently been made with offshore DAS measurements (Lindsey et al., 2019; Williams et al., 2019). These studies have been used to make observations of in situ microseism generation, and to identify additional oceanic effects including infragravity waves and internal waves (Lindsey et al., 2019; Williams et al., 2019).

Ground motions driven by atmospheric processes are usually regarded as a source of noise on seismic recordings, with sensors conventionally emplaced in boreholes or vaults to reduce this contribution to ground motion. With increasing network densities, it becomes impractical to deploy seismometers in this manner. Further, nodal arrays and DAS systems are designed to be deployed quickly and easily, and thus are typically more sensitive to atmospheric ground motions than conventional broadband seismic deployments (C. W. Johnson et al., 2019). These ground motions include contributions due to meteorological drivers and air-to-ground coupled acoustic and infrasound waves (Arrowsmith et al., 2010). Early studies of The USArray Transportable Array (TA) identified air-to-ground coupled infrasound on seismic recordings, providing unique information on atmospheric gravity waves (Hedlin et al., 2010). These studies motivated the addition of barometric pressure sensors, microbarometers, and meteorological sensors to TA sites (Tytell et al., 2016). These data have been used to directly measure gravity waves in the atmosphere (de Groot-Hedlin et al., 2014), phenomena that occur at sufficiently fine spatiotemporal scales that they are not resolved in state-of-the-art atmospheric models. The observation of atmospheric ground motions thus can provide direct information on the state of the atmosphere, through measurements of gravity waves and of infrasound waves, which can be used to probe the atmosphere at altitudes that are poorly sampled with other technologies (Assink et al., 2019). At very dense spatial resolutions, the study of a nodal seismic data set at Sage Brush Flats in California has related spatial variations in wind-generated ground motions to effects of vegetation and structures (C. W. Johnson et al., 2019). DAS sensors have recently been used to study thunder-induced ground motions (Figure 12), and have been able to detect and locate thunder events (Hone & Zhu, 2021; T. Zhu & Stensrud, 2019).

3.3.2. Urban Seismology

Traditional applications for seismology have avoided urban environments, which are characterized by high levels of noise and restrictive land access. However, the new types of data-dense sensor deployments described in Section 2 offer unique opportunities to study urban sources and near-surface seismic properties (discussed in the Imaging section), with various practical and scientific applications. Community sensors are more commonly located in urban settings and DAS can exploit dark fiber in urban environments. These opportunities are aided by new algorithms, in particular the use of ML for signal detection and classification.

Urban environments contain a wealth of sources of ground motion, commonly referred to as "cultural noise" and containing frequencies above 1 Hz (Bonnefoy-Claudet et al., 2006). New data-dense networks in urban environments—including community sensors, nodes, and DAS—have shown the potential of seismology for studying many of these sources. The recent restrictions associated with the COVID-19 pandemic have also generated significant interest in urban seismology, with data from classical and community sensors (Lecocq et al., 2020) and DAS (Lindsey, Yuan, et al., 2020) showing clear reductions in cultural noise associated with shutdown measures. These studies have highlighted the potential to use seismology for city-scale monitoring, without the privacy concerns associated with using cellphone data (Lindsey, Yuan, et al., 2020). Data from dense networks have recently been used to study traffic, trains, aircraft, and footsteps. Recent studies of data from the Long Beach nodal array explored spatiotemporal characteristics of traffic, railroad, airport noise, and oil production (Riahi & Gerstoft, 2015, 2017). These studies have highlighted that seismic data could be useful for traffic monitoring (Figure 13). Even in remote tectonic areas, aircraft-generated ground motions have been found to contribute more to ground motions than earthquakes (H. Meng & Ben-Zion, 2018a). Supervised ML techniques have been leveraged to detect ground motions generated by footsteps (Jakkampudi et al., 2020) and airplanes (X. Zhang

10.1029/2021RG000769

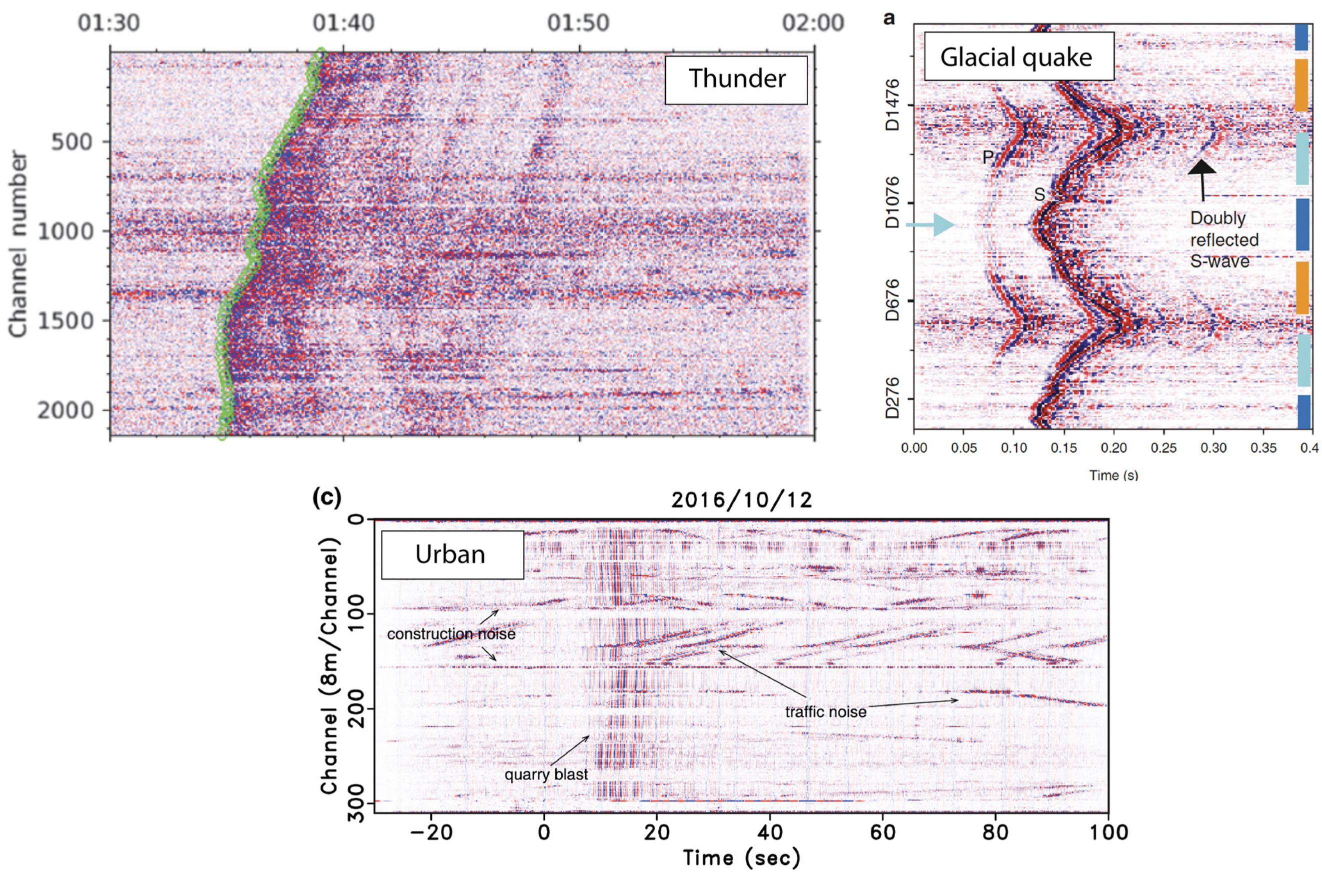

Figure 12. A selection of examples of Distributed Acoustic Sensing measurements from the literature including thunder (Hone & Zhu, 2021), basal ice quakes (F. Walter et al., 2020), and urban noise (Fang et al., 2020).

et al., 2022). The role for unsupervised ML for grouping time periods of distinct cultural noise into clusters has been demonstrated using data from the Long Beach nodal array (Snover et al., 2021).

3.3.3. Volcano Seismology

Seismology has been used to understand and monitor volcanoes for many years (McNutt, 2005). Major challenges associated with volcano seismology include the variety of different types of source processes as well as the complex medium they propagate through. Here, we explore the application of Big Data Seismology to addressing these challenges including recent applications of ML for classifying volcanic signals, and deployments of nodal and DAS sensors on volcanoes to improve our understanding of volcanic processes. Imaging of volcanic interiors is discussed in the Imaging section.

There are several challenges of monitoring volcanic seismicity that are distinct from most other applications of seismology. First, there is a very large corpus of seismic signals that are produced by different physical processes at volcanoes (McNutt, 2005). Second, the onset times of distinct P and S waves from volcanic earthquakes and tremors are often hard to identify, which means that volcano seismologists often rely on arrays for locating events (Nishimura et al., 2021). Third, volcanic edifices are highly heterogeneous 3D structures that also include significant topography, posing further challenges for event location. The era of Big Data Seismology—especially the introduction of new data-dense networks and ML techniques—has potential to help resolve many of these challenges. The use of dense networks on volcanoes has been shown to help reduce the detection threshold, enabling much improved characterization of seismicity, which is a key diagnostic in volcano monitoring (Hansen & Schmandt, 2015). Dense nodal and DAS deployments are also inherently suitable for array processing and have been shown to enable much more precise event location capabilities in volcano environments (Hansen & Schmandt, 2015; Nishimura et al., 2021; S.-M. Wu et al., 2020). In addition, since instruments are often

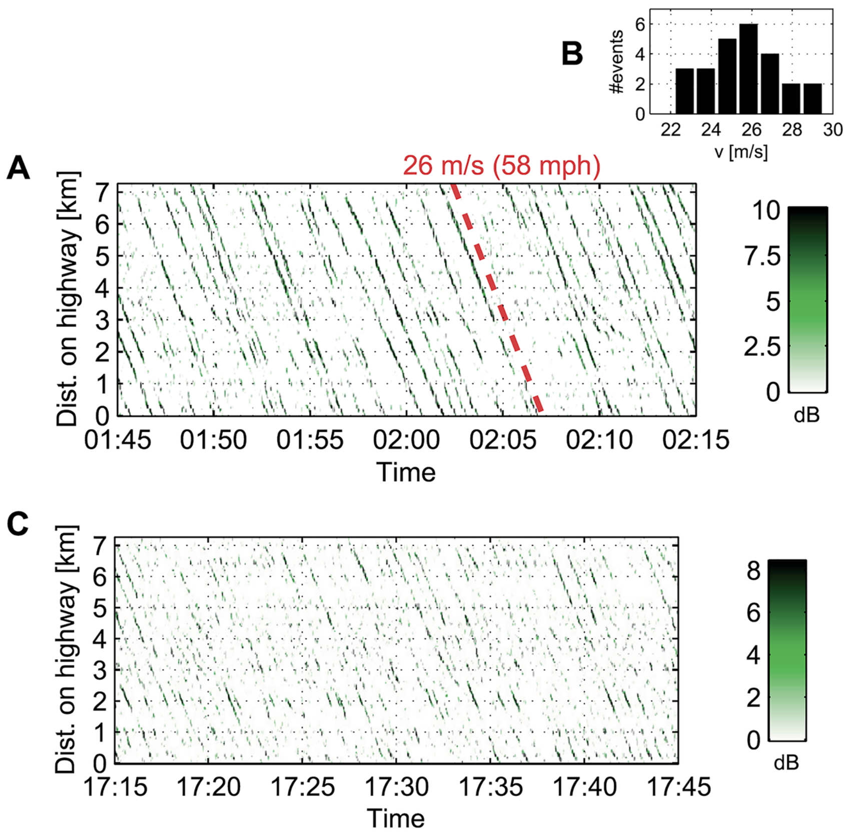

Figure 13. Figure showing seismic power as a function of time at different locations along a 7.3 km section of the I-405 highway in Long Beach, CA. The power is estimated from 97 nodal seismic receivers and shows signals from vehicles moving eastward. Panel (a) shows the measurements at night between 01:45 and 02:15, where individual vehicles can be observed. Panel (c) shows measurements during rush hour in the afternoon, between 17:15 and 17:45, where the data are more cluttered. Panel (b) shows the histogram of velocities of sources in (a). From Riahi and Gerstoft (2015).

damaged or lost in hazardous volcano environments, nodal and DAS deployments have a much lower risk of costly damage. The deployment of nodal arrays on Mount St. Helens (Hansen & Schmandt, 2015), Kiluaea (S.-M. Wu et al., 2020), and Piton de la Fournaise (Brenguier, Kowalski, et al., 2016), and of DAS arrays on Azuma volcano (Nishimura et al., 2021) and Etna (Currenti et al., 2021) speak to the opportunity that dense deployments may ultimately provide in understanding volcanoes. Machine learning will also play an important role in understanding volcano seismicity, with recent work exploring supervised ML (Malfante et al., 2018) as well as unsupervised ML (Duque et al., 2020) to classify volcano seismic signatures.

3.3.4. Forensic Seismology

As described in Section 1, the first true global seismic network was designed and funded for the purpose of monitoring nuclear explosion testing, and regional networks have proven critical for the detection of ever-smaller sources (Koper, 2019). Today, nuclear explosion monitoring, and other applications of forensic seismology that include the analysis of accidental explosions, are beginning to exploit the new types of network and algorithms associated with Big Data Seismology.

With increasing data volumes and the subsequent reduction in event detection thresholds, nuclear explosion monitoring workflows must sift through an increasingly large volume of earthquakes to find events of interest. A critical part of the workflow is the discrimination between explosions and earthquakes (Richards & Kim, 2009). With growing time-histories of data and associated event catalogs, recent work using data from regional networks in Utah and the Netherlands has demonstrated that machine learning can play an increasingly important role in this problem (Linville et al., 2019; Trani et al., 2021). At local scales, data from the Mount Saint Helens nodal array and from a dense network in Wyoming have been used to explore the use of classical *P*/*S* amplitude ratios for explosion discrimination (O'Rourke et al., 2016; R. Wang et al., 2020). Using the dense data deployments, both studies identified strong site effects that varied over relatively short distances. While these empirical discrimination techniques can help in routine data processing, improved models of the physics of explosions are necessary to confidently identify explosions in new settings. For this problem, DAS measurements—which have a large dynamic range—have been used to capture strong ground motions in the near-field from chemical explosions in Nevada, which are invaluable for validating source-physics models (Mellors et al., 2021). Apart from the nuclear explosion monitoring problem, forensic seismology has an important role to play in understanding accidental explosions. Dense data deployments can provide improved fidelity on key properties such as the origin time and explosive yield of accidental events. Two events captured on the dense AlpArray—the Baumgarten gas explosion (Schneider et al., 2018) and Ingolstadt refinery explosion (Fuchs et al., 2019)—have been interpreted to provide accurate origin times (within 1 s). In a similar vein, the accidental explosion of a chemical plant in Xiangshui, China generated seismic observations on a dense sensor network, which were used to estimate the explosive yield (Song et al., 2021).

4. Challenges and Opportunities

This section explores various important challenges and opportunities of Big Data Seismology. These include opportunities for hazard monitoring, and challenges associated with data accessibility and education, which could have implications for diversity and inclusion.

4.1. Streaming, Real-Time Network Data, and Implications for Hazard

Earthquake monitoring is effectively the public face of seismology, as it is the regime in which fundamental science can most directly impact society. Historically, earthquake hazard analysis has been approached from a long-term perspective, for example, through the development of hazard maps to inform building design codes and identify buildings for structural retrofit (Baker, 2013; Gerstenberger et al., 2020). As advancements in computing and telecommunication technology have now opened the door for more rapid responses to damaging earthquakes, Big Data Seismology will play a fundamental role in these near real-time analyses.

Since the late 1800s (Omori, 1894), it has been well understood that large earthquakes do not occur in isolation, but instead trigger aftershocks along nearby fault systems, some of which can pose significant hazards. Thus, in the aftermath of a large earthquake, the hazard is greatly elevated over the long-term background level. Quantifying this hazard in real time and communicating this information to the public is sometimes known as operational earthquake forecasting (Jordan & Jones, 2010; Jordan et al., 2011), and has become increasingly prominent in recent years. While the science of aftershock forecasting is not new (Reasenberg & Jones, 1989, 1994), the challenges with the task are still apparent, as accurate characterizations of large events may take time, and the physics of stress transfer that control aftershock sequence dynamics remain poorly understood. The performance of aftershock forecasts, whether statistical or physics-based, will inevitably depend on the quality of the input data (Field & Working Group on California Earthquake Probabilities, 2018; Field et al., 2020; Mancini et al., 2019; Segou, 2020). Because of this, there is an exigent need for new algorithms and approaches to leverage large volumes of real-time seismic network data to rapidly detect, locate, and characterize events as earthquake sequences evolve (J. I. Walter et al., 2020). Even state-of-the-art, high-resolution catalogs suffer from incompleteness and inconsistencies that can prove problematic in hazard applications (Herrmann & Marzocchi, 2020).

Nowhere is this challenge more apparent as it is in Earthquake Early Warning (EEW), where the mission is to provide seconds-to-minutes warning to nearby population centers that strong shaking is imminent (R. M. Allen & Melgar, 2019; R. M. Allen et al., 2009; Brown et al., 2011; Satriano et al., 2011). EEW systems operate on the principal that the first-arriving P-waves at near-source stations can provide enough information about the earthquake source (or the intensity of its ground motion) that this information can be rapidly broadcast throughout the region before the damaging S-waves and surface waves arrive. While EEW is simple in concept, the implementation of an effective system is surprisingly difficult in practice (Chung et al., 2020; Kohler et al., 2020; McGuire et al., 2021; Meier et al., 2020; Minson et al., 2018; Stubailo et al., 2020; Trugman et al., 2019; Wald, 2020), requiring an upfront investment in network densification, novel algorithms to analyze real-time data streams to rapidly and accurately characterize earthquakes, telecommunication and engineering infrastructure to broadcast alerts to citizens in danger, and buy-in from the community to prepare for and follow alert messages at a moment's notice. Many of the themes that pervade this article are essential to the EEW problem, from the application of dense arrays, to the adaptation of new technologies like smartphone-based sensor networks (Brooks et al., 2021; Inbal et al., 2019; Kong, Inbal, et al., 2019; Kong et al., 2020a, 2020b), to the fusion of distinct types of data such as seismic and geodetic measurements (Murray et al., 2018). EEW and Big Data Seismology are inherently intertwined, and it is likely that the two disciplines will co-evolve in tandem in the years to come.

4.2. Data Accessibility and Sharing

Seismology has a long history and reputation of open science (Agnew, 2002; Ben-Menahem, 1995). With some limited exceptions for security and proprietary reasons, most regional and global data sets are readily accessible to the general public. This accessibility makes it easy to share scientific results and the data sets that underpin them between groups at different institutions. With the emergence of Big Data Seismology, access will likely remain possible in principle but increasingly difficult in practice. Data processing that historically was feasible on a personal laptop or local computing cluster will become out of reach for those without ready access to high performance or cloud computing resources. This transition has already begun to occur; a recent survey (Quinteros et al., 2021) highlighted the disconnect between the present workflow of typical Big Data Seismology users and existing frameworks for data retrieval, storage, and processing.

Maintaining accessibility to seismological data sets in the years to come presents a significant challenge that will need to be addressed for the science to advance. Studies ranging from simple replication efforts to optimize and test algorithm performance all the way to large-scale observational analyses tracking spatiotemporal changes in subsurface properties will hinge upon the ability for multiple users to access and process large volumes of waveform data. Cloud computing workflows provide a compelling solution (MacCarthy et al., 2020), with data stored in perpetuity on cloud servers and users accessing on a case-by-case basis. However, not all users will have the financial means to process routinely on the cloud, and even fewer have the requisite skills to confidently implement such cloud-based analysis workflows without additional education and training efforts. The advent and standardization of HDF5-based data formats (Hess et al., 2018; Krischer et al., 2016) should optimize file compression and storage, but will only dampen and not avert the need for more creative solutions. Communication between data managers and users will become increasingly important, and the need for targeted education and outreach efforts will be crucial.

A related issue concerns "data rescue," or the resuscitation of legacy or otherwise long-forgotten data sets (Diviacco et al., 2015). These data sets are challenging to access and analyze, due a combination of proprietary restrictions, outdated or analog formats, mechanical degradation, and simple loss of institutional knowledge. Nevertheless, legacy data can also contain unique insights into historical or time-dependent geoscience problems, and thus a concerted effort to preserve, maintain, and enable public access to these data sets may well prove worthwhile (Griffin, 2015; Hwang et al., 2020). If these efforts are successful, the large-scale analyses of legacy data will provide an additional frontier in Big Data Seismology. Some of the new techniques highlighted in this article may even prove useful in this regard. For example, K. Wang et al. (2018) adopted ML-based imaging processing algorithms to rescue analog seismograms from the classic Rangely Earthquake Control Experiment, thus providing new observational constraints for one of the seminal studies of induced seismicity.

4.3. Education

Learning the knowledge and skills required to conduct research in Big Data Seismology is challenging as it is necessary to cross traditional academic disciplines. Existing seismology textbooks, which form the basis of most college seismology classes, cover foundational theory and many also introduce the practice of passive or exploration seismology. However, to conduct cutting-edge research in Big Data Seismology also requires expertise with concepts traditionally covered in the disciplines of computer science, engineering, and statistics.

10.1029/2021RG000769

The breadth of skills required poses a challenge for educating graduate students in seismology in traditional Earth Science degree programs. For example, does the need to educate students in computer science and statistics come at the expense of providing a broad Earth Science education? Data science and computational skills clearly cannot replace physics-based knowledge and intuition (Emanuel, 2020), but are becoming increasingly essential nonetheless (Fleming et al., 2021; McGovern & Allen, 2021). A related challenge exists for practicing seismologists, who may have no formal education in computer science or statistics, but will need to load up on concepts from these fields to tackle seismic research with big data sets. How does one get started acquiring the necessary knowledge and skills in these realms?

A variety of excellent resources have been developed that provide stepping-stones for seismologists into the world of Big Data Seismology. To develop and improve skills in computer programming—a fundamental skill for most research in this area—a valuable resource is seismo-live (Krischer, Aiman, et al., 2018). Seismo-live provides a set of Jupyter notebooks that illustrate how to implement various classical seismological tasks in a clean and reproducible way. Since a fundamental tenet of science is that it should be reproducible, and Big Data Seismology algorithms often exploit a large stack of dependencies, providing supporting codes on GitHub is becoming increasingly important. This trend is of course not unique to seismology, and various disciplines are grappling with issues of scientific reproducibility (Gil et al., 2016; Somers, 2018; Toelch & Ostwald, 2018).

Thinking critically about diversity, equity, and inclusion will also be crucial for the future of seismology. It is becoming widely appreciated that despite substantive efforts in recent years, there has been little measurable progress on diversity in the geosciences (Berhe et al., 2021; Bernard & Cooperdock, 2018; Dutt, 2020). The emergence of Big Data Seismology presents both challenges and opportunities for progress. On the one hand, the existence of cutting edge and societally impactful research at the interface between geoscience and data science opens up the potential to reach a broader range of communities. On the other hand, the need for advanced technical training to enact this research could pose significant obstacles to diversity if geoscience or computational education is not uniformly available and equitably distributed. As educational efforts evolve in the age of Big Data Seismology, these dueling considerations will be important to keep in mind.

5. Conclusions

Our fundamental understanding of earthquakes and the solid Earth have been enabled by seismic data. As data volumes have increased with time, seismologists have been able to make new inquiries into these phenomena. This paper is motivated by the emergence of a new type of "big data" inquiry that has only recently become available to seismologists. Rather than simply asking similar questions using more data, we believe that the research highlighted in this paper suggests that this emergent type of inquiry will allow seismologists to seek answers to entirely new questions. As a concrete example, one area in which Big Data Seismology is already asking new types of questions from traditional seismology is through using seismic data to study the hydrosphere, cryosphere, atmosphere, and anthroposphere. Seismic data can also now be used to explore interactions between different components of the Earth System. For example, seismic imaging research is being used to study changes in groundwater in the critical zone at meter-scale spatial resolutions and hourly scale temporal resolution. Many of these new types of investigation are inherently multidisciplinary, requiring seismologists to collaborate with specialists in other disciplines. In addition to opening up new frontiers, Big Data Seismology will play an increasing role in tackling questions that have long eluded traditional seismological investigations, such as improving our understanding of earthquake nucleation and triggering.

The new type of "big data" inquiry that characterizes Big Data Seismology is only possible due the confluence of three drivers: the growth of large data volumes, the development of new data-hungry algorithms, and advances in computing infrastructure. The growth of large data volumes is primarily a result of innovative sensing systems such as nodal arrays and DAS systems. The ongoing and rapid development of ML techniques, which thrive on large volumes of data, is a notable driver from a data exploitation perspective. Finally, the shift in "big data" computing beyond HPC to include a large-scale data-intensive infrastructure and software ecosystems on commodity hardware is a notable driver in computing infrastructure.

The new inquiries made possible by Big Data Seismology represent remarkable opportunities but also present significant challenges. Notable challenges include the continued accessibility and sharing of data as volumes grow, and the challenge of educating seismologists on skills that are not currently part of a traditional geoscience education. Since the same drivers are affecting diverse fields beyond seismology, many of the same challenges are being met today by other fields, but practical and workable solutions have yet to be realized. With these challenges in mind, we hope that this paper can provide a useful resource for seismologists who are interested in exploring "big data" inquiries themselves.

Glossary

Amazon Web Services (AWS): One of the first and most commonly used commercial cloud providers, a subsidiary of Amazon, Inc.

Amazon Simple Storage (S3): Amazon's cloud-based object storage service, in which data are stored as binary objects with minimal metadata, referenced using a globally unique identifier, and accessed using a web page or web-based programming interface.

Ambient Seismic Noise (ASN): Temporally persistent, background signals without an implicit source but that are captured by sensitive seismometers and can be used for tomography and imaging.

Apache Spark: A computational engine from Apache Software designed for large-scale computing on both single-node and distributed systems.

Central Processing Unit (CPU): The core component of a computing device that acts as in main processor that executes instructions. Most modern computing systems contain one or more CPUs.

Commercial Cloud (CC): A set of commercially available computing services, including computing servers, data storage, and managed software.

Common Mid Point (CMP): A widely utilized technique for stacking active source shot gathers in exploration seismology by the horizontal midpoint from source to receiver such that all traces image (approximately) the same reflector.

Community Seismic Network (CSN): A community-driven, academic-coordinated network of low-cost seismometers distributed at public and private locations in Los Angeles used for research, monitoring, and rapid response efforts.

Component: As it applies to seismology, refers to direction of motion a seismometer records. One-component sensors (1C) typically record only vertical motions, while three-component (3C) sensors record both vertical and horizontal motions. Four-component (4C) sensors combine a 3C and 1C sensors of different sensitivities.

Convolutional Neural Network (CNN): A machine learning algorithm commonly applied to image data sets in which a sequence of convolution and activation operations are used to extract useful features directly from the underlying data.

Dask: A Python-specific open source library specializing in parallel computing solutions across distributed hardware systems like computer clusters.

Deep Neural Network (DNN): A flexible class of machine learning models that stack multiple layers, each representing a nonlinear transformation of the data, between the input and output layers.

Direction of Arrival (DOA): The plane wave direction vector from source to receiver that can be inferred through array seismology analysis.

Distributed Acoustic Sensing (DAS): The seismological application of optical fiber cables to measure along-axis strain through optical backscatter.

Docker: A specific "platform as a service" product to deliver software packages within a uniform computing environment known as a container.

Earthquake Early Warning (EEW): An earthquake hazard and monitoring framework in which waveform data from a dense network of seismometers is rapidly and automatically analyzed in order to alert the nearby population of imminent strong shaking immediately following the occurrence of a significant earthquake.

Features: In machine learning, the input variables used to perform a specific task. Often feature variables are hand-selected by human experts or are otherwise mathematically transformed from raw data.

Federation of Digital Seismograph Stations (FDSN): A global organization whose members install, operate, and maintain regional, national, and international seismic networks and who provide public, open access to quality-controlled network data.

Floating Point Operations Per Second (FLOPS): A measure of the processing power of a computational system; a key performance metric for HPC clusters where CPU-intensive simulations are common.

Full Waveform Inversion (FWI): In tomography, the objective of matching synthetic seismograms with observations "wiggle for wiggle," rather than just implicitly through the computation of travel times.

Function as a Service (FaaS): A computing model, commonly offered by commercial cloud providers, in which user-defined computing functions that conform to a pre-defined interface are executed in parallel over a potentially large set of inputs, on computing resources that the user doesn't maintain or specify. Commonly referred to as *serverless* computing.

Generative Adversarial Network (GAN): A machine learning technique in which two neural networks, a generator and a discriminator, are trained competitively and in tandem to accomplish some task. The generator learns to synthesize progressively more realistic sample data, while the discriminator learns to distinguish real from synthetic data samples.

Global Seismographic Network (GSN): A global, digital network of more than 150 seismic stations that includes very broadband, high-dynamic range seismometers and additional instrumentation.

Google Cloud Platform (GCP): A commercial cloud provider operated by Google.

Google File System (GFS): A proprietary filing system from Google that is designed to provide rapid and reliable access to data distributed across multiple large computing clusters.

Graphical Processing Unit (GPU): A specialized form of computing hardware with a highly parallelized structure with numerous cores. GPUs ideal for certain classes of computing or machine learning algorithms.

Hadoop Distributed Filing System (HDFS): A fault-tolerant system for distributed file storage, designed to give high throughput data access while still performing on lower cost computer hardware.

Hierarchical Data Format (e.g., HDF5): A generic data format used in seismology and other fields to efficiently store large data sets in a standardized, hierarchical system.

High-dimensional Data: In machine learning and data science, a generic term describing data sets in which the number of relevant features are much larger than can be easily visualized or in which the large number of features makes calculations or learning difficult. Note that low-dimensional data sets can still be quite large; seismic waveforms and images are often 2D, 3D, or 4D (where nD is refers to the number of dimensions) but can be quite voluminous nonetheless. In statistics, high-dimensional data refers to the case in which the number of features is greater than the number of observations.

High-performance Computing (HPC): Broadly defined, the usage of large-scale and often centralized computing clusters to perform calculations that are not feasible on traditional laptops, desktops, or local computing clusters. Commonly executes CPU-intensive highly coupled computations.

High-throughput Computing (HTC): Similar to HPC, but characterized by the use of *de*centralized computing clusters to perform loosely coupled data-intensive analyses.

Imaging Magma Under St. Helens (iMUSH): A 4 year, multidisciplinary geophysical experiment designed to provide a high resolution portrait of the Mount St. Helens volcano.

Incorporated Research Institutions for Seismology (IRIS): A consortium of more than 100 universities in the United States tasked with coordination of operation of the scientific facilities for waveform data acquisition, management, storage, and distribution.

10.1029/2021RG000769

IRIS Data Management Center (IRIS-DMC): An IRIS facility responsible for archiving and distribution of seismic waveform and metadata to the research community and general public.

IRIS Program for Array Seismic Studies of the Continental Lithosphere (IRIS-PASSCAL): An IRIS facility that provides research-grade instrumentation for local, regional, and global seismic experiments.

International Monitoring System (IMS): A global network of instruments for monitoring nuclear testing that is operated by the Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization and includes a network of seismic stations.

Internet Protocol (IP): The set of rules governing flow of information across an network, often used in reference to the IP address (present internet location) of a given computer.

Interrogator Unit (IU): The central, driving device in DAS technology responsible for emitting laser pulses and recording backscattered light.

Kubernetes: An open-source software container organization system originally designed by Google but now developed and maintained by the Cloud Native Computing Foundation. Commercial cloud companies including Google, Amazon, Microsoft, and Oracle currently offer Kubernetes platforms as part of their services.

Labeled Data: In machine learning, refers to data sets with an explicit connection to one or more associated target variables. Machine learning analyses of labeled data are often referred to as "supervised learning" problems and emphasize the development of predictive models where input data are used to predict target labels.

Large-N: In seismology, typically refers to the temporary deployment of hundreds to thousands of geophones or seismometers over a compact spatial footprint.

LArge-N Seismic Survey in Oklahoma (LASSO): A month-long experiment led by the US Geological Survey to deploy more than 1,800 nodal seismometers in area of northern Oklahoma with intensive hydrocarbon production and associated induced seismicity.

Long Short-Term Memory (LSTM): A particular form of recurrent neural network often used in time series prediction problems. A common architecture is to have multiple units composed of a cell, an input gate, an output gate and a forget gate, with information flowing through and regulated by each unit in sequence.

Machine Learning (ML): An automated form of model building in which an algorithm learns by example from training data how to best to make predictions or otherwise accomplish a specified task.

MapReduce A parallel computing paradigm that uses a "map" procedure to organize calculations on different computing units and a "reduce" procedure to aggregate and summarize results from each unit.

Micro Electro Mechanical System (MEMS): A sensor manufactured using microelectronic fabrication techniques. In seismology, MEMS accelerometers are an increasingly prevalent form of low-cost sensor, sometimes used in temporary or large-scale deployments at multipurpose station locations.

Ocean Bottom Cable (OBC): Optical cables used as receivers in modern marine seismic surveys.

Ocean Bottom Node (OBN): Geophone sensors used as receivers in modern marine seismic surveys.

PyTorch: A deep learning focused tensor computing library developed and maintained by Facebook.

Random Access Memory (RAM): The portion of computer memory that can be actively and rapidly operated on during calculation, as opposed to long term "disk" storage memory.

Recurrent Neural Network (RNN): A machine learning algorithm commonly applied to time series or data with temporal structure in which data flows through recurrent loops in which gating mechanisms allow for limited temporal memory in making predictions.

Remotely Operated Vehicle (ROV): A vehicle without onboard human control operated from a distance, sometimes used in marine seismic surveying.

10.1029/2021RG000769

Representation Learning: Sometimes called feature learning, this refers to a class of machine learning techniques that aim to automatically discover the appropriate mathematical representations or transformations of raw input data to perform a specific task.

Resilient Distributed Data Set (RDD): A fundamental data structure in Apache Spark in which a collection of objects is partitioned and distributed across different nodes of a cluster.

Reverse Time Migration (RTM): An advanced imaging technique that exploits an adjoint wave equation operator to sharply image subsurface reflectors and scattering points.

Scikit-Learn: An open source machine learning library written in Python that emphasizes simplicity, ease-of-use, and flexibility for small-to-mid-scale model testing and development.

Seismic Array: A suite of seismometers at closely spaced sites such that signals are coherent at the different sites and effectively provide a single high-quality observation of the seismic wavefield. Arrays are often designed with the goal of enhancing signals of interest while suppressing sources of noise that may contaminate measurements.

Seismic Network: A generic term for group of affiliated seismic stations, often administered by a single entity. Individual stations typically provide distinct observations of the seismic wavefield, and are not a-priori optimized to detect or suppress a specific signal in particular.

Short-Term Average/Long-Term Average (STA/LTA): A classic event detection algorithm in which the signal amplitude (often filtered) measured over a short time span is compared to a longer term measurement. When this ratio is higher than a specified threshold, an event is declared.

Signal-to-Noise Ratio (SNR): The ratio of waveform amplitudes between the signal of interest and the background noise level, often used as a measure of data quality in seismological research. The SNR of an individual trace often dictates the range of potential analyses than can be usefully performed.

Standard for the Exchange of Earthquake Data (SEED): A uniform, international format designed to ease the distribution and sharing of seismic waveform and metadata, as well as associated processing algorithms, across institutions and research groups.

Structured Query Language (SQL): A widely used framework for managing relational databases using a standardized programming paradigm and set of commands.

TensorFlow: A deep learning focused tensor computing library developed and maintained by Google.

Tony Creek Dual Microseismic Experiment (ToC2ME): An academic-industry partnership and jointly coordinated field experiment in Alberta, Canada that used a mixture of broadband seismometers, borehole arrays, and accelerometers to study hydraulic fracturing induced earthquakes.

Transportable Array (TA): An alias for USArray, highlighting its mobile design as it rolled out across the United States.

U-Net: A specific form of convolutional neural network originally used for image segmentation but later generalized for other problems. The distinguishing feature of the U-Net architecture is its symmetric combination of downsampling (pooling) and upsampling (expansionary) layers.

Unlabeled Data: In machine learning, refers to data sets without an explicit connection to one or more associated target variables. Machine learning analyses of unlabeled data are often referred to as "unsupervised learning" problems and emphasize pattern and group identification (clustering) or dimensionality reduction.

USArray: An National Science Foundation Earthscope led project in which a spatially dense and uniform array of seismometers was progressively rolled out from west to east across the continental United States and then into Alaska.

Very-Broadband (VBB): High-quality seismometers that form the backbone of global monitoring networks and that are capable of precise recordings in a broad frequency band (~0.0028–50.0 Hz).

World Wide Standardized Seismographic Network (WWSSN): The first global-scale seismic network that comprised standardized equipment and facilities for sharing waveform data.

Data Availability Statement

As a review, this paper does not rely on new data collections. The data used to produce Figure 1 is available from the IRIS Data Management Center (https://www.iris.edu/hq/). The latest data volumes for the IRIS DMC are available at the IRIS data statistics page (https://ds.iris.edu/data/distribution/). All other figures shown in the paper summarize published work, and the data availability is described in the associated references.

Acknowledgments

The authors thank Jerry Carter of IRIS Data Services for providing data on the IRIS-DMC archive used in Figure 5. The authors thank Martha Savage, Peggy Hellweg, and an anonymous reviewer for their constructive comments, which helped significantly improve the paper. The authors thank Cathy Chickering for her assistance with editing, Brian Stump for providing an informal review, and John Etgen and Hejun Zhu for their technical comments on early drafts of the manuscript. This work was partially supported by the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development. Los Alamos National Laboratory completed this work under the auspices of the U.S. Department of Energy under contract 89233218CNA000001. This article is approved for unlimited release as LA-UR-22-23779.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). TensorFlow: A system for large-scale machine learning. In *Paper presented at 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (pp. 265–283).

Abercrombie, R. E., & Mori, J. (1996). Occurrence patterns of foreshocks to large earthquakes in the western United States. *Nature*, *381*(6580), 303–307. https://doi.org/10.1038/381303a0

Addair, T., Dodge, D. A., Walter, W., & Ruppert, S. D. (2014). Large-scale seismic signal analysis with Hadoop. *Computers & Geosciences*, *66*, 145–154. https://doi.org/10.1016/j.cageo.2014.01.014

Agnew, D. C. (2002). History of seismology. In *International handbook of earthquake and engineering seismology* (Vol. 81, pp. 3–11). Elsevier. https://doi.org/10.1016/S0074-6142(02)80203-0

Ahrens, J., Hendrickson, B., Long, G., Miller, S., Ross, R., & Williams, D. (2011). Data-intensive science in the US DOE: Case studies and future challenges. *Computing in Science & Engineering*, *13*(6), 14–24. https://doi.org/10.1109/mcse.2011.77

Aiken, C., & Peng, Z. (2014). Dynamic triggering of microearthquakes in three geothermal/volcanic regions of California. *Journal of Geophysical Research: Solid Earth*, *119*(9), 6992–7009. https://doi.org/10.1002/2014JB011218

Ajo-Franklin, J. B., Dou, S., Lindsey, N. J., Monga, I., Tracy, C., Robertson, M., et al. (2019). Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-018-36675-8

Allen, R. M., Brown, H., Hellweg, M., Khainovski, O., Lombard, P., & Neuhauser, D. (2009). Real-time earthquake detection and hazard assessment by ElarmS across California. *Geophysical Research Letters*, *36*(5), L00B08. https://doi.org/10.1029/2008GL036766

Allen, R. M., & Melgar, D. (2019). Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences*, *47*(1), 361–388. https://doi.org/10.1146/annurev-earth-053018-060457

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, *68*(5), 1521–1532. https://doi.org/10.1785/bssa0680051521

Ammon, C. J., Lay, T., & Simpson, D. W. (2010). Great earthquakes and global seismic networks. *Seismological Research Letters*, *81*(6), 965–971. https://doi.org/10.1785/gssrl.81.6.965

Anstey, N. A. (1991). *Vibroseis*. Prentice Hall.

Anthony, R. E., Aster, R., & McGrath, D. (2017). Links between atmosphere, ocean, and cryosphere from two decades of microseism observations on the Antarctic Peninsula. *Journal of Geophysical Research: Earth Surface*, *122*(1), 153–166. https://doi.org/10.1002/2016jf004098

Anthony, R. E., Ringler, A. T., Wilson, D. C., & Wolin, E. (2019). Do low-cost seismographs perform well enough for your network? An overview of laboratory tests and field observations of the OSOP Raspberry Shake 4D. *Seismological Research Letters*, *90*(1), 219–228. https://doi.org/10.1785/0220180251

Apache Software Foundation. (2010). *Hadoop*.

Arnulf, A., Biemiller, J., Lavier, L., Wallace, L. M., Bassett, D., Henrys, S., et al. (2021). Physical conditions and frictional properties in the source region of a slow-slip event. *Nature Geoscience*, *14*(5), 334–340. https://doi.org/10.1038/s41561-021-00741-0

Arnulf, A., Harding, A., Kent, G., & Wilcock, W. (2018). Structure, seismicity, and accretionary processes at the hot spot-influenced Axial Seamount on the Juan de Fuca Ridge. *Journal of Geophysical Research: Solid Earth*, *123*(6), 4618–4646. https://doi.org/10.1029/2017jb015131

Arora, R. (2016). An introduction to big data, high performance computing, high-throughput computing, and Hadoop. In R. Arora (Ed.), *Conquering Big Data with high performance computing* (pp. 1–12). Springer International Publishing. https://doi.org/10.1007/978-3-319-33742-5_1

Arrowsmith, S., Johnson, J. B., Drob, D. P., & Hedlin, M. A. (2010). The seismoacoustic wavefield: A new paradigm in studying geophysical phenomena. *Reviews of Geophysics*, *48*(4), RG4003. https://doi.org/10.1029/2010rg000335

Arrowsmith, S., Negraru, P., & Johnson, G. (2021). Bolide energetics and infrasound propagation: Exploring the 18 December 2018 Bering Sea event to identify limitations of empirical and numerical models. *The Seismic Record*, *1*(3), 164–171. https://doi.org/10.1785/0320210034

Arrowsmith, S., Young, C., & Pankow, K. (2018). Implementation of the Waveform Correlation Event Detection System (WCEDS) method for regional seismic event detection in Utah. *Bulletin of the Seismological Society of America*, *108*(6), 3548–3561. https://doi.org/10.1785/0120180097

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, *34*(6), 26–38. https://doi.org/10.1109/msp.2017.2743240

Asano, Y., Obara, K., & Ito, Y. (2008). Spatiotemporal distribution of very-low frequency earthquakes in Tokachi-oki near the junction of the Kuril and Japan trenches revealed by using array signal processing. *Earth, Planets and Space*, *60*(8), 871–875. https://doi.org/10.1186/BF03352839

Assink, J., Smets, P., Marcillo, O., Weemstra, C., Lalande, J.-M., Waxler, R., & Evers, L. (2019). Advances in infrasonic remote sensing methods. In *Infrasound monitoring for atmospheric studies* (pp. 605–632). Springer.

Aster, R., Beaudoin, B., Hole, J., Fouch, M., Fowler, J., & James, D. (2005). IRIS Seismology Program marks 20 years of discovery. *Eos, Transactions American Geophysical Union*, *86*(17), 171–172. https://doi.org/10.1029/2005eo170002

Aster, R., McNamara, D. E., & Bromirski, P. D. (2008). Multidecadal climate-induced variability in microseisms. *Seismological Research Letters*, *79*(2), 194–202. https://doi.org/10.1785/gssrl.79.2.194

Aster, R., McNamara, D. E., & Bromirski, P. D. (2010). Global trends in extremal microseism intensity. *Geophysical Research Letters*, *37*(14). https://doi.org/10.1029/2010gl043472

Aster, R., & Winberry, J. P. (2017). Glacial seismology. *Reports on Progress in Physics*, *80*(12), 126801. https://doi.org/10.1088/1361-6633/aa8473

Avedik, F., Hirn, A., Renard, V., Nicolich, R., Olivet, J., & Sachpazi, M. (1996). "Single-bubble" marine source offers new perspectives for lithospheric exploration. *Tectonophysics*, *267*(1–4), 57–71. https://doi.org/10.1016/s0040-1951(96)00112-6

Ayala-Garcia, D., Curtis, A., & Branicki, M. (2021). *Seismic interferometry from correlated noise sources*. arXiv preprint arXiv:2105.07250.

Baillard, C., Crawford, W. C., Ballu, V., Hibert, C., & Mangeney, A. (2014). An automatic kurtosis-based P-and S-phase picker designed for local seismic networks. *Bulletin of the Seismological Society of America*, *104*(1), 394–409. https://doi.org/10.1785/0120120347

Baker, J. W. (2013). *An introduction to probabilistic seismic hazard analysis*. *White paper version 2.0* (pp. 1–79).

Bakulin, A., Silvestrov, I., & Pevzner, R. (2020). Surface seismics with DAS: An emerging alternative to modern point-sensor acquisition. *The Leading Edge*, *39*(11), 808–818. https://doi.org/10.1190/tle39110808.1

Beaudoin, G., & Ross, A. A. (2007). Field design and operation of a novel deepwater, wide-azimuth node seismic survey. *The Leading Edge*, *26*(4), 494–503. https://doi.org/10.1190/1.2723213

Bedle, H., Lou, X., & Van der Lee, S. (2021). Continental tectonics inferred from high-resolution imaging of the mantle beneath the United States, through the combination of USArray data types. *Geochemistry, Geophysics, Geosystems*, *22*(10), e2021GC009674. https://doi.org/10.1029/2021gc009674

Bednar, J. B. (2005). A brief history of seismic migration. *Geophysics*, *70*(3), 3MJ–20MJ. https://doi.org/10.1190/1.1926579

Benfield, N., Rambaran, V., Dowlath, J., Sinclair, T., Evans, M., Richardson, J., et al. (2017). Extracting geologic information directly from high-resolution full-waveform inversion velocity models—A case study from offshore Trinidad. *The Leading Edge*, *36*(1), 67–74. https://doi.org/10.1190/tle36010067.1

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. https://doi.org/10.1109/tpami.2013.50

Ben-Menahem, A. (1995). A concise history of mainstream seismology: Origins, legacy, and perspectives. *Bulletin of the Seismological Society of America*, *85*(4), 1202–1225.

Ben-Zion, Y., Vernon, F. L., Ozakin, Y., Zigone, D., Ross, Z. E., Meng, H., et al. (2015). Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. *Geophysical Journal International*, *202*(1), 370–380. https://doi.org/10.1093/gji/ggv142

Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433). https://doi.org/10.1126/science.aau0323

Berhe, A. A., Barnes, R. T., Hastings, M. G., Mattheis, A., Schneider, B., Williams, B. M., & Marín-Spiotta, E. (2021). Scientists from historically excluded groups face a hostile obstacle course. *Nature Geoscience*, *15*, 1–3. https://doi.org/10.1038/s41561-021-00868-0

Bernard, R. E., & Cooperdock, E. H. G. (2018). No progress on diversity in 40 years. *Nature Geoscience*, *11*(5), 292–295. https://doi.org/10.1038/s41561-018-0116-6

Blandford, R. (1974). An automatic event detector at the Tonto Forest seismic observatory. *Geophysics*, *39*(5), 633–643. https://doi.org/10.1190/1.1440453

Bolton, D. C., Shokouhi, P., Rouet-Leduc, B., Hulbert, C., Rivière, J., Marone, C., & Johnson, P. A. (2019). Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismological Research Letters*, *90*(3), 1088–1098. https://doi.org/10.1785/0220180367

Bonnefoy-Claudet, S., Cotton, F., & Bard, P.-Y. (2006). The nature of noise wavefield and its applications for site effects studies: A literature review. *Earth-Science Reviews*, *79*(3–4), 205–227. https://doi.org/10.1016/j.earscirev.2006.07.004

Booth, A. D., Christoffersen, P., Schoonman, C., Clarke, A., Hubbard, B., Law, R., et al. (2020). Distributed acoustic sensing of seismic properties in a borehole drilled on a fast-flowing Greenlandic outlet glacier. *Geophysical Research Letters*, *47*(13), e2020GL088148. https://doi.org/10.1029/2020GL088148

Brenguier, F., Boué, P., Ben-Zion, Y., Vernon, F., Johnson, C., Mordret, A., et al. (2019). Train traffic as a powerful noise source for monitoring active faults with seismic interferometry. *Geophysical Research Letters*, *46*(16), 9529–9536. https://doi.org/10.1029/2019gl083438

Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R. M., & Larose, E. (2008). Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations. *Science*, *321*(5895), 1478–1481. https://doi.org/10.1126/science.1160943

Brenguier, F., Campillo, M., Takeda, T., Aoki, Y., Shapiro, N., Briand, X., et al. (2014). Mapping pressurized volcanic fluids from induced crustal seismic velocity drops. *Science*, *345*(6192), 80–82. https://doi.org/10.1126/science.1254073

Brenguier, F., Kowalski, P., Ackerley, N., Nakata, N., Boué, P., Campillo, M., et al. (2016). Toward 4D noise-based seismic probing of volcanoes: Perspectives from a large-N experiment on Piton de la Fournaise Volcano. *Seismological Research Letters*, *87*(1), 15–25. https://doi.org/10.1785/0220150173

Brenguier, F., Rivet, D., Obermann, A., Nakata, N., Boué, P., Lecocq, T., et al. (2016). 4-D noise-based seismology at volcanoes: Ongoing efforts and perspectives. *Journal of Volcanology and Geothermal Research*, *321*, 182–195. https://doi.org/10.1016/j.jvolgeores.2016.04.036

Brenguier, F., Shapiro, N. M., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O., & Nercessian, A. (2008). Towards forecasting volcanic eruptions using seismic noise. *Nature Geoscience*, *1*(2), 126–130. https://doi.org/10.1038/ngeo104

Brink, M., & Spackman, J. (2004). Solid streamers and single hydrophones. In *SEG technical program expanded abstracts 2004* (pp. 41–44). Society of Exploration Geophysicists.

Brodsky, E. E. (2019). The importance of studying small earthquakes. *Science*, *364*(6442), 736–737. https://doi.org/10.1126/science.aax2490

Bromirski, P. D., & Duennebier, F. K. (2002). The near-coastal microseism spectrum: Spatial and temporal wave climate relationships. *Journal of Geophysical Research*, *107*(B8), ESE–5. https://doi.org/10.1029/2001jb000265

Brooks, B. A., Protti, M., Ericksen, T., Bunn, J., Vega, F., Cochran, E. S., et al. (2021). Robust earthquake early warning at a fraction of the cost: ASTUTI Costa Rica. *AGU Advances*, *2*(3), e2021AV000407. https://doi.org/10.1029/2021AV000407

Brown, H. M., Allen, R. M., Hellweg, M., Khainovski, O., Neuhauser, D., & Souf, A. (2011). Development of the ElarmS methodology for earthquake early warning: Realtime application in California and offline testing in Japan. *Soil Dynamics and Earthquake Engineering*, *31*(2), 188–200. https://doi.org/10.1016/j.soildyn.2010.03.008

Buehler, J. S., & Shearer, P. M. (2017). Uppermost mantle seismic velocity structure beneath USArray. *Journal of Geophysical Research: Solid Earth*, *122*(1), 436–448. https://doi.org/10.1002/2016JB013265

Burdick, S., Van der Hilst, R. D., Vernon, F. L., Martynov, V., Cox, T., Eakins, J., et al. (2014). Model update January 2013: Upper mantle heterogeneity beneath North America from travel-time tomography with global and USArray Transportable Array data. *Seismological Research Letters*, *85*(1), 77–81. https://doi.org/10.1785/0220130098

Burdick, S., Vernon, F. L., Martynov, V., Eakins, J., Cox, T., Tytell, J., et al. (2017). Model update May 2016: Upper-mantle heterogeneity beneath North America from travel-time tomography with global and USArray data. *Seismological Research Letters*, *88*(2A), 319–325. https://doi.org/10.1785/0220160186

Busby, R., Woodward, R., Hafner, K., Vernon, F., & Frassetto, A. (2018). The design and implementation of EarthScope's USArray Transportable Array in the conterminous United States and Southern Canada. *Earth Scope, Technical Report* (pp. 26–35).

Butler, R., Lay, T., Creager, K., Earl, P., Fischer, K., Gaherty, J., et al. (2004). The global seismographic network surpasses its design goal. *Eos, Transactions American Geophysical Union*, *85*(23), 225–229. https://doi.org/10.1029/2004eo230001

Cabrera, L., Poli, P., & Frank, W. B. (2022). Tracking the spatio-temporal evolution of foreshocks preceding the Mw 6.3 2009 L'Aquila Earthquake. *Journal of Geophysical Research: Solid Earth*, *127*(3), e2021JB023888. https://doi.org/10.1029/2021JB023888

Calvert, A., Sawyer, E., Davis, W., & Ludden, J. (1995). Archaean subduction inferred from seismic images of a mantle suture in the Superior Province. *Nature*, *375*(6533), 670–674. https://doi.org/10.1038/375670a0

Calvert, R. (2005). *Insights and methods for 4D reservoir monitoring and characterization*. Society of Exploration Geophysicists and European Association of Geoscientists and Engineers.

Campillo, M., & Paul, A. (2003). Long-range correlations in the diffuse seismic coda. *Science*, *299*(5606), 547–549. https://doi.org/10.1126/science.1078551

Castellanos, J. C., & Clayton, R. W. (2021). The fine-scale structure of Long Beach, California, and its impact on ground motion acceleration. *Journal of Geophysical Research: Solid Earth*, *126*(12), e2021JB022462. https://doi.org/10.1029/2021jb022462

Catchings, R. D., Goldman, M. R., Steidl, J. H., Chan, J. H., Allam, A. A., Criley, C. J., et al. (2020). Nodal seismograph recordings of the 2019 Ridgecrest earthquake sequence. *Seismological Research Letters*, *91*(6), 3622–3633. https://doi.org/10.1785/0220200203

Cattania, C., & Segall, P. (2021). Precursory slow slip and foreshocks on rough faults. *Journal of Geophysical Research: Solid Earth*, *126*(4), e2020JB020430. https://doi.org/10.1029/2020JB020430

Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., et al. (2020). Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical Research Letters*, *47*(16), e2020GL088651. https://doi.org/10.1029/2020gl088651

Chamberlain, C. J., Frank, W. B., Lanza, F., Townend, J., & Warren-Smith, E. (2021). Illuminating the pre-co-and post-seismic phases of the 2016 M7.8 Kaikōura earthquake with 10 years of seismicity. *Journal of Geophysical Research: Solid Earth*, *126*(8), e2021JB022304. https://doi.org/10.1029/2021JB022304

Chapelle, O., Schölkopf, B., & Zien, A. (2010). *Semi-supervised learning*. MIT Press.

Chaput, J., Campillo, M., Aster, R., Roux, P., Kyle, P., Knox, H., & Czoski, P. (2015). Multiple scattering from icequakes at Erebus volcano, Antarctica: Implications for imaging at glaciated volcanoes. *Journal of Geophysical Research: Solid Earth*, *120*(2), 1129–1141. https://doi.org/10.1002/2014jb011278

Chen, J., & Ran, X. (2019). Deep learning with edge computing: A review. *Proceedings of the IEEE*, *107*(8), 1655–1674. https://doi.org/10.1109/jproc.2019.2921977

Chen, X., & Shearer, P. M. (2016). Analysis of foreshock sequences in California and implications for earthquake triggering. *Pure and Applied Geophysics*, *173*(1), 133–152. https://doi.org/10.1007/s00024-015-1103-0

Cheng, F., Chi, B., Lindsey, N. J., Dawe, T. C., & Ajo-Franklin, J. B. (2021). Utilizing distributed acoustic sensing and ocean bottom fiber optic cables for submarine structural characterization. *Scientific Reports*, *11*(1), 1–14. https://doi.org/10.1038/s41598-021-84845-y

Cheng, Y., Ben-Zion, Y., Brenguier, F., Johnson, C. W., Li, Z., Share, P.-E., et al. (2020). An automated method for developing a catalog of small earthquakes using data of a dense seismic array and nearby stations. *Seismological Research Letters*, *91*(5), 2862–2871. https://doi.org/10.1785/0220200134

Chevrot, S., Villaseñor, A., Sylvander, M., Benahmed, S., Beucler, E., Cougoulat, G., et al. (2014). High-resolution imaging of the Pyrenees and Massif Central from the data of the PYROPE and IBERARRAY portable array deployments. *Journal of Geophysical Research: Solid Earth*, *119*(8), 6399–6420. https://doi.org/10.1002/2014jb010953

Christeson, G. L., McIntosh, K. D., & Karson, J. A. (2007). Inconsistent correlation of seismic layer 2a and lava layer thickness in oceanic crust. *Nature*, *445*(7126), 418–421. https://doi.org/10.1038/nature05517

Christie-Blick, N., & Driscoll, N. (1995). Sequence stratigraphy. *Annual Review of Earth and Planetary Sciences*, *23*(1), 451–478. https://doi.org/10.1146/annurev.ea.23.050195.002315

Chung, A. I., Meier, M.-A., Andrews, J., Böse, M., Crowell, B. W., McGuire, J. J., & Smith, D. E. (2020). ShakeAlert earthquake early warning system performance during the 2019 Ridgecrest earthquake sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1904–1923. https://doi.org/10.1785/0120200032

Clayton, R. W., Heaton, T., Kohler, M., Chandy, M., Guy, R., & Bunn, J. (2015). Community seismic network: A dense array to sense earthquake strong motion. *Seismological Research Letters*, *86*(5), 1354–1363. https://doi.org/10.1785/0220150094

Cochran, E. S. (2018). To catch a quake. *Nature Communications*, *9*(1), 1–4. https://doi.org/10.1038/s41467-018-04790-9

Cochran, E. S., Vidale, J. E., & Tanaka, S. (2004). Earth tides can trigger shallow thrust fault earthquakes. *Science*, *306*(5699), 1164–1166. https://doi.org/10.1126/science.1103961

Cochran, E. S., Wickham-Piotrowski, A., Kemna, K. B., Harrington, R. M., Dougherty, S. L., & Peña Castro, A. F. (2020). Minimal clustering of injection-induced earthquakes observed with a large-N seismic array. *Bulletin of the Seismological Society of America*, *110*(5), 2005–2017. https://doi.org/10.1785/0120200101

Contreras-Reyes, E., Grevemeyer, I., Flueh, E. R., & Reichert, C. (2008). Upper lithospheric structure of the subduction zone offshore of southern Arauco peninsula, Chile, at ∼38°S. *Journal of Geophysical Research*, *113*(B7), B07303. https://doi.org/10.1029/2007jb005569

Cook, F. A., Clowes, R. M., Snyder, D. B., van der Velden, A. J., Hall, K. W., Erdmer, P., & Evenchick, C. A. (2004). Precambrian crust beneath the Mesozoic northern Canadian Cordillera discovered by Lithoprobe seismic reflection profiling. *Tectonics*, *23*(2). https://doi.org/10.1029/2002tc001412

Corbi, F., Sandri, L., Bedford, J., Funiciello, F., Brizzi, S., Rosenau, M., & Lallemand, S. (2019). Machine learning can predict the timing and size of analog earthquakes. *Geophysical Research Letters*, *46*(3), 1303–1311. https://doi.org/10.1029/2018GL081251

Currenti, G., Jousset, P., Napoli, R., Krawczyk, C., & Weber, M. (2021). On the comparison of strain measurements from fibre optics with a dense seismometer array at Etna volcano (Italy). *Solid Earth*, *12*(4), 993–1003. https://doi.org/10.5194/se-12-993-2021

Dahlman, O., Mackby, J., Mykkeltveit, S., & Haak, H. (2011). *Detect and deter: Can countries verify the nuclear test ban?* Springer Science & Business Media.

Daley, T. M., Freifeld, B. M., Ajo-Franklin, J., Dou, S., Pevzner, R., Shulakova, V., et al. (2013). Field testing of fiber-optic distributed acoustic sensing (DAS) for subsurface seismic monitoring. *The Leading Edge*, *32*(6), 699–706. https://doi.org/10.1190/tle32060699.1

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107–113. https://doi.org/10.1145/1327452.1327492

Dean, T., Tulett, J., & Barnwell, R. (2018). Nodal land seismic acquisition: The next generation. *First Break*, *36*(1), 47–52. https://doi.org/10.3997/1365-2397.n0061

de Groot-Hedlin, C. D., Hedlin, M. A., & Walker, K. T. (2014). Detection of gravity waves across the USArray: A case study. *Earth and Planetary Science Letters*, *402*, 346–352. https://doi.org/10.1016/j.epsl.2013.06.042

Delorey, A. A., Chao, K., Obara, K., & Johnson, P. A. (2015). Cascading elastic perturbation in Japan due to the 2012 M_w 8.6 Indian Ocean earthquake. *Science Advances*, *1*(9), e1500468. https://doi.org/10.1126/sciadv.1500468

Delorey, A. A., van der Elst, N. J., & Johnson, P. A. (2017). Tidal triggering of earthquakes suggests poroelastic behavior on the San Andreas Fault. *Earth and Planetary Science Letters*, *460*, 164–170. https://doi.org/10.1016/j.epsl.2016.12.014

Dewey, J., & Byerly, P. (1969). The early history of seismometry (to 1900). *Bulletin of the Seismological Society of America*, *59*(1), 183–227.

Díaz, J., Villaseñor, A., Gallart, J., Morales, J., Pazos, D., Córdoba, D., et al. (2008). The IBERARRAY broadband seismic network: A new tool to investigate the deep structure beneath Iberia. *ORFEUS Newletter*, *8*, 1–6.

Diviacco, P., Wardell, N., Forlin, E., Sauli, C., Burca, M., Busato, A., et al. (2015). Data rescue to extend the value of vintage seismic data: The OGS-SNAP experience. *GeoResJ*, *6*, 44–52. https://doi.org/10.1016/j.grj.2015.01.006

Dodge, D. A., Beroza, G. C., & Ellsworth, W. L. (1995). Foreshock sequence of the 1992 Landers, California, earthquake and its implications for earthquake nucleation. *Journal of Geophysical Research*, *100*(B6), 9865–9880. https://doi.org/10.1029/95JB00871

Dodge, D. A., Beroza, G. C., & Ellsworth, W. L. (1996). Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *Journal of Geophysical Research*, *101*(B10), 22371–22392. https://doi.org/10.1029/96JB02269

Dou, S., Lindsey, N., Wagner, A. M., Daley, T. M., Freifeld, B., Robertson, M., et al. (2017). Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-11986-4

Dougherty, S. L., Cochran, E. S., & Harrington, R. M. (2019). The LArge-n Seismic Survey in Oklahoma (LASSO) Experiment. *Seismological Research Letters*, *90*(5), 2051–2057. https://doi.org/10.1785/0220190094

Draelos, T. J., Ballard, S., Young, C. J., & Brogan, R. (2015). A new method for producing automated seismic bulletins: Probabilistic event detection, association, and location. *Bulletin of the Seismological Society of America*, *105*(5), 2453–2467. https://doi.org/10.1785/0120150099

Dragoset, W. H. (1984). A comprehensive method for evaluating the design of airguns and airgun arrays. In *Offshore Technology Conference*. OnePetro.

Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in Geophysics*, *61*, 1–55. https://doi.org/10.1016/bs.agph.2020.08.002

Drew, J., White, R. S., Tilmann, F., & Tarasewicz, J. (2013). Coalescence microseismic mapping. *Geophysical Journal International*, *195*(3), 1773–1785. https://doi.org/10.1093/gji/ggt331

Dumont, V., Tribaldos, V. R., Ajo-Franklin, J., & Wu, K. (2020). Deep learning for surface wave identification in distributed acoustic sensing data. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 1293–1300). IEEE.

Duque, A., González, K., Pérez, N., Benítez, D., Grijalva, F., Lara-Cueva, R., & Ruiz, M. (2020). Exploring the unsupervised classification of seismic events of Cotopaxi volcano. *Journal of Volcanology and Geothermal Research*, *403*, 107009. https://doi.org/10.1016/j.jvolgeores.2020.107009

Durand, V., Bentz, S., Kwiatek, G., Dresen, G., Wollin, C., Heidbach, O., et al. (2020). A two-scale preparation phase preceded an M_w 5.8 earthquake in the Sea of Marmara offshore Istanbul, Turkey. *Seismological Research Letters*, *91*(6), 3139–3147. https://doi.org/10.1785/0220200110

Dutt, K. (2020). Race and racism in the geosciences. *Nature Geoscience*, *13*(1), 2–3. https://doi.org/10.1038/s41561-019-0519-z

Dziewonski, A. M. (2018). The global seismographic network: Progress and promise. In *Observatory seismology: A centennial symposium for the Berkeley seismographic stations* (p. 64). University of California Press.

Dziewonski, A. M., & Woodhouse, J. H. (1983). An experiment in systematic study of global seismicity: Centroid-moment tensor solutions for 201 moderate and large earthquakes of 1981. *Journal of Geophysical Research*, *88*(B4), 3247–3271. https://doi.org/10.1029/jb088ib04p03247

Earle, P. S., Bowden, D. C., & Guy, M. (2011). Twitter earthquake detection: Earthquake monitoring in a social world. *Annals of Geophysics*, *54*(6). https://doi.org/10.4401/ag-5364

Eaton, D. W., Igonin, N., Poulin, A., Weir, R., Zhang, H., Pellegrino, S., & Rodriguez, G. (2018). Induced seismicity characterization during hydraulic-fracture monitoring with a shallow-wellbore geophone array and broadband sensors. *Seismological Research Letters*, *89*(5), 1641–1651. https://doi.org/10.1785/0220180055

Ellsworth, W. L., & Beroza, G. C. (1995). Seismic evidence for an earthquake nucleation phase. *Science*, *268*(5212), 851–855. https://doi.org/10.1126/science.268.5212.851

Ellsworth, W. L., & Bulut, F. (2018). Nucleation of the 1999 Izmit earthquake by a triggered cascade of foreshocks. *Nature Geoscience*, *1*(7), 531–535. https://doi.org/10.1038/s41561-018-0145-1

Emanuel, K. (2020). The relevance of theory for contemporary research in atmospheres, oceans, and climate. *AGU Advances*, *1*(2), e2019AV000129. https://doi.org/10.1029/2019AV000129

Fan, W., Barbour, A. J., Cochran, E. S., & Lin, G. (2021). Characteristics of frequent dynamic triggering of microearthquakes in Southern California. *Journal of Geophysical Research: Solid Earth*, *126*(1), e2020JB020820. https://doi.org/10.1029/2020JB020820

Fan, W., & Shearer, P. M. (2015). Detailed rupture imaging of the 25 April 2015 Nepal earthquake using teleseismic *P* waves. *Geophysical Research Letters*, *42*(14), 5744–5752. https://doi.org/10.1002/2015GL064587

Fan, W., & Shearer, P. M. (2017). Investigation of backprojection uncertainties with *M*6 earthquakes. *Journal of Geophysical Research: Solid Earth*, *122*(10), 7966–7986. https://doi.org/10.1002/2017JB014495

Fang, G., Li, Y. E., Zhao, Y., & Martin, E. R. (2020). Urban near-surface seismic monitoring using distributed acoustic sensing. *Geophysical Research Letters*, *47*(6), e2019GL086115. https://doi.org/10.1029/2019gl086115

Feng, T., Wu, J., Fang, L., Guo, X., Cai, Y., & Wang, W. (2021). Foreshocks of the 2018 ML 4.0 Shimian earthquake in the Anninghe fault and its implications for earthquake nucleation. *Seismological Research Letters*, *92*(3), 1937–1949. https://doi.org/10.1785/0220200332

Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H.-P. (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, *179*(3), 1703–1725. https://doi.org/10.1111/j.1365-246X.2009.04368.x

Field, E. H., Milner, K. R., & Page, M. T. (2020). Generalizing the inversion-based PSHA source model for an interconnected fault system. *Bulletin of the Seismological Society of America*, *111*(1), 371–390. https://doi.org/10.1785/0120200219

Field, E. H., & Working Group on California Earthquake Probabilities. (2018). Improving earthquake rupture forecasts using California as a guide. *Seismological Research Letters*, *89*(6), 2337–2346. https://doi.org/10.1785/0220180151

Finlay, T. S., Worthington, L. L., Schmandt, B., Ranasinghe, N. R., Bilek, S. L., & Aster, R. (2020). Teleseismic scattered-wave imaging using a large-N array in the Albuquerque Basin, New Mexico. *Seismological Research Letters*, *91*(1), 287–303. https://doi.org/10.1785/0220190146

Fleming, S. W., Watson, J. R., Ellenson, A., Cannon, A. J., & Vesselinov, V. C. (2021). Machine learning in Earth and environmental science requires education and research policy reforms. *Nature Geoscience*, *14*(12), 878–880. https://doi.org/10.1038/s41561-021-00865-3

Frank, W. B., & Abercrombie, R. E. (2018). Adapting the matched-filter search to a wide-aperture network: An aftershock sequence and an earthquake swarm in Connecticut. *Bulletin of the Seismological Society of America*, *108*(1), 524–532. https://doi.org/10.1785/0120170190

Freed, A. M. (2005). Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annual Review of Earth and Planetary Sciences*, *33*(1), 335–367. https://doi.org/10.1146/annurev.earth.33.092203.122505

Fuchs, F., Schneider, F. M., Kolínskỳ, P., Serafin, S., & Bokelmann, G. (2019). Rich observations of local and regional infrasound phases made by the AlpArray seismic network after refinery explosion. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-019-49494-2

Gaiser, J., & Strudley, A. (2005). Acquisition and application of multicomponent vector wavefields: Are they practical? *First Break*, *23*(6). https://doi.org/10.3997/1365-2397.2005012

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 1–37. https://doi.org/10.1145/2523813

Gardonio, B., Schubnel, A., Das, S., Lyon-Caen, H., Marsan, D., Bouchon, M., & Kato, A. (2020). The preseismic and postseismic phases of the 700-km deep Mw 7.9 Bonin Islands earthquake, Japan. *Geophysical Research Letters*, *47*(1), e2019GL085589. https://doi.org/10.1029/2019GL085589

Gerstenberger, M. C., Marzocchi, W., Allen, T., Pagani, M., Adams, J., Danciu, L., et al. (2020). Probabilistic seismic hazard analysis at regional and national scales: State of the art and future challenges. *Reviews of Geophysics*, *58*(2), e2019RG000653. https://doi.org/10.1029/2019RG000653

Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles* (pp. 29–43).

Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, *165*(1), 149–166. https://doi.org/10.1111/j.1365-246X.2006.02865.x

Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., et al. (2016). Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science*, *3*(10), 388–415. https://doi.org/10.1002/2015EA000136

Giles, B. F. (1968). Pneumatic acoustic energy source. *Geophysical Prospecting*, *16*(1), 21–53. https://doi.org/10.1111/j.1365-2478.1968.tb01959.x

Gimbert, F., Nanni, U., Roux, P., Helmstetter, A., Garambois, S., Lecointre, A., et al. (2021). A multi-physics experiment with a temporary dense seismic array on the Argentière glacier, French Alps: The RESOLVE project. *Seismological Society of America*, *92*(2A), 1185–1201. https://doi.org/10.1785/0220200280

Glasgow, M. E., Schmandt, B., & Hansen, S. M. (2018). Upper crustal low-frequency seismicity at Mount St. Helens detected with a dense geophone array. *Journal of Volcanology and Geothermal Research*, *358*, 329–341. https://doi.org/10.1016/j.jvolgeores.2018.06.006

Gomberg, J., & Davis, S. (1996). Stress/strain changes and triggered seismicity at the Geysers, California. *Journal of Geophysical Research*, *101*(B1), 733–749. https://doi.org/10.1029/95JB03250

Gomes, H. M., Read, J., Bifet, A., Barddal, J. P., & Gama, J. (2019). Machine learning for streaming data: State of the art, challenges, and opportunities. *ACM SIGKDD Explorations Newsletter*, *21*(2), 6–22. https://doi.org/10.1145/3373464.3373470

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.

Goutorbe, B., & Combier, V. (2010). On the reconstruction of the shape of a seismic streamer. *Geophysics*, *75*(1), H1–H6. https://doi.org/10.1190/1.3267855

Griffin, E. (2015). When are old data new data? *GeoResJ*, *6*, 92–97. https://doi.org/10.1016/j.grj.2015.02.004

Grigoli, F., Cesca, S., Amoroso, O., Emolo, A., Zollo, A., & Dahm, T. (2014). Automated seismic event location by waveform coherence analysis. *Geophysical Journal International*, *196*(3), 1742–1753. https://doi.org/10.1093/gji/ggt477

Hainzl, S. (2016). Rate-dependent incompleteness of earthquake catalogs. *Seismological Research Letters*, *87*(2A), 337–344. https://doi.org/10.1785/0220150211

Hammond, J. O., England, R., Rawlinson, N., Curtis, A., Sigloch, K., Harmon, N., & Baptie, B. (2019). The future of passive seismic acquisition. *Astronomy and Geophysics*, *60*(2), 2–37. https://doi.org/10.1093/astrogeo/atz102

Han, J., Vidale, J. E., Houston, H., Schmidt, D. A., & Creager, K. C. (2018). Deep long-period earthquakes beneath Mount St. Helens: Their relationship to tidal stress, episodic tremor and slip, and regular earthquakes. *Geophysical Research Letters*, *45*(5), 2241–2247. https://doi.org/10.1002/2018GL077063

Han, L., Hole, J. A., Stock, J. M., Fuis, G. S., Kell, A., Driscoll, N. W., et al. (2016). Continental rupture and the creation of new crust in the Salton Trough rift, Southern California and northern Mexico: Results from the Salton Seismic Imaging Project. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7469–7489. https://doi.org/10.1002/2016jb013139

Hansen, S. M., & Schmandt, B. (2015). Automated detection and location of microseismicity at Mount St. Helens with a large-N geophone array. *Geophysical Research Letters*, *42*(18), 7390–7397. https://doi.org/10.1002/2015gl064848

Hardebeck, J. L. (2014). The impact of static stress change, dynamic stress change, and the background stress on aftershock focal mechanisms. *Journal of Geophysical Research: Solid Earth*, *119*(11), 8239–8266. https://doi.org/10.1002/2014JB011533

Hardebeck, J. L., Nazareth, J. J., & Hauksson, E. (1998). The static stress change triggering model: Constraints from two southern California aftershock sequences. *Journal of Geophysical Research*, *103*(B10), 24427–24437. https://doi.org/10.1029/98JB00573

Harris, D. B. (2006). *Subspace detectors: Theory* (Technical Report). Lawrence Livermore National Lab. (LLNL)

Harris, D. B., & Kvaerna, T. (2010). Superresolution with seismic arrays using empirical matched field processing. *Geophysical Journal International*, *182*(3), 1455–1477. https://doi.org/10.1111/j.1365-246x.2010.04684.x

Hayashida, Y., Matsumoto, S., Iio, Y., Sakai, S., & Kato, A. (2020). Non-double-couple microearthquakes in the focal area of the 2000 Western Tottori earthquake (M 7.3) via hyperdense seismic observations. *Geophysical Research Letters*, *47*(4), e2019GL084841. https://doi.org/10.1029/2019GL084841

Hayes, G. P. (2017). The finite, kinematic rupture properties of great-sized earthquakes since 1990. *Earth and Planetary Science Letters*, *468*, 94–100. https://doi.org/10.1016/j.epsl.2017.04.003

Heaton, T. H. (1975). Tidal triggering of earthquakes. *Geophysical Journal International*, *43*(2), 307–326. https://doi.org/10.1111/j.1365-246X.1975.tb00637.x

Hedlin, M. A., Drob, D., Walker, K., & de Groot-Hedlin, C. (2010). A study of acoustic propagation from a large bolide in the atmosphere with a dense seismic network. *Journal of Geophysical Research*, *115*(B11), B11312. https://doi.org/10.1029/2010jb007669

Herrmann, M., & Marzocchi, W. (2020). Inconsistencies and lurking pitfalls in the magnitude–frequency distribution of high-resolution earthquake catalogs. *Seismological Research Letters*, *92*(2A), 909–922. https://doi.org/10.1785/0220200337

Hess, D., Falco, N., Rsdeazevedo, D., & Jacobs, K. (2018). PIC-IRIS/PH5: V4.1.2 [Dataset]. . https://doi.org/10.5281/ZENODO.1284569

Hetényi, G., Molinari, I., Clinton, J., Bokelmann, G., Bondár, I., Crawford, W. C., et al. (2018). The AlpArray seismic network: A large-scale European experiment to image the Alpine orogen. *Surveys in Geophysics*, *39*(5), 1009–1033. https://doi.org/10.1007/s10712-018-9472-4

Hightower, K., Burns, B., & Beda, J. (2017). *Kubernetes: Up and running dive into the future of infrastructure* (1st ed.). O'Reilly Media, Inc.

Hillers, G., Campillo, M., Ben-Zion, Y., & Roux, P. (2014). Seismic fault zone trapped noise. *Journal of Geophysical Research: Solid Earth*, *119*(7), 5786–5799. https://doi.org/10.1002/2014jb011217

Hirata, N. (2009). An outline of the Special Project for Earthquake Disaster Mitigation in the Tokyo Metropolitan Area-Subproject I: Characterization of the plate structure and source faults in and around the Tokyo metropolitan area. *Bulletin of the Earthquake Research Institute, University of Tokyo*, *84*, 41–56.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Holbrook, W. S., Páramo, P., Pearse, S., & Schmitt, R. W. (2003). Thermohaline fine structure in an oceanographic front from seismic reflection profiling. *Science*, *301*(5634), 821–824. https://doi.org/10.1126/science.1085116

Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field. *Science Advances*, *4*(5), eaao2929. https://doi.org/10.1126/sciadv.aao2929

Honda, R., Yukutake, Y., Ito, H., Harada, M., Aketagawa, T., Yoshida, A., et al. (2011). A complex rupture image of the 2011 off the Pacific coast of Tohoku Earthquake revealed by the MeSO-net. *Earth, Planets and Space*, *63*(7), 13–588. https://doi.org/10.5047/eps.2011.05.034

Hone, S., & Zhu, T. (2021). Seismic observations of four thunderstorms using an underground fiber-optic array. *Seismological Society of America*, *92*(4), 2389–2398. https://doi.org/10.1785/0220200264

Hudson, T. S., Baird, A. F., Kendall, J.-M., Kufner, S.-K., Brisbourne, A. M., Smith, A. M., et al. (2021). Distributed Acoustic Sensing (DAS) for natural microseismicity studies: A case study from Antarctica. *Journal of Geophysical Research: Solid Earth*, *126*(7), e2020JB021493. https://doi.org/10.1029/2020jb021493

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, *12*(1), 69–74. https://doi.org/10.1038/s41561-018-0272-8

Humphreys, E. D., & Dueker, K. G. (1994). Western US upper mantle structure. *Journal of Geophysical Research*, *99*(B5), 9615–9634. https://doi.org/10.1029/93jb01724

Hutchison, A. A., & Ghosh, A. (2017). Ambient tectonic tremor in the San Jacinto fault, near the Anza Gap, detected by multiple mini seismic arrays. *Bulletin of the Seismological Society of America*, *107*(5), 1985–1993. https://doi.org/10.1785/0120160385

Hwang, L. J., Ahern, T., Ebinger, C. J., Ellsworth, W. L., Euler, G. G., Okal, E. A., et al. (2020). Rescuing legacy seismic data FAIR'ly. *Seismological Research Letters*, *91*(3), 1339–1340. https://doi.org/10.1785/0220200027

Igonin, N., Verdon, J. P., Kendall, J.-M., & Eaton, D. W. (2021). Large-scale fracture systems are permeable pathways for fault activation during hydraulic fracturing. *Journal of Geophysical Research: Solid Earth*, *126*(3), e2020JB020311. https://doi.org/10.1029/2020JB020311

Igonin, N., Zecevic, M., & Eaton, D. W. (2018). Bilinear magnitude-frequency distributions and characteristic earthquakes during hydraulic fracturing. *Geophysical Research Letters*, *45*(23), 12866–12874. https://doi.org/10.1029/2018GL079746

Inbal, A., Ampuero, J.-P., & Avouac, J.-P. (2017). Locally and remotely triggered aseismic slip on the central San Jacinto Fault near Anza, CA, from joint inversion of seismicity and strainmeter data. *Journal of Geophysical Research: Solid Earth*, *122*(4), 3033–3061. https://doi.org/10.1002/2016JB013499

Inbal, A., Kong, Q., Savran, W., & Allen, R. M. (2019). On the feasibility of using the dense MyShake smartphone array for earthquake location. *Seismological Research Letters*, *90*(3), 1209–1218. https://doi.org/10.1785/0220180349

Isacks, B., Oliver, J., & Sykes, L. R. (1968). Seismology and the new global tectonics. *Journal of Geophysical Research*, *73*(18), 5855–5899. https://doi.org/10.1029/jb073i018p05855

Ishii, M., Shearer, P. M., Houston, H., & Vidale, J. E. (2007). Teleseismic *P* wave imaging of the 26 December 2004 Sumatra-Andaman and 28 March 2005 Sumatra earthquake ruptures using the Hi-net array. *Journal of Geophysical Research*, *112*(B11), B11307. https://doi.org/10.1029/2006JB004700

Jacobs, T. (2018). Shell, bluware team up on seismic with video game tech. *Journal of Petroleum Technology*, *70*(10), 42–43. https://doi.org/10.2118/1018-0042-jpt

Jakkampudi, S., Shen, J., Li, W., Dev, A., Zhu, T., & Martin, E. R. (2020). Footstep detection in urban seismic data with a convolutional neural network. *The Leading Edge*, *39*(9), 654–660. https://doi.org/10.1190/tle39090654.1

Jha, D., Choudhary, K., Tavazza, F., Liao, W.-K., Choudhary, A., Campbell, C., & Agrawal, A. (2019). Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-13297-w

Jia, Z., & Clayton, R. W. (2021). Determination of near surface shear-wave velocities in the central Los Angeles Basin with dense arrays. *Journal of Geophysical Research: Solid Earth*, *126*(5), e2020JB021369. https://doi.org/10.1029/2020jb021369

Jiang, W., Zhang, J., & Bell, L. (2019). 3D seismic geometry quality control and corrections by applying machine learning. *Geophysics*, *84*(6), P87–P96. https://doi.org/10.1190/geo2018-0617.1

Johnson, C. W., Kilb, D., Baltay, A., & Vernon, F. (2020). Peak ground velocity spatial variability revealed by dense seismic array in southern California. *Journal of Geophysical Research: Solid Earth*, *125*(6), e2019JB019157. https://doi.org/10.1029/2019JB019157

Johnson, C. W., Meng, H., Vernon, F., & Ben-Zion, Y. (2019). Characteristics of ground motion generated by wind interaction with trees, structures, and other surface obstacles. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8519–8539. https://doi.org/10.1029/2018jb017151

Johnson, P. A., Ferdowsi, B., Kaproth, B. M., Scuderi, M., Griffa, M., Carmeliet, J., et al. (2013). Acoustic emission and microslip precursors to stick-slip failure in sheared granular material. *Geophysical Research Letters*, *40*(21), 5627–5631. https://doi.org/10.1002/2013GL057848

Jones, L., & Molnar, P. (1976). Frequency of foreshocks. *Nature*, *262*(5570), 677–679. https://doi.org/10.1038/262677a0

Jordan, T. H., Chen, Y.-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., et al. (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Annals of Geophysics*, *54*(4). https://doi.org/10.4401/ag-5350

Jordan, T. H., & Jones, L. M. (2010). Operational earthquake forecasting: Some thoughts on why and how. *Seismological Research Letters*, *81*(4), 571–574. https://doi.org/10.1785/gssrl.81.4.571

Jousset, P., Reinsch, T., Ryberg, T., Blanck, H., Clarke, A., Aghayev, R., et al. (2018). Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features. *Nature Communications*, *9*(1), 1–11. https://doi.org/10.1038/s41467-018-04860-y

Kao, H., & Shan, S.-J. (2004). The source-scanning algorithm: Mapping the distribution of seismic sources in time and space. *Geophysical Journal International*, *157*(2), 589–594. https://doi.org/10.1111/j.1365-246x.2004.02276.x

Karcher, J. C. (1987). The reflection seismograph: Its invention and use in the discovery of oil and gas fields. *The Leading Edge*, *6*(11), 10–19. https://doi.org/10.1190/1.1439341

Karplus, M., & Schmandt, B. (2018). Preface to the focus section on geophone array seismology. *Seismological Research Letters*, *89*(5), 1597–1600. https://doi.org/10.1785/0220180212

Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., & Hirata, N. (2012). Propagation of slow slip leading up to the 2011 M_w 9.0 Tohoku-Oki earthquake. *Science*, *335*(6069), 705–708. https://doi.org/10.1126/science.1215141

Kay, S. M. (1993). *Fundamentals of statistical signal processing: Estimation theory*. Prentice-Hall, Inc.

Kemna, K. B., Peña Castro, A. F., Harrington, R. M., & Cochran, E. S. (2020). Using a large-*n* seismic array to explore the robustness of spectral estimations. *Geophysical Research Letters*, *47*(21), e2020GL089342. https://doi.org/10.1029/2020GL089342

Kerr, R. A. (2013). *Geophysical exploration linking deep Earth and backyard geology*. American Association for the Advancement of Science.

Kilb, D., Gomberg, J., & Bodin, P. (2000). Triggering of earthquake aftershocks by dynamic stresses. *Nature*, *408*(6812), 570–574. https://doi.org/10.1038/35046046

Kim, D., Lekić, V., Ménard, B., Baron, D., & Taghizadeh-Popp, M. (2020). Sequencing seismograms: A panoptic view of scattering in the core-mantle boundary region. *Science*, *368*(6496), 1223–1228. https://doi.org/10.1126/science.aba8972

King, G. C. P., Stein, R. S., & Lin, J. (1994). Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America*, *84*(3), 935–953.

Kingma, D. P., & Welling, M. (2013). *Auto-encoding variational Bayes*. arXiv preprint arXiv:1312.6114.

Kiser, E., & Ishii, M. (2012). The March 11, 2011 Tohoku-Oki earthquake and cascading failure of the plate interface. *Geophysical Research Letters*, *39*(7). https://doi.org/10.1029/2012GL051170

Kiser, E., & Ishii, M. (2017). Back-projection imaging of earthquakes. *Annual Review of Earth and Planetary Sciences*, *45*(1), 271–299. https://doi.org/10.1146/annurev-earth-063016-015801

Kiser, E., Palomeras, I., Levander, A., Zelt, C., Harder, S., Schmandt, B., et al. (2016). Magma reservoirs from the upper crust to the Moho inferred from high-resolution Vp and Vs models beneath Mount St. Helens, Washington State, USA. *Geology*, *44*(6), 411–414. https://doi.org/10.1130/g37591.1

Kohler, M. D., Smith, D. E., Andrews, J., Chung, A. I., Hartog, R., Henson, I., et al. (2020). Earthquake early warning ShakeAlert 2.0: Public rollout. *Seismological Research Letters*, *91*(3), 1763–1775. https://doi.org/10.1785/0220190245

Komatitsch, D., Erlebacher, G., Göddeke, D., & Michéa, D. (2010). High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster. *Journal of Computational Physics*, *229*(20), 7692–7714. https://doi.org/10.1016/j.jcp.2010.06.024

Konečnỳ, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). *Federated learning: Strategies for improving communication efficiency*. arXiv preprint arXiv:1610.05492.

Kong, Q., Allen, R. M., Schreier, L., & Kwon, Y.-W. (2016). MyShake: A smartphone seismic network for earthquake early warning and beyond. *Science Advances*, *2*(2). https://doi.org/10.1126/sciadv.1501055

Kong, Q., Inbal, A., Allen, R. M., Lv, Q., & Puder, A. (2019). Machine learning aspects of the MyShake global smartphone seismic network. *Seismological Research Letters*, *90*(2A), 546–552. https://doi.org/10.1785/0220180309

Kong, Q., Martin-Short, R., & Allen, R. M. (2020a). Toward global earthquake early warning with the MyShake smartphone seismic network, Part 1: Simulation platform and detection algorithm. *Seismological Research Letters*, *91*(4), 2206–2217. https://doi.org/10.1785/0220190177

Kong, Q., Martin-Short, R., & Allen, R. M. (2020b). Toward global earthquake early warning with the MyShake smartphone seismic network, Part 2: Understanding MyShake performance around the world. *Seismological Research Letters*, *91*(4), 2218–2233. https://doi.org/10.1785/0220190178

Kong, Q., Trugman, D., Ross, Z., Bianco, M., Meade, B., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, *90*(1), 3–14. https://doi.org/10.1785/0220180259

Koper, K. D. (2019). The importance of regional seismic networks in monitoring Nuclear Test-Ban Treaties. *Seismological Research Letters*, *91*(2A), 573–580. https://doi.org/10.1785/0220190160

Koper, K. D., Hutko, A. R., Lay, T., Ammon, C. J., & Kanamori, H. (2011). Frequency-dependent rupture process of the 2011 Mw 9.0 Tohoku Earthquake: Comparison of short-period P wave backprojection images and broadband seismic rupture models. *Earth, Planets and Space*, *63*(7), 16–602. https://doi.org/10.5047/eps.2011.05.026

Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K. S., & Igel, C. (2017). Big universe, big data: Machine learning and image analysis for astronomy. *IEEE Intelligent Systems*, *32*(2), 16–22. https://doi.org/10.1109/MIS.2017.40

Krischer, L., Aiman, Y. A., Bartholomaus, T., Donner, S., van Driel, M., Duru, K., et al. (2018). Seismo-Live: An educational online library of Jupyter notebooks for seismology. *Seismological Research Letters*, *89*(6), 2413–2419. https://doi.org/10.1785/0220180167

Krischer, L., Fichtner, A., Boehm, C., & Igel, H. (2018). Automated large-scale full seismic waveform inversion for North America and the North Atlantic. *Journal of Geophysical Research: Solid Earth*, *123*(7), 5902–5928. https://doi.org/10.1029/2017JB015289

Krischer, L., Smith, J., Lei, W., Lefebvre, M., Ruan, Y., de Andrade, E. S., et al. (2016). An adaptable seismic data format. *Geophysical Journal International*, *207*(2), 1003–1011. https://doi.org/10.1093/gji/ggw319

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*.

Lampert, C. H., Nickisch, H., & Harmeling, S. (2009). Learning to detect unseen object classes by between-class attribute transfer. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 951–958). IEEE.

Langet, N., Maggi, A., Michelini, A., & Brenguier, F. (2014). Continuous kurtosis-based migration for seismic event detection and location, with application to Piton de la Fournaise Volcano, La Reunion. *Bulletin of the Seismological Society of America*, *104*(1), 229–246. https://doi.org/10.1785/0120130107

Larose, E., Carrière, S., Voisin, C., Bottelin, P., Baillet, L., Guéguen, P., et al. (2015). Environmental seismology: What can we learn on Earth surface processes with ambient noise? *Journal of Applied Geophysics*, *116*, 62–74. https://doi.org/10.1016/j.jappgeo.2015.02.001

Larson, K. M. (2019). Unanticipated uses of the global positioning system. *Annual Review of Earth and Planetary Sciences*, *47*(1), 19–40. https://doi.org/10.1146/annurev-earth-053018-060203

Lawyer, L. C., Bates, C. C., Rice, R. B., & Bates, C. C. (2001). *Geophysics in the affairs of mankind: A personalized history of exploration geophysics* (2nd ed., Vol. 10). Society of Exploration Geophysicists.

Lay, T., Kanamori, H., Ammon, C. J., Nettles, M., Ward, S. N., Aster, R., et al. (2005). The great Sumatra-Andaman earthquake of 26 December 2004. *Science*, *308*(5725), 1127–1133. https://doi.org/10.1126/science.1112250

Lecocq, T., Hicks, S. P., Van Noten, K., Van Wijk, K., Koelemeijer, P., De Plaen, R. S., et al. (2020). Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures. *Science*, *369*(6509), 1338–1343. https://doi.org/10.1126/science.abd2438

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lellouch, A., Lindsey, N. J., Ellsworth, W. L., & Biondi, B. L. (2020). Comparison between distributed acoustic sensing and geophones: Downhole microseismic monitoring of the FORGE geothermal experiment. *Seismological Research Letters*, *91*(6), 3256–3268. https://doi.org/10.1785/0220200149

Lellouch, A., Schultz, R., Lindsey, N. J., Biondi, B. L., & Ellsworth, W. L. (2021). Low-magnitude seismicity with a downhole distributed acoustic sensing array—Examples from the FORGE geothermal experiment. *Journal of Geophysical Research: Solid Earth*, *126*(1), e2020JB020462. https://doi.org/10.1029/2020JB020462

L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with Big Data: Challenges and approaches. *IEEE Access*, *5*, 7776–7797. https://doi.org/10.1109/access.2017.2696365

Li, J., Shillington, D. J., Saffer, D. M., Bécel, A., Nedimović, M. R., Kuehn, H., et al. (2018). Connections between subducted sediment, pore-fluid pressure, and earthquake behavior along the Alaska megathrust. *Geology*, *46*(4), 299–302. https://doi.org/10.1130/g39557.1

Li, Z., Peng, Z., Hollis, D., Zhu, L., & McClellan, J. (2018). High-resolution seismic event detection using local similarity for large-N arrays. *Scientific Reports*, *8*(1), 1646. https://doi.org/10.1038/s41598-018-19728-w

Li, Z., Shen, Z., Yang, Y., Williams, E., Wang, X., & Zhan, Z. (2021). Rapid response to the 2019 Ridgecrest earthquake with distributed acoustic sensing. *AGU Advances*, *2*(2), e2021AV000395. https://doi.org/10.1029/2021AV000395

Li, Z., & Zhan, Z. (2018). Pushing the limit of earthquake detection with distributed acoustic sensing and template matching: A case study at the Brady geothermal field. *Geophysical Journal International*, *215*(3), 1583–1593. https://doi.org/10.1093/gji/ggy359

Lin, F.-C., Li, D., Clayton, R. W., & Hollis, D. (2013). High-resolution 3D shallow crustal structure in Long Beach, California: Application of ambient noise tomography on a dense seismic array. *Geophysics*, *78*(4), Q45–Q56. https://doi.org/10.1190/geo2012-0453.1

Lin, F.-C., Ritzwoller, M. H., Yang, Y., Moschetti, M. P., & Fouch, M. J. (2011). Complex and variable crustal and uppermost mantle seismic anisotropy in the western United States. *Nature Geoscience*, *4*(1), 55–61. https://doi.org/10.1038/ngeo1036

Lindseth, R. O. (1968). *Digital processing of geophysical data: A review*. Society of Exploration Geophysicists.

Lindsey, N. J., Dawe, T. C., & Ajo-Franklin, J. B. (2019). Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing. *Science*, *366*(6469), 1103–1107. https://doi.org/10.1126/science.aay5881

Lindsey, N. J., & Martin, E. R. (2021). Fiber-optic seismology. *Annual Review of Earth and Planetary Sciences*, *49*(1), 309–336. https://doi.org/10.1146/annurev-earth-072420-065213

Lindsey, N. J., Martin, E. R., Dreger, D. S., Freifeld, B., Cole, S., James, S. R., et al. (2017). Fiber-optic network observations of earthquake wavefields. *Geophysical Research Letters*, *44*(23), 11792–11799. https://doi.org/10.1002/2017GL075722

Lindsey, N. J., Rademacher, H., & Ajo-Franklin, J. B. (2020). On the broadband instrument response of fiber-optic DAS arrays. *Journal of Geophysical Research: Solid Earth*, *125*(2), e2019JB018145. https://doi.org/10.1029/2019jb018145

Lindsey, N. J., Yuan, S., Lellouch, A., Gualtieri, L., Lecocq, T., & Biondi, B. (2020). City-scale dark fiber DAS measurements of infrastructure use during the COVID-19 pandemic. *Geophysical Research Letters*, *47*(16), e2020GL089931. https://doi.org/10.1029/2020gl089931

Linville, L., Pankow, K., & Draelos, T. (2019). Deep learning models augment analyst decisions for event discrimination. *Geophysical Research Letters*, *46*(7), 3643–3651. https://doi.org/10.1029/2018gl081119

Lior, I., Mercerat, E., Rivet, D., Sladen, A., & Ampuero, J.-P. (2021). *Imaging an underwater basin and its resonance modes using optical fiber distributed acoustic sensing*.

Long, M. D., Levander, A., & Shearer, P. M. (2014). *An introduction to the special issue of Earth and Planetary Science Letters on USArray science*. Elsevier.

Lubbers, N., Bolton, D. C., Mohd-Yusof, J., Marone, C., Barros, K., & Johnson, P. A. (2018). Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophysical Research Letters*, *45*(24), 13269–13276. https://doi.org/10.1029/2018GL079712

Lumley, D. (2001). Time-lapse seismic reservoir monitoring. *Geophysics*, *66*(1), 50–53. https://doi.org/10.1190/1.1444921

Lumley, D. (2010). 4D seismic monitoring of CO_2 sequestration. *The Leading Edge*, *29*(2), 150–155. https://doi.org/10.1190/1.3304817

Lumley, D., Landrø, M., Vasconcelos, I., Eisner, L., Hatchell, P., Li, Y., et al. (2015). Advances in time-lapse geophysics—Introduction. *Society of Exploration Geophysicists*, *80*(2), WAi–WAii. https://doi.org/10.1190/2015-0120-spseintro.1

Lythgoe, K. H., Ong Su Qing, M., & Wei, S. (2020). Large-scale crustal structure beneath Singapore using receiver functions from a dense urban nodal array. *Geophysical Research Letters*, *47*(7), e2020GL087233. https://doi.org/10.1029/2020gl087233

Ma, X., & Lowry, A. R. (2017). USArray imaging of continental crust in the conterminous United States. *Tectonics*, *36*(12), 2882–2902. https://doi.org/10.1002/2017TC004540

MacCarthy, J., Marcillo, O., & Trabant, C. (2020). Seismology in the cloud: A new streaming workflow. *Seismological Research Letters*, *91*(3), 1804–1812. https://doi.org/10.1785/0220190357

Mackenzie, G., Thybo, H., & Maguire, P. (2005). Crustal velocity structure across the Main Ethiopian Rift: Results from two-dimensional wide-angle seismic modelling. *Geophysical Journal International*, *162*(3), 994–1006. https://doi.org/10.1111/j.1365-246x.2005.02710.x

Magana-Zook, S., Gaylord, J. M., Knapp, D. R., Dodge, D. A., & Ruppert, S. D. (2016). Large-scale seismic waveform quality metric calculation using Hadoop. *Computers & Geosciences*, *94*, 18–30. https://doi.org/10.1016/j.cageo.2016.05.012

Makovsky, Y., Klemperer, S. L., Ratschbacher, L., Brown, L. D., Li, M., Zhao, W., & Meng, F. (1996). INDEPTH wide-angle reflection observation of P-wave-to-S-wave conversion from crustal bright spots in Tibet. *Science*, *274*(5293), 1690–1691. https://doi.org/10.1126/science.274.5293.1690

Malfante, M., Dalla Mura, M., Mars, J. I., Métaxian, J.-P., Macedo, O., & Inza, A. (2018). Automatic classification of volcano seismic signatures. *Journal of Geophysical Research: Solid Earth*, *123*(12), 10–645. https://doi.org/10.1029/2018jb015470

Malin, P. E., Bohnhoff, M., Blümle, F., Dresen, G., Martínez-Garzón, P., Nurlu, M., et al. (2018). Microearthquakes preceding a M4.2 earthquake offshore Istanbul. *Scientific Reports*, *8*(1), 16176. https://doi.org/10.1038/s41598-018-34563-9

Mancini, S., Segou, M., Werner, M. J., & Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016–2017 central Italy earthquake cascade. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8626–8643. https://doi.org/10.1029/2019JB017874

Manin, M., Boucquaert, F., Regnaudin, J., Regnault, A., & Thevenot, A. (1988). Recent developments in source and streamer positioning. *First Break*, *6*(6). https://doi.org/10.3997/1365-2397.1988009

Mao, S., Campillo, M., van der Hilst, R. D., Brenguier, F., Stehly, L., & Hillers, G. (2019). High temporal resolution monitoring of small variations in crustal strain by dense seismic arrays. *Geophysical Research Letters*, *46*(1), 128–137. https://doi.org/10.1029/2018gl079944

Marcillo, O. E., & MacCarthy, J. (2020). Mapping seismic tonal noise in the contiguous United States. *Seismological Research Letters*, *91*(3), 1707–1716. https://doi.org/10.1785/0220190355

Marone, F., Gung, Y., & Romanowicz, B. (2007). Three-dimensional radial anisotropic structure of the North American upper mantle from inversion of surface waveform data. *Geophysical Journal International*, *171*(1), 206–222. https://doi.org/10.1111/j.1365-246x.2007.03465.x

Marra, G., Clivati, C., Luckett, R., Tampellini, A., Kronjäger, J., Wright, L., et al. (2018). Ultrastable laser interferometry for earthquake detection with terrestrial and submarine cables. *Science*, *361*(6401), 486–490. https://doi.org/10.1126/science.aat4458

Martin, J., Özbek, A., Combee, L., Lunde, N., Bittleston, S., & Kragh, E. (2000). Acquisition of marine point receiver seismic data with a towed streamer. In *SEG Technical Program Expanded Abstracts 2000* (pp. 37–40). Society of Exploration Geophysicists.

Matoza, R. S., Shearer, P. M., Lin, G., Wolfe, C. J., & Okubo, P. G. (2013). Systematic relocation of seismicity on Hawaii Island from 1992 to 2009 using waveform cross correlation and cluster analysis. *Journal of Geophysical Research: Solid Earth*, *118*(5), 2275–2288. https://doi.org/10.1002/jgrb.50189

Mayne, W. H. (1962). Common reflection point horizontal data stacking techniques. *Geophysics*, *27*(6), 927–938. https://doi.org/10.1190/1.1439118

McBrearty, I. W., Delorey, A. A., & Johnson, P. A. (2019). Pairwise association of seismic arrivals with convolutional neural networks. *Seismological Research Letters*, *90*(2A), 503–509. https://doi.org/10.1785/0220180326

McCormack, M. D. (1991). Neural computing in geophysics. *The Leading Edge*, *10*(1), 11–15. https://doi.org/10.1190/1.1436771

McGovern, A., & Allen, J. (2021). Training the next generation of physical data scientists. *EoS*, *102*. https://doi.org/10.1029/2021EO210536

McGuire, J. J., Minson, S. E., Murray, J. R., & Brooks, B. A. (2021). Commentary: The role of geodetic algorithms for earthquake early warning in Cascadia. *Geophysical Research Letters*, *48*(6), e2020GL092324. https://doi.org/10.1029/2020GL092324

McLaskey, G. C. (2019). Earthquake initiation from laboratory observations and implications for foreshocks. *Journal of Geophysical Research: Solid Earth*, *124*(12), 12882–12904. https://doi.org/10.1029/2019JB018363

McLaskey, G. C., & Lockner, D. A. (2014). Preslip and cascade processes initiating laboratory stick slip. *Journal of Geophysical Research: Solid Earth*, *119*(8), 6323–6336. https://doi.org/10.1002/2014JB011220

McNutt, S. R. (2005). Volcanic seismology. *Annual Review of Earth and Planetary Sciences*, *32*(1), 461–491. https://doi.org/10.1146/annurev.earth.33.092203.122459

Meier, M.-A., Kodera, Y., Böse, M., Chung, A., Hoshiba, M., Cochran, E., et al. (2020). How often can earthquake early warning systems alert sites with high-intensity ground motion? *Journal of Geophysical Research: Solid Earth*, *125*(2), e2019JB017718. https://doi.org/10.1029/2019JB017718

Melbourne, T. I., Szeliga, W. M., Marcelo Santillan, V., & Scrivner, C. W. (2021). Global navigational satellite system seismic monitoring. *Bulletin of the Seismological Society of America*, *111*(3), 1248–1262. https://doi.org/10.1785/0120200356

Mellors, R. J., Abbott, R., Steedman, D., Podrasky, D., & Pitarka, A. (2021). Modeling subsurface explosions recorded on a distributed fiber optic sensor. *Journal of Geophysical Research: Solid Earth*, *126*(12), e2021JB022690. https://doi.org/10.1029/2021jb022690

Meng, H., & Ben-Zion, Y. (2018a). Characteristics of airplanes and helicopters recorded by a dense seismic array near Anza California. *Journal of Geophysical Research: Solid Earth*, *123*(6), 4783–4797. https://doi.org/10.1029/2017jb015240

Meng, H., & Ben-Zion, Y. (2018b). Detection of small earthquakes with dense array data: Example from the San Jacinto fault zone, southern California. *Geophysical Journal International*, *212*(1), 442–457. https://doi.org/10.1093/gji/ggx404

Meng, H., & Fan, W. (2021). Immediate foreshocks indicating cascading rupture developments for 527 M 0.9 to 5.4 Ridgecrest earthquakes. *Geophysical Research Letters*, *48*(19), e2021GL095704. https://doi.org/10.1029/2021GL095704

Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, *57*, 115–129. https://doi.org/10.1016/j.inffus.2019.12.001

Merkel, D. (2014). Docker: Lightweight Linux containers for consistent development and deployment. *Linux Journal*, *2014*(239), 2.

Mesimeri, M., Zhang, H., & Pankow, K. L. (2020). Backprojection imaging of the 2020 Mw 5.5 Magna, Utah, earthquake using a local dense strong-motion network. *Seismological Research Letters*, *92*(2A), 640–646. https://doi.org/10.1785/0220200326

Mestayer, J., Cox, B., Wills, P., Kiyashchenko, D., Lopez, J., Costello, M., et al. (2011). Field trials of distributed acoustic sensing for geophysical monitoring. In *SEG Technical Program Expanded Abstracts 2011* (pp. 4253–4257). Society of Exploration Geophysicists.

Mignan, A. (2014). The debate on the prognostic value of earthquake foreshocks: A meta-analysis. *Scientific Reports*, *4*(1), 4099. https://doi.org/10.1038/srep04099

Mignan, A., & Woessner, J. (2012). *Estimating the magnitude of completeness for earthquake catalogs*. Community Online Resource for Statistical Seismicity Analysis. https://doi.org/10.5078/corssa-00180805

Minson, S. E., Brooks, B. A., Glennie, C. L., Murray, J. R., Langbein, J. O., Owen, S. E., et al. (2015). Crowdsourced earthquake early warning. *Science Advances*, *1*(3), e1500036. https://doi.org/10.1126/sciadv.1500036

Minson, S. E., Meier, M.-A., Baltay, A. S., Hanks, T. C., & Cochran, E. S. (2018). The limits of earthquake early warning: Timeliness of ground motion estimates. *Science Advances*, *4*(3), eaaq0504. https://doi.org/10.1126/sciadv.aaq0504

Miyazawa, M., Brodsky, E. E., & Guo, H. (2021). Dynamic earthquake triggering in southern California in high resolution: Intensity, time decay, and regional variability. *AGU Advances*, *2*(2), e2020AV000309. https://doi.org/10.1029/2020AV000309

Monk, D. J. (2020). *Survey design and seismic acquisition for land, marine, and in-between in light of new technology and techniques*. Society of Exploration Geophysicists.

Mordret, A., Roux, P., Boué, P., & Ben-Zion, Y. (2019). Shallow three-dimensional structure of the San Jacinto fault zone revealed from ambient noise imaging with a dense seismic array. *Geophysical Journal International*, *216*(2), 896–905. https://doi.org/10.1093/gji/ggy464

Moschetti, M., Ritzwoller, M., Lin, F., & Yang, Y. (2010). Seismic evidence for widespread western-US deep-crustal deformation caused by extension. *Nature*, *464*(7290), 885–889. https://doi.org/10.1038/nature08951

Mosher, D., & Simpkin, P. (1999). Environmental marine geoscience 1. Status and trends of marine high-resolution seismic reflection profiling: Data acquisition. *Geoscience Canada*, *26*(4), 174–188.

Mougenot, D., & Thorburn, N. (2004). MEMS-based 3D accelerometers for land seismic acquisition: Is it time? *The Leading Edge*, *23*(3), 246–250. https://doi.org/10.1190/1.1690897

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 3952. https://doi.org/10.1038/s41467-020-17591-w

Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019). STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI. *IEEE Access*, *7*, 179464–179476. https://doi.org/10.1109/ACCESS.2019.2947848

Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C. (2019). CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific Reports*, *9*(1), 10267. https://doi.org/10.1038/s41598-019-45748-1

Moutote, L., Marsan, D., Lengliné, O., & Duputel, Z. (2021). Rare occurrences of non-cascading foreshock activity in Southern California. *Geophysical Research Letters*, *48*(7), e2020GL091757. https://doi.org/10.1029/2020GL091757

Murray, J. R., Crowell, B. W., Grapenthin, R., Hodgkinson, K., Langbein, J. O., Melbourne, T., et al. (2018). Development of a geodetic component for the U.S. West Coast earthquake early warning system. *Seismological Research Letters*, *89*(6), 2322–2336. https://doi.org/10.1785/0220180162

Mustafazade, A., Pandit, M., Zhao, C., Sobreviela, G., Du, Z., Steinmann, P., et al. (2020). A vibrating beam MEMS accelerometer for gravity and seismic measurements. *Scientific Reports*, *10*(1), 1–8. https://doi.org/10.1038/s41598-020-67046-x

Mutter, J. C. (1985). Seaward dipping reflectors and the continent-ocean boundary at passive continental margins. *Tectonophysics*, *114*(1–4), 117–131. https://doi.org/10.1016/0040-1951(85)90009-5

Nagaoka, Y., Nishida, K., Aoki, Y., Takeo, M., & Ohminato, T. (2012). Seismic imaging of magma chamber beneath an active volcano. *Earth and Planetary Science Letters*, *333*, 1–8. https://doi.org/10.1016/j.epsl.2012.03.034

Nakata, N., Chang, J. P., Lawrence, J. F., & Boué, P. (2015). Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry. *Journal of Geophysical Research: Solid Earth*, *120*(2), 1159–1173. https://doi.org/10.1002/2015jb011870

Nakata, N., Gualtieri, L., & Fichtner, A. (Eds.), (2019). *Seismic ambient noise*. Cambridge University Press.

Nanni, U., Gimbert, F., Roux, P., & Lecointre, A. (2021). Observing the subglacial hydrology network and its dynamics with a dense seismic array. *Proceedings of the National Academy of Sciences*, *118*(28). https://doi.org/10.1073/pnas.2023757118

Nayak, A., Ajo-Franklin, J., & the Imperial Valley Dark Fiber Team. (2021). Distributed acoustic sensing using dark fiber for array detection of regional earthquakes. *Seismological Research Letters*, *92*(4), 2441–2452. https://doi.org/10.1785/0220200416

Nelson, K., Arnow, J., McBride, J., Willemin, J., Huang, J., Zheng, L., et al. (1985). New COCORP profiling in the southeastern United States. Part I: Late Paleozoic suture and Mesozoic rift basin. *Geology*, *13*(10), 714–718. https://doi.org/10.1130/0091-7613(1985)13<714:ncpits>2.0.co;2

Ng, A. (2011). *Machine learning*. MOOC, Coursera. Retrieved from https://www.coursera.org/learn/machine-learning

Nishimura, T., Emoto, K., Nakahara, H., Miura, S., Yamamoto, M., Sugimura, S., et al. (2021). Source location of volcanic earthquakes and subsurface characterization using fiber-optic cable and distributed acoustic sensing system. *Scientific Reports*, *11*(1), 1–12. https://doi.org/10.1038/s41598-021-85621-8

Noda, H., Nakatani, M., & Hori, T. (2013). Large nucleation before large earthquakes is sometimes skipped due to cascade-up—Implications from a rate and state simulation of faults with hierarchical asperities. *Journal of Geophysical Research: Solid Earth*, *118*(6), 2924–2952. https://doi.org/10.1002/jgrb.50211

Nolet, G., Hello, Y., van der Lee, S., Bonnieux, S., Ruiz, M. C., Pazmino, N. A., et al. (2019). Imaging the Galápagos mantle plume with an unconventional application of floating seismometers. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-018-36835-w

Nur, A. (1989). Four-dimensional seismology and (true) direct detection of hydrocarbons: The petrophysical basis. *The Leading Edge*, *8*(9), 30–36. https://doi.org/10.1190/1.1439661

Oakley, D. O., Forsythe, B., Gu, X., Nyblade, A. A., & Brantley, S. L. (2021). Seismic ambient noise analyses reveal changing temperature and water signals to 10s of meters depth in the critical zone. *Journal of Geophysical Research: Earth Surface*, *126*(2), e2020JF005823. https://doi.org/10.1029/2020jf005823

Ojo, A. O., Kao, H., Visser, R., & Goerzen, C. (2021). Spatiotemporal changes in seismic velocity associated with hydraulic fracturing-induced earthquakes near Fox Creek, Alberta, Canada. *Journal of Petroleum Science and Engineering*, *208*, 109390. https://doi.org/10.1016/j.petrol.2021.109390

Oldham, R. D. (1906). The constitution of the interior of the Earth, as revealed by earthquakes. *Quarterly Journal of the Geological Society*, *62*(1–4), 456–475. https://doi.org/10.1144/gsl.jgs.1906.062.01-04.21

Oliver, J., & Murphy, L. (1971). WWNSS: Seismology's global network of observing stations. *Science*, *174*(4006), 254–261. https://doi.org/10.1126/science.174.4006.254

Omori, F. (1894). Investigation of aftershocks. *Reports of the Imperial Earthquake Investigation Committee*, *2*, 103–139.

Oncken, O., Asch, G., Haberland, C., Metchie, J., Sobolev, S., Stiller, M., et al. (2003). Seismic imaging of a convergent continental margin and plateau in the central Andes (Andean Continental Research Project 1996 (ANCORP'96)). *Journal of Geophysical Research*, *108*(B7). https://doi.org/10.1029/2002jb001771

O'Rourke, C. T., Baker, G. E., & Sheehan, A. F. (2016). Using *P*/*S* amplitude ratios for seismic discrimination at local distances. *Bulletin of the Seismological Society of America*, *106*(5), 2320–2331. https://doi.org/10.1785/0120160035

Paffrath, M., Friederich, W., Schmid, S. M., Handy, M. R., & the AlpArray and AlpArray-SWATH D Working Group. (2021). Imaging structure and geometry of slabs in the greater Alpine area–A P-wave travel-time tomography using AlpArray Seismic Network data. *Solid Earth*, *12*(11), 2671–2702. https://doi.org/10.5194/se-12-2671-2021

Paitz, P., Edme, P., Gräff, D., Walter, F., Doetsch, J., Chalari, A., et al. (2020). Empirical investigations of the instrument response for distributed acoustic sensing (DAS) across 17 octaves. *Bulletin of the Seismological Society of America*, *111*(1), 1–10. https://doi.org/10.1785/0120200185

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/tkde.2009.191

Park, J., Butler, R., Anderson, K., Berger, J., Davis, P., Benz, H., et al. (2005). Performance review of the global seismographic network for the Sumatra-Andaman megathrust earthquake. *Seismological Research Letters*, *76*(3), 331–343. https://doi.org/10.1785/gssrl.76.3.331

Park, S., & Ishii, M. (2015). Inversion for rupture properties based upon 3-D directivity effect and application to deep earthquakes in the Sea of Okhotsk region. *Geophysical Journal International*, *203*(2), 1011–1025. https://doi.org/10.1093/gji/ggv352

Parsekian, A., Singha, K., Minsley, B. J., Holbrook, W. S., & Slater, L. (2015). Multiscale geophysical imaging of the critical zone. *Reviews of Geophysics*, *53*(1), 1–26. https://doi.org/10.1002/2014rg000465

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*.

Peardon, L., & Cameron, N. (1986). Acoustic and mechanical design considerations for digital streamers. In *SEG Technical Program Expanded Abstracts 1986* (pp. 291–294). Society of Exploration Geophysicists.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peña Castro, A. F., Dougherty, S. L., Harrington, R. M., & Cochran, E. S. (2019). Delayed dynamic triggering of disposal-induced earthquakes observed by a dense array in northern Oklahoma. *Journal of Geophysical Research: Solid Earth*, *124*(4), 3766–3781. https://doi.org/10.1029/2018JB017150

Perol, T., Gharbi, M., & Denolle, M. A. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), e1700578. https://doi.org/10.1126/sciadv.1700578

Pesciullesi, G., Schwaller, P., Laino, T., & Reymond, J.-L. (2020). Transfer learning enables the molecular transformer to predict regio-and stereoselective reactions on carbohydrates. *Nature Communications*, *11*(1), 1–8. https://doi.org/10.1038/s41467-020-18671-7

Pham, T.-S., Tkalčić, H., Sambridge, M., & Kennett, B. L. (2018). Earth's correlation wavefield: Late coda correlation. *Geophysical Research Letters*, *45*(7), 3035–3042. https://doi.org/10.1002/2018gl077244

Pinzon-Rincon, L., Lavoué, F., Mordret, A., Boué, P., Brenguier, F., Dales, P., et al. (2021). Humming trains in seismology: An opportune source for probing the shallow crust. *Seismological Society of America*, *92*(2A), 623–635. https://doi.org/10.1785/0220200248

Podolskiy, E. A., & Walter, F. (2016). Cryoseismology. *Reviews of Geophysics*, *54*(4), 708–758. https://doi.org/10.1002/2016rg000526

Porritt, R. W., Becker, T. W., Boschi, L., & Auer, L. (2021). Multiscale, radially anisotropic shear wave imaging of the mantle underneath the contiguous United States through joint inversion of usarray and global data sets. *Geophysical Journal International*, *226*(3), 1730–1746. https://doi.org/10.1093/gji/ggab185

Poulin, A., Weir, R., Eaton, D., Igonin, N., Chen, Y., Lines, L., & Lawton, D. (2019). Focal-time analysis: A new method for stratigraphic depth control of microseismicity and induced seismic events. *Geophysics*, *84*(6), KS173–KS182. https://doi.org/10.1190/geo2019-0046.1

Prieux, V., Brossier, R., Operto, S., & Virieux, J. (2013). Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: Imaging compressional wave speed, density and attenuation. *Geophysical Journal International*, *194*(3), 1640–1664. https://doi.org/10.1093/gji/ggt177

Quinteros, J., Carter, J. A., Schaeffer, J., Trabant, C., & Pedersen, H. A. (2021). Exploring approaches for large data in seismology: User and data repository perspectives. *Seismological Society of America*, *92*(3), 1531–1540. https://doi.org/10.1785/0220200390

Raghu, M., & Schmidt, E. (2020). *A survey of deep learning for scientific discovery*. arXiv preprint arXiv:2003.11755.

Raina, R., Madhavan, A., & Ng, A. Y. (2009). Large-scale deep unsupervised learning using graphics processors. In *Proceedings of the 26th annual international conference on machine learning* (pp. 873–880).

Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine*, *34*(6), 96–108. https://doi.org/10.1109/msp.2017.2738401

Reasenberg, P. A., & Jones, L. M. (1989). Earthquake hazard after a mainshock in California. *Science*, *243*(4895), 1173–1176. https://doi.org/10.1126/science.243.4895.1173

Reasenberg, P. A., & Jones, L. M. (1994). Earthquake aftershocks: Update. *Science*, *265*(5176), 1251–1252. https://doi.org/10.1126/science.265.5176.1251

Riahi, N., & Gerstoft, P. (2015). The seismic traffic footprint: Tracking trains, aircraft, and cars seismically. *Geophysical Research Letters*, *42*(8), 2674–2681. https://doi.org/10.1002/2015gl063558

Riahi, N., & Gerstoft, P. (2017). Using graph clustering to locate sources within a dense sensor array. *Signal Processing*, *132*, 110–120. https://doi.org/10.1016/j.sigpro.2016.10.001

Richards, P. G., & Kim, W.-Y. (2009). Monitoring for nuclear explosions. *Scientific American*, *300*(3), 70–77. https://doi.org/10.1038/scientificamerican0309-70

Richards-Dinger, K., Stein, R. S., & Toda, S. (2010). Decay of aftershock density with distance does not indicate triggering by dynamic stress. *Nature*, *467*(7315), 583–586. https://doi.org/10.1038/nature09402

Ringler, A., Anthony, R., Aster, R., Ammon, C., Arrowsmith, S., Benz, H., et al. (2022). Achievements and prospects of global broadband seismographic networks after 30 years of geophysical observations. *Reviews of Geophysics*, Submitted.

Ritsema, J., & Lekić, V. (2020). Heterogeneity of seismic wave velocity in Earth's mantle. *Annual Review of Earth and Planetary Sciences*, *48*(1), 377–401. https://doi.org/10.1146/annurev-earth-082119-065909

Rocklin, M. (2015). Dask: Parallel computation with blocked algorithms and task scheduling. In *Proceedings of the 14th Python in Science Conference* (Vol. 130, p. 136).

Rodríguez Tribaldos, V., & Ajo-Franklin, J. B. (2021). Aquifer monitoring using ambient seismic noise recorded with distributed acoustic sensing (DAS) deployed on dark fiber. *Journal of Geophysical Research: Solid Earth*, *126*(4), e2020JB021004. https://doi.org/10.1029/2020jb021004

Romanowicz, B., Cara, M., Fel, J. F., & Rouland, D. (1984). GEOSCOPE: A French initiative in long-period three-component global seismic networks. *Eos, Transactions American Geophysical Union*, *65*(42), 753–753. https://doi.org/10.1029/eo065i042p00753-01

Romanowicz, B., & Giardini, D. (2001). *The future of permanent seismic networks*. American Association for the Advancement of Science.

Romanowicz, B., McGill, P., Neuhauser, D., & Dolenc, D. (2009). Acquiring real time data from the broadband ocean bottom seismic observatory at Monterey Bay (MOBB). *Seismological Research Letters*, *80*(2), 197–202. https://doi.org/10.1785/gssrl.80.2.197

Rong, K., Yoon, C. E., Bergen, K. J., Elezabi, H., Bailis, P., Levis, P., & Beroza, G. C. (2018). *Locality-sensitive hashing for earthquake detection: A case study of scaling data-driven science*. arXiv preprint arXiv:1803.09835.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234–241).

Ross, Z. E., Cochran, E. S., Trugman, D. T., & Smith, J. D. (2020). 3D fault architecture controls the dynamism of earthquake swarms. *Science*, *368*(6497), 1357–1361. https://doi.org/10.1126/science.abb0779

Ross, Z. E., Idini, B., Jia, Z., Stephenson, O. L., Zhong, M., Wang, X., et al. (2019). Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence. *Science*, *366*(6463), 346–351. https://doi.org/10.1126/science.aaz0109

Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *108*(5A), 2894–2901. https://doi.org/10.1785/0120180080

Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in Southern California. *Science*, *364*(6442), 767–771. https://doi.org/10.1126/science.aaw6888

Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2019). PhaseLink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, *124*(1), 856–869. https://doi.org/10.1029/2018JB016674

Rost, S., & Thomas, C. (2002). Array seismology: Methods and applications. *Reviews of Geophysics*, *40*(3), 2-1–2-27. https://doi.org/10.1029/2000rg000100

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*(18), 9276–9282. https://doi.org/10.1002/2017GL074677

Roux, P., Lecointre, A., Moreau, L., Campillo, M., Ben-Zion, Y., & Vernon, F. (2015). High-resolution imaging of the San Jacinto fault zone with a dense seismic array and local seismic noise. *Journal of the Acoustical Society of America*, *138*(3), 1764–1764. https://doi.org/10.1121/1.4933576

Sager, K., Tsai, V. C., Sheng, Y., Brenguier, F., Boué, P., Mordret, A., & Igel, H. (2021). Modelling P waves in seismic noise correlations: Advancing fault monitoring using train traffic sources. *Geophysical Journal International*, *228*(3), 1556–1567. https://doi.org/10.1093/gji/ggab389

Sakai, S., & Hirata, N. (2009). Distribution of the metropolitan seismic observation network. *Bulletin of the Earthquake Research Institute, University of Tokyo*, *84*, 57–69.

Sánchez-Reyes, H., Essing, D., Beaucé, E., & Poli, P. (2021). The imbricated foreshock and aftershock Activities of the Balsorano (Italy) Mw 4.4 normal fault earthquake and implications for earthquake initiation. *Seismological Research Letters*, *92*(3), 1926–1936. https://doi.org/10.1785/0220200253

Sanderson, R. W., Matoza, R. S., Fee, D., Haney, M. M., & Lyons, J. J. (2020). Remote detection and location of explosive volcanism in Alaska with the EarthScope Transportable Array. *Journal of Geophysical Research: Solid Earth*, *125*(4), e2019JB018347. https://doi.org/10.1029/2019jb018347

Satriano, C., Wu, Y.-M., Zollo, A., & Kanamori, H. (2011). Earthquake early warning: Concepts, methods and physical grounds. *Soil Dynamics and Earthquake Engineering*, *31*(2), 106–118. https://doi.org/10.1016/j.soildyn.2010.07.007

Schaeffer, A., & Lebedev, S. (2014). Imaging the North American continent using waveform inversion of global and USArray data. *Earth and Planetary Science Letters*, *402*, 26–41. https://doi.org/10.1016/j.epsl.2014.05.014

Schneider, F. M., Fuchs, F., Kolínskỳ, P., Caffagni, E., Serafin, S., Dorninger, M., et al. (2018). Seismo-acoustic signals of the Baumgarten (Austria) gas explosion detected by the AlpArray seismic network. *Earth and Planetary Science Letters*, *502*, 104–114. https://doi.org/10.1016/j.epsl.2018.08.034

Schoenball, M., & Ellsworth, W. L. (2017). Waveform-relocated earthquake catalog for Oklahoma and southern Kansas illuminates the regional fault network. *Seismological Research Letters*, *88*(5), 1252–1258. https://doi.org/10.1785/0220170083

Segou, M. (2020). The physics of earthquake forecasting. *Seismological Research Letters*, *91*(4), 1936–1939. https://doi.org/10.1785/0220200127

Seif, S., Zechar, J. D., Mignan, A., Nandan, S., & Wiemer, S. (2019). Foreshocks and their potential deviation from general seismicity. *Bulletin of the Seismological Society of America*, *109*(1), 1–18. https://doi.org/10.1785/0120170188

Sergeant, A., Chmiel, M., Lindner, F., Walter, F., Roux, P., Chaput, J., et al. (2020). On the Green's function emergence from interferometry of seismic wave fields generated in high-melt glaciers: Implications for passive imaging and monitoring. *The Cryosphere*, *14*(3), 1139–1171. https://doi.org/10.5194/tc-14-1139-2020

Seydoux, L., Balestriero, R., Poli, P., de Hoop, M., Campillo, M., & Baraniuk, R. (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-17841-x

Shapiro, N. M., Campillo, M., Stehly, L., & Ritzwoller, M. H. (2005). High-resolution surface-wave tomography from ambient seismic noise. *Science*, *307*(5715), 1615–1618. https://doi.org/10.1126/science.1108339

Share, P.-E., Ben-Zion, Y., Ross, Z. E., Qiu, H., & Vernon, F. L. (2017). Internal structure of the San Jacinto fault zone at Blackburn Saddle from seismic data of a linear array. *Geophysical Journal International*, *210*(2), 819–832. https://doi.org/10.1093/gji/ggx191

Shearer, P., Hauksson, E., & Lin, G. (2005). Southern California hypocenter relocation with waveform cross-correlation, Part 2: Results using source-specific station terms and cluster analysis. *Bulletin of the Seismological Society of America*, *95*(3), 904–915. https://doi.org/10.1785/0120040168

Shearer, P. M. (1990). Seismic imaging of upper-mantle structure with new evidence for a 520-km discontinuity. *Nature*, *344*(6262), 121–126. https://doi.org/10.1038/344121a0

Shearer, P. M. (2012). Space-time clustering of seismicity in California and the distance dependence of earthquake triggering. *Journal of Geophysical Research*, *117*(B10). https://doi.org/10.1029/2012JB009471

Shearer, P. M., & Buehler, J. (2019). Imaging upper-mantle structure under USArray using long-period reflection seismology. *Journal of Geophysical Research: Solid Earth*, *124*(9), 9638–9652. https://doi.org/10.1029/2019JB017326

Shearer, P. M., & Masters, T. G. (1992). Global mapping of topography on the 660-km discontinuity. *Nature*, *355*(6363), 791–796. https://doi.org/10.1038/355791a0

Shelly, D. R. (2020). A high-resolution seismic catalog for the initial 2019 Ridgecrest earthquake sequence: Foreshocks, aftershocks, and faulting complexity. *Seismological Research Letters*, *91*(4), 1971–1978. https://doi.org/10.1785/0220190309

Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, *446*(7133), 305–307. https://doi.org/10.1038/nature05666

Shelly, D. R., Ellsworth, W. L., & Hill, D. P. (2016). Fluid-faulting evolution in high definition: Connecting fault structure and frequency-magnitude variations during the 2014 Long Valley Caldera, California, earthquake swarm. *Journal of Geophysical Research: Solid Earth*, *121*(3), 1776–1795. https://doi.org/10.1002/2015JB012719

Shen, H., & Shen, Y. (2021). Array-based convolutional neural networks for automatic detection and 4D localization of earthquakes in Hawai'i. *Seismological Research Letters*, *92*(5), 2961–2971. https://doi.org/10.1785/0220200419

Shen, W., & Ritzwoller, M. H. (2016). Crustal and uppermost mantle structure beneath the United States. *Journal of Geophysical Research: Solid Earth*, *121*(6), 4306–4342. https://doi.org/10.1002/2016jb012887

Sheng, Y., Kong, Q., & Beroza, G. C. (2021). Network analysis of earthquake ground motion spatial correlation: A case study with the San Jacinto seismic nodal array. *Geophysical Journal International*, *225*(3), 1704–1713. https://doi.org/10.1093/gji/ggab058

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298. https://doi.org/10.1109/tmi.2016.2528162

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48. https://doi.org/10.1186/s40537-019-0197-0

Shragge, J., Bourget, J., Lumley, D., Giraud, J., Wilson, T., Iqbal, A., et al. (2019). The Western Australia Modeling project—Part 1: Geomodel building. *Interpretation*, *7*(4), T773–T791. https://doi.org/10.1190/int-2018-0217.1

Shragge, J., Lumley, D., Bourget, J., Potter, T., Miyoshi, T., Witten, B., et al. (2019). The Western Australia Modeling project—Part 2: Seismic validation. *Interpretation*, *7*(4), T793–T807. https://doi.org/10.1190/int-2018-0218.1

Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The Hadoop Distributed File System. In *Paper presented at 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)* (pp. 1–10). IEEE. https://doi.org/10.1109/MSST.2010.5496972

Singh, S., Collier, J., Harding, A., Kent, G., & Orcutt, J. (1999). Seismic evidence for a hydrothermal layer above the solid roof of the axial magma chamber at the southern East Pacific Rise. *Geology*, *27*(3), 219–222. https://doi.org/10.1130/0091-7613(1999)027<0219:sefahl>2.3.co;2

Sirgue, L., Barkved, O., Dellinger, J., Etgen, J., Albertin, U., & Kommedal, J. (2010). Thematic set: Full waveform inversion: The next leap forward in imaging at Valhall. *First Break*, *28*(4). https://doi.org/10.3997/1365-2397.2010012

Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2016). An efficient repeating signal detector to investigate earthquake swarms. *Journal of Geophysical Research: Solid Earth*, *121*(8), 5880–5897. https://doi.org/10.1002/2016jb012981

Skoumal, R. J., Brudzinski, M. R., Currie, B. S., & Levy, J. (2014). Optimizing multi-station earthquake template matching through re-examination of the Youngstown, Ohio, sequence. *Earth and Planetary Science Letters*, *405*, 274–280. https://doi.org/10.1016/j.epsl.2014.08.033

Skoumal, R. J., Kaven, J. O., & Walter, J. I. (2019). Characterizing seismogenic fault structures in Oklahoma using a relocated template-matched catalog. *Seismological Research Letters*, *90*(4). https://doi.org/10.1785/0220190045

Snover, D., Johnson, C. W., Bianco, M. J., & Gerstoft, P. (2021). Deep clustering to identify sources of urban seismic noise in Long Beach, California. *Seismological Society of America*, *92*(2A), 1011–1022. https://doi.org/10.1785/0220200164

Solano, C. P., & Plessix, R.-É. (2019). Velocity-model building with enhanced shallow resolution using elastic waveform inversion—An example from onshore Oman. *Geophysics*, *84*(6), R977–R988. https://doi.org/10.1190/geo2018-0736.1

Somers, J. (2018). The scientific paper is obsolete. *The Atlantic*, *4*.

Song, Y., Zhao, L.-F., Xie, X.-B., Ma, X., Du, G., Tian, X., & Yao, Z.-X. (2021). Seismological observations on the 2019 March 21 accidental explosion at Xiangshui chemical plant in Jiangsu, China. *Geophysical Journal International*, *228*(1), 538–550. https://doi.org/10.1093/gji/ggab356

Spica, Z. J., Nishida, K., Akuhara, T., Pétrélis, F., Shinohara, M., & Yamada, T. (2020). Marine sediment characterized by ocean-bottom fiber-optic seismology. *Geophysical Research Letters*, *47*(16), e2020GL088360. https://doi.org/10.1029/2020GL088360

Spica, Z. J., Perton, M., Martin, E. R., Beroza, G. C., & Biondi, B. (2020). Urban seismic site characterization by fiber-optic seismology. *Journal of Geophysical Research: Solid Earth*, *125*(3), e2019JB018656. https://doi.org/10.1029/2019JB018656

Stein, R. S. (1999). The role of stress transfer in earthquake occurrence. *Nature*, *402*(6762), 605–609. https://doi.org/10.1038/45144

Stubailo, I., Alvarez, M., Biasi, G., Bhadha, R., & Hauksson, E. (2020). Latency of waveform data delivery from the southern California seismic network during the 2019 Ridgecrest earthquake sequence and its effect on ShakeAlert. *Seismological Research Letters*, *92*(1), 170–186. https://doi.org/10.1785/0220200211

Suárez, G., van Eck, T., Giardini, D., Ahern, T., Butler, R., & Tsuboi, S. (2008). The International Federation of Digital Seismograph Networks (FDSN): An integrated system of seismological observatories. *IEEE Systems Journal*, *2*(3), 431–438. https://doi.org/10.1109/jsyst.2008.2003294

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 843–852).

Sweet, J. R., Anderson, K. R., Bilek, S., Brudzinski, M., Chen, X., DeShon, H., et al. (2018). A community experiment to record the full seismic wavefield in Oklahoma. *Seismological Research Letters*, *89*(5), 1923–1930. https://doi.org/10.1785/0220180079

Sweet, J. R., Creager, K. C., & Houston, H. (2014). A family of repeating low-frequency earthquakes at the downdip edge of tremor and slip. *Geochemistry, Geophysics, Geosystems*, *15*(9), 3713–3721. https://doi.org/10.1002/2014GC005449

Sweet, J. R., Creager, K. C., Houston, H., & Chestler, S. R. (2019). Variations in Cascadia low-frequency earthquake behavior with downdip distance. *Geochemistry, Geophysics, Geosystems*, *20*(2), 1202–1217. https://doi.org/10.1029/2018GC007998

Sykes, L. R. (1967). Mechanism of earthquakes and nature of faulting on the mid-oceanic ridges. *Journal of Geophysical Research*, *72*(8), 2131–2153. https://doi.org/10.1029/jz072i008p02131

Takagi, R., Uchida, N., Nakayama, T., Azuma, R., Ishigami, A., Okada, T., et al. (2019). Estimation of the orientations of the S-net cabled ocean-bottom sensors. *Seismological Research Letters*, *90*(6), 2175–2187. https://doi.org/10.1785/0220190093

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In *International Conference on Artificial Neural Networks* (pp. 270–279).

Tan, Y. J., Waldhauser, F., Ellsworth, W. L., Zhang, M., Zhu, W., Michele, M., et al. (2021). Machine-learning-based high-resolution earthquake catalog reveals how complex fault structures were activated during the 2016–2017 central Italy sequence. *The Seismic Record*, *1*(1), 11–19. https://doi.org/10.1785/0320210001

Tao, K., Grand, S. P., & Niu, F. (2018). Seismic structure of the upper mantle beneath eastern Asia from full waveform seismic tomography. *Geochemistry, Geophysics, Geosystems*, *19*(8), 2732–2763. https://doi.org/10.1029/2018gc007460

Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2009). Adjoint tomography of the southern California crust. *Science*, *325*(5943), 988–992. https://doi.org/10.1126/science.1175298

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, *49*(8), 1259–1266. https://doi.org/10.1190/1.1441754

Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. In *Scattering and attenuations of seismic waves, Part 1* (pp. 365–399). Springer.

Thomas, T. W., Vidale, J. E., Houston, H., Creager, K. C., Sweet, J. R., & Ghosh, A. (2013). Evidence for tidal triggering of high-amplitude rapid tremor reversals and tremor streaks in northern Cascadia. *Geophysical Research Letters*, *40*(16), 4254–4259. https://doi.org/10.1002/grl.50832

Thompson, B. D., Young, R. P., & Lockner, D. A. (2009). Premonitory acoustic emissions and stick-slip in natural and smooth-faulted Westerly granite. *Journal of Geophysical Research*, *114*(B2), B02205. https://doi.org/10.1029/2008JB005753

Tkalčić, H. (2017). *The Earth's inner core: Revealed by observational seismology*.

Toelch, U., & Ostwald, D. (2018). Digital open science—Teaching digital tools for reproducible and transparent research. *PLoS Biology*, *16*(7), e2006022. https://doi.org/10.1371/journal.pbio.2006022

Trani, L., Pagani, G. A., Zanetti, J. P. P., Chapeland, C., & Evers, L. (2021). DeepQuake—An application of CNN for seismo-acoustic event classification in The Netherlands. *Computers & Geosciences*, *159*, 104980. https://doi.org/10.1016/j.cageo.2021.104980

Tromp, J. (2020). Seismic wavefield imaging of Earth's interior across scales. *Nature Reviews Earth & Environment*, *1*(1), 40–53. https://doi.org/10.1038/s43017-019-0003-8

Trugman, D. T., Chu, S. X., & Tsai, V. C. (2021). Earthquake source complexity controls the frequency dependence of near-source radiation patterns. *Geophysical Research Letters*, *48*(17), e2021GL095022. https://doi.org/10.1029/2021GL095022

Trugman, D. T., McBrearty, I. W., Bolton, D. C., Guyer, R. A., Marone, C., & Johnson, P. A. (2020). The spatiotemporal evolution of granular microslip precursors to laboratory earthquakes. *Geophysical Research Letters*, *47*(16), e2020GL088404. https://doi.org/10.1029/2020GL088404

Trugman, D. T., Page, M. T., Minson, S. E., & Cochran, E. S. (2019). Peak ground displacement saturates exactly when expected: Implications for earthquake early warning. *Journal of Geophysical Research: Solid Earth*, *124*(5), 4642–4653. https://doi.org/10.1029/2018JB017093

Trugman, D. T., & Ross, Z. E. (2019). Pervasive foreshock activity across southern California. *Geophysical Research Letters*, *46*(15), 8772–8781. https://doi.org/10.1029/2019GL083725

Trugman, D. T., Ross, Z. E., & Johnson, P. A. (2020). Imaging stress and faulting complexity through earthquake waveform similarity. *Geophysical Research Letters*, *47*(1), e2019GL085888. https://doi.org/10.1029/2019GL085888

Trugman, D. T., & Shearer, P. M. (2017). GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences. *Seismological Research Letters*, *88*(2A), 379–391. https://doi.org/10.1785/0220160188

Tytell, J., Vernon, F., Hedlin, M., de Groot Hedlin, C., Reyes, J., Busby, B., et al. (2016). The USArray Transportable Array as a platform for weather observation and research. *Bulletin of the American Meteorological Society*, *97*(4), 603–619. https://doi.org/10.1175/bams-d-14-00204.1

Ueno, T., Maeda, T., Obara, K., Asano, Y., & Takeda, T. (2010). Migration of low-frequency tremors revealed from multiple-array analyses in western Shikoku, Japan. *Journal of Geophysical Research*, *115*(B9), B00A26. https://doi.org/10.1029/2008JB006051

Ulberg, C. W., Creager, K. C., Moran, S. C., Abers, G. A., Thelen, W. A., Levander, A., et al. (2020). Local source Vp and Vs tomography in the Mount St. Helens region with the iMUSH broadband array. *Geochemistry, Geophysics, Geosystems*, *21*(3), e2019GC008888. https://doi.org/10.1029/2019gc008888

USArray Data at the DMC. (2015). Retrieved from http://ds.iris.edu/ds/nodes/dmc/earthscope/usarray

Vail, P. R. (1987). *Seismic stratigraphy interpretation using sequence stratigraphy: Part 1: Seismic stratigraphy interpretation procedure*.

Vail, P. R., Mitchum, R., Jr., & Thompson, S., III. (1977). *Seismic stratigraphy and global changes of sea level: Part 4. Global cycles of relative changes of sea level: Section 2. Application of seismic reflection configuration to stratigraphic interpretation*.

van den Ende, M., Lior, I., Ampuero, J.-P., Sladen, A., Ferrari, A., & Richard, C. (2021). *A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data*. EarthArXiv.

van den Ende, M. P. A., & Ampuero, J.-P. (2020). On the statistical significance of foreshock sequences in Southern California. *Geophysical Research Letters*, *47*(3), e2019GL086224. https://doi.org/10.1029/2019GL086224

van der Elst, N. J., & Brodsky, E. E. (2010). Connecting near-field and far-field earthquake triggering to dynamic strain. *Journal of Geophysical Research*, *115*(B7), B07311. https://doi.org/10.1029/2009JB006681

van der Hilst, R., Kennett, B., Christie, D., & Grant, J. (1994). Project Skippy explores lithosphere and mantle beneath Australia. *EOS, Transactions American Geophysical Union*, *75*(15), 177–181. https://doi.org/10.1029/94eo00857

van Gestel, J.-P. (2021). Ten years of time-lapse seismic on Atlantis Field. *The Leading Edge*, *40*(7), 494–501. https://doi.org/10.1190/tle40070494.1

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Vera, E., Mutter, J., Buhl, P., Orcutt, J., Harding, A., Kappus, M., et al. (1990). The structure of 0-to 0.2-my-old oceanic crust at 9°N on the East Pacific Rise from expanded spread profiles. *Journal of Geophysical Research*, *95*(B10), 15529–15556. https://doi.org/10.1029/jb095ib10p15529

Vigh, D., & Starr, E. W. (2008). 3D prestack plane-wave, full-waveform inversion. *Geophysics*, *73*(5), VE135–VE144. https://doi.org/10.1190/1.2952623

Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1–WCC26. https://doi.org/10.1190/1.3238367

Wald, D. J. (2020). Practical limitations of earthquake early warning. *Earthquake Spectra*, *36*(3), 1412–1447. https://doi.org/10.1177/8755293020911388

Waldhauser, F. (2009). Near-real-time double-difference event location using long-term seismic archives, with application to northern California. *Bulletin of the Seismological Society of America*, *99*(5), 2736–2748. https://doi.org/10.1785/0120080294

Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bulletin of the Seismological Society of America*, *90*(6), 1353–1368. https://doi.org/10.1785/0120000006

Walker, K. T., Ishii, M., & Shearer, P. M. (2005). Rupture details of the 28 March 2005 Sumatra Mw 8.6 earthquake imaged with teleseismic P waves. *Geophysical Research Letters*, *32*(24), L24303. https://doi.org/10.1029/2005GL024395

Walker, K. T., Pichon, A. L., Kim, T. S., de Groot-Hedlin, C., Che, I.-Y., & Garcés, M. (2013). An analysis of ground shaking and transmission loss from infrasound generated by the 2011 Tohoku earthquake. *Journal of Geophysical Research: Atmospheres*, *118*(23), 12–831. https://doi.org/10.1002/2013jd020187

Walker, K. T., & Shearer, P. M. (2009). Illuminating the near-sonic rupture velocities of the intracontinental Kokoxili Mw 7.8 and Denali Fault Mw 7.9 strike-slip earthquakes with global P wave back projection imaging. *Journal of Geophysical Research*, *114*(B2), B02304. https://doi.org/10.1029/2008JB005738

Walter, F., Gräff, D., Lindner, F., Paitz, P., Köpfli, M., Chmiel, M., & Fichtner, A. (2020). Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-15824-6

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., & Woelfel, I. (2020). EasyQuake: Putting machine learning to work for your regional seismic network or local earthquake study. *Seismological Research Letters*, *92*(1), 555–563. https://doi.org/10.1785/0220200226

Wang, B., Harrington, R. M., Liu, Y., Kao, H., & Yu, H. (2018). Remote dynamic triggering of earthquakes in three unconventional Canadian hydrocarbon regions based on a multiple-station matched-filter approach. *Bulletin of the Seismological Society of America*, *109*(1), 372–386. https://doi.org/10.1785/0120180164

Wang, H., Burtz, O., Routh, P., Wang, D., Violet, J., Lu, R., & Lazaratos, S. (2021). Anisotropic 3D elastic full-wavefield inversion to directly estimate elastic properties and its role in interpretation. *The Leading Edge*, *40*(4), 277–286. https://doi.org/10.1190/tle40040277.1

Wang, H. F., Zeng, X., Miller, D. E., Fratta, D., Feigl, K. L., Thurber, C. H., & Mellors, R. J. (2018). Ground motion response to an ML 4.3 earthquake using co-located distributed acoustic sensing and seismometer arrays. *Geophysical Journal International*, *213*(3), 2020–2036. https://doi.org/10.1093/gji/ggy102

Wang, J., Templeton, D. C., & Harris, D. B. (2015). Discovering new events beyond the catalogue—Application of empirical matched field processing to Salton Sea geothermal field seismicity. *Geophysical Journal International*, *203*(1), 22–32. https://doi.org/10.1093/gji/ggv260

Wang, K., Ellsworth, W. L., Beroza, G. C., Williams, G., Zhang, M., Schroeder, D., & Rubinstein, J. (2018). Seismology with dark data: Image-based processing of analog records using machine learning for the rangely earthquake control experiment. *Seismological Research Letters*, *90*(2A), 553–562. https://doi.org/10.1785/0220180298

Wang, K., Lu, L., Maupin, V., Ding, Z., Zheng, C., & Zhong, S. (2020). Surface wave tomography of northeastern Tibetan Plateau using beamforming of seismic noise at a dense array. *Journal of Geophysical Research: Solid Earth*, *125*(4), e2019JB018416. https://doi.org/10.1029/2019jb018416

Wang, Q.-Y., Brenguier, F., Campillo, M., Lecointre, A., Takeda, T., & Aoki, Y. (2017). Seasonal crustal seismic velocity changes throughout Japan. *Journal of Geophysical Research: Solid Earth*, *122*(10), 7987–8002. https://doi.org/10.1002/2017jb014307

Wang, R., Schmandt, B., & Kiser, E. (2020). Seismic discrimination of controlled explosions and earthquakes near Mount St. Helens using *P*/*S* ratios. *Journal of Geophysical Research: Solid Earth*, *125*(10), e2020JB020338. https://doi.org/10.1029/2020jb020338

Wang, S., & Zhang, J. (2020). Automatic detection of amplitude-distorted samples from clipped seismic waveforms. *Seismological Society of America*, *91*(6), 3563–3573. https://doi.org/10.1785/0220200011

Wang, T., Trugman, D., & Lin, Y. (2021). SeismoGen: Seismic waveform synthesis using GAN with application to seismic data augmentation. *Journal of Geophysical Research: Solid Earth*, *126*(4), e2020JB020077. https://doi.org/10.1029/2020JB020077

Wang, W., Cai, G., Lai, G., Chen, M., & Zhang, L. (2021). Ambient noise level in eastern north China from ChinArray and its response to COVID-19. *Seismological Research Letters*, *92*(6), 3586–3601. https://doi.org/10.1785/0220200261

Wang, X., Zhan, Z., Zhong, M., Persaud, P., & Clayton, R. W. (2021). Urban basin structure imaging based on dense arrays and Bayesian array-based coherent receiver functions. *Journal of Geophysical Research: Solid Earth*, *126*(9), e2021JB022279. https://doi.org/10.1029/2021jb022279

Wang, Y., Allam, A., & Lin, F.-C. (2019). Imaging the fault damage zone of the San Jacinto fault near Anza with ambient noise tomography using a dense nodal array. *Geophysical Research Letters*, *46*(22), 12938–12948. https://doi.org/10.1029/2019GL084835

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., et al. (2013). Anisotropic 3D full-waveform inversion. *Geophysics*, *78*(2), R59–R80. https://doi.org/10.1190/geo2012-0338.1

Wielandt, E., & Steim, J. (1986). A digital very-broad-band seismograph. *Annales Geophysicae. Series B*, *4*(3), 227–232.

Williams, E. F., Fernández-Ruiz, M. R., Magalhaes, R., Vanthillo, R., Zhan, Z., González-Herráez, M., & Martins, H. F. (2019). Distributed sensing of microseisms and teleseisms with submarine dark fibers. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-13262-7

Williams, E. F., Fernández-Ruiz, M. R., Magalhaes, R., Vanthillo, R., Zhan, Z., González-Herráez, M., & Martins, H. F. (2021). Scholte wave inversion and passive source imaging with ocean-bottom DAS. *The Leading Edge*, *40*(8), 576–583. https://doi.org/10.1190/tle40080576.1

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., & Trujillo, J. (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the Seismological Society of America*, *88*(1), 95–106. https://doi.org/10.1785/bssa0880010095

Woessner, J., & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, *95*(2), 684–698. https://doi.org/10.1785/0120040007

Woodhouse, J. H., & Dziewonski, A. M. (1984). Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms. *Journal of Geophysical Research*, *89*(B7), 5953–5986. https://doi.org/10.1029/jb089ib07p05953

Wu, S.-M., Lin, F.-C., Farrell, J., Shiro, B., Karlstrom, L., Okubo, P., & Koper, K. (2020). Spatiotemporal seismic structure variations associated with the 2018 Kilauea eruption based on temporary dense geophone arrays. *Geophysical Research Letters*, *47*(9), e2019GL086668. https://doi.org/10.1029/2019gl086668

Wu, W., Zhan, Z., Peng, S., Ni, S., & Callies, J. (2020). Seismic ocean thermometry. *Science*, *369*(6510), 1510–1515. https://doi.org/10.1126/science.abb9519

Xiao, Z., Fuji, N., Iidaka, T., Gao, Y., Sun, X., & Liu, Q. (2020). Seismic structure beneath the Tibetan Plateau from iterative finite-frequency tomography based on ChinArray: New insights into the Indo-Asian collision. *Journal of Geophysical Research: Solid Earth*, *125*(2), e2019JB018344. https://doi.org/10.1029/2019jb018344

Xu, Y., Koper, K. D., Sufri, O., Zhu, L., & Hutko, A. R. (2009). Rupture imaging of the Mw 7.9 12 May 2008 Wenchuan earthquake from back projection of teleseismic P waves. *Geochemistry, Geophysics, Geosystems*, *10*(4). https://doi.org/10.1029/2008GC002335

Yao, D., Huang, Y., Peng, Z., & Castro, R. R. (2020). Detailed investigation of the foreshock sequence of the 2010 Mw7.2 El Mayor-Cucapah earthquake. *Journal of Geophysical Research: Solid Earth*, *125*(6), e2019JB019076. https://doi.org/10.1029/2019JB019076

Yeck, W. L., Patton, J. M., Johnson, C. E., Kragness, D., Benz, H. M., Earle, P. S., et al. (2019). GLASS3: A standalone multiscale seismic detection associator. *Bulletin of the Seismological Society of America*, *109*(4), 1469–1478. https://doi.org/10.1785/0120180308

Yilmaz, Ö. (2001). *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. Society of Exploration Geophysicists.

Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances*, *1*(11), e1501057. https://doi.org/10.1126/sciadv.1501057

Yoon, C. E., Yoshimitsu, N., Ellsworth, W. L., & Beroza, G. C. (2019). Foreshocks and mainshock nucleation of the 1999 Mw 7.1 Hector Mine, California, earthquake. *Journal of Geophysical Research: Solid Earth*, *124*(2), 1569–1582. https://doi.org/10.1029/2018JB016383

York, D. G., Adelman, J., Anderson, J. E., Jr., Anderson, S. F., Annis, J., Bahcall, N. A., et al. (2000). The sloan digital sky survey: Technical summary. *The Astronomical Journal*, *120*(3), 1579.

Yu, C., Zhan, Z., Lindsey, N. J., Ajo-Franklin, J. B., & Robertson, M. (2019). The potential of DAS in teleseismic studies: Insights from the goldstone experiment. *Geophysical Research Letters*, *46*(3), 1320–1328. https://doi.org/10.1029/2018GL081195

Yu, E., Bhaskaran, A., Chen, S.-L., Ross, Z. E., Hauksson, E., & Clayton, R. W. (2021). Southern California earthquake data now available in the AWS Cloud. *Seismological Society of America*, *92*(5), 3238–3247. https://doi.org/10.1785/0220210039

Yu, S., & Ma, J. (2021). Deep learning for geophysics: Current and future trends. *Reviews of Geophysics*, *59*(3), e2021RG000742. https://doi.org/10.1029/2021RG000742

Yuan, H., French, S., Cupillard, P., & Romanowicz, B. (2014). Lithospheric expression of geological units in central and eastern North America from full waveform tomography. *Earth and Planetary Science Letters*, *402*, 176–186. https://doi.org/10.1016/j.epsl.2013.11.057

Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster computing with working sets. *HotCloud*, *10*(10–10), 95.

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., et al. (2016). Apache Spark: A unified engine for Big Data processing. *Communications of the ACM*, *59*(11), 56–65. https://doi.org/10.1145/2934664

Zeng, S., Zheng, Y., Niu, F., & Ai, S. (2020). Measurements of seismometer orientation of the first phase CHINArray and their implications on vector-recording-based seismic studies. *Bulletin of the Seismological Society of America*, *111*(1), 36–49. https://doi.org/10.1785/0120200129

Zhan, Z. (2020). Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas. *Seismological Research Letters*, *91*(1), 1–15. https://doi.org/10.1785/0220190112

Zhan, Z., Helmberger, D. V., Kanamori, H., & Shearer, P. M. (2014). Supershear rupture in a Mw 6.7 aftershock of the 2013 Sea of Okhotsk earthquake. *Science*, *345*(6193), 204–207. https://doi.org/10.1126/science.1252717

Zhang, X., Arrowsmith, S., Tsongas, S., Hayward, C., Meng, H., & Ben-Zion, Y. (2022). A data-driven framework for automated detection of aircraft-generated signals in seismic array data using machine learning. *Seismological Society of America*, *93*(1), 226–240. https://doi.org/10.1785/0220210198

Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/tkde.2021.3070203

Zhang, Y., & Zhao, Y. (2015). Astronomy in the Big Data Era. *Data Science Journal*, *14*, 11. https://doi.org/10.5334/dsj-2015-011

Zhao, W., Nelson, K., Che, J., Quo, J., Lu, D., Wu, C., & Liu, X. (1993). Deep seismic reflection evidence for continental underthrusting beneath southern Tibet. *Nature*, *366*(6455), 557–559. https://doi.org/10.1038/366557a0

Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. O'Reilly Media, Inc.

Zhu, H., Bozdağ E., Peter, D., & Tromp, J. (2012). Structure of the European upper mantle revealed by adjoint tomography. *Nature Geoscience*, *5*(7), 493–498. https://doi.org/10.1038/ngeo1501

Zhu, H., Komatitsch, D., & Tromp, J. (2017). Radial anisotropy of the North American upper mantle based on adjoint tomography with USArray. *Geophysical Journal International*, *211*(1), 349–377. https://doi.org/10.1093/gji/ggx305

Zhu, H., & Tromp, J. (2013). Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean. *Science*, *341*(6148), 871–875. https://doi.org/10.1126/science.1241335

Zhu, H., Yang, J., & Li, X. (2020). Azimuthal anisotropy of the North American upper mantle based on full waveform inversion. *Journal of Geophysical Research: Solid Earth*, *125*(2). https://doi.org/10.1029/2019JB018432

Zhu, T., & Stensrud, D. J. (2019). Characterizing thunder-induced ground motions using fiber-optic distributed acoustic sensing array. *Journal of Geophysical Research: Atmospheres*, *124*(23), 12810–12823. https://doi.org/10.1029/2019jd031453

Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273. https://doi.org/10.1093/gji/ggy423

10.1029/2021RG000769

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9476–9488. https://doi.org/10.1109/tgrs.2019.2926772