

QuakeFlow: a scalable machine-learning-based earthquake monitoring workflow with cloud computing

Weiqiang Zhu,^{1,2} Alvin Brian Hou,³ Robert Yang,³ Avoy Datta,⁴ S. Mostafa Mousavi,² William L. Ellsworth² and Gregory C. Beroza²

¹Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

²Department of Geophysics, Stanford University, Stanford, CA 94305, USA. E-mail: beroza@stanford.edu

³Computer Science Department, Stanford University, Stanford, CA 94305, USA

⁴Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA

Accepted 2022 August 31. Received 2022 August 30; in original form 2022 July 26

SUMMARY

Earthquake monitoring workflows are designed to detect earthquake signals and to determine source characteristics from continuous waveform data. Recent developments in deep learning seismology have been used to improve tasks within earthquake monitoring workflows that allow the fast and accurate detection of up to orders of magnitude more small events than are present in conventional catalogues. To facilitate the application of machine-learning algorithms to large-volume seismic records at scale, we developed a cloud-based earthquake monitoring workflow, QuakeFlow, which applies multiple processing steps to generate earthquake catalogues from raw seismic data. QuakeFlow uses a deep learning model, PhaseNet, for picking P/S phases and a machine learning model, GaMMA, for phase association with approximate earthquake location and magnitude. Each component in QuakeFlow is containerized, allowing straightforward updates to the pipeline with new deep learning/machine learning models, as well as the ability to add new components, such as earthquake relocation algorithms. We built QuakeFlow in Kubernetes to make it auto-scale for large data sets and to make it easy to deploy on cloud platforms, which enables large-scale parallel processing. We used QuakeFlow to process three years of continuous archived data from Puerto Rico within a few hours, and found more than a factor of ten more events that occurred on much the same structures as previously known seismicity. We applied QuakeFlow to monitoring earthquakes in Hawaii and found over an order of magnitude more events than are in the standard catalogue, including many events that illuminate the deep structure of the magmatic system. We also added Kafka and Spark streaming to deliver real-time earthquake monitoring results. QuakeFlow is an effective and efficient approach both for improving real-time earthquake monitoring and for mining archived seismic data sets.

Key words: Machine learning; Cloud computing; Computational seismology; Earthquake source observations.

1 INTRODUCTION

Continuous seismic waveforms are recorded across seismic networks and processed by earthquake monitoring workflows to detect, locate and characterize seismic events. The resulting earthquake catalogues illuminate the 3D geometry of seismically active structures and reveal the spatio-temporal evolution of seismicity. A comprehensive earthquake catalogue provides crucial information for understanding complex earthquake sequences and quantifying earthquake hazard. The increasing number of dense seismic networks and the rapidly accumulating amount of seismic waveform

data over time pose a challenge for mining seismic data sets to realize the full benefit of more extensive instrumentation. The rapid progress of deep learning algorithms when coupled with cloud computing provides a promising pathway to address the big data challenge in earthquake monitoring to generate comprehensive and accurate earthquake catalogues.

Deep learning seismology (Mousavi & Beroza 2022) has dramatically improved earthquake monitoring performance particularly in earthquake detection and phase picking (Perol *et al.* 2018; Ross *et al.* 2018; Zhu & Beroza 2019; Mousavi *et al.* 2020). Earthquake monitoring workflows using deep-learning-based phase pickers have

been applied to studying, dense earthquake sequences (Liu *et al.* 2020; Ross *et al.* 2020; Tan *et al.* 2021), induced seismicity (Chai *et al.* 2020; Park *et al.* 2020, *in press*; Wang *et al.* 2020; Zhou *et al.* 2021), marine seismicity (Gong *et al.* 2022; Jiang *et al.* 2022), and magmatic systems (Retailleau *et al.* 2022a). These studies have demonstrated that deep learning can: detect up to orders of magnitude more small earthquakes than conventional algorithms, provide a more complete accounting of seismicity, and enable new insight into earthquake behaviour. Deep-learning-based earthquake monitoring workflows are now feasible for adoption in operational earthquake monitoring systems and that transition is proceeding apace (Yeck (Walter *et al.* 2021; Yeck *et al.* 2021; Retailleau *et al.* 2022b; Shi *et al.* 2022; Zhang *et al.* 2022).

Massive amounts of continuous seismic data have accumulated over the last several decades. As of April, 2022 the archived data volume at the Incorporated Research Institutions for Seismology (IRIS) has reached about 800 TiB¹. There are many times this much data from local and regional earthquake monitoring networks that are stored elsewhere. With the advent of distributed acoustic sensing (DAS) for earthquake monitoring, the rate of data accumulation is poised to accelerate dramatically (Zhan 2020; Lindsey & Martin 2021). The current data access model, in which individual users download all data of interest for local processing, would seem to have a limited future. Cloud computing has the potential to address the coupled big data challenges posed by the need to access, and the need to process, massive seismic data sets. A recent survey of international seismological data users (Quinteros *et al.* 2021) expressed the need for cloud-ready seismic processing software as part of a future vision for cloud-based seismic waveform data repositories. Cloud computing can significantly decrease the wall time required for data processing by using thousands of computational nodes provided by cloud computing platforms, such as AWS (Amazon Web Services), GCP (Google Cloud Platform), and Microsoft Azure. Cloud-native computing technologies enable building and deploying scalable applications in the cloud. For example, Docker² is widely used to package an application and its dependencies in a virtual container that can flexibly run on different computing platforms. Kubernetes³ is a container orchestration system that deploys, maintains and scales applications based on computational workloads such as CPU, memory, and storage. Kubeflow⁴ is a cloud-native framework to run machine learning pipelines on Kubernetes clusters. Kafka⁵ is a distributed message streaming service that allows writing, reading, and storing data streams. Spark streaming⁶ is a scalable stream processing engine that performs analytics on streaming data from messaging services such as Kafka. FastAPI⁷ is a web framework for developing RESTful APIs, where REST stands for Representational State Transfer, a software architectural style that describes a uniform interface in a client-server architecture, and API stands for Application Programming Interface. These rapidly developing cloud-native software, particularly when coupled with cloud-based data sets, such as the SCEDC AWS public data set⁸, make cloud-based processing particularly efficient.

In this study, we combined two machine learning algorithms for earthquake detection with cloud computing for parallel processing, to build an earthquake monitoring workflow that we call ‘QuakeFlow’. QuakeFlow can be applied to either mining massive archived data sets or to processing real-time streamed waveforms. For the initial application of QuakeFlow we used the deep neural network model, PhaseNet (Zhu & Beroza 2019), to pick P- and S-phases and the Gaussian mixture model, GaMMA (Zhu *et al.* 2022), to associate picks and estimate approximate earthquake locations and magnitudes. We added Kafka and Spark Streaming services to support real-time earthquake monitoring, and deployed the QuakeFlow system in Kubernetes, making it platform-independent and scalable to thousands of cloud computing nodes for large-scale seismic data mining. QuakeFlow is set up to combine state-of-the-art machine learning models and cloud computing techniques for earthquake monitoring.

2 EARTHQUAKE MONITORING WORKFLOW

An earthquake monitoring workflow consists of a sequence of tasks: phase detection/picking, association, location, and characterization, to detect earthquake signals and estimate source parameters. In this proof-of-concept study, we focused on two of these tasks: phase picking and phase association, which have been significantly improved by machine learning models. Models such as these are not yet widely used by operational earthquake monitoring systems. We built the QuakeFlow system to explore the potential of machine learning algorithms and cloud computing infrastructure for efficient earthquake monitoring. Fig. 1 shows an overview of QuakeFlow. Two key modules used by QuakeFlow are a deep learning model, PhaseNet, for picking P- and S-phase arrival times and a Gaussian mixture model, GaMMA, for associating phases and estimating approximate earthquake locations and magnitudes. Both models are containerized (meaning that the software is packaged with all necessary dependencies) using Docker, and deployed in Kubernetes. These models can be flexibly replaced by other machine learning models, such as GPD (Ross *et al.* 2018) or EQ-Transformer (Mousavi *et al.* 2020) for phase picking and REAL (Zhang *et al.* 2019) or PhaseLink (Ross *et al.* 2019) for phase association, once containerized. We enable auto-scaling for parallel processing using Kubernetes, and built a batch prediction pipeline that can mine archived data sets in an embarrassingly parallel manner, which means that the data set can be distributed into a number of parallel processes without communication between them, using KubeFlow (Fig. 1a). We used ObsPy (Beyreuther *et al.* 2010) to retrieve continuous seismic waveforms from seismic data centres. These waveforms are processed using PhaseNet and GaMMA in parallel, and then relocated using HypoDD, a joint earthquake relocation algorithm (Waldhauser 2001). In addition to the default batch processing for mining archived data sets, we added Kafka and Spark Streaming services to support stream processing for real-time monitoring (Fig. 1b). We received real-time seismic waveforms from seismic networks using the SeedLink API⁹ and continuously pushed these waveforms to the Kafka messaging service. We used the Spark Streaming service to build an ETL (extract, transform, load) pipeline, which applies a sequence of MapReduce transformations, such as windowing, grouping, filtering, and aggregation. The pre-processed data is then sent to PhaseNet and GaMMA APIs that

¹<http://ds.iris.edu/data/distribution/>

²<https://www.docker.com/>

³<https://kubernetes.io/>

⁴<https://www.kubeflow.org/>

⁵<https://kafka.apache.org/>

⁶<https://spark.apache.org/>

⁷<https://fastapi.tiangolo.com/>

⁸<https://scedc.caltech.edu/data/cloud.html>

⁹<http://ds.iris.edu/ds/nodes/dmc/services/seedlink/>

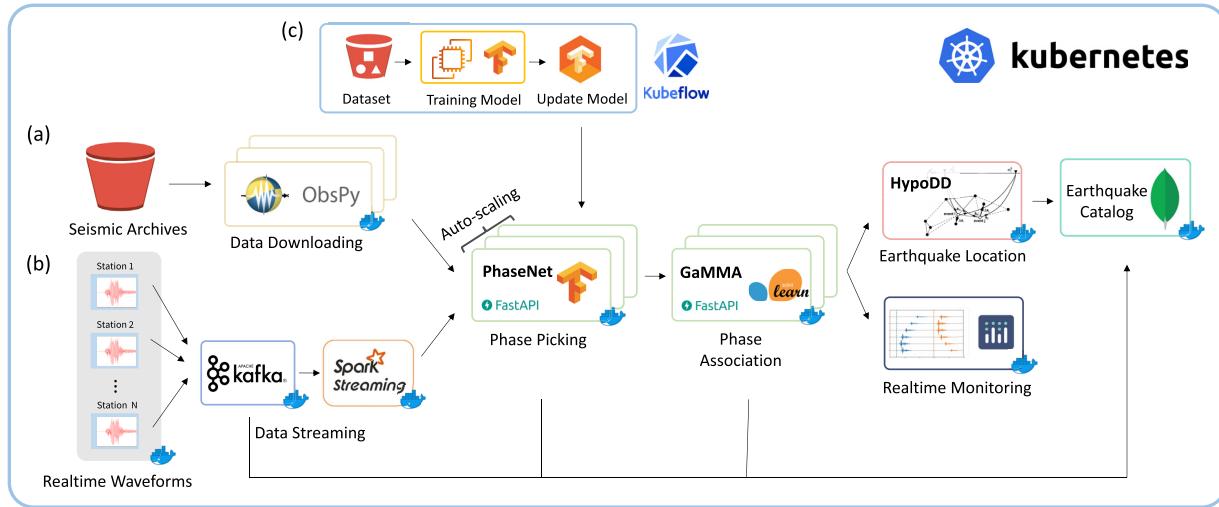


Figure 1. QuakeFlow diagram: (a) batch prediction for mining seismic archives; (b) stream prediction for real-time monitoring; (c) training pipeline for updating deep learning models. Each component of QuakeFlow is containerized using Docker and orchestrated using Kubernetes with auto-scaling. We implement both PhaseNet and GaMMA models and support both large-scale parallel processing of archived seismic waveforms and stream processing of real-time seismic waveforms.

are exposed as RESTful services using FastAPI. The detected earthquakes with approximate locations and magnitudes are saved to an earthquake catalogue using the MongoDB database¹⁰ and broadcast through Kafka to a web app to display real-time waveform and earthquake information. We also built a training and inference pipeline based on the Kubeflow framework for training and updating machine learning models on the cloud (Fig. 1c). We note that QuakeFlow is not limited to its current implementation. New models and more components such as earthquake location algorithms (Lomax *et al.* 2000; Klein 2002; Smith *et al.* 2022) can be containerized and added to this earthquake monitoring workflow.

2.1 Machine learning models

Phase picking and phase association are two key tasks for earthquake monitoring (Fig. 2). New AI-based phase pickers can detect and pick the arrival times of seismic phases, that is *P*-phase and *S*-phase (Fig. 2a) at each seismic station; phase association assembles these phase picks into associated arrivals from causative earthquakes, or unassociated picks that are from either insufficiently well recorded earthquakes or from non-earthquake noise sources (Fig. 2b). In QuakeFlow, we replace conventional phase picking and phase association algorithms with a deep neural network model, PhaseNet (Fig. 2c), and a Gaussian mixture model, GaMMA (Fig. 2d). The two models together extract time and amplitude information of *P* and *S* phases, and detect earthquakes with approximate earthquake locations and magnitudes.

2.1.1 PhaseNet

We used the pre-trained PhaseNet model (Zhu & Beroza 2019) to pick the arrival times of *P* and *S* phases from continuous seismic waveforms (Fig. 2c). PhaseNet is a convolutional neural network (CNN) model that effectively predicts two Gaussian-shaped characteristic functions for *P* and *S* phases, from which we can extract

accurate arrival times. By training on more than 700k examples labeled by human analysts, PhaseNet achieves a much better picking performance than conventional algorithms—typically detecting an order of magnitude more S-picks with high precision and low bias.

2.1.2 GaMMA

We use the unsupervised GaMMA algorithm (Zhu *et al.* 2022), which stands for **G**aussian **M**ixture **M**odel **A**sso**C**iation, to associate picked phase arrivals from multiple stations across a seismic network. GaMMA treats earthquake phase association as an unsupervised clustering problem in a probabilistic framework, where phases are clustered following an approximately hyperbolic move-out of phase travel times (Fig. 2d). The probabilistic framework flexibly considers multiple types of phase information, including arrival time, amplitude, phase type, and quality score, to associate phases from a dense earthquake sequence effectively.

2.2 Cloud computing

2.2.1 Data streaming

Apache Kafka is a fault-tolerant, highly scalable, distributed messaging system for streaming applications (Kreps *et al.* 2011). In QuakeFlow, Kafka acts as the central hub for real-time streaming of waveform data and model prediction results. Our Kafka settings are as follows: there are three pre-defined Kafka topics: `waveform_raw`, `phasenet_picks` and `gamma_events`. The monitoring stations continuously send fragments of seismic waveforms to topic `waveform_raw`. We then use the scalable, fault-tolerant Spark Streaming processing system, which supports both batch and streaming workloads (Zaharia *et al.* 2013), as an ETL pipeline to apply data transformations and pre-processing to the streaming data. Spark Streaming supports operations to aggregate streaming data over a sliding window. We group the streaming data in a specified window size (e.g. 30 s) using a sequence of MapReduce operations to prepare a structured data format for subsequent processing of PhaseNet and GaMMA. The outputs of these machine learning

¹⁰<https://www.mongodb.com/>

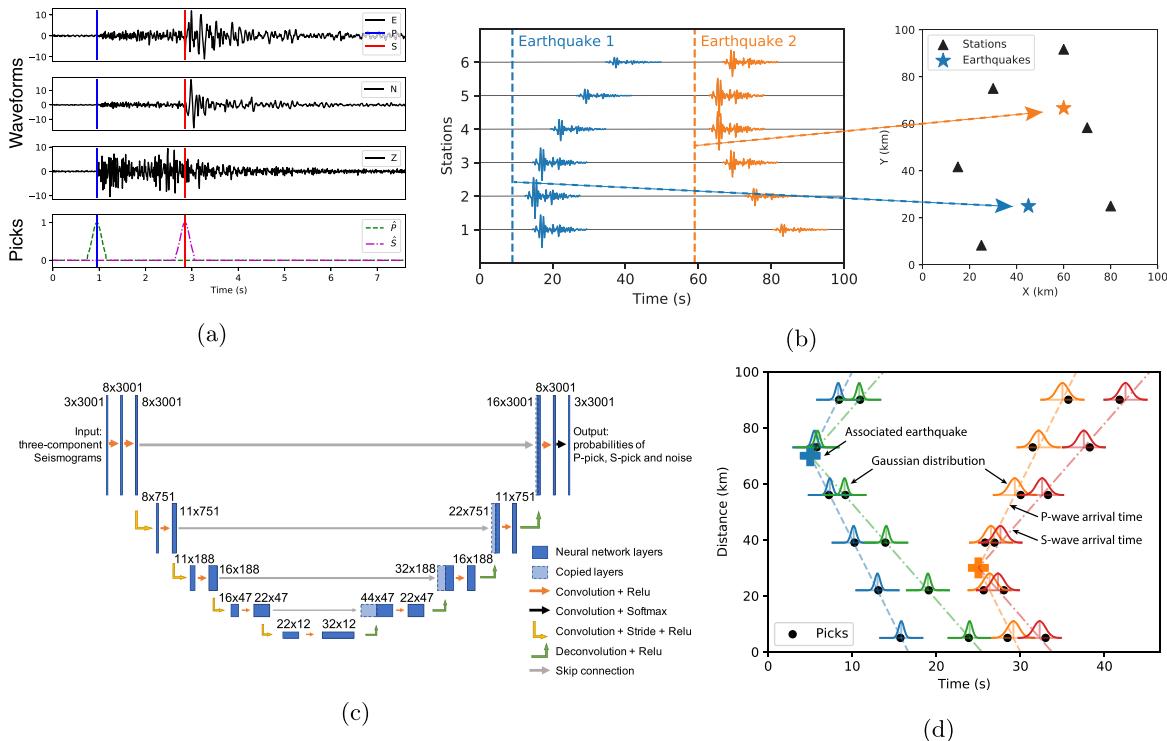


Figure 2. Upper two panels depict two tasks in the earthquake monitoring workflow: (a) Picking P and S phases from waveforms for each seismic station. (b) Associating P and S phases picked across multiple stations and determining earthquake locations and magnitudes. Lower two panels show (c) the PhaseNet model used for phase picking and (d) the GaMMA model used for phase association.

models are broadcast to the `phasenet_picks` and `gamma_events` topics, respectively. Finally, the earthquake detection results can be saved and visualized by subscribing to the corresponding Kafka topics.

2.2.2 Auto-scaling

We deployed QuakeFlow in the Kubernetes system, making it platform-independent and applicable to both on-premise servers and any cloud-platforms with Kubernetes services. Kubernetes automatically orchestrates different components of QuakeFlow to make it run on the cloud efficiently. We used both the horizontal pod auto-scaling provided by Kubernetes, as well as the node auto-provision provided by cloud-platforms, such as Google Cloud Platform (GCP), to match computational resources automatically with computational load. We carried out a simple pressure test on GCP using a maximum of eight computational nodes of machine type ‘n2-standard-2’¹¹ (2 vCPU and 8GB of memory) to evaluate the speedup using auto-scaling when processing a large data volume. Fig. 3(a) shows that the computational time with auto-scaling is significantly reduced compared to the computational time without it. Data throughput, that is the number of waveform-hours processed per second, linearly increases with the data volume when auto-scaling is enabled (Fig. 3b). This means that we can apply auto-scaling for embarrassingly parallel large-scale seismic data mining.

3 APPLICATIONS

We applied QuakeFlow to study two cases: tectonic earthquakes in Puerto Rico and volcanic earthquakes in Hawaii.

3.1 Earthquake detection in Puerto Rico

An earthquake sequence in Puerto Rico started on December 28, 2019 and continued through 2021. The largest earthquake ($M 6.4$) to date occurred on 2020 January 7, causing many injuries and widespread damage (Vanacore *et al.* 2022). The sequence was rich in seismicity (Vičić *et al.* 2022) and involved both strike-slip and normal faulting (ten Brink *et al.* 2022) with much of the activity on the Punta Montalva and Guayanilla Canyon Faults. Deformation in this area is transtensional and diffuse, with unmapped faults capable of generating moderate-to-large earthquakes (Viltres *et al.* 2022).

We focus on testing QuakeFlow’s detection performance and processing speed. We applied QuakeFlow to three years of archived data from 2018-05-01 to 2021-05-01 using 70 stations within a region 65°W – 68°W and 17°N – 19°N (Fig. 4a) for approximately 210 station-years of three-component continuous data from the Puerto Rico Seismic Network (University of Puerto Rico 1986) and the US Geological Survey Networks (Albuquerque Seismological Laboratory (ASL)/USGS 1980). We ran QuakeFlow on GCP with auto-scaling using a maximum of 60 computational nodes of machine type ‘n2-standard-2’ (2 vCPU and 8GB of memory). Downloading waveform data through the IRIS data centre¹² using ObsPy took approximately 3.5 hr, depending on internet conditions and data centre server load. Picking P and S phase arrival times using PhaseNet took

¹¹<https://cloud.google.com/compute/docs/machine-types>

¹²<http://ds.iris.edu/ds/nodes/dmc/data/>

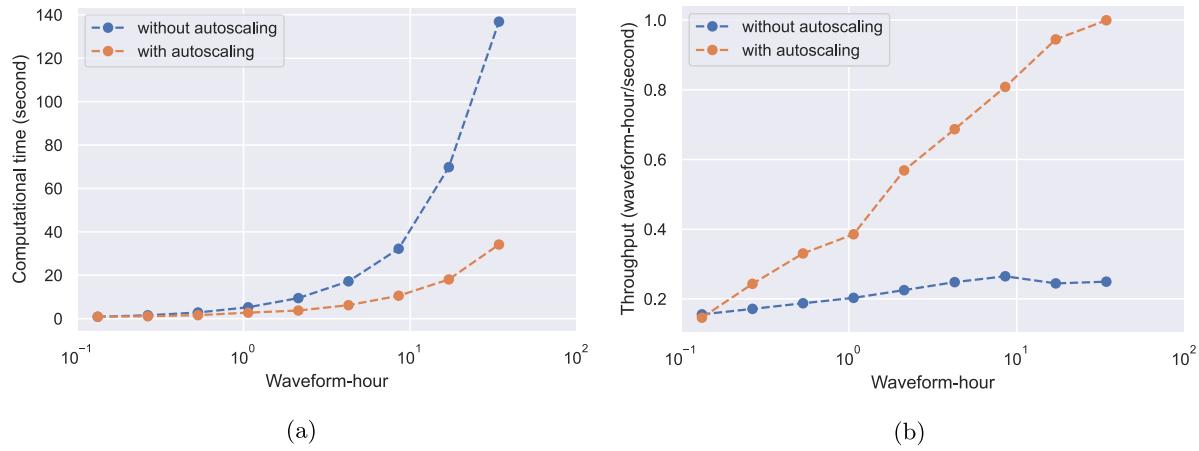


Figure 3. Speedup by auto-scaling: (a) computational time; (b) data throughput.

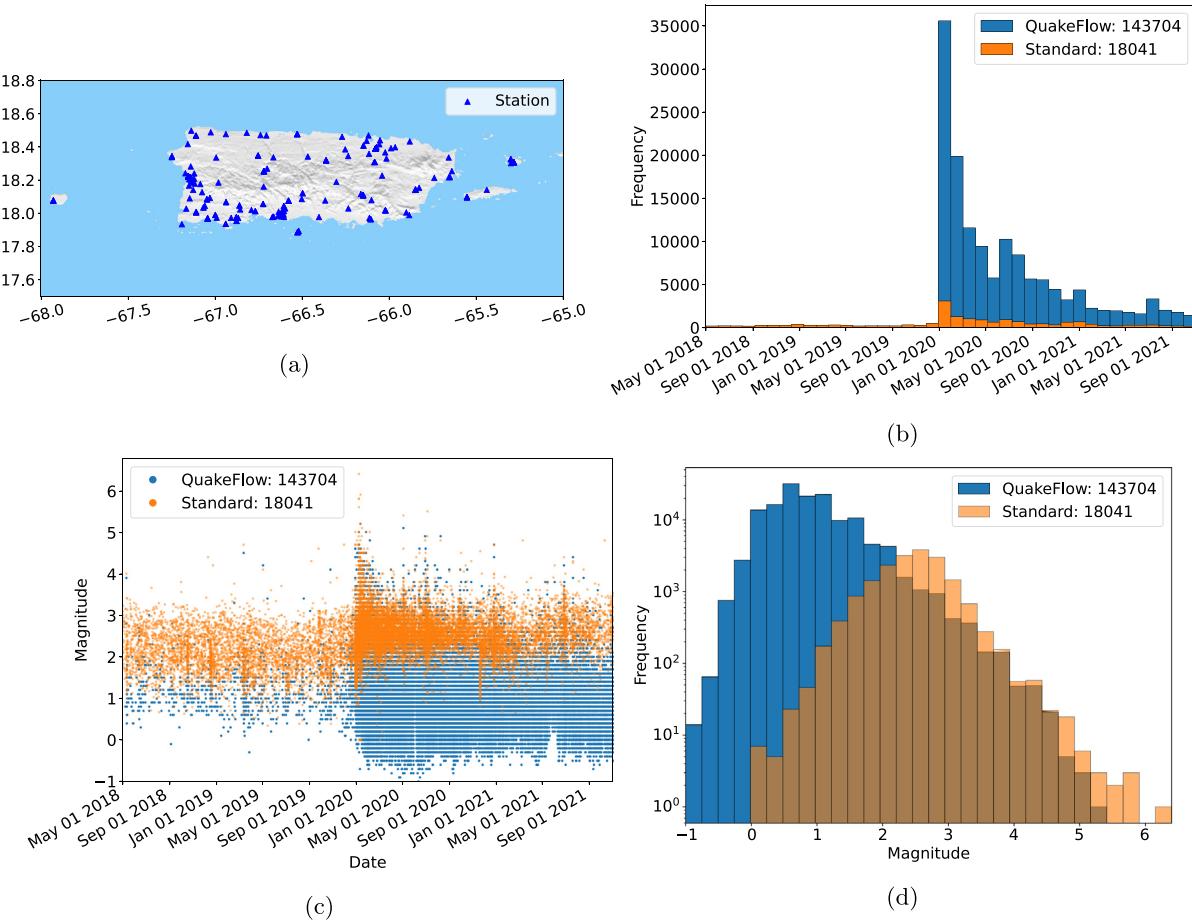


Figure 4. Results for Puerto Rico: (a) seismic station locations; (b) earthquake frequency; (c) earthquake magnitude; (d) earthquake magnitude-frequency distribution. Blue indicates QuakeFlow results. Orange indicates the standard catalogue. Note that earthquake magnitudes are estimated approximately during phase association using GaMMA.

approximately 3 hr, and associating phases using GaMMA took an additional 30 min, or in other words, fast enough to run overnight. The total cost of this processing was around \$40 based on a price of \$0.07/hour per computational node¹³. The earthquake detection

results are shown in Fig. 4 b-d. Compared with the standard catalogue generated by the Puerto Rico Seismic Network (University of Puerto Rico 1986), QuakeFlow detected over an order of magnitude more small earthquakes, particularly during active aftershock periods after 2020-01-01. The exact number of earthquakes will vary depending on the hyperparameters, and there exists a trade-off between false positives and false negatives in both the standard

¹³<https://cloud.google.com/compute/all-pricing>

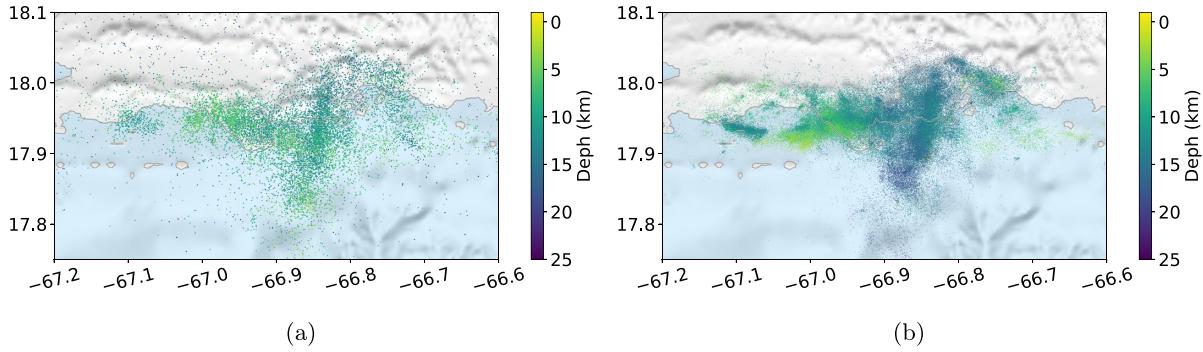


Figure 5. Earthquake locations for Puerto Rico: (a) standard catalogue; (b) QuakeFlow catalogue.

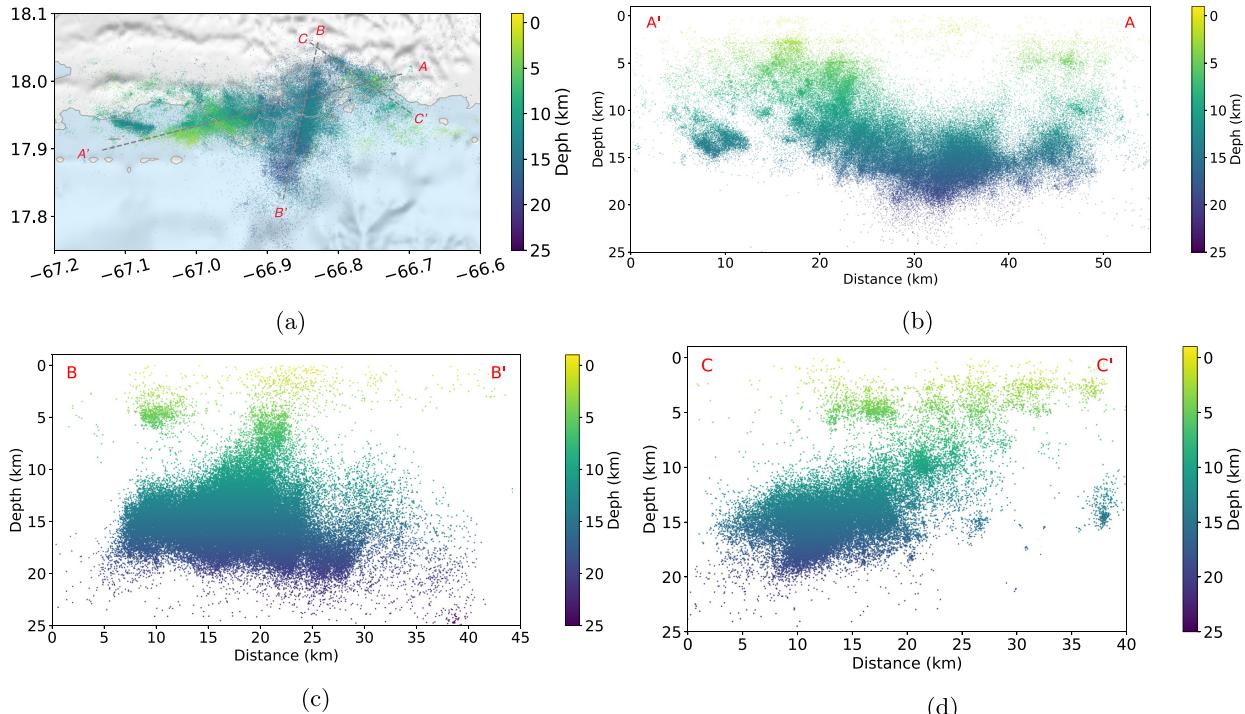


Figure 6. Cross-sections of the earthquake catalogue for Puerto Rico: (a) map view of three cross-sections; (b) A-A' cross-section; (b) B-B' cross-section; (c) C-C' cross-section. We plot events within 10 km of each of the cross-section lines in (a).

catalogue and the QuakeFlow catalogue. Detailed comparison between these catalogues, and quantification of this trade-off, is an important direction for future research. The magnitudes in this workflow are based on simplified ground motion prediction equations (Picozzi *et al.* 2018; Zhu *et al.* 2022). More accurate magnitudes could be determined by adding direct magnitude estimation to the QuakeFlow processing pipeline. The improved earthquake catalogue provides important information to characterize aftershock activity, reveal detailed fault structure, and potentially improve aftershock forecasting (Figs 5 and 6).

3.2 Earthquake detection in Hawaii

We applied QuakeFlow to study volcanic earthquakes on the big island of Hawaii, which has seen a surge of eruptive activity, including the collapse of Kilauea Caldera (Fig. 7a). Most volcanic earthquakes have small magnitudes, which make them an ideal target for QuakeFlow. The active magmatic system in Hawaii has

prodigious seismicity, particularly during eruptions (Klein *et al.* 1987; Matoza *et al.* 2021). Detecting and locating volcanic earthquakes can help illuminate the magma reservoirs and magmatic plumbing systems (Gillard *et al.* 1996; Wech & Thelen 2015). We retrieved seismic waveform data for 66 stations from the Hawaii Volcano Observatory Network (HVO) (USGS Hawaiian Volcano Observatory (HVO) 1956). With only a few hours of cloud computing, we obtain significantly improved resolution of seismicity over a broad depth range. As in Puerto Rico we detect over a factor of 10 more earthquakes than in the standard catalogue reported by Hawaii Volcano Observatory Network over the same time period (Figs 7b–d), which can help illuminate the magmatic system.

We find many deep events (below 30 km) in the Pahala Mantle feature that showed a surge of activity since 2015 and is thought to be caused by the emplacement of new magma (Burgess & Roman 2021; Figs 8 and 9). We also find both deep and shallow events forming lineations: one stretching to the northeast towards Kilauea (Fig. 9b), which is known as the mantle fault zone (Wolfe *et al.* 2003), and the other near the coastline (Fig. 9c), which corresponds

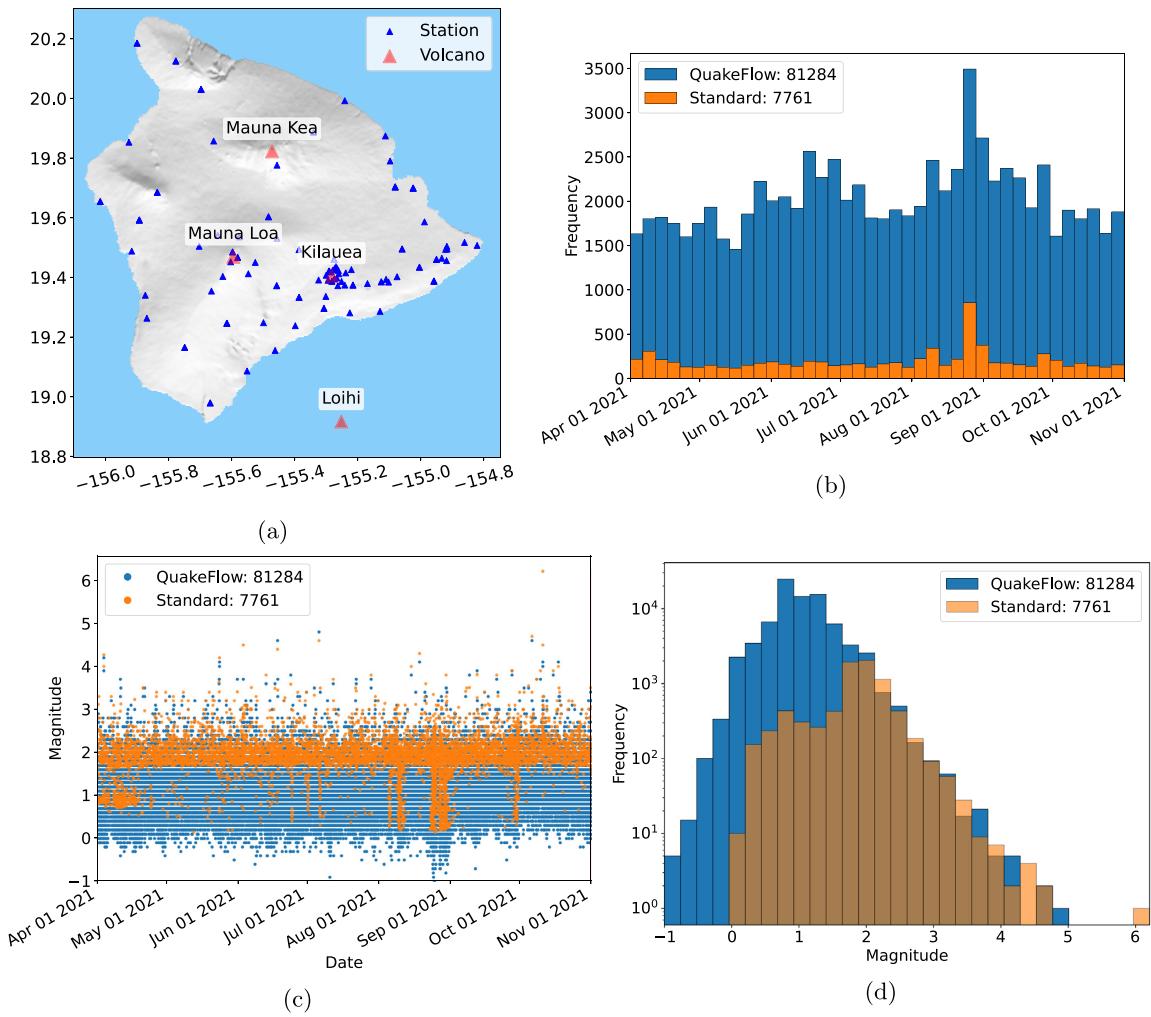


Figure 7. Results for Hawaii: (a) seismic station locations; (b) earthquake frequency; (c) earthquake magnitude; (d) earthquake magnitude-frequency distribution. Blue indicates QuakeFlow results. Orange indicates the standard catalogue. Note that earthquake magnitudes are approximately estimated during phase association using GAMMA.

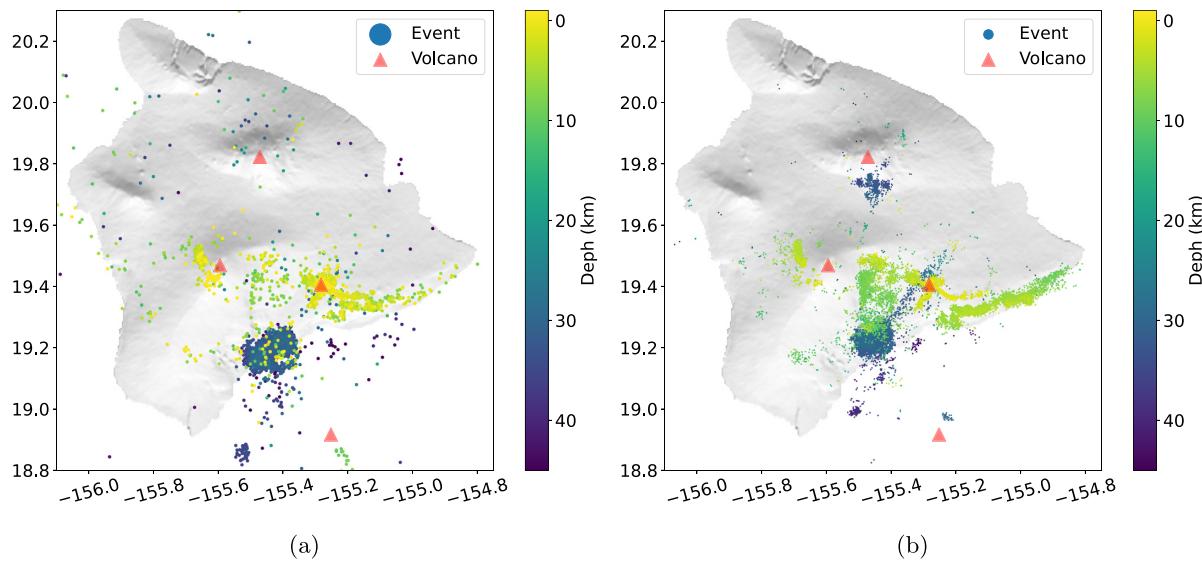


Figure 8. Earthquake locations for Hawaii: (a) standard catalogue; (b) QuakeFlow catalogue.

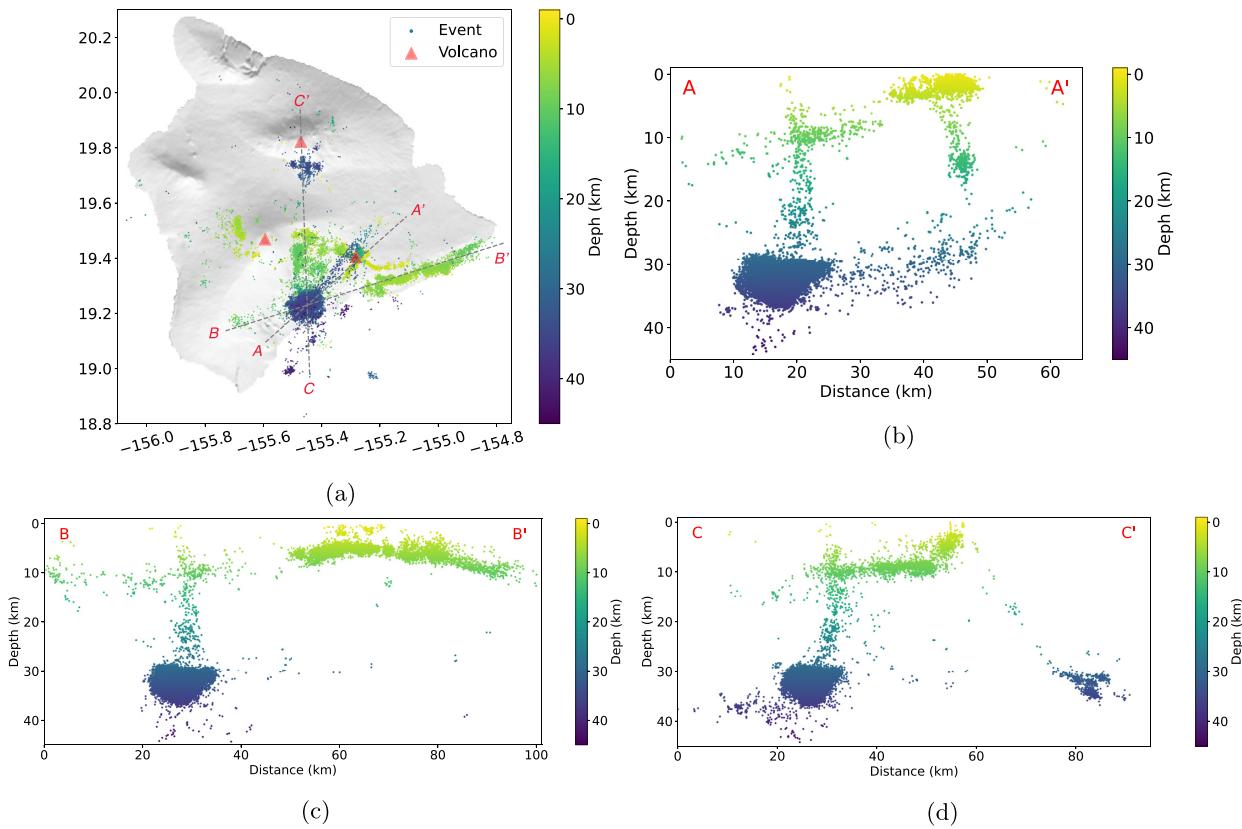


Figure 9. Cross-sections of the earthquake catalogue for Hawaii: (a) a map view of three cross-sections; (b) A-A' cross-section; (b) B-B' cross-section; (c) C-C' cross-section. Events within 5 km of the cross-section lines in (a) are plotted.

to the decollement at about 10 km depth (Denlinger & Okubo 1995). More importantly, the improved catalogue provides a clear picture of the connections between deep and shallow events, illuminating two potential magma transport paths from the deep Pahala cluster to the Kilauea volcano (Fig. 9b): one migrating upwards from the Pahala cluster and the other migrating horizontally towards Kilauea along the mantle fault zone (Wright & Klein 2006). There is also prodigious activity along the rift system, as expected, including additional shallow activity near the summit of Mauna Loa (Matoza *et al.* 2021), and deep events under Mauna Kea (Wech *et al.* 2020) as seen in cross-section C-C' (Fig. 9d). We note that all of these earthquakes are located with PhaseNet arrival time measurements. By adding cross-correlation-based arrival times, it would be possible to more clearly illuminate the full extent and geometric detail of active structures, to track temporal evolution, and to understand its relationship to eruptive activity.

4 DISCUSSION AND CONCLUSIONS

Machine learning—and especially deep learning—methods have developed rapidly in the last few years. Applications to several earthquake sequences [e.g. 2016–2017 the Central Apennines, Italy sequence (Tan *et al.* 2021)] have demonstrated that machine learning models trained on large labeled data sets can significantly outperform conventional approaches, resulting in an earthquake catalogue with unprecedented spatial-temporal resolution. Applying these machine learning methods to revisit archived seismic data sets is a rewarding, but computationally challenging, task. Cloud computing addresses the computational challenge using almost unlimited

computing nodes to parallelize seismic data mining workloads efficiently. As detailed here, we developed the QuakeFlow project to combine the impressive earthquake detection performance of machine learning algorithms with the powerful parallel processing capability of cloud computing to improve earthquake monitoring workflows. We built QuakeFlow based on the container-orchestration system, Kubernetes, and the Kubeflow project to run machine learning models in parallel on the cloud. QuakeFlow is made using containerized components, facilitating updates to deep learning/machine learning models current with the state-of-the-art. Quakeflow also facilitates benchmarking and comparison of the performance of competing algorithms by allowing swappable modules while holding data and other algorithms fixed. In its current implementation, QuakeFlow contains a deep neural network model for phase picking and a Gaussian mixture model for phase association to detect many more small earthquakes than conventional methods. Additional steps, such as denoising (Zhu *et al.* 2019) or additional processing, such as precise earthquake location, magnitude, and focal mechanism determination, could be added to QuakeFlow to improve the catalogue output.

QuakeFlow explores a new approach to the earthquake monitoring workflow based on machine learning and cloud computing. We applied QuakeFlow to study both tectonic earthquakes in Puerto Rico and volcanic earthquakes in Hawaii. The results of these experiments demonstrate that cloud computing enables a flexible and efficient implementation of machine learning models for earthquake monitoring workflows. QuakeFlow runs on most cloud platforms with Kubernetes services to process huge amounts of seismic data in parallel with auto-scaling. It can be applied to many seismic

networks and data sets to improve earthquake detection and reveal details of earthquake occurrence.

ACKNOWLEDGMENTS

We thank Martijn van den Ende, Jannes Münchmeyer and Margarita Segou for their constructive reviews and suggestions. We thank Miao Zhang, Yongsoo Park and Ian McBrearty for helpful discussions. The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used for access to waveforms, related metadata and/or derived products used in this study. This work was supported by AFRL under contract number FA9453-19-C-0073.

DATA AVAILABILITY

The seismic waveforms and earthquake catalogues used in this study come from the Puerto Rico Seismic Network, the US Geological Survey Networks, and the Hawaiian Volcano Observatory Network. The QuakeFlow codebase is accessible at DOI:[10.5281/zenodo.7023970](https://doi.org/10.5281/zenodo.7023970)

REFERENCES

- Albuquerque Seismological Laboratory (ASL)/USGS, 1980. *US Geological Survey Networks [Data set]*, International Federation of Digital Seismograph Networks, doi:[10.7914/SN/GS](https://doi.org/10.7914/SN/GS).
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. Obspy: a python toolbox for seismology, *Seismol. Res. Lett.*, **81**(3), 530–533.
- Burgess, M.K. & Roman, D.C., 2021. Ongoing (2015-) magma surge in the upper mantle beneath the island of Hawaii, *Geophys. Res. Lett.*, **48**(7), e2020GL091096.
- Chai, C. et al., 2020. Using a deep neural network and transfer learning to bridge scales for seismic phase picking, *Geophys. Res. Lett.*, **47**(16), e2020GL088651.
- Denlinger, R.P. & Okubo, P., 1995. Structure of the mobile south flank of Kilauea volcano, Hawaii, *J. geophys. Res.*, **100**(B12), 24 499–24 507.
- Gillard, D., Rubin, A.M. & Okubo, P., 1996. Highly concentrated seismicity caused by deformation of Kilauea's deep magma system, *Nature*, **384**(6607), 343–346.
- Gong, J., Fan, W. & Parnell-Turner, R., 2022. Microseismicity indicates atypical small-scale plate rotation at the Quebrada transform fault system, East Pacific Rise, *Geophys. Res. Lett.*, **49**(3), e2021GL097000.
- Jiang, C., Zhang, P., White, M.C., Pickle, R. & Miller, M.S., 2022. A detailed earthquake catalog for Banda arc–Australian plate collision zone using machine-learning phase picker and an automated workflow, *Seismic Record*, **2**(1), 1–10.
- Klein, F.W., 2002. User's guide to hypoinverse-2000, a fortran program to solve for earthquake locations and magnitudes, Tech. Rep., US Geological Survey.
- Klein, F.W., Koyanagi, R.Y., Nakata, J.S. & Tanigawa, W.R., 1987. The seismicity of Kilauea's magma system, in *Volcanism in Hawaii*, Vol. 2, pp. 1019–1185, eds Decker, R.W., Wright, T.L. & Stauffer, P.H., U.S. Geol. Survey.
- Kreps, J., Narkhede, N. & Rao, J., 2011. Kafka: a distributed messaging system for log processing, in *Proceedings of the NetDB*, Vol. 11, pp. 1–7, Athens, Greece.
- Lindsey, N.J. & Martin, E.R., 2021. Fiber-optic seismology, *Annu. Rev. Earth Planet. Sci.*, **49**, 309–336.
- Liu, M., Zhang, M., Zhu, W., Ellsworth, W.L. & Li, H., 2020. Rapid characterization of the July 2019 Ridgecrest, California, earthquake sequence from raw seismic data using machine-learning phase picker, *Geophys. Res. Lett.*, **47**(4), e2019GL086189.
- Lomax, A., Virieux, J., Volant, P. & Berge-Thierry, C., 2000. Probabilistic earthquake location in 3D and layered models, in *Advances in Seismic Event Location*, pp. 101–134, eds Thurber, C.H. & Rabinowitz, N., Springer.
- Matoza, R.S., Okubo, P.G. & Shearer, P.M., 2021. Comprehensive high-precision relocation of seismicity on the island of Hawaii 1986–2018, *Earth Space Sci.*, **8**(1), e2020EA001253.
- Mousavi, S.M. & Beroza, G.C., 2022. Deep-learning seismology, *Science*, **377**(6607), eabm4470.
- 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, *Nat. Commun.*, **11**(1), 1–12.
- Park, Y., Mousavi, S.M., Zhu, W., Ellsworth, W.L. & Beroza, G.C., 2020. Machine-learning-based analysis of the guy-greenbrier, Arkansas earthquakes: a tale of two sequences, *Geophys. Res. Lett.*, **47**(6), e2020GL087032.
- Park, Y., Beroza, G.C. & Ellsworth, W.L., 2022. Basement fault activation before larger earthquakes in Oklahoma and Kansas, *Seismic Record*, **2**(3), 197–206.
- Perol, T., Gharbi, M. & Denolle, M., 2018. Convolutional neural network for earthquake detection and location, *Sci. Adv.*, **4**(2), e1700578.
- Picozzi, M., Bindi, D., Spallarossa, D., Di Giacomo, D. & Zollo, A., 2018. A rapid response magnitude scale for timely assessment of the high frequency seismic radiation, *Sci. Rep.*, **8**(1), 1–10.
- 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, *Seismol. Res. Lett.*, **92**(3), 1531–1540.
- Retaillau, L. et al., 2022a. Automatic detection for a comprehensive view of mayotte seismicity, *C. R. Géosci.*, **354**(S2), 1–18.
- Retaillau, L. et al., 2022b. A wrapper to use a machine-learning-based algorithm for earthquake monitoring, *Seismol. Res. Lett.*, **93**(3), 1673–1682.
- Ross, Z.E., Meier, M.-A., Hauksson, E. & Heaton, T.H., 2018. Generalized seismic phase detection with deep learning, *Bull. seism. Soc. Am.*, **108**(5A), 2894–2901.
- Ross, Z.E., Yue, Y., Meier, M.-A., Hauksson, E. & Heaton, T.H., 2019. Phaselink: a deep learning approach to seismic phase association, *J. geophys. Res.*, **124**(1), 856–869.
- 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.
- Shi, P., Grigoli, F., Lanza, F., Beroza, G.C., Scarabello, L. & Wiemer, S., 2022. MALMI: an automated earthquake detection and location workflow based on machine learning and waveform migration, *Seismol. Res. Lett.*, **93**, 2467–2483.
- Smith, J.D., Ross, Z.E., Azizzadenesheli, K. & Muir, J.B., 2022. HypoSVI: Hypocentre inversion with stein variational inference and physics informed neural networks, *Geophys. J. Int.*, **228**(1), 698–710.
- Tan, Y. et al., 2021. Machine-learning-based high-resolution earthquake catalog reveals how complex fault structures were activated during the 2016–2017 central Italy sequence, *Seismic Record*, **1**(1), 11–19.
- ten Brink, U.S., Vanacore, E., Fielding, E.J., Chaytor, J.D., López-Venegas, A.M., Baldwin, W.E., Foster, D.S. & Andrews, B.D., 2022. Mature diffuse tectonic block boundary revealed by the 2020 southwestern Puerto Rico seismic sequence, *Tectonics*, **41**(3), e2021TC006896.
- University of Puerto Rico, 1986. *Puerto Rico Seismic Network & Puerto Rico Strong Motion Program*, International Federation of Digital Seismograph Networks.
- USGS Hawaiian Volcano Observatory (HVO), 1956. Hawaiian volcano observatory network [Data set], International Federation of Digital Seismograph Networks, doi:[10.7914/SN/HV](https://doi.org/10.7914/SN/HV).
- Vanacore, E., von Hillebrandt Andrade, C. & McNamara, D.E., 2022. Preface to the SRL focus section on the 2020 southwestern Puerto Rico M_w 6.4 earthquake and seismic sequence, *Seismol. Res. Lett.*, **93**(2A), 531–532.
- Vičić, B., Momeni, S., Borghi, A., Lomax, A. & Aoudia, A., 2022. The 2019–2020 southwest Puerto Rico earthquake sequence: seismicity and faulting, *Seismol. Res. Lett.*, **93**(2A), 533–543.

- Viltres, R., Nobile, A., Vasyura-Bathke, H., Tripanera, D., Xu, W. & Jónsson, S., 2022. Transtensional rupture within a diffuse plate boundary zone during the 2020 M_w 6.4 Puerto Rico earthquake, *Seismol. Res. Lett.*, **93**(2A), 567–583.
- Waldhauser, F., 2001. hypoDD—a program to compute double-difference hypocenter locations. Open-File Report 2001-113, USGS.
- Walter, J.I., Ogwari, P., Thiel, A., Ferrer, F. & Woelfel, I., 2021. Easyquake: putting machine learning to work for your regional seismic network or local earthquake study, *Seismol. Res. Lett.*, **92**(1), 555–563.
- Wang, R., Schmandt, B., Zhang, M., Glasgow, M., Kiser, E., Rysanek, S. & Stairs, R., 2020. Injection-induced earthquakes on complex fault zones of the raton basin illuminated by machine-learning phase picker and dense nodal array, *Geophys. Res. Lett.*, **47**(14), e2020GL088168.
- Wech, A.G. & Thelen, W.A., 2015. Linking magma transport structures at KīLauea volcano, *Geophys. Res. Lett.*, **42**(17), 7090–7097.
- Wech, A.G., Thelen, W.A. & Thomas, A.M., 2020. Deep long-period earthquakes generated by second boiling beneath Mauna Kea volcano, *Science*, **368**(6492), 775–779.
- Wolfe, C.J., Okubo, P.G. & Shearer, P.M., 2003. Mantle fault zone beneath Kilauea volcano, Hawaii, *Science*, **300**(5618), 478–480.
- Wright, T.L. & Klein, F.W., 2006. Deep magma transport at Kilauea volcano, Hawaii, *Lithos*, **87**(1-2), 50–79.
- Yeck, W.L. et al., 2021. Leveraging deep learning in global 24/7 real-time earthquake monitoring at the national earthquake information center, *Seismol. Res. Lett.*, **92**(1), 469–480.
- Zaharia, M., Das, T., Li, H., Hunter, T., Shenker, S. & Stoica, I., 2013. Discretized streams: fault-tolerant streaming computation at scale, in *Proceedings of the 24th ACM Symp. Operating Systems Principles*, Farmington, Pennsylvania, USA, pp. 423–438.
- Zhan, Z., 2020. Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas, *Seismol. Res. Lett.*, **91**(1), 1–15.
- Zhang, M., Ellsworth, W.L. & Beroza, G.C., 2019. Rapid earthquake association and location, *Seismol. Res. Lett.*, **90**(6), 2276–2284.
- Zhang, M., Liu, M., Feng, T., Wang, R. & Zhu, W., 2022. Loc-flow: an end-to-end machine learning-based high-precision earthquake location workflow, *Seismol. Res. Lett.*, **93**(5), 2426–2438.
- Zhou, P., Ellsworth, W., Yang, H., Tan, Y., Beroza, G., Sheng, M. H. & Chu, R., 2021. Machine-learning-facilitated earthquake and anthropogenic source detections near the Weiyuan Shale Gas Blocks, Sichuan, China, *Earth Planet. Phys.*, **5**(6), 501–519.
- Zhu, W. & Beroza, G.C., 2019. Phasenet: a deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.*, **216**(1), 261–273.
- Zhu, W., Mousavi, S.M. & Beroza, G.C., 2019. Seismic signal denoising and decomposition using deep neural networks, *IEEE Trans. Geosci. Remote Sens.*, **57**(11), 9476–9488.
- Zhu, W., McBrearty, I.W., Mousavi, S.M., Ellsworth, W.L. & Beroza, G.C., 2022. Earthquake phase association using a Bayesian Gaussian mixture model, *J. geophys. Res.*, **127**(5), e2021JB023249.