Rapid Estimation of Single-Station Earthquake Magnitudes with Machine Learning on a Global Scale

Sydney N. Dybing*[1], William L. Yeck[2], Hank M. Cole[2], and Diego Melgar[1]

ABSTRACT

The foundation of earthquake monitoring is the ability to rapidly detect, locate, and estimate the size of seismic sources. Earthquake magnitudes are particularly difficult to rapidly characterize because magnitude types are only applicable to specific magnitude ranges, and location errors propagate to substantial magnitude errors. We developed a method for rapid estimation of single-station earthquake magnitudes using raw three-component *P* waveforms observed at local to teleseismic distances, independent of prior size or location information. We used the MagNet regression model architecture (Mousavi and Beroza, 2020b), which combines convolutional and recurrent neural networks. We trained our model using ∼2.4 million *P*-phase arrivals labeled by the authoritative magnitude assigned by the U.S. Geological Survey. We tested input data parameters (e.g., window length) that could affect the performance of our model in near-real-time monitoring applications. At the longest waveform window length of 114 s, our model (Artificial Intelligence Magnitude [AIMag]) is accurate (median estimated magnitude within ±0.5 magnitude units from catalog magnitude) between M 2.3 and 7.6. However, magnitudes above M ∼7 are more underestimated as true magnitude increases. As the windows are shortened down to 1 s, the point at which higher magnitudes begin to be underestimated moves toward lower magnitudes, and the degree of underestimation increases. The over and underestimation of magnitudes for the smallest and largest earthquakes, respectively, are potentially related to the limited number of events in these ranges within the training data, as well as magnitude saturation effects related to not capturing the full source time function of large earthquakes. Importantly, AIMag can determine earthquake magnitudes with individual stations' waveforms without instrument response correction or knowledge of an earthquake's source-station distance. This work may enable monitoring agencies to more rapidly recognize large, potentially tsunamigenic global earthquakes from few stations, allowing for faster event processing and reporting. This is critical for timely warnings for seismic-related hazards.

KEY POINTS

- Earthquake magnitudes are difficult to accurately determine rapidly for real-time earthquake monitoring.
- We developed a machine learning method for global single-station magnitude estimation using raw *P* waveforms.
- This may allow monitoring agencies to more rapidly process seismic data, enabling timelier hazard warnings.

Supplemental Material

INTRODUCTION

Monitoring agencies worldwide ensure that earthquake information is shared in near-real time as quickly and as accurately as possible. Earthquake information products are key for organizing emergency response efforts after substantial shaking, as well as anticipating and modeling earthquake-associated hazards that can also have catastrophic effects on people and property. For example, the U.S. Tsunami Warning Centers monitor the globe for large earthquakes that could cause tsunamis, and the U.S. Geological Survey (USGS) analyzes and publishes

1. Department of Earth Sciences, University of Oregon, Eugene, Oregon, U.S.A., https://orcid.org/0000-0002-9274-6568 (SND); https://orcid.org/0000-0001-6259-1852 (DM); 2. U.S. Geological Survey, Geologic Hazards Science Center, Golden, Colorado, U.S.A., https://orcid.org/0000-0002-2801-8873 (WLY); https://orcid.org/0000-0003-1684-9116 (HMC)

*Corresponding author: sdybing@uoregon.edu

Cite this article as Dybing, S. N., W. L. Yeck, H. M. Cole, and D. Melgar (2024). Rapid Estimation of Single-Station Earthquake Magnitudes with Machine Learning on a Global Scale, *Bull. Seismol. Soc. Am.* **114**, 1523–1538, doi: 10.1785/0120230171

information about earthquake economic and loss of life effects, as well as landslide and liquefaction risk. In addition, national and international security groups (e.g., the Air Force Technical Applications Center in the United States) conduct global monitoring to detect explosive seismic sources that are not earthquakes, which may indicate violations of the Comprehensive Test Ban Treaty, which bans nuclear explosions worldwide (Ringler *et al.*, 2022). Foundational to all of these missions is the ability to rapidly detect, locate, and characterize seismic sources.

In the United States, the USGS National Earthquake Information Center (NEIC) is tasked with 24/7 rapid response to earthquakes both nationally and across the globe. Automatic processing software and human analysts at the NEIC work to accurately assess the character of earthquakes within minutes to aid in estimates of economic and life loss, and provide information to emergency responders, the media, governments, and the public (Benz, 2017). Key earthquake characteristics include location, magnitude, and rupture extent and pattern, all of which can inform of the potential for associated hazards such as strong shaking and tsunamis. Of these, magnitude is a particularly important source parameter because it is the foundation of subsequent hazard estimates (e.g., shaking intensities) and is the fundamental parameter used by the public to interpret the severity of an earthquake. However, magnitude estimation can be a complicated process, and more simplistic methods may fail to adequately characterize an event. This results in variability of the magnitude based on how it is calculated, and it is prone to errors propagated from location and station metadata inaccuracies.

Magnitude estimation is particularly complex at the NEIC because it monitors earthquakes both domestically and globally in all tectonic environments. For small domestic earthquakes, the NEIC relies on amplitude-based measurements, preferring either M_L (local magnitude) or m_{b_lg} (short-period surface-wave magnitude) depending on the tectonic environment, or measurements derived from waveform modeling (M_{wr}—regional magnitude) when sufficient data are available. For larger global earthquakes, the NEIC relies on different amplitude-based magnitudes (primarily m_b—short-period body-wave magnitude) or magnitudes calculated with waveform modeling (primarily M_{ww}—moment *W*-phase magnitude or M_{wb}—body-wave magnitude).

Each of these magnitude types is only appropriate for specific magnitude ranges, and the appropriateness of each can depend on the properties of the waveforms. As an example, although the NEIC uses M_{ww} as its preferred magnitude, M_{ww} can perform poorly after larger earthquakes due to increased long-period noise levels. Amplitude-based magnitudes are often the first magnitude produced internally in the NEIC's processing system, but these are prone to substantial errors because of mislocation due to their heavy reliance on attenuation relationships. Therefore, magnitude processing that can span all settings and is independent of location information has the potential to aid internal processing systems, as well as rapidly alert analysts of potential significant events.

MACHINE LEARNING FOR EARTHQUAKE SOURCE PARAMETER ESTIMATION

In recent years, research into applications of machine learning in seismology and geophysics has grown substantially, with several studies demonstrating that deep learning models improve upon traditional data processing methods. A review of the history of machine learning use in the seismology field and its current applications and future directions can be found in Mousavi and Beroza (2023). Deep learning typically requires large amounts of data to perform well, and the standard of maintaining publicly accessible data within the geophysics community (e.g., the NEIC's earthquake catalog, the EarthScope waveform data archives) lends itself well to this methodology. The utility of deep learning techniques has been established for many seismology applications, including detection and location of earthquakes and expansion of earthquake catalogs through accurate *P*- and *S*-phase picking for improved locations and velocity models (e.g., Perol *et al.*, 2018; Ross *et al.*, 2018; Zhu and Beroza, 2019; Mousavi and Beroza, 2020a; Mousavi *et al.*, 2020; Zhang *et al.*, 2020).

Another application, the estimation of earthquake magnitudes, has been approached with a variety of deep learning techniques. As reviewed by Mousavi and Beroza (2023), these include seismic single-station classification and regression approaches (Lomax *et al.*, 2019; Mousavi and Beroza, 2020b; Ristea and Radoi, 2021), as well as multistation methods, which in addition to seismic data (van den Ende and Ampuero, 2020; Münchmeyer *et al.*, 2021) can include the use of Global Navigation Satellite Systems (GNSS; Lin *et al.*, 2021) and elastogravity data (Licciardi *et al.*, 2022). Mousavi and Beroza (2020b) demonstrated that a regression model built with the machine learning framework TensorFlow (Abadi *et al.*, 2015) as a combination of a convolutional neural network and a recurrent neural network could accurately estimate local magnitudes from single waveforms with no instrument response removal. The Mousavi and Beroza (2020b) model, known as MagNet, was trained and tested on data from global earthquakes (primarily of $\mathbf{M} < 5$) from the STanford EArthquake Dataset (Mousavi *et al.*, 2019) with epicentral distances of less than one degree. The MagNet model shows great promise, but its applicability as currently trained is limited to small scales and, therefore, is not appropriate for global earthquake monitoring.

Cole *et al.* (2023) constructed the Machine Learning Asset Aggregation of the Preliminary Determination of Epicenters (MLAAPDE, pronounced "millipede") to aid in the development of machine learning tools at the NEIC. MLAAPDE is a data set and query tool built from the Preliminary Determination of Epicenters (PDE) bulletin published by the NEIC (Sipkin *et al.*, 2000). MLAAPDE is formatted such that users will readily be

able to use the >5.1 million waveforms and associated labels in model development, training, and testing. As such, it is this data set that we utilize for this study. Detailed information about MLAAPDE and how to access the code and default data set is available in Cole *et al.* (2023). In this study, we leverage the previous work on the MagNet model with the MLAAPDE data set to test the applicability of the approach to global earthquake monitoring.

METHODS

MLAAPDE data access

The default MLAAPDE catalog contains 120-s-long waveforms, centered on an analyst-verified phase pick and in units of raw counts with a 40 Hz sampling rate. Our goal was to create a model that can be applied in rapid response settings, so we chose to only sample from the available first arrival *P*-phase picks in the catalog (either *P*, *Pn*, or *Pg*). We stream-normalized input waveforms by dividing each observation by the maximum value of all three channel traces for each station. We did not apply instrument response correction or otherwise filter the waveform data. However, the waveforms were decimated to a sampling rate of 20 Hz to reduce computation time and memory required for the code to run.

A common practice in machine learning is to split data sets into training, validation, and testing partitions with a convenient ratio (e.g., 80% samples for training, 20% for testing). Our training data set was composed of earthquakes with origin times spanning five years and two months from 1 August 2013 to 30 September 2018, and earthquake magnitudes between **M** 0.0 and 8.3, containing 2,431,341 waveforms from 2121 stations in the MLAAPDE data set. The validation data set was composed of earthquakes with origin times spanning one year and three months from 1 October 2018 to 31 December 2019, and earthquake magnitudes between **M** 0.7 and 8.0, containing 489,268 waveforms from 1615 stations in the MLAAPDE data set. The testing data set was composed primarily of 324,365 MLAAPDE waveforms from 1536 stations, with earthquake origin times spanning one year from 1 January 2020 to 31 December 2020, and earthquake magnitudes between **M** 0.7 and 7.8. The full data set contains a variety of different magnitude types, but the majority are either m_b, M_L, or M_{ww} (refer to Cole *et al.*, 2023, their fig. 2, for the full list and distribution of magnitude types in MLAAPDE). We ensured that no earthquakes in the training data set were in the validation or testing data sets to reduce leakage and the likelihood that the model would learn the earthquake magnitudes from the waveform similarity of a single event. The separation between validation and testing data sets was made to ensure that the final model would not be biased by the testing process because the best model is chosen in training based on the lowest validation loss achieved.

We also manually expanded the testing data set beyond MLAAPDE to test the model on larger earthquakes, even though it was not trained on these events. MLAAPDE currently only draws from the most modern version of the PDE, which has been used since mid-2013. As a result, the training data set used in this study does not contain any earthquakes above M 8.3 because this was the largest event that occurred between 1 August 2013 and 30 September 2018. To supplement the testing data set, we used ObsPy (Beyreuther *et al.*, 2010) to download 13,151 waveforms from the EarthScope Data Management Center from an additional 78 earthquakes **M** > 7.5 from 1 January 2000 to 31 July 2013. This data set includes several great earthquakes, such as the 2004 **M** 9.1 Sumatra earthquake, 2010 **M** 8.8 Maule earthquake, and 2011 **M** 9.1 Tōhoku earthquake. Waveforms for the earthquakes in this extended testing data set came from 1736 stations.

In total, the full data set was split to be ~75% training, 15% validation, and 10% testing (Fig. S1, available in the supplemental material to this article). Maps of the earthquakes and stations from which waveforms were drawn for this study can be found in Figure 1.

Training data processing and augmentation

The speed at which an earthquake magnitude can be determined in a real-time environment depends on the arrival time and telemetry delay of waveform data, the latency to load the waveform into monitoring applications, the window of waveform data required, and the speed of the magnitude calculation. For this reason, it is important to find the minimum effective window for magnitude estimation. We tested our model across a range of 1–114 s of waveform data. MLAAPDE stores waveforms at a length of 120 s (Fig. 2a), which we shortened further in our data augmentation process (Fig. 2b–f).

Because MLAAPDE is a well-curated catalog of human-reviewed picks, we manually augmented the data to simulate the challenges of real-time processing because we wanted our model to be resilient to issues that can manifest in seismic monitoring. Data augmentation involves synthetically expanding the number of samples in a data set, as well as their variety, to prevent overfitting in machine learning models and improve generalization (Zhu *et al.*, 2020).

We built our data augmentation using a TensorFlow Keras Sequence data generator (Abadi *et al.*, 2015) that prepared the arrays to be fed into the model for training. By implementing the augmentation "on the fly," the data were shuffled and augmented differently with every training epoch, therefore increasing variability.

We augmented the data from MLAAPDE in four different ways following methods tested in Zhu *et al.* (2020). We first implemented a random shift in the location of the phase pick so that it was no longer in the exact center of each waveform. The phase picks in MLAAPDE are all human-reviewed, and in many cases, repicked for accuracy. In real-time monitoring settings, phase picks are performed automatically, often using short-term average/long-term average methods (Lee and

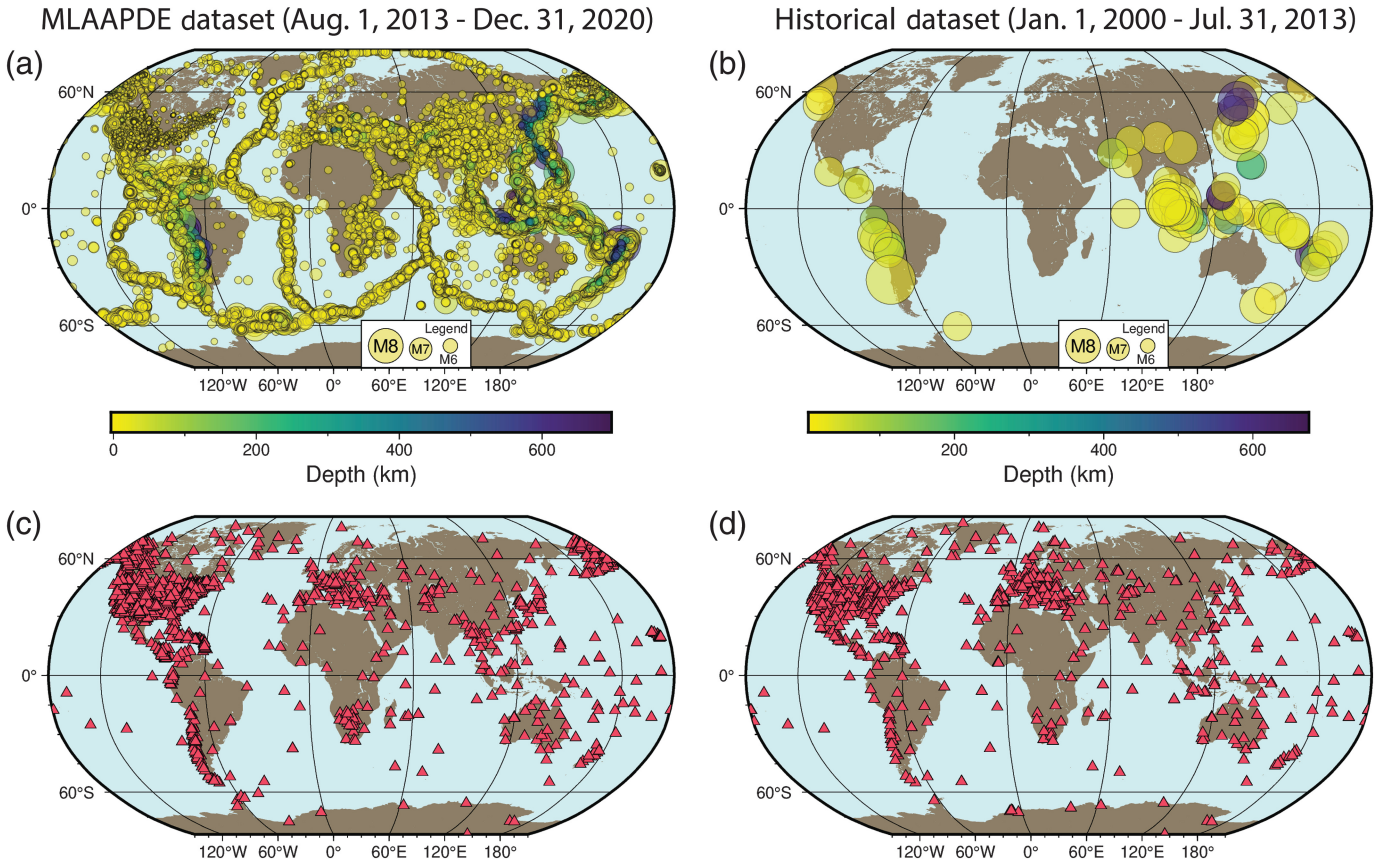

Figure 1. Maps of the earthquakes and seismometers that recorded them used in this study. (a) The 189,475 earthquakes in the combined training, validation, and testing data sets from the Machine Learning Asset Aggregation of the Preliminary Determination of Epicenters (MLAAPDE) catalog with origin times from 1 August 2013 to 31 December 2020. (b) The 78 earthquakes **M** > 7.5 in the expanded historical testing data set with origin times from 1 January 2000 to 31 July 2013. (c) The 2422 stations from which waveforms were drawn for the MLAAPDE earthquake catalog shown in panel (a). (d) The 1736 stations from which waveforms were drawn for the historical earthquake catalog shown in panel (b). These maps were constructed using PyGMT (Uieda *et al.*, 2023).

Stewart, 1981). Automatic picks are often not as accurate as manual picks from analysts. Therefore, to allow our model to learn to estimate earthquake magnitudes on real-time data that may not have perfectly accurate phase arrival times, we implemented a random shift of up to 3 s (forward or backward) into each waveform from MLAAPDE (Fig. 2c), which further shortened the window length by 6 s. This random shift was applied to all waveforms in the training, validation, and testing data sets.

Next, noise was added to some of the waveforms only in the training data set. Because the data generator looped through the waveforms, ∼50% of the time, additional noise was introduced. For each point in the waveform, a value drawn from a normal distribution was added, with the standard deviation of each of those normal distributions being drawn from a uniform distribution between 0.01 and 0.15 (Fig. 2d). Many of the waveforms in the MLAAPDE data set already have low signal-to-noise ratios (SNRs), so this augmentation simply increased the number of "different" waveforms that the model could learn about in training.

The third augmentation involved swapping the two horizontal components ∼50% of the time in the training data (Fig. 2e). This simulated potential station orientation issues and changed the source-station path, again introducing more data variety.

The fourth and final augmentation dropped each channel ∼5% of the time, with a limitation to prevent all three from being dropped simultaneously (in this case, the third or vertical channel was kept). An example of this is shown in Figure 2f. In a real-time setting, it is very possible that some instruments or channels may not be recording or transmitting data due to malfunctions or telemetry problems. Introducing dropped channels into the training data allowed for the machine learning model to be resilient to these issues when making magnitude estimations.

In the training process, we used the same Adam optimizer (Kingma and Ba, 2015) as Mousavi and Beroza (2020b) as well as a mean squared error loss function, and we utilized the same early stopping and learning rate reduction to prevent overfitting and stop the training when the loss rate plateaued. Training was repeated for each different window length of data, resulting in a total of 26 trained models.

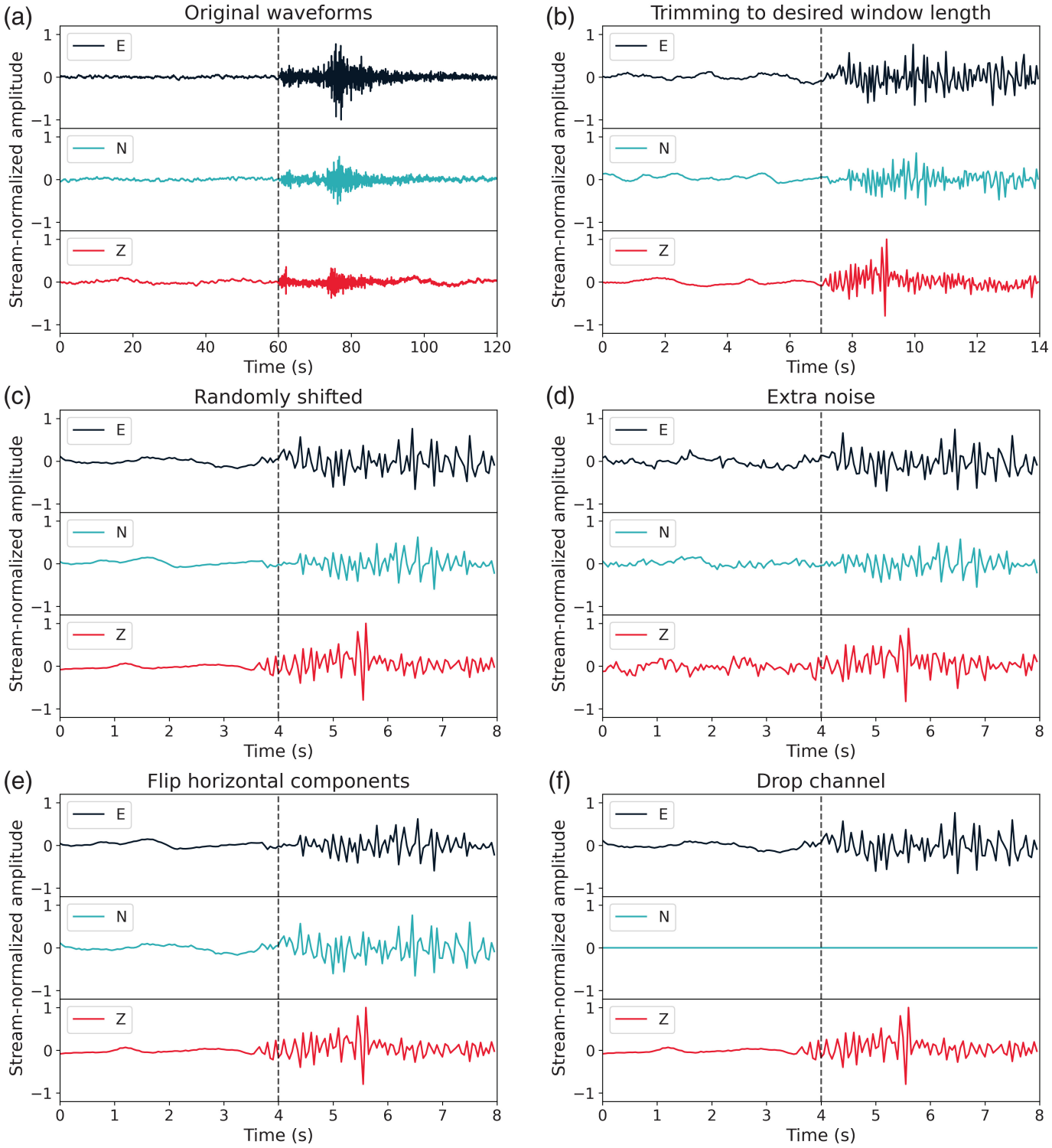

Figure 2. Time-series plots demonstrating the different types of augmentation implemented in our data generator. (a) The original three-component 120-s waveforms loaded from the MLAAPDE data set, stream-normalized with no instrument response removed. The black trace indicates the east–west component, the blue trace the north–south component, and the red trace the vertical component. The phase pick is in the center of the figure at 60 s. The center of the figure is indicated by the black dashed line. (b) The same waveforms as in panel (a), cut down to a smaller window length (14 s in this example). (c) The same waveforms as in panel (b), but the window length has been cut down an additional 6 s to accommodate an applied random shift of the location of the phase pick by up to 3 s forward or backward. The center of the figure (which is no longer aligned with the phase pick) is still indicated by the black dashed line. (d) The same waveforms as in panel (c) with additional noise applied. For each point in the waveform, a value drawn from a normal distribution was added, with the standard deviation of each of those normal distributions being drawn from a uniform distribution between 0.01 and 0.15. The amount of added noise is not consistent between components. (e) The same waveforms as in panel (c), but the east–west and north–south components (black and blue) have been switched with each other. (f) The same waveforms as in panel (c), but in this instance the north–south (blue) component has been "dropped" by zeroing-out the trace.

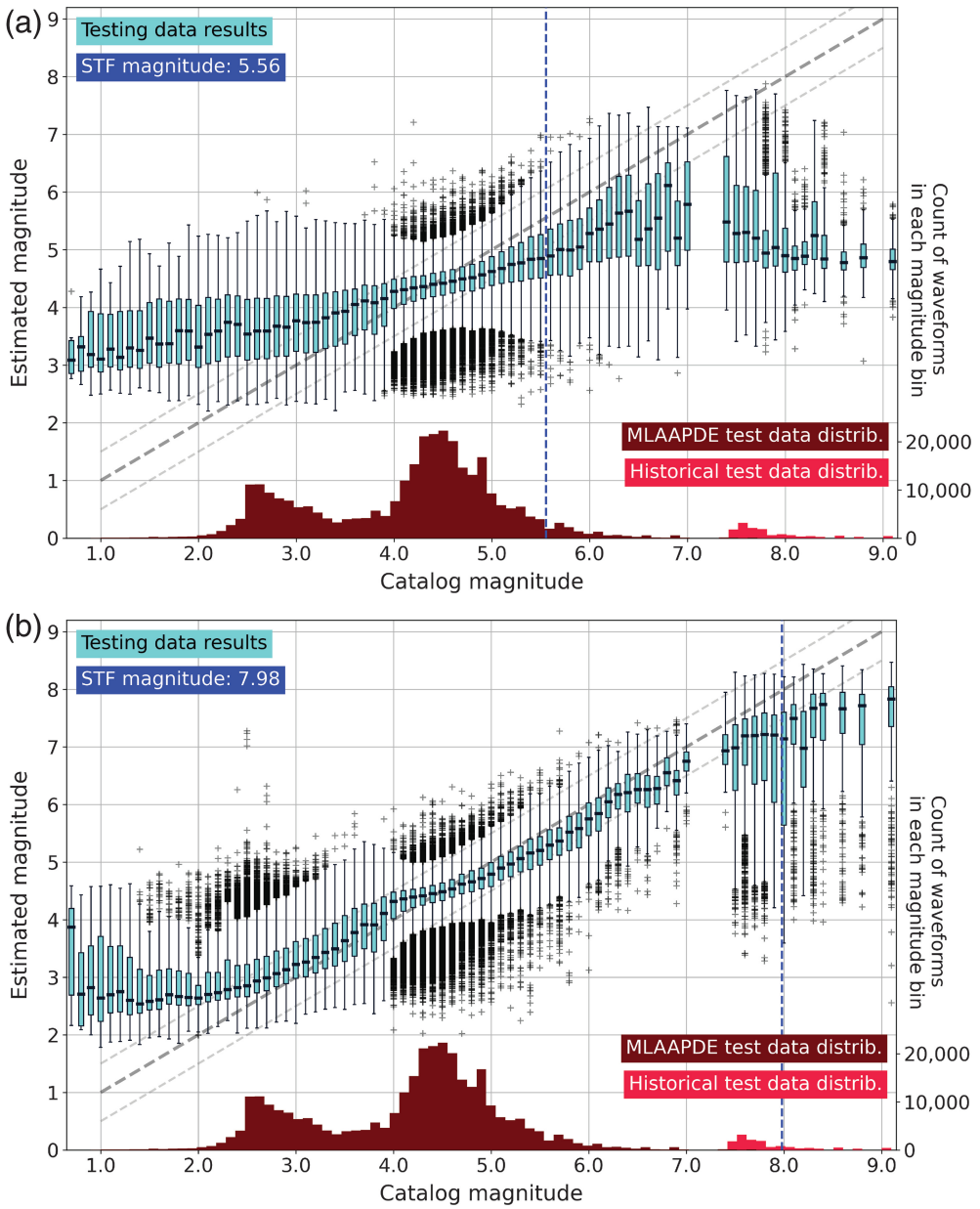

Figure 3. Boxplots plotted using Matplotlib (Hunter, 2007) showing the model's estimated earthquake magnitudes for the testing data set (both MLAAPDE and historical data) for two example time windows, with the catalog magnitude for the earthquakes on the *x* axis, and the model's estimated magnitudes for those same earthquakes on the left *y* axis. An animation showing the boxplots for all time windows can be found in Video S1. (a) Model magnitude estimations for data with a waveform window length of 7 s. (b) Model magnitude estimations for data with the longest tested waveform window length of 114 s. In each of these panels, the median magnitude estimation for each catalog magnitude bin on the *x* axis is shown by the horizontal black line in the center of each light blue box. The lower and upper bounds of the light blue boxes indicate the first and third quartiles of the estimations in each bin, respectively, and the whiskers extend to 1.5 times the interquartile range. The individual points beyond the whiskers indicate outliers. The diagonal bold gray dashed line represents a one-to-one magnitude estimation line, with the lighter gray dashed lines above and below it ± half a magnitude unit. The vertical blue dashed line indicates the magnitude of a theoretical earthquake whose source time function (STF) would entirely fit within half of the window length for that particular plot, as the *P* wave should arrive at approximately halfway through the window. Finally, the red histogram at the bottom of each panel shows the magnitude distribution for the testing data set, with the testing data from the MLAAPDE data set in the year 2020 in dark red, and the historical testing data set in light red. A scale for the counts for each of the histogram bins is shown on the right *y* axis.

RESULTS AND DISCUSSION

Model performance by magnitude

After training, our machine learning model (which will be referred to as Artificial Intelligence Magnitude [AIMag]) was used to make estimations on the previously unseen testing data set. The testing process was, like the training, repeated with each of the data sets of different window length for a total of 26 iterations.

Because most of the earthquakes in the data set have observations from multiple stations, and AIMag makes magnitude estimations for individual stations, there are multiple estimated magnitudes for each earthquake in the catalog, and the statistics for these estimations can be more easily visualized using a boxplot plotting method. Results shown as boxplots are included for two selected iterations in Figure 3, for which Figure 3a shows the results for the data with a shorter window length of 7 s, and Figure 3b shows the results for the longest window length of 114 s. The full set of figures for all window lengths can be found in Video S1. As with the default waveforms, the phase arrivals are approximately in the center of these waveforms with the random shift, so approximately half of the window length of these waveforms are noise and half are signal.

By examining the changes between Figure 3a,b, it is possible to observe that as the window length increases, the median estimated magnitudes and their corresponding blue boxes converge toward a one-to-one relationship with the catalog magnitudes, as

indicated by the diagonal bold gray dashed line. The lighter gray dashed lines above and below the bold gray line show half a magnitude above and below the one-to-one relationship.

At the longest window length of 114 s as shown in Figure 3b, the smallest earthquake magnitudes are overestimated by AIMag, and the largest earthquake magnitudes are underestimated. Some of this effect is likely attributable to limited data availability at the lowest and highest magnitudes, as shown in the histograms in each panel of Figure 3. The most populated magnitude bins in the catalog are around **M** 2.5 and 4.5. This is because the NEIC's PDE catalog publication criteria include earthquakes **M** 2.5 and larger or felt in the United States (excluding California, which is published for **M** 3.0 and larger), and M 4.0 and larger or felt in the rest of the world, resulting in a catalog that is complete to about **M** 4.5 globally. Therefore, fewer earthquakes **M** < 2.5 are in the catalog, and naturally fewer high-magnitude earthquakes are included because of the Gutenberg–Richter law (Gutenberg and Richter, 1949). This data paucity in training likely affects the model's ability to accurately assess the magnitudes of the smallest and largest earthquakes. Where data are sparse, the model tends to hedge the magnitude estimates to be more toward the mean of the magnitudes in the data set (4.4 for the training data set, 4.2 for the validation data set, and 4.3 for the testing data set). For example, at the 1 s time window (Video S1) the median estimated magnitudes are all close to 4.3. As the time window grows, the estimates shift to fall more along the one-to-one line, but they never move low enough for the magnitudes in the lowest ranges—likely due to the undersampling.

Another feature of these plots is the vertical darker blue dashed line, which moves to the right as the window length increases from Figure 3a to Figure 3b. This line indicates the magnitude of a theoretical earthquake for which the source time function would entirely fit within half of the window length for that particular plot (i.e., the length of earthquake signal in the window). This theoretical magnitude was calculated by dividing the window length by two and using that as the total rupture duration (τ_t) in the following equation:

$$M_0 = \tau_t^3 \times 0.625 \times 10^{23}, \quad (1)$$

in which M_0 gives the seismic moment in dyn·cm and τ_t is the total rupture duration. Singh *et al.* (2000) gave a range of 0.25×10^{23} to 1.0×10^{23} dyn · cm/s^3 for the value of M_0/τ_t^3 for shallow earthquakes, so the midpoint of this range was used for the approximation of earthquake moment in this study. Using that calculated M_0, we then converted this to the moment magnitude M_w following Kanamori (1977) using the following equation:

$$M_w = \frac{2}{3}\log_{10}(M_0) - 10.73, \quad (2)$$

in which M_0 again gives the seismic moment in dyn · cm. This M_w is what was plotted as the vertical dark blue dashed line in Figure 3. We expected that AIMag would be more accurately able to estimate the magnitude of an earthquake that can be entirely seen within the window of data provided (i.e., for which the magnitude bin is to the left of the blue dashed line), and that it would struggle with earthquakes larger than this (where the magnitude bin is to the right of the line). This effect can be seen most clearly in Video S1, in which all tested time windows are shown.

At the longest time window of 114 s (Fig. 3b), the theoretical earthquake magnitude is ~8, and it is true that the earthquakes above this magnitude are underestimated by AIMag. However, some of the larger earthquakes still to the left of the line are also underestimated. In this higher magnitude range, the model is likely limited by the small number of waveforms from large magnitude earthquakes in the training data set. The largest earthquake in the training data set is **M** 8.3, and the entire training data set only contains 31 earthquakes **M** 7.5 and greater for a total of 4486 waveforms. Compared to the 170,553 training waveforms available for just **M** 4.5 alone, for example, this is a very limited sample from which AIMag can learn. Future work could explore the possibility of including synthetic large magnitude earthquakes in the training data set to improve the model's exposure to these larger events during training. This could be achieved using codes such as FakeQuakes (Melgar *et al.*, 2016), which is capable of generating synthetic teleseismic data. Another potential avenue for increasing the proportion of large earthquakes in the data set would be to upsample the existing ones, as was done in Münchmeyer *et al.* (2021). We also explored the possibility of improving the performance of the models at the extreme ends of the magnitude scale by implementing weighting into the training based on the relative abundance of samples in the data set by magnitude. We split the data set into bins by 0.5 magnitude intervals and defined the weight for each bin as 1—(samples in bin divided by the total number of samples in the training data set). This way, a magnitude bin with many samples would have a small weight in the training process, and a bin with few samples would have a large weight. These weights were then applied to the waveform samples as they were processed through the data generator in the training process. We found slight improvements for some time windows with this method (Video S2; Fig. S2), which indicates that for future work exploring a more sophisticated weighting scheme may lead to improved results.

We included the historical catalog of larger earthquakes in the testing data set to determine whether AIMag was capable of accurately estimating their magnitudes, despite not being trained on events **M** > 8.3. Although the magnitudes are underestimated, the model can at least determine that they are greater than **M** 7 in many cases at the longest window length of data, shown by the positions of the medians for those magnitude bins in Figure 3b. This result makes AIMag useful for flagging large magnitude earthquake detections in monitoring operations.

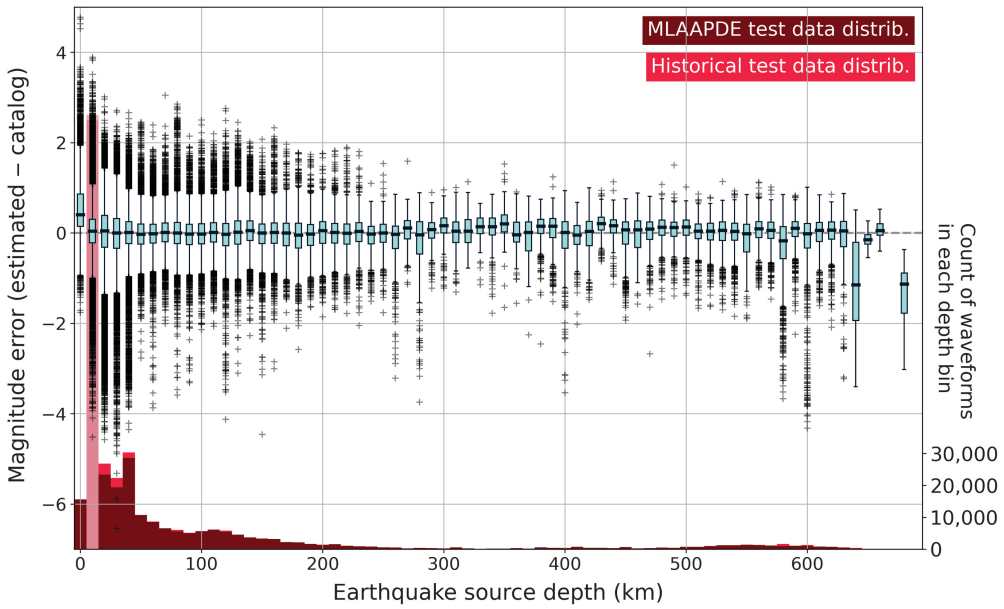

Figure 4. A boxplot showing the error in the magnitude estimation (estimated magnitude–catalog magnitude) made by the model on the left *y* axis plotted against the source depth of the earthquake in kilometers on the *x* axis, corresponding to each estimation for the 114 s time window. An animation showing the boxplots for all time windows can be found in Video S4. The characteristics of the boxplot in this figure are the same as those for Figure 3, with the median magnitude estimation error for each earthquake source depth bin shown by the horizontal black line in the center of each light blue box. On the right *y* axis, we have plotted a histogram showing the distribution of earthquake source depths corresponding to the waveforms in the testing data set. The distribution of source depths in the MLAAPDE data set is shown in dark red, with the distribution of source depths in the historical data set stacked on top of those in light red. The transparency of the second (10 km) source depth histogram bin was reduced to ease visualization of the overlaid boxplot data.

Another potential reason that the highest magnitude earthquakes are underestimated by AIMag is because of the complications of magnitude saturation in seismic data. Magnitudes calculated using the *P* waves in seismic waveforms tend to saturate at the high end of the scale (i.e., large earthquakes become indistinguishable from even larger ones). The relationship between log peak ground displacement and magnitude is linear for moderate magnitude earthquakes, but transitions to a scaling that is independent of magnitude at larger magnitudes (with this transition point moving to higher magnitudes as the data window length increases; refer to Trugman *et al.*, 2019 their fig. 4). In addition, traditional seismic data processing requires high-pass filtering to remove baseline drift effects that appear when data are double integrated to displacement, and this filtering removes low-frequency content that may change the relationship between measured displacement and earthquake magnitude (Colombelli *et al.*, 2012). In this study, although we do not remove the instrument response, integrate, or filter any data before training AIMag on it, by virtue of the diversity of the NEIC's contributor networks we are using a variety of different types of instruments, some of which likely have a smaller range of periods at which they have a flat response. This means that some longer period information may be suppressed, which could contribute to the model's underestimations at high magnitudes. In future studies, performance could potentially be improved for large earthquakes by the use or inclusion of displacement data as an input in the training process. Machine learning methods have previously been shown to be effective for real-time earthquake magnitude estimation using GNSS displacement data (e.g., Lin *et al.*, 2021).

Although this article focuses on AIMag's single-station magnitude estimation results, in operational practice, magnitudes are typically determined and reported by averaging the magnitudes calculated from many different stations. We calculated event-averaged magnitudes from the single-station results described previously (Video S3) and observed that this approach had the expected effect of suppressing many of the outlier single-station magnitudes. This effect could also be seen when the standard deviations of the magnitude estimation errors for the event-averaged magnitude method were compared to the standard deviations of the errors for the single-station method (Fig. S3), as the standard deviations for the event-averaged magnitude method were lower.

Model performance by waveform and earthquake characteristics

We were interested in determining whether AIMag had any biases toward particular characteristics of earthquakes or waveforms. We analyzed the performance of the model by earthquake source depth, hypocentral distance (the distance between the hypocenter and the receiver station), and the SNR of the waveforms. For Figures 4–6 as discussed in the following section, we calculated the magnitude estimation error by subtracting the model's estimated magnitude from the true catalog magnitude.

In Figure 4, we plotted the magnitude estimation errors by the source depth of the earthquakes in the testing data set. Only the plot for the 114-s window is shown, as the results do not change substantially across time windows for this parameter. The remaining plots can be seen in an animation in Video S4. For this figure, we have also plotted a histogram showing the distribution of earthquake source depths corresponding to the

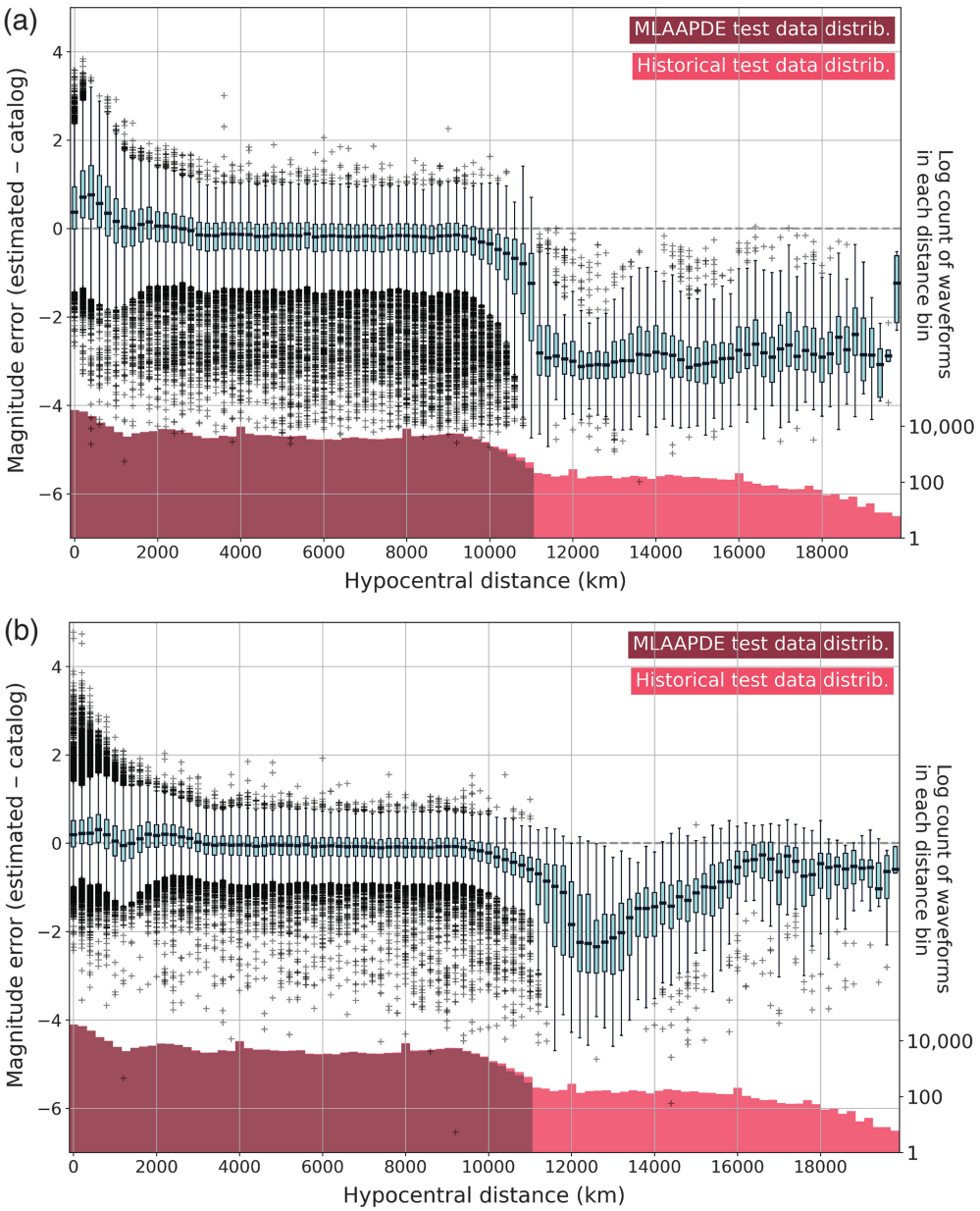

Figure 5. Boxplots for (a) the 7 s window and (b) the 114 s window showing the error in the magnitude estimation (estimated magnitude–catalog magnitude) made by the model on the left *y* axis plotted against the hypocentral distance in kilometers on the *x* axis corresponding to each estimation. An animation showing the boxplots for all time windows can be found in Video S5. The characteristics of the boxplot in this figure are the same as those for Figures 3 and 4, with the median magnitude estimation error for each hypocentral distance bin shown by the horizontal black line in the center of each light blue box. On the right *y* axis, we have plotted a logarithmic histogram showing the distribution of hypocentral distances corresponding to the waveforms in the testing data set. The distribution of hypocentral distances in the MLAAPDE dataset is shown in dark red, with the distribution of hypocentral distances in the historical data set stacked on top of those in light red. The MLAAPDE data set in the format used in this study only contains waveforms with hypocentral distances of ∼100° or less, so the only waveforms with further hypocentral distances come from the historical data set.

waveforms in the testing data set. The very populous second bin is as such because it encompasses the 10-km-depth range, and 10 km is the default depth assigned by the NEIC when a shallow earthquake's depth cannot be precisely constrained. In general, Figure 4 shows that AIMag is not biased in its estimations by the depth of the source earthquake, with the median error for most catalog magnitude bins being near zero. It only shows more variability in error at the far end of the depth range, as very few 650 km and greater deep earthquakes are available either for training or testing.

We next looked at the magnitude estimation error by the hypocentral distance for each waveform. These results can be found in Figure 5, in which boxplots for the 7 s window (Fig. 5a) and 114 s window (Fig. 5b) are shown. The remaining plots for all other time windows can be seen in the animation in Video S5. We have also plotted in these panels a logarithmic histogram showing the distribution of hypocentral distances corresponding to the waveforms in the testing data set. The MLAAPDE data set only contains waveforms with maximum hypocentral distances of ∼100° (or ∼11,110 km), so the only waveforms with further hypocentral distances come from the historical data set. This is why we elected to use a logarithmic histogram for this plot—the relatively smaller size of the historical data set made the histogram bars nearly disappear with a nonlogarithmic plot.

In comparing Figure 5a to Figure 5b across the time window range, we observe that the main difference in performance is with the far hypocentral distances greater than 10,000 km where only the historical data exist. In Figure 5a with a 7 s window, AIMag greatly underestimates the magnitudes of the earthquakes in the waveforms with large hypocentral distances, and in Figure 5b at the 114 s window length, the performance is improved (although still underestimated). We would expect

AIMag to perform poorly at these far distances because the training data set includes no waveforms with hypocentral distances over ∼11,110 km, and because only a small portion of the *P* wave will exist in waveforms with such large hypocentral distances. Of note, however, the model with a longer window length of data was able to estimate the magnitudes more closely, and it performs better at the very longest distance range of 16,000 km and greater compared to a more intermediate range of 11,000–16,000 km. From Figure 5, we can generally conclude that AIMag, when tested with data with a similar hypocentral distance range to that on which it was trained, is not substantially biased by hypocentral distance. We therefore recommend using this model as currently trained only for hypocentral distances of ∼100° or less.

Finally, we examined the performance of AIMag by the SNR in decibels (dB) of the waveform. We calculated waveform SNRs for the MLAAPDE and historical data sets according to the following equation:

$$\mathrm{SNR_{dB}} = 20 \times \log_{10}\left(\frac{S_{\mathrm{rms}}}{N_{\mathrm{rms}}}\right), \tag{3}$$

in which S_{rms} is the root mean square of the earthquake signal time series from the *P*-wave arrival to 30 s after, and N_{rms} is the root mean square of the background noise from 60 s before the *P*-wave arrival up until 1 s before its arrival (for a total of 59 s of noise).

Figure 6 shows boxplots for the magnitude estimation error plotted against the waveform SNR corresponding to each estimation for the 7 s window (Fig. 6a) and the 114 s window (Fig. 6b). The remaining plots for all other time windows can be seen in the animation in Video S6. We have again also included a logarithmic histogram showing the distribution of SNRs corresponding to the waveforms in the testing data set. We observe that at the full waveform window length of 114 s in Figure 6b, the median errors have converged around zero, indicating that there is not a bias in AIMag's performance by SNR. However, with the 7 s window in Figure 6a, the model has more spread in the estimations and appears to underestimate the magnitudes of waveforms with an SNR above ∼10 dB, but it performs better below this range.

This range of underestimations aligns with where the historical data are present in greatest number, as seen in the histogram in Figure 6a. To verify that the data type was causing this bias above ∼10 dB, we plotted the magnitude error versus SNR for the MLAAPDE and historical testing waveform data sets separately (Fig. S4). This showed that the underestimations seen in Figure 6 came primarily from the historical waveforms because the MLAAPDE waveforms have median estimation errors of around zero even at 1 s window lengths. It is, therefore, likely that some properties of the historical waveforms that were not present in the MLAAPDE waveforms led to the magnitude underestimations by AIMag at short time windows, as the model was not trained on these types of waveforms. In Figure 6a, at the short 7 s time window, the model appears to perform better below the ∼10 dB range because there are many more MLAAPDE data samples in those bins, and the MLAAPDE data magnitudes do not tend to be underestimated by AIMag (Fig. S4b). In the higher SNR bins where underestimation occurs, a larger proportion of the samples come from the historical data that were not included in the training data set. At longer time windows (Fig. 6b), this gap closes, as the magnitude errors are lower for the historical data than at shorter time windows (Fig. S4c). However, AIMag still underestimates the magnitudes of earthquakes from the historical testing data set from waveforms with lower SNRs, even at the longest 114 s time window.

Model performance by window length

To assess the performance of AIMag by window length, for each time window we calculated the error (estimated magnitude–catalog magnitude) for each model estimation, followed by the median and standard deviation of those errors. We then plotted the median and standard deviation of the errors for each time window (Fig. 7). Generally, the median error is close to zero across the entire range of tested time windows, but the standard deviation of the errors tends to decrease from a maximum of 1.18 at the 1 s time window to a minimum of 0.55 at the 114 s time window. The standard deviation decreases most quickly at the smaller time windows through 14 s, and decreases more slowly after that.

Because the potential utility of this study is for real-time earthquake analysis, we were interested in investigating the trade-off between estimation accuracy and speed. The estimation speed will be limited by how long of a waveform AIMag needs to make the estimation, which is why we iterated our training and testing over a range of time windows. To assess this trade-off, we took a closer look at the estimation errors within each time window using the median (50th percentile) estimated magnitude from the boxplots for each time window as shown in Figure 3 and Video S1. Because AIMag's performance varies greatly based on earthquake magnitude and our interest is primarily on larger events, we assessed the quality of our models by disaggregating their performance by event magnitude. We set our estimation error tolerance to 0.5 (i.e., if AIMag's median estimation for a particular catalog magnitude bin is within 0.5 of the correct catalog magnitude, we call this an acceptable estimation). We could then plot the median errors across every time window to determine what the acceptable estimation range was for each time window. Figure 8 shows that by the time the window length reaches 29 s, the high-magnitude range by which we can achieve an acceptable estimation error performance is already essentially maximized, with limited improvements above this window. For the 29 s window, magnitude estimate median errors are acceptable between **M** 2.7 and 7.6, excluding **M** 7.5 for which the errors

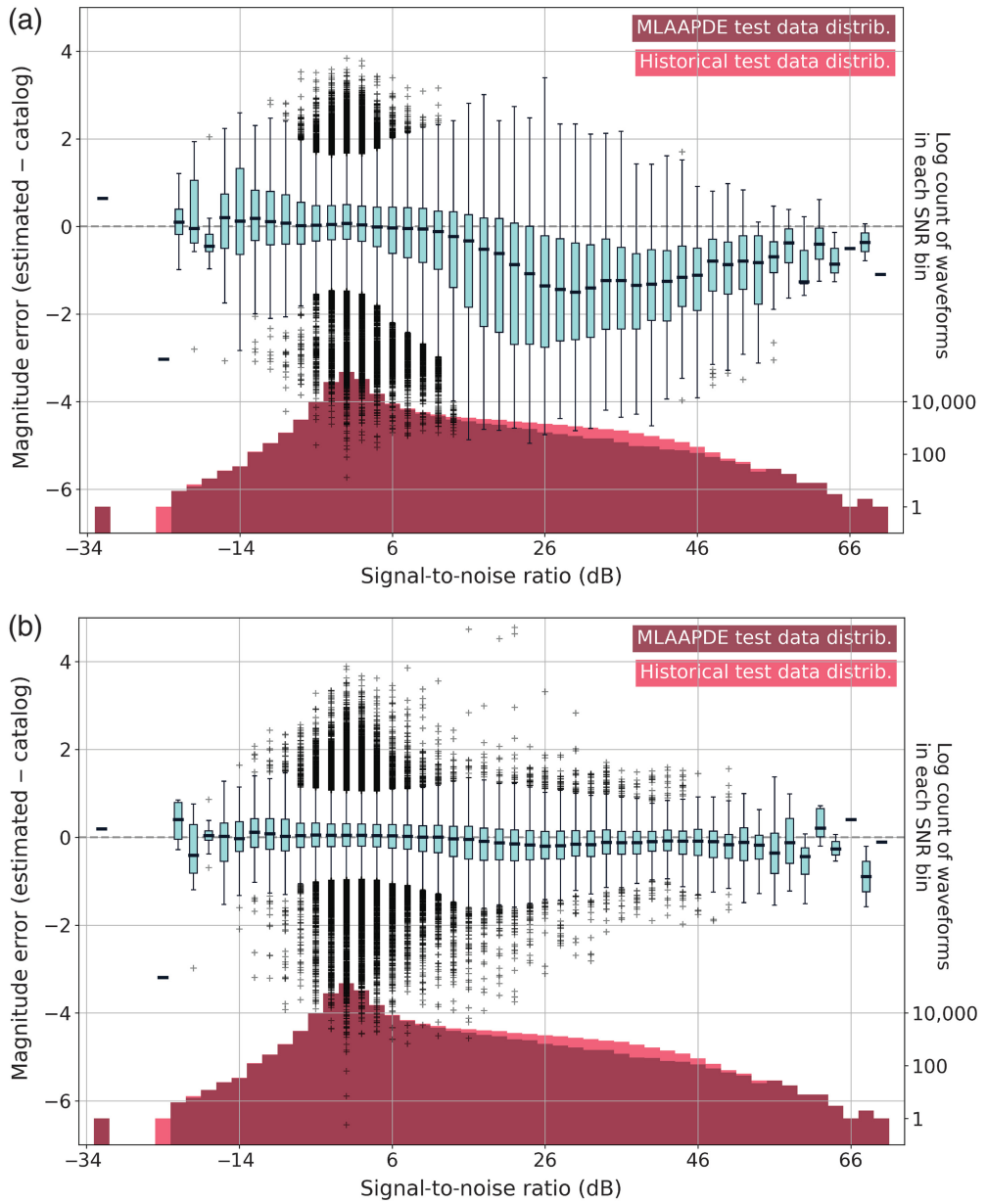

Figure 6. Boxplots for (a) the 7 s window and (b) the 114 s window showing the error in the magnitude estimation (estimated magnitude–catalog magnitude) made by the model on the left *y* axis plotted against the waveform signal-to-noise ratio (SNR) in decibels (dB) on the *x* axis corresponding to each estimation. An animation showing the boxplots for all time windows can be found in Video S6. The characteristics of the boxplot in this figure are the same as those for Figures 3–5, with the median magnitude estimation error for each SNR bin shown by the horizontal black line in the center of each light blue box. On the right *y* axis, we have plotted a logarithmic histogram showing the distribution of SNRs corresponding to the waveforms in the testing data set. The distribution of SNRs in the MLAAPDE data set is shown in dark red, with the distribution of SNRs in the historical data set stacked on top of those in light red.

are slightly too high. For the longest 114 s window, magnitude estimate median errors are acceptable between **M** 2.3 and 7.6, also excluding **M** 7.5 for the same reason.

We also calculated and plotted the absolute magnitude estimation errors for the 95th percentile of the data (Fig. S5) and found that the range over which the absolute error was less than or equal to 0.5 magnitude units extends from **M** 4.4 to 8.6 for the 114 s window. For the 29 s window discussed in the previous paragraph, the 95th percentile absolute error is less than or equal to 0.5 for the magnitude range **M** 4.4–8.3, excluding **M** 8.2 for which the errors are slightly too high. Using the 95th percentile to choose our acceptable performance range instead of the median (50th percentile) would exclude the data below **M** 4.4, as they tend to be overestimated more often. However, it would also result in including the ∼**M** 7.5 to 8.6 data that we know tend to be underestimated. Which metric to use for determining an acceptable operating magnitude range over which to implement this model is likely best left as a user choice based on the monitoring goal.

As discussed in the Model performance by magnitude section, AIMag tends to underestimate the magnitudes of large earthquakes, likely primarily due to the short time windows we use. In tradition, the magnitudes of large earthquakes are characterized with long-period observations. The *W*-phase is particularly important—it is a long-period (∼100–1000 s) phase that arrives in displacement seismograms between the *P* and *S* waves and represents a superposition of other body-wave phases (Kanamori, 1993; Kanamori and Rivera, 2008). *W*-phase moment magnitudes (M_{ww}) provide accurate measures of earthquake size that work for moderate (Hayes *et al.*, 2009) as well as great earthquakes (e.g., the 2011 Tōhoku earthquake as demonstrated in Duputel *et al.*, 2011). M_{ww} can be calculated relatively rapidly and is the preferred magnitude used by the NEIC.

In this study, we used raw waveforms (stream-normalized in units of counts) rather than displacement waveforms and

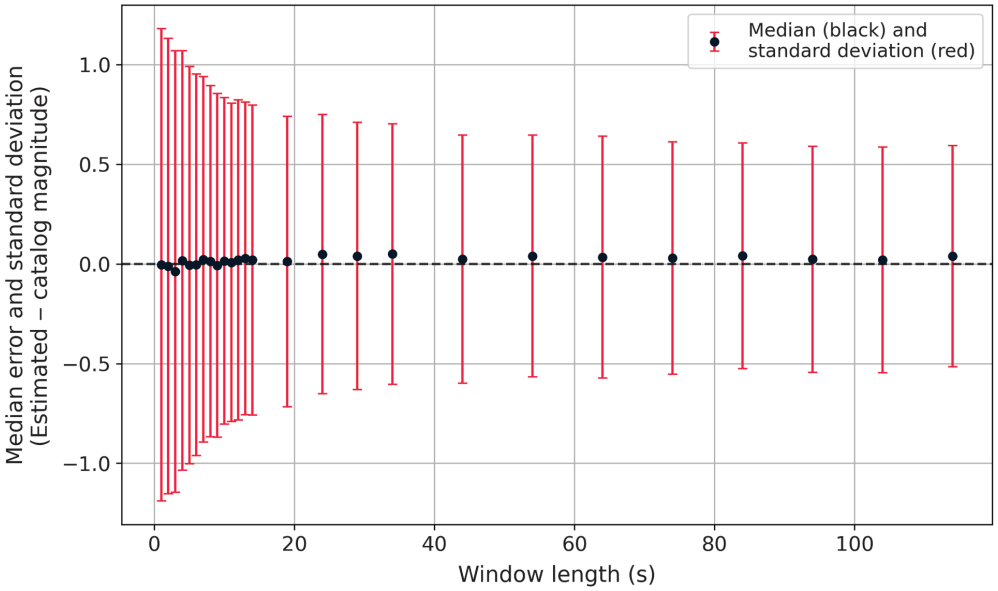

Figure 7. The median and standard deviation of the model's estimation error on the testing data set (estimated magnitude–catalog magnitude) plotted for each of the tested time windows. The black dots show the median error for each time window, and the red bars extend to one standard deviation above the median and one standard deviation below.

window lengths only up to 114 s (only half of which contains earthquake signal). The *W*-phase is, therefore, not fully captured in the observations because it will not have arrived in its entirety yet. It is possible that training and testing with longer waveforms that include the *S* wave (and therefore the entire *W*-phase) or data scaling that enhances long period data would improve AIMag's performance on large earthquakes. The drawback of that approach would be an increase in the complexity of data preprocessing and potentially the amount of time required to obtain a good magnitude estimate. For comparison, how fast M_{ww} can be calculated depends on earthquake location and station density, and in a best-case scenario the NEIC can calculate M_{ww} in as few as ~7 min and typically within about 20 min (Hayes *et al.*, 2009). This is much longer than the 29 s time window beyond which we saw limited returns in this study (Fig. 8). Although that 29 s does not account for the time it takes for the signal to be picked up by the seismometer, as ours is a single-station method, the additional delay time only depends on the travel time to the single closest station to the hypocenter.

A closer analog to our magnitude estimation method is likely that of the broadband *P*-wave moment magnitude (M_{wp}). Similar to our model, M_{wp} only relies on the *P* waves from earthquakes, but similar to M_{ww} it is also an inversion method that takes integrated displacement waveforms. It performs well on as little as 120 s of waveform data, but it is of limited use for earthquakes smaller than **M** ~ 5.5 when long period signals are weaker and SNR is low because it may overestimate the magnitude (Tsuboi *et al.*, 1999). In addition, when data are used from seismometers that are less sensitive to longer periods, the M_{wp} method underestimates the magnitudes of large earthquakes (**M** ~ 8 and greater). It is also less reliable for data from stations in the near field because the method was based on far field *P*-waveform displacement, and this leads to underestimations of the earthquake magnitude (Hirshorn *et al.*, 2013).

We have shown that although the magnitudes of larger earthquakes may be underestimated by AIMag, similar to the M_{wp} method, our model can still tell that they are large (Fig. 3). In addition, it also works for much smaller magnitude earthquakes than M_{wp} (Fig. 3), performs well in the near field (Fig. 5b), and can provide estimates more quickly than M_{ww}. AIMag can feasibly be useful for flagging earthquakes that a monitoring agency would be concerned are large and potentially tsunamigenic within a much shorter time window than is currently possible with the M_{ww} inversion method and without some of the drawbacks of the M_{wp} method.

Model performance on edge cases

With a model designed to generalize across the globe, we wanted to check AIMag's performance for specific earthquakes with several unusual characteristics. We browsed the testing data catalog for more atypical earthquakes, both from the historical data recorded between 2000 and 2013 and the MLAAPDE data recorded in the year 2020, and hand-picked seven events to examine more closely.

This group of seven included two tsunami earthquakes, which are earthquakes that produce tsunamis that are much larger than expected based on their magnitude (Kanamori, 1972). They are typically depleted in high-frequency seismic radiation, and likely rupture all the way to the trench at slow rupture velocities (Lay *et al.*, 2012). The two earthquakes we tested were the 2006 M 7.7 Pangandaran earthquake off the coast of Java, Indonesia (Ammon *et al.*, 2006) and the 2010 **M** 7.8 Mentawai earthquake off the coast of Sumatra, Indonesia (Yue *et al.*, 2014). We also looked at three unusually deep earthquakes because most of the waveforms come from shallow events as seen in the histogram in Figure 4. These were a 598-km-deep **M** 8.3, a 624-km-deep **M** 6.9, and a 641-km-deep **M** 7.5 earthquake. The final two earthquakes in the group of seven were unusually "complex" events to determine how the

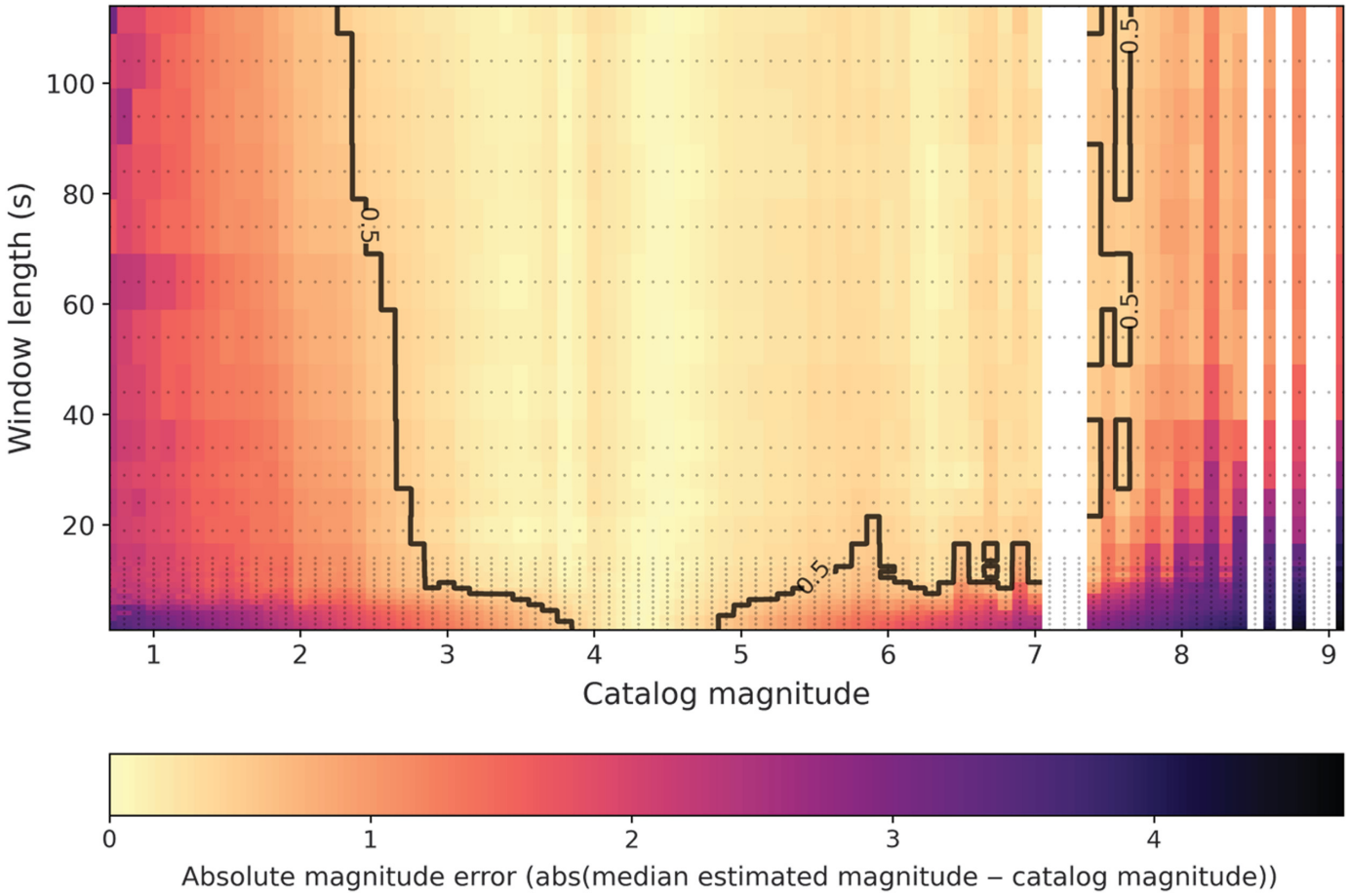

Figure 8. Plot showing the absolute error of the median of the estimations made by the model for each catalog magnitude bin (abs(median_$M_{estimated}$–$M_{catalog}$)) across the range of window lengths tested (1–114 s) for the testing data set. The catalog magnitude is on the *x* axis and the waveform window lengths are on the *y* axis. The color spectrum shows the absolute error, with lighter colors indicating a smaller error and darker colors indicating a larger error. White gaps indicate magnitudes for which there were no earthquakes in the testing data set. The black line is the contour of an absolute error of 0.5 magnitude units.

model could handle this. We chose the 2012 **M** 8.6 Wharton basin earthquake off the coast of Sumatra, which is the largest strike-slip earthquake ever recorded and ruptured multiple faults (Meng *et al.*, 2012; Yue *et al.*, 2012), and the 2020 **M** 7.6 Sand Point earthquake on the Aleutian subduction zone, which generated an unusually large tsunami for the reported strike-slip mechanism of the event (Herman and Furlong, 2021; Santellanes *et al.*, 2022).

We again produced a boxplot of AIMag's estimated earthquake magnitudes versus the catalog magnitudes for the testing data set for the 114 s window (Fig. 9), but in this figure we plotted the seven selected edge case earthquakes in light red and all other earthquakes in light blue. All of the estimations for these edge case events, with the exception of the 641-km-deep event (marked B in Fig. 9), fall within the typical range of estimations for the other earthquakes in the data set for their respective magnitude bins, indicating that in many cases the model performs well even with unusual earthquakes.

The underestimated 641-km-deep **M** 7.5 (B) is one of a series of three deep earthquakes that occurred west of the Philippines on 23 July 2010 (Hayes *et al.*, 2017). With magnitudes of **M** 7.3, 7.5, and 7.6, the first of this series was not included in the expanded historical testing data set because we only downloaded data for earthquakes **M** 7.5 and greater. To explore why AIMag may have underestimated the magnitude of the 641-km-deep event, we also pulled out and plotted the estimations made for the **M** 7.6 earthquake that occurred at a depth of 578 km in this same area. These two earthquakes can be seen in Figure S6. Although the **M** 7.6 is not underestimated as much as the **M** 7.5, it is still underestimated compared to the other edge cases visualized in Figure 9. This suggests that the underestimation for these specific events could be related to the tectonic setting in this area, or potentially other spatial effects such as the geometry of the seismic stations that recorded the waveforms from these events.

CONCLUSIONS

In this study, we trained a machine learning model for global single-station earthquake magnitude estimation based solely on the recorded waveform and requiring no additional information

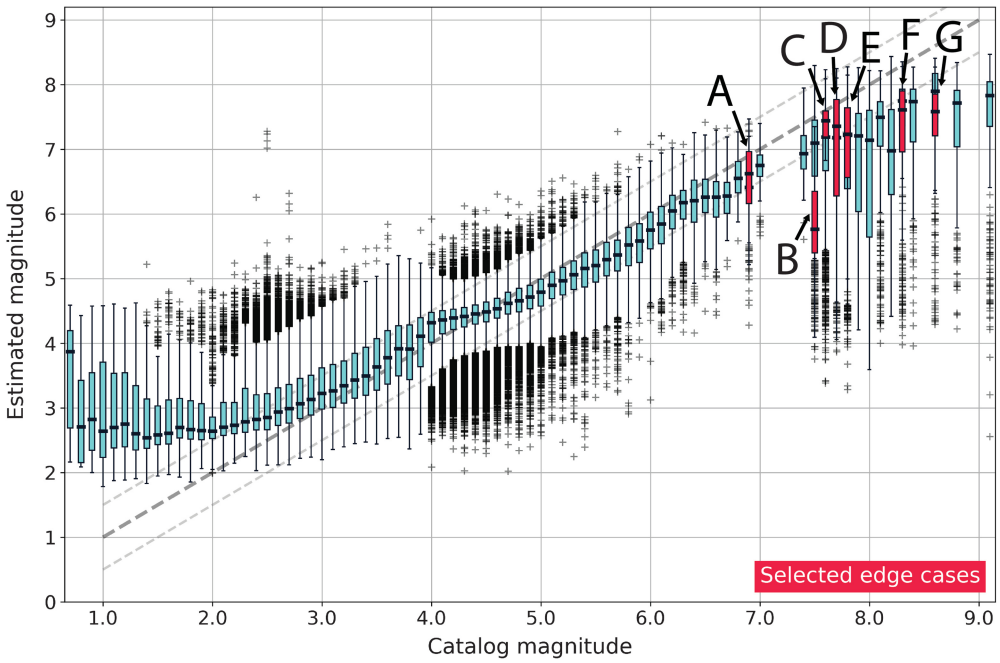

Figure 9. Plot of the model's estimated earthquake magnitudes for the testing data set (both MLAAPDE data and historical) for the 114 s window (similar to Fig. 3b), with the catalog magnitude for the earthquakes on the *x* axis, and the model's estimated magnitudes for those same earthquakes on the left *y* axis. The seven selected edge case earthquakes as described in the body of the article are shown in light red, whereas all other earthquakes are plotted in light blue. The earthquake marked A on the plot is the 624-km-deep event, B is the 641-km-deep event, C is the Sand Point complex earthquake, D is the Pangandaran tsunami earthquake, E is the Mentawai tsunami earthquake, F is the 598-km-deep event, and G is the Wharton basin complex earthquake.

about the event or station involved. We demonstrate here the effectiveness and utility of the MLAAPDE data set described in Cole *et al.* (2023), which was designed to make the USGS NEIC's PDE catalog more easily accessible for machine learning research. Our training process implements data augmentation to improve and generalize model performance.

At its best, our model, AIMag, can successfully determine earthquake magnitudes with an absolute error margin in median estimated magnitude of 0.5 magnitude units or less for a range of earthquake magnitudes from **M** 2.3 to 7.6. It can achieve success across this magnitude range with a minimum window length of 29 s of data (with the phase arrival at halfway through this window). AIMag is not biased in its magnitude estimations by either earthquake source depth, hypocentral distance, or waveform SNR, and it also performs well on most tested "edge case" earthquakes that are unique in one aspect or another (e.g., tsunami earthquakes, deep earthquakes, and complex ruptures).

AIMag can aid near-real-time global earthquake monitoring operations by quickly and accurately assessing the magnitudes of earthquake recordings by avoiding reliance on earthquake location. Future work may include testing the performance of this model within the NEIC's real-time monitoring system and expanding its capability to handle variable-length input data in real time, as well as combining it with other work on a machine-learning-based pick association method in development at the NEIC.

DATA AND RESOURCES

The codes used to train and test the models and construct the figures described in this article are available at https://github.com/UO-Geophysics/AIMag. As of the time of publication, this repository was last updated in December 2023. The Machine Learning Asset Aggregation of the Preliminary Determination of Epicenters (MLAAPDE) module used in this work is available in Cole and Yeck (2022). The facilities of EarthScope Consortium were used for access to waveforms, related metadata, and/or derived products used in this study. These services are funded through the Seismological Facility for the Advancement of Geoscience (SAGE) Award of the National Science Foundation (NSF) under Cooperative Support Agreement EAR-1851048. We make use of data from the following networks: AC, AE, AF, AG, AI, AK, AT, AU, AV, AY, AZ, BC, BE, BK, BL, BX, C, C0, C1, CA, CC, CH, CI, CM, CN, CO, CW, CY, CZ, DK, DR, EC, EO, EP, ET, G, GB, GE, GM, GO, GR, GS, GT, HK, HL, HT, HV, IE, II, IM, IN, IO, IU, JP, KC, KG, KN, KO, KR, KS, KY, KZ, LB, LD, LO, LX, MB, MG, MI, MM, MN, MP, MU, MX, MY, N4, NA, NC, NI, NM, NN, NP, NQ, NU, NV, NY, NZ, O2, OE, OH, OK, OO, OV, OX, PA, PE, PL, PM, PO, PR, PS, PT, RM, RO, RV, SB, SC, SE, SN, SS, SV, TA, TC, TJ, TM, TW, TX, UO, US, UU, UW, VU, WC, WI, WM, WU, WY, YX, ZC, ZD, and ZW. We appreciate the hard work of the engineers and network operators who make this data available in near real-time. We have included the International Federation of Digital Seismograph Network (FDSN) references in the supplemental material. The supplemental material for this article contains six figures as well as six multiframe figure animations to provide extra information about the data sets used in training and testing, as well as additional testing results for all time windows as described in this article.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

This work was conducted as part of a U.S. Geological Survey (USGS) Pathways Internship at the National Earthquake Information Center (NEIC) in Golden, Colorado. It was also supported by National

Science Foundation (NSF) Grant Number OAC-1835661 and National Aeronautics and Space Administration (NASA) Grant Numbers 80NSSC19K0360 and 80NSSC22K0458. The authors thank two anonymous reviewers, Adam Ringler, Brian Shiro, and two additional internal USGS reviewers for their careful reading and feedback, which helped improve this article.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES

Abadi, M., A. Ashish, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.* (2015). TensorFlow: Large-scale machine learning on heterogeneous systems, 19 pp., doi: 10.5281/zenodo.4724125, software available from tensorflow.org (last accessed November 2023).

Ammon, C. J., H. Kanamori, T. Lay, and A. A. Velasco (2006). The 17 July 2006 Java tsunami earthquake, *Geophys. Res. Lett.* **33,** no. 24, L24308, doi: 10.1029/2006GL028005.

Benz, H. (2017). Building a National Seismic Monitoring Center: NEIC from 2000 to the present, *Seismol. Res. Lett.* **88,** no. 2B, 457–461, doi: 10.1785/0220170034.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81,** no. 3, 530–533, doi: 10.1785/gssrl.81.3.530.

Cole, H. M., and W. L. Yeck (2022). Global earthquake machine learning dataset: Machine Learning Asset Aggregation of the PDE (MLAAPDE), *U.S. Geol. Surv. Data Release*, doi: 10.5066/P96FABIB.

Cole, H. M, W. L. Yeck, and H. M. Benz (2023). MLAAPDE: A machine learning dataset for determining global earthquake source parameters, *Seismol. Res. Lett.* **94,** no. 5, 2489–2499, doi: 10.1785/0220230021.

Colombelli, S., A. Zollo, G. Festa, and H. Kanamori (2012). Early magnitude and potential damage zone estimates for the great Mw 9 Tohoku-Oki earthquake, *Geophys. Res. Lett.* **39,** no. 22, L22306, doi: 10.1029/2012GL053923.

Duputel, Z., L. Rivera, H. Kanamori, G. P. Hayes, B. Hirshorn, and S. Weinstein (2011). Real-time W phase inversion during the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planets Space* **63,** 5, doi: 10.5047/eps.2011.05.032.

Gutenberg, B., and C. F. Richter (1949). *Seismicity of the Earth and Associated Phenomena*, Princeton University Press, Princeton, New Jersey, 273 pp.

Hayes, G. P., E. K. Myers, J. W. Dewey, R. W. Briggs, P. S. Earle, H. M. Benz, G. M. Smoczyk, H. E. Flamme, W. D. Barnhart, R. D. Gold, *et al.* (2017). Tectonic summaries of magnitude 7 and greater earthquakes from 2000 to 2015, *U.S. Geol. Surv. Open-File Rept. 2016-1192*, 148 pp., doi: 10.3133/ofr20161192.

Hayes, G. P., L. Rivera, and H. Kanamori (2009). Source inversion of the W-Phase: Real-time implementation and extension to low magnitudes, *Seismol. Res. Lett.* **80,** no. 5, 817–822, doi: 10.1785/gssrl.80.5.817.

Herman, M. W, and K. P. Furlong (2021). Triggering an unexpected earthquake in an uncoupled subduction zone, *Sci. Adv.* **7,** no. 13, eabf7590, doi: 10.1126/sciadv.abf7590.

Hirshorn, B., S. Weinstein, and S. Tsuboi (2013). On the application of Mwp in the Near Field and the March 11, 2011 Tohoku earthquake, *Pure Appl. Geophys.* **170,** 975–991, doi: 10.1007/s00024-012-0495-3.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.* **9,** no. 3, 90–95, doi: 10.1109/MCSE.2007.55.

Kanamori, H. (1972). Mechanism of tsunami earthquakes, *Phys. Earth Planet. In.* **6,** no. 5, 346–359, doi: 10.1016/0031-9201(72)90058-1.

Kanamori, H. (1977). The energy release in great earthquakes, *J. Geophys. Res.* **82,** no. 20, 2981–2987, doi: 10.1029/JB082i020p02981.

Kanamori, H. (1993). W phase, *Geophys. Res. Lett.* **20,** no. 16, 1691–1694, doi: 10.1029/93GL01883.

Kanamori, H., and L. Rivera (2008). Source inversion of W phase: Speeding up seismic tsunami warning, *Geophys. J. Int.* **175,** no. 1, 222–238, doi: 10.1111/j.1365-246X.2008.03887.x.

Kingma, D. P., and J. Ba (2015). Adam: A method for stochastic optimization, *Third International Conf. on Learning Representations*, San Diego, California, 15 pp., doi: 10.48550/arXiv.1412.6980.

Lay, T., H. Kanamori, C. J. Ammon, K. D. Koper, A. R. Hutko, L. Ye, H. Yue, and T. M. Rushing (2012). Depth-varying rupture properties of subduction zone megathrust faults, *J. Geophys. Res.* **117,** no. B4, doi: 10.1029/2011JB009133.

Lee, W. H. K., and S. W. Stewart (1981). *Advances in Geophysics, Supplement 2—Principles and Applications of Microearthquake Networks*, Academic Press, New York, New York, 293 pp.

Licciardi, A., Q. Bletery, B. Rouet-Leduc, J.-P. Ampuero, and K. Juhel (2022). Instantaneous tracking of earthquake growth with elastogravity signals, *Nature* **606,** 319–324, doi: 10.1038/s41586-022-04672-7.

Lin, J.-T., D. Melgar, A. M. Thomas, and J. Searcy (2021). Early warning for great earthquakes from characterization of crustal deformation patterns with deep learning, *J. Geophys. Res.* **126,** no. 10, e2021JB022703, doi: 10.1029/2021JB022703.

Lomax, A., A. Michelini, and D. Jozinović (2019). An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network, *Seismol. Res. Lett.* **90,** no. 2A, 517–529, doi: 10.1785/0220180311.

Melgar, D., R. J. LeVeque, D. S. Dreger, and R. M. Allen (2016). Kinematic rupture scenarios and synthetic displacement data: An example application to the Cascadia subduction zone, *J. Geophys. Res.* **121,** no. 9, 6658–6674, doi: 10.1002/2016JB013314.

Meng, L., J. P. Ampuero, J. Stock, Z. Duputel, Y. Luo, and V. C. Tsai (2012). Earthquake in a maze: Compressional rupture branching during the 2012 M(w) 8.6 Sumatra earthquake, *Science* **337,** no. 6095, 724–726, doi: 10.1126/science.1224030.

Mousavi, S. M., and G. C. Beroza (2020a). Bayesian-deep-learning estimation of earthquake location from single-station observations, *IEEE Trans. Geosci. Remote Sens.* **58,** no. 11, 8211–8224, doi: 10.1109/TGRS.2020.2988770.

Mousavi, S. M., and G. C. Beroza (2020b). A machine-learning approach for earthquake magnitude estimation, *Geophys. Res. Lett.* **47,** no. 1, e2019GL085976, doi: 10.1029/2019GL085976.

Mousavi, S. M., and G. C. Beroza (2023). Machine learning in earthquake seismology. *Ann. Rev. Earth Planet. Sci.* **51,** no. 1, 105–129, doi: 10.1146/annurev-earth-071822-100323.

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

Mousavi, S. M., Y. Sheng, W. Zhu, and G. C. Beroza (2019). STanford EArthquake Dataset (STEAD): A global data set of seismic signals

for AI, *IEEE Access* **7,** 179,464–179,476, doi: 10.1109/ACCESS.2019.2947848.

Münchmeyer, J., D. Bindi, U. Leser, and F. Tilmann (2021). Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network, *Geophys. J. Int.* **226,** no. 2, 1086–1104, doi: 10.1093/gji/ggab139.

Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Sci. Adv.* **4,** no. 2, e1700578, doi: 10.1126/sciadv.1700578.

Ringler, A. T., R. E. Anthony, R. C. Aster, C. J. Ammon, S. Arrowsmith, H. Benz, C. Ebeling, A. Frassetto, W.-Y. Kim, P. Koelemeijer, *et al.* (2022). Achievements and prospects of global broadband seismographic networks after 30 years of continuous geophysical observations, *Rev. Geophys.* **60,** no. 3, e2021RG000749, doi: 10.1029/2021RG000749.

Ristea, N.-C., and A. Radoi (2021). Complex neural networks for estimating epicentral distance, depth, and magnitude of seismic waves, *IEEE Trans. Geosci. Remote Sens.* **19,** 7502305, doi: 10.1109/LGRS.2021.3059422.

Ross, Z. E., M.-A. Meier, E. Hauksson, and T. H. Heaton (2018). Generalized seismic phase detection with deep learning, *Bull. Seismol. Soc. Am.* **108,** no. 5A, 2894–2901, doi: 10.1785/0120180080.

Santellanes, S. R., D. Melgar, B. W. Crowell, and J.-T. Lin (2022). Potential megathrust co-seismic slip during the 2020 Sand Point, Alaska strike-slip earthquake, *ESS Open Archive*, doi: 10.1002/essoar.10508002.2.

Singh, S. K., J. Pacheco, M. Ordaz, and V. Kostoglodov (2000). Source time function and duration of Mexican earthquakes, *Bull. Seismol. Soc. Am.* **90,** no. 2, 468–482, doi: 10.1785/0119990081.

Sipkin, S. A., W. J. Person, and B. W. Presgrave (2000). Earthquake bulletins and catalogs at the USGS NEIC, *IRIS Newsletter*, 2–4, available at http://ftp.iris.edu/news/IRISnewsletter/volume-12000no1/page-02-04.htm (last accessed November 2023).

Trugman, D. T., M. T. Page, S. E. Minson, and E. S. Cochran (2019). Peak ground displacement saturates exactly when expected: Implications for earthquake early warning, *J. Geophys. Res.* **124,** no. 5, 4642–4653, doi: 10.1029/2018JB017093.

Tsuboi, S., P. M. Whitmore, and T. J. Sokolowski (1999). Application of Mwp to deep and teleseismic earthquakes, *Bull. Seismol. Soc. Am.* **89,** no. 5, 1345–1351, doi: 10.1785/BSSA0890051345.

Uieda, L., D. Tian, W. J. Leong, W. Schlitzer, M. Grund, M. Jones, Y. Fröhlich, L. Toney, J. Yao, Y. Magen, *et al.* (2023). PyGMT: A Python interface for the Generic Mapping Tools (v0.9.0), *Zenodo*, doi: 10.5281/zenodo.3781524.

van den Ende, M. P. A., and J.-P. Ampuero (2020). Automated seismic source characterization using deep graph neural networks, *Geophys. Res. Lett.* **47,** no. 17, e2020GL088690, doi: 10.1029/2020GL088690.

Yue, H., T. Lay, and K. D. Koper (2012). En échelon and orthogonal fault ruptures of the 11 April 2012 great intraplate earthquakes, *Nature* **490,** 245–249, doi: 10.1038/nature11492.

Yue, H., T. Lay, L. Rivera, Y. Bai, Y. Yamazaki, K. F. Cheung, E. M. Hill, K. Sieh, W. Kongko, and A. Muhari (2014). Rupture process of the 2010 Mw7.8 Mentawai tsunami earthquake from joint inversion of near-field hr-GPS and teleseismic body wave recordings constrained by tsunami observations, *J. Geophys. Res.* **119,** no. 7, 5574–5593, doi: 10.1002/2014JB011082.

Zhang, X., J. Zhang, C. Yuan, S. Liu, Z. Chen, and W. Li (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method, *Sci. Rep.* **10,** 1941, doi: 10.1038/s41598-020-58908-5.

Zhu, W., and G. C. Beroza (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.* **216,** no. 1, 261–273, doi: 10.1093/gji/ggy423.

Zhu, W., S. M. Mousavi, and G. C. Beroza (2020). Seismic signal augmentation to improve generalization of deep neural networks, *Adv. Geophys.* **61,** 151–177, doi: 10.1016/bs.agph.2020.07.003.

Manuscript received 7 July 2023
Published online 2 January 2024