Supplementary Materials: Curated Pacific Northwest AI-ready Seismic Dataset

Yiyu Ni [*1], **Alexander R. Hutko** [1,2], **Francesca Skene** [1], **Marine A. Denolle** [1], **Stephen D. Malone** [1,2], **Paul Bodin** [1,2], **J. Renate Hartog** [1,2], **Amy K. Wright** [1,2]

[1]Department of Earth and Space Sciences, University of Washington, Seattle, WA, [2]Pacific Northwest Seismic Network, Seattle, WA

UTC Time	CISN Event ID	PNSN Event ID	Magnitude
2022-12-20 10:34:24	nc73821036	uw61899256	Mw 6.4
2022-12-20 10:52:39	nc73821106	uw61899266	Ml 3.1
2022-12-20 11:42:54	nc73821226	uw61899301	Ml 3.0
2022-12-20 12:21:43	nc73821346	uw61899311	Ml 3.0
2022-12-20 13:35:06	nc73821486	uw61899326	Ml 3.4
2022-12-20 13:53:19	nc73821516	uw61899331	Ml 3.3
2022-12-20 15:09:05	nc73821636	uw61899336	Mw 4.0
2022-12-20 15:30:01	nc73821656	uw61899346	Ml 3.0
2022-12-20 16:33:41	nc73821761	uw61899381	Ml 3.2
2022-12-20 22:06:34	nc73822026	uw61890602	Mw 4.4
2022-12-21 01:07:33	nc73822146	uw61890647	Ml 3.0
2022-12-21 07:17:15	nc73822341	uw61890697	Ml 3.0
2022-12-21 16:28:16	nc73822556	uw61890767	Ml 3.3
2022-12-22 08:47:13	nc73822961	uw61890997	Ml 3.3
2022-12-22 11:49:55	nc73823036	uw61891007	Mw 3.8
2022-12-24 19:33:44	nc73824236	uw61900186	Mw 4.2
2022-12-25 19:40:29	nc73824826	uw61900336	Md 3.0
2022-12-29 00:16:18	nc73826156	uw61900806	Md 3.0
2023-01-01 18:35:04	nc73827571	uw61901146	Mw 5.4
2023-01-06 12:27:19	nc73829331	uw61901521	Ml 3.3

Table S1 The events selected from the 20 December 2022 Northern California earthquake sequence that are included in the data set. Source origin time, event ID, and magnitude are reported by the California Integrated Seismic Network (CISN), with corresponding PNSN event ID.

*Corresponding author: niyiyu@uw.edu

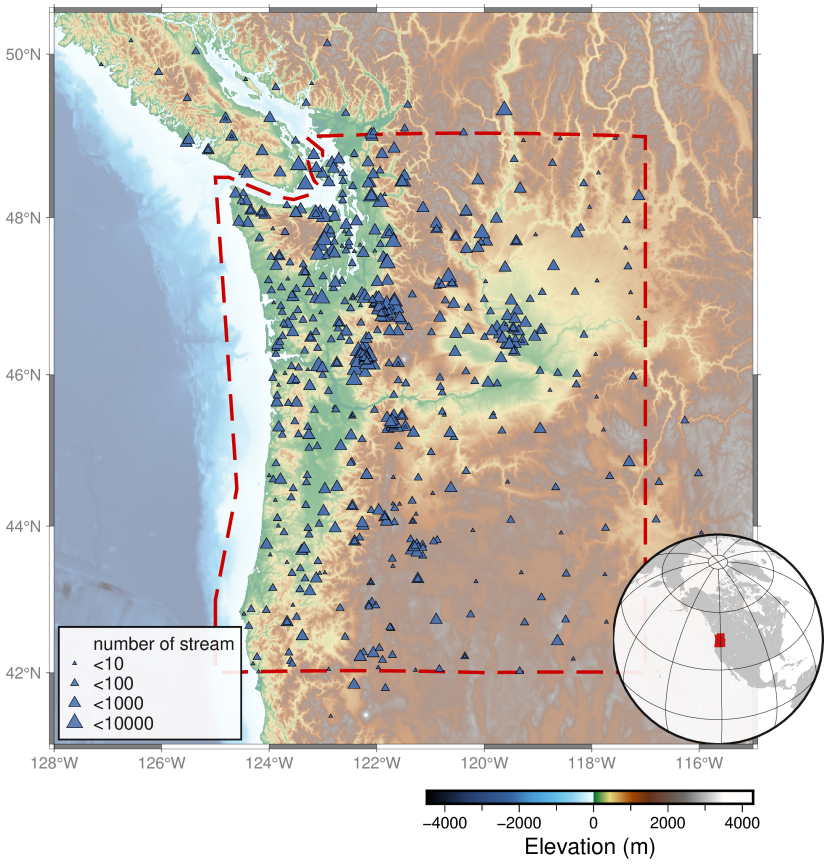

Figure S1 Number of streams from the ComCat event catalog per station. The red dashed polygon denotes the authoritative region boundaries of PNSN.

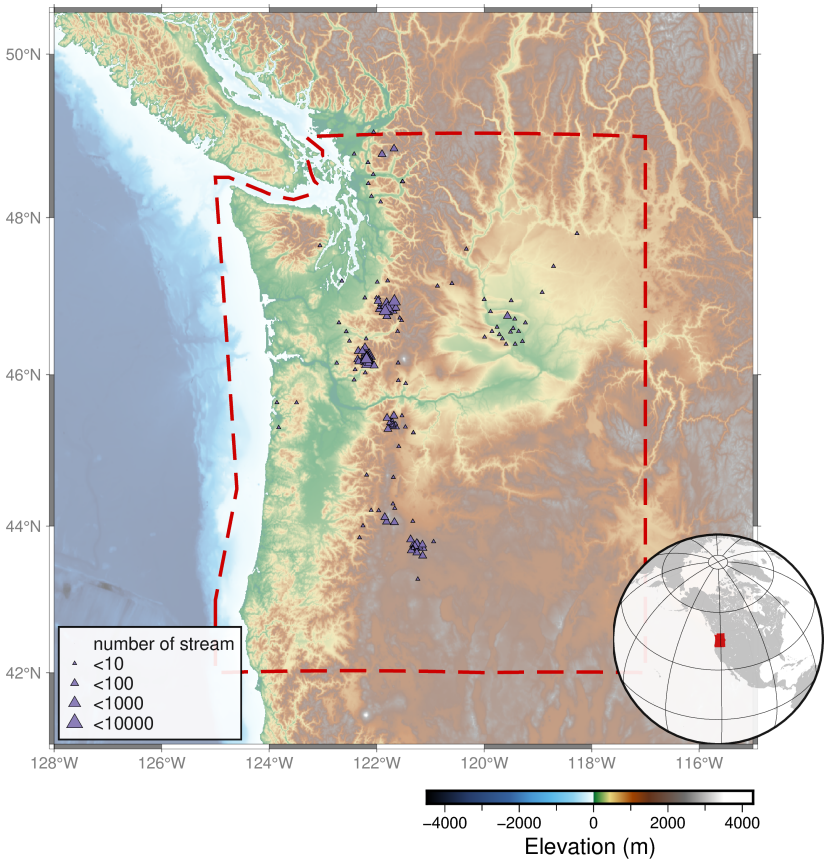

Figure S2 Number of streams from the Exotic event catalog per station. The red dashed polygon denotes the authoritative region boundaries of PNSN. Note that 96% of the exotic events are surface events.

AQMS event type use by the PNSN	ComCat label	Description
eq	**earthquake**	earthquake
le		local earthquake
re		regional earthquake
ts		teleseism
se		slow earthquake
lp		long period volcanic earthquake
lf		low-frequency event
ex	**explosion**	generic chemical blast
px		unconfirmed blast or explosion
sh		refraction/reflection survey shot
su	**surface event**	surface event
th	**thunder**	thunder
sn	**sonic boom**	sonic shockwave
pc	**plane crash**	plane crash
qb	quarry blast	quarry blast
nt	nuclear explosion	nuclear test
ve	volcanic eruption	volcanic eruption
co	mine collapse	mine/tunnel collapse
df	debris avalanche	debris flow/avalanche
av	snow avalanche	snow/ice avalanche
ls	landslide	landslide
rb	rock burst	rockburst
rs	rockslide	rockslide
bc	building collapse	building collapse/demolition
mi	meteor impact	meteor/comet impact
uk	unknown	unknown type

Table S2 The event types and labels used in PNSN's ANSS Quake Monitoring System (AQMS). Several event types are merged into one when being reported to the ComCat.

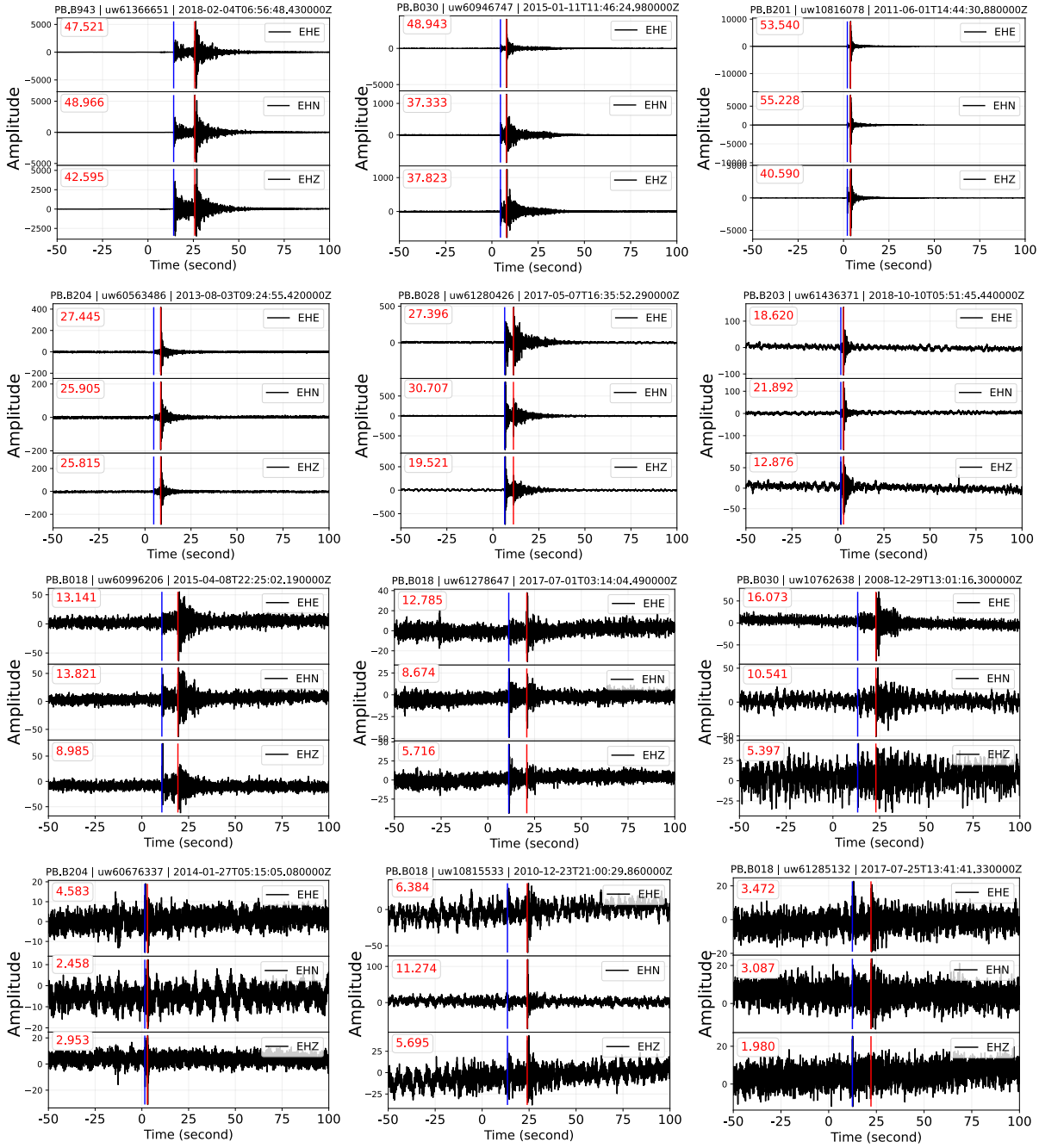

Figure S3 Randomly selected waveform samples of ComCat earthquake events from short-period three-component EH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts.

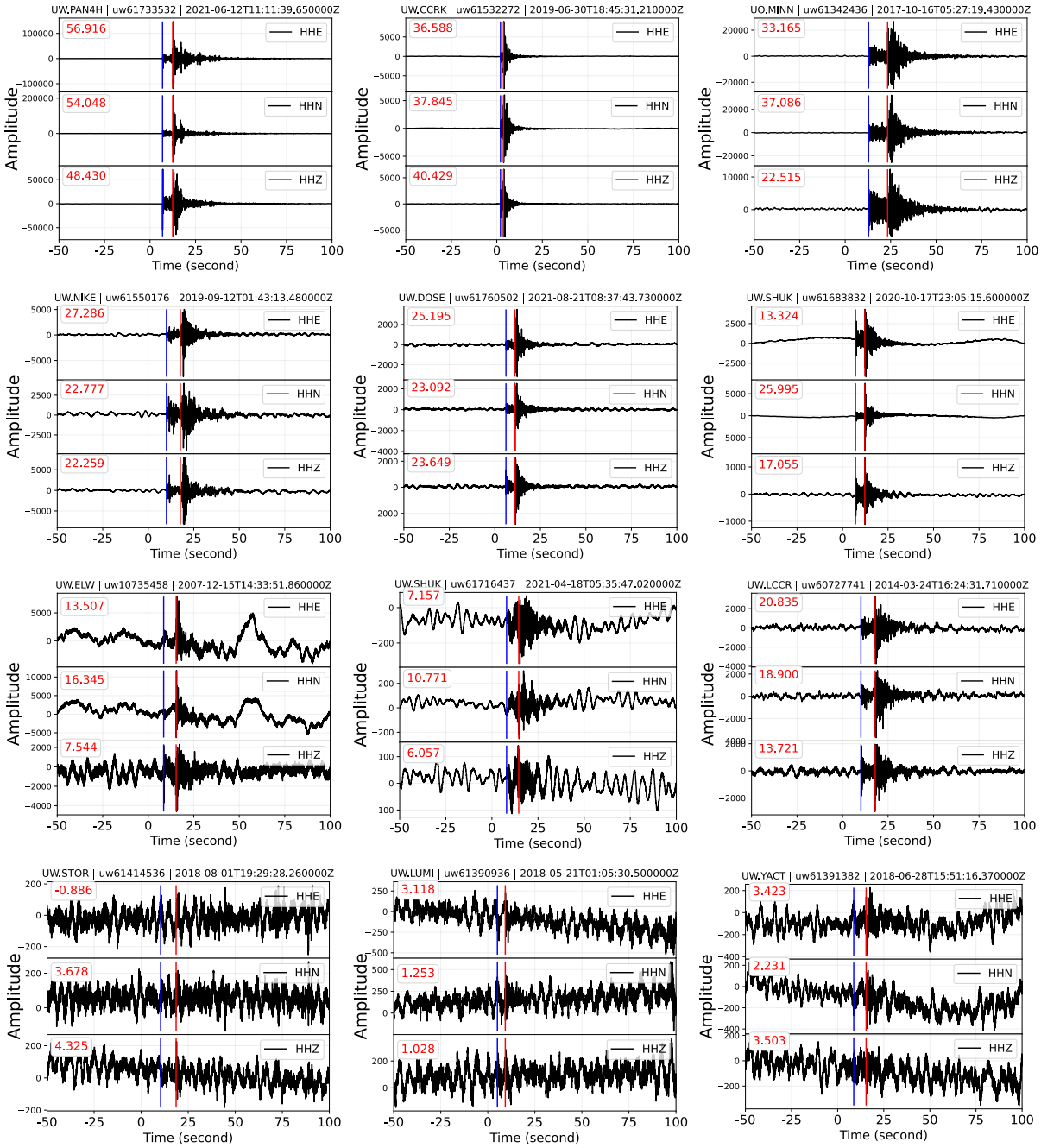

Figure S4 Randomly selected waveform samples of ComCat earthquake events from broad-band three-component HH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts but are detrended.

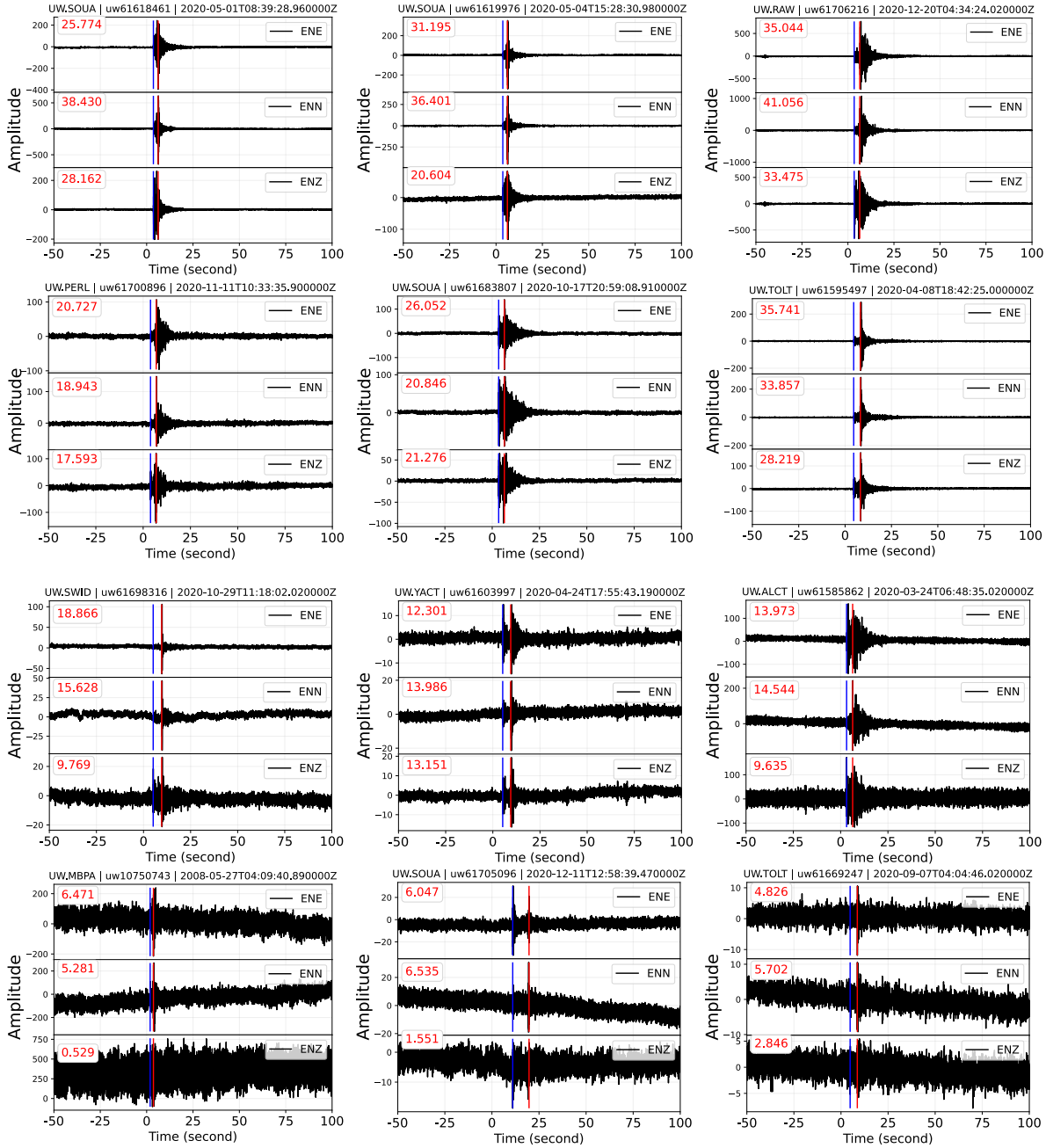

Figure S5 Randomly selected waveform samples of ComCat earthquake events from strong-motion EN? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts but are detrended.

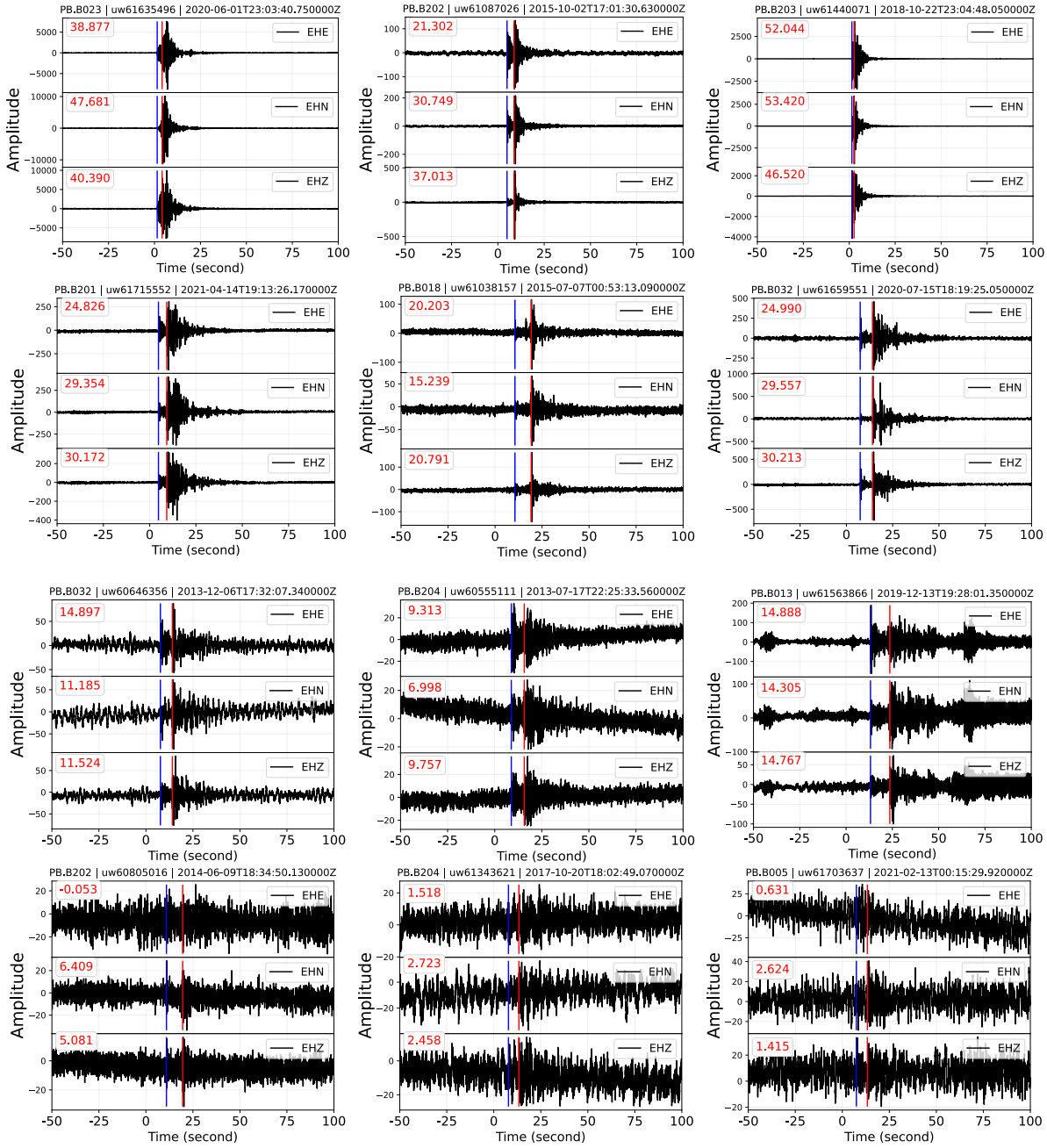

Figure S6 Randomly selected waveform samples of ComCat explosion events from short-period three-component EH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts but are detrended.

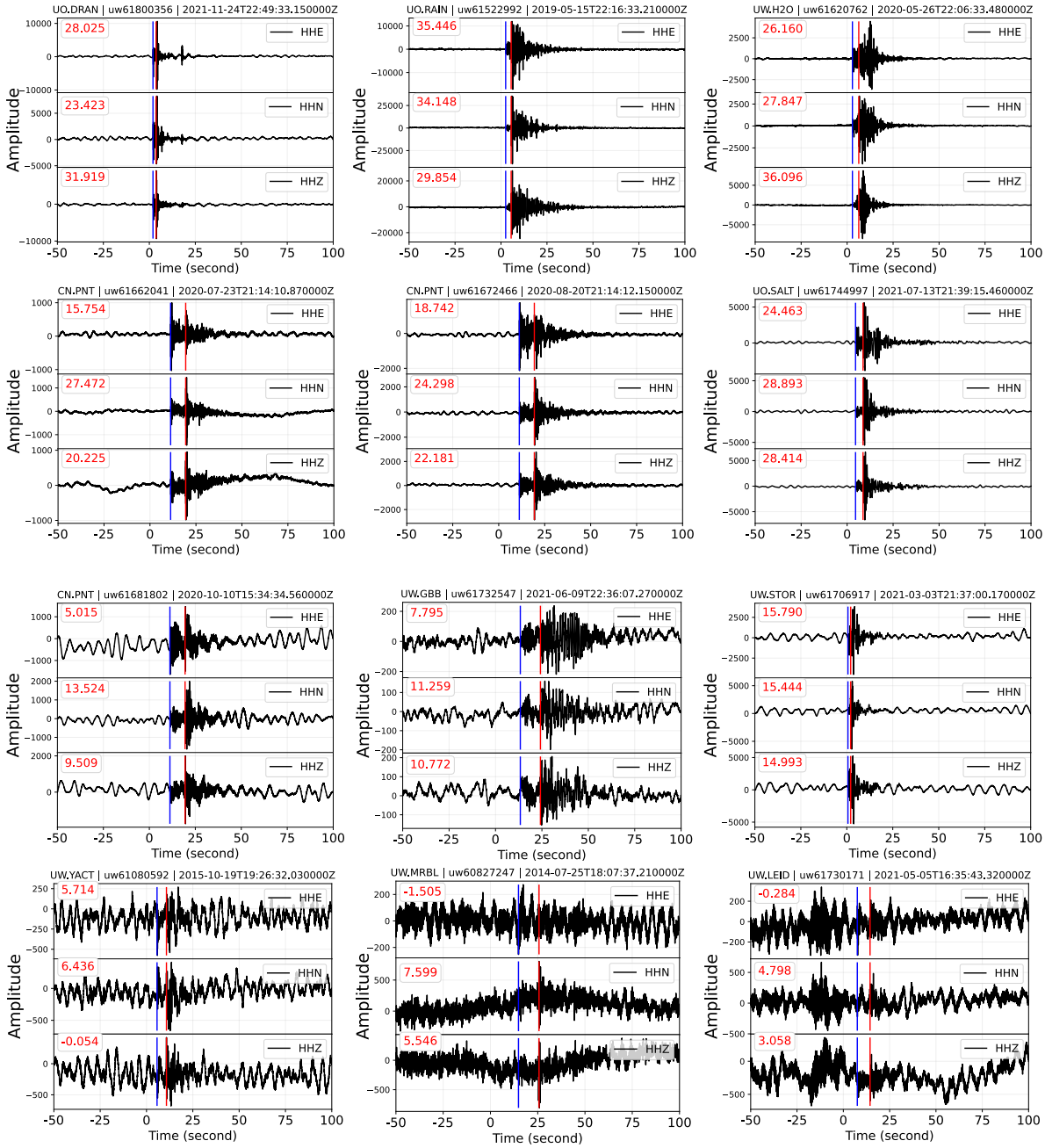

Figure S7 Randomly selected waveform samples of ComCat explosion events from broad-band three-component HH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts but are detrended.

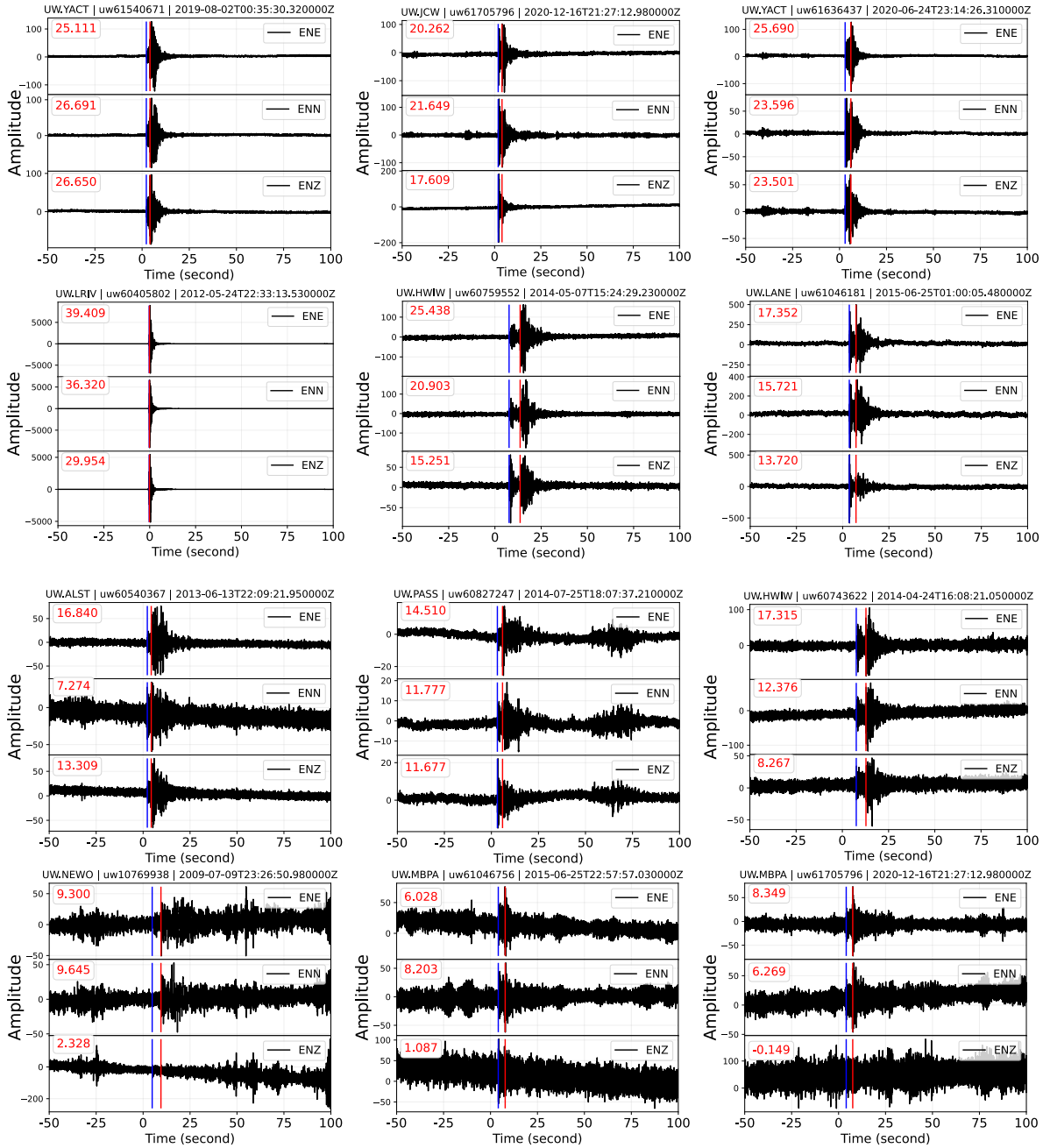

Figure S8 Randomly selected waveform samples of ComCat explosion events from strong-motion EN? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts but are detrended.

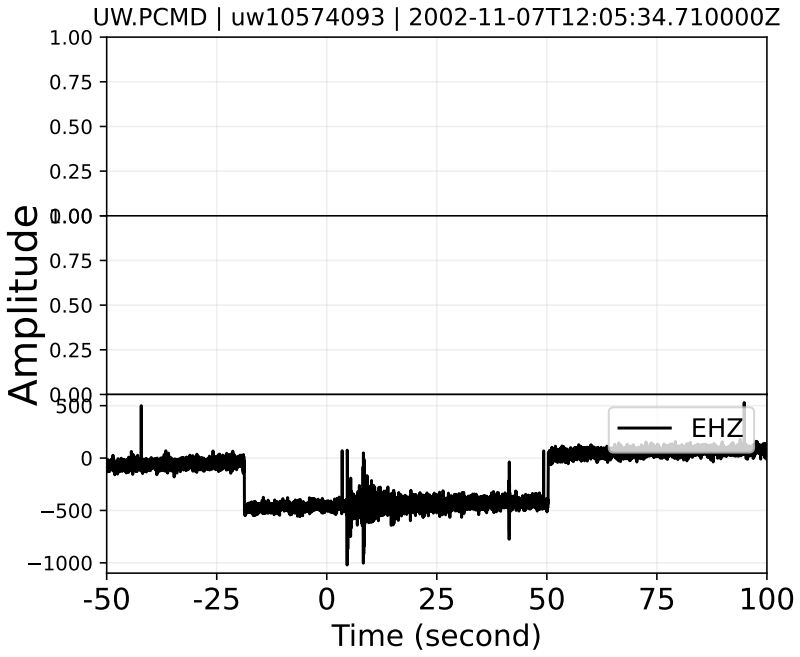

Figure S9 Example of a stream with a step offset from short-period EH channel. FDSN network and station code, ComCat event ID, and source origin time are labeled on the top.

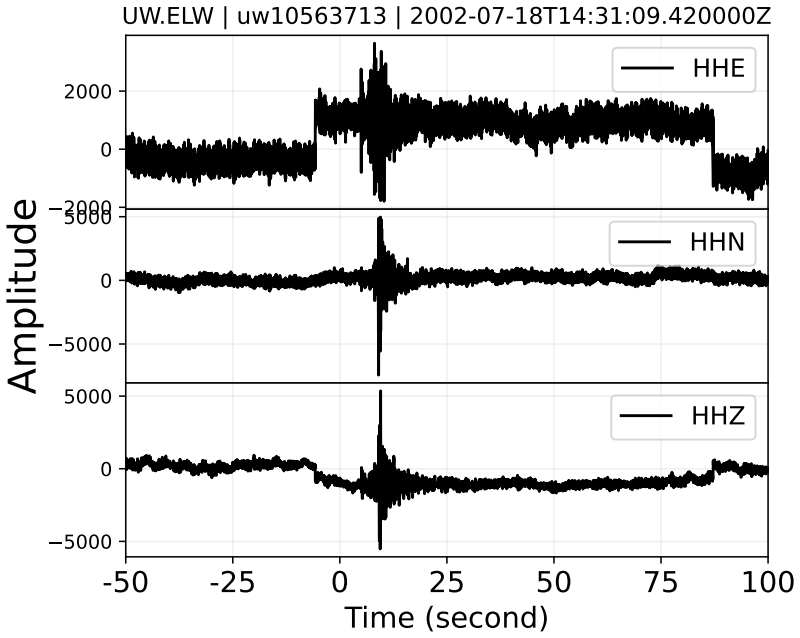

Figure S10 Example of a stream with a step offset from broad-band HH channel. FDSN network and station code, ComCat event ID, and source origin time are labeled on the top.

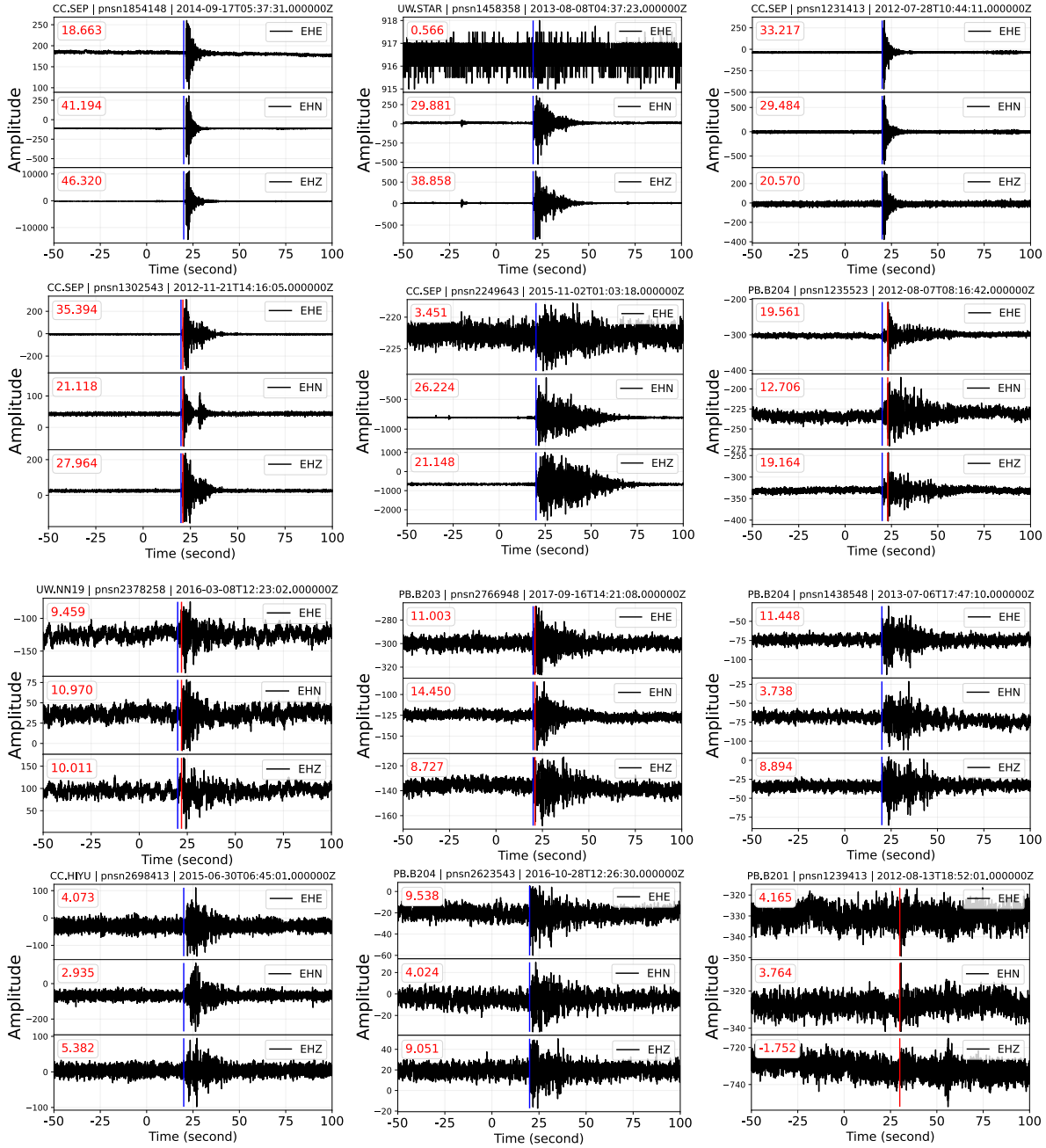

Figure S11 Randomly selected waveform samples of exotic surface events from short-period three-component EH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts.

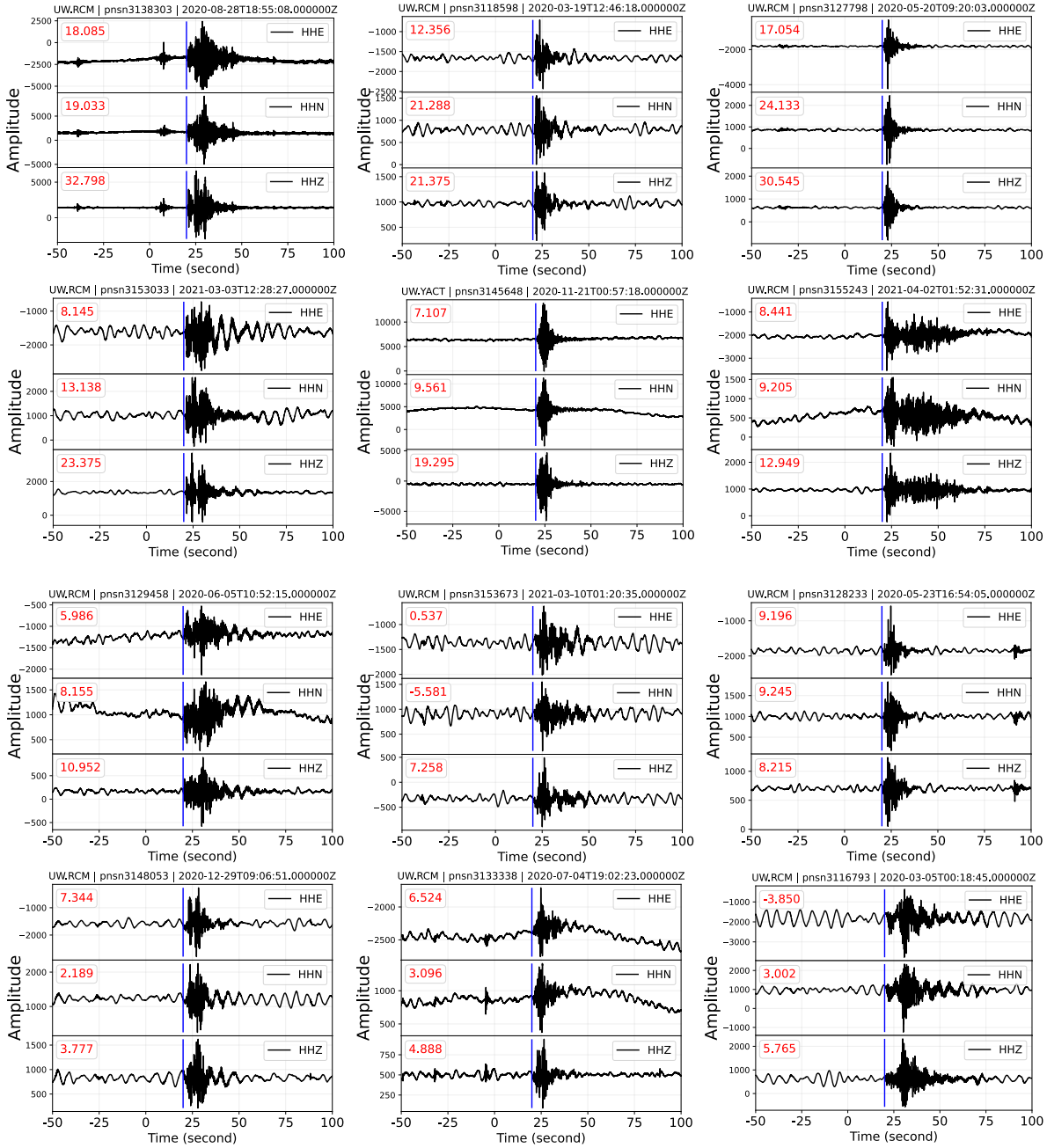

Figure S12 Randomly selected waveform samples of exotic surface events from broad-band three-component HH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts.

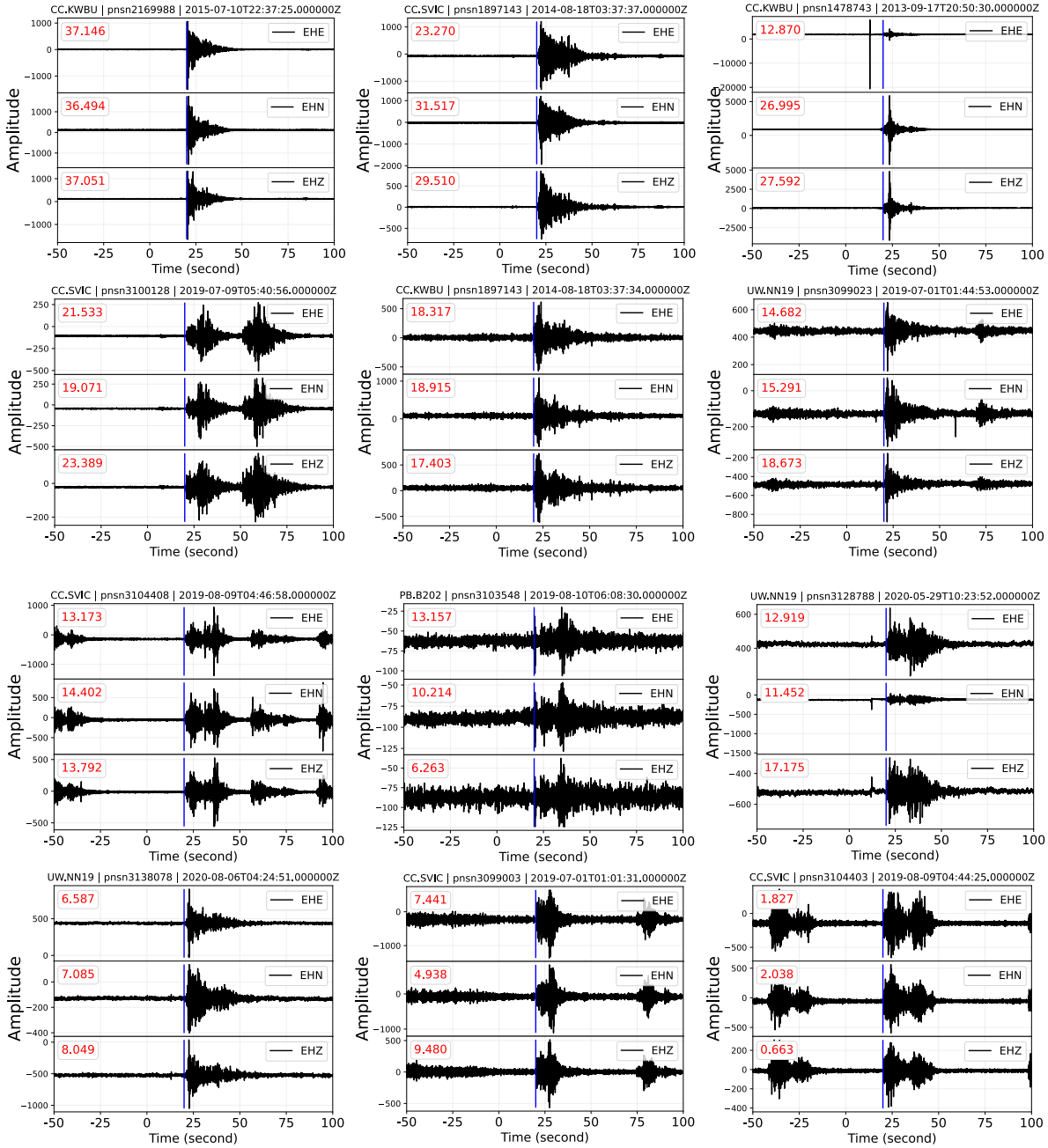

Figure S13 Randomly selected waveform samples of exotic thunder events from short-period three-component EH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts.

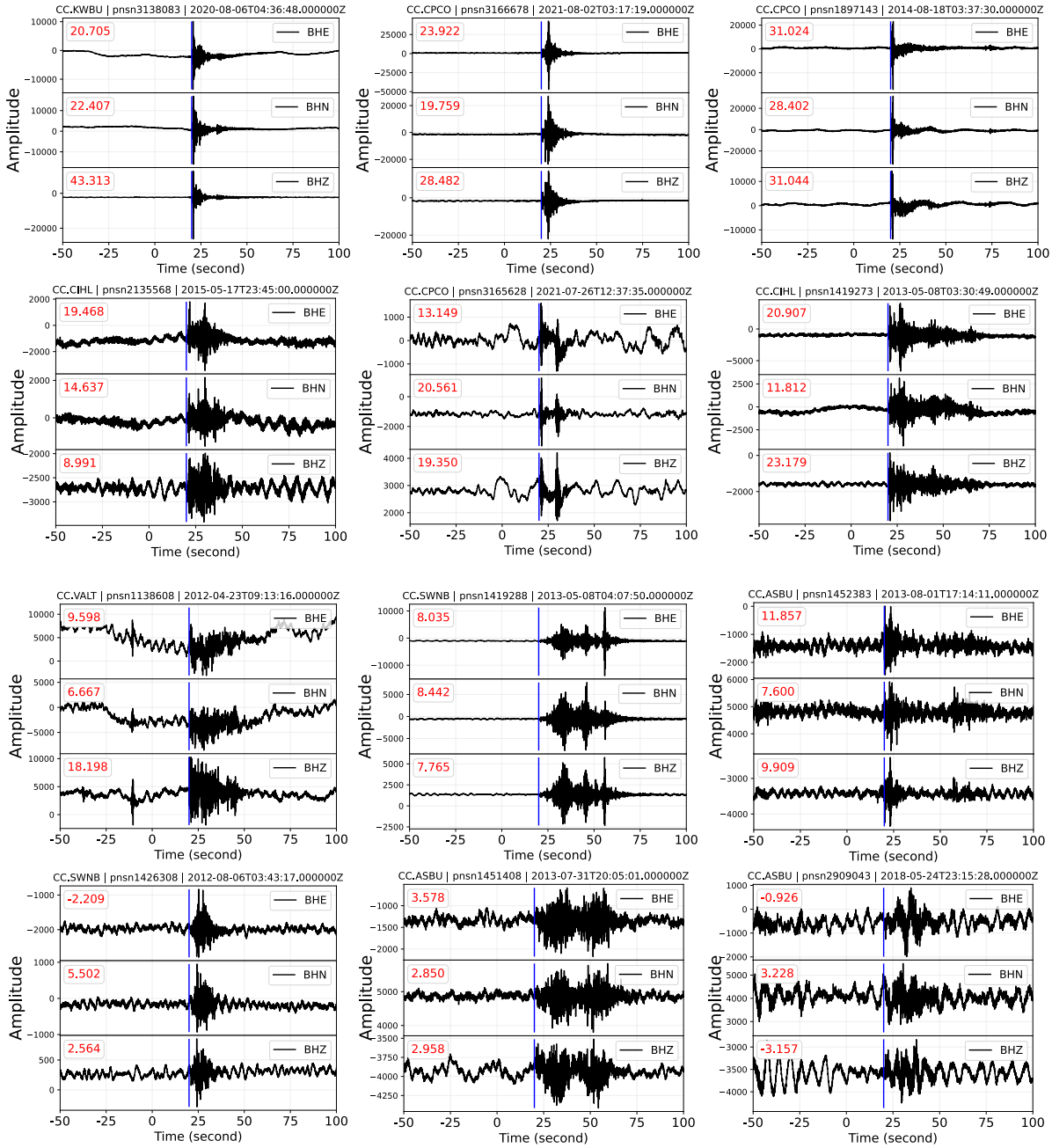

Figure S14 Randomly selected waveform samples of exotic thunder events from broad-band three-component BH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts.

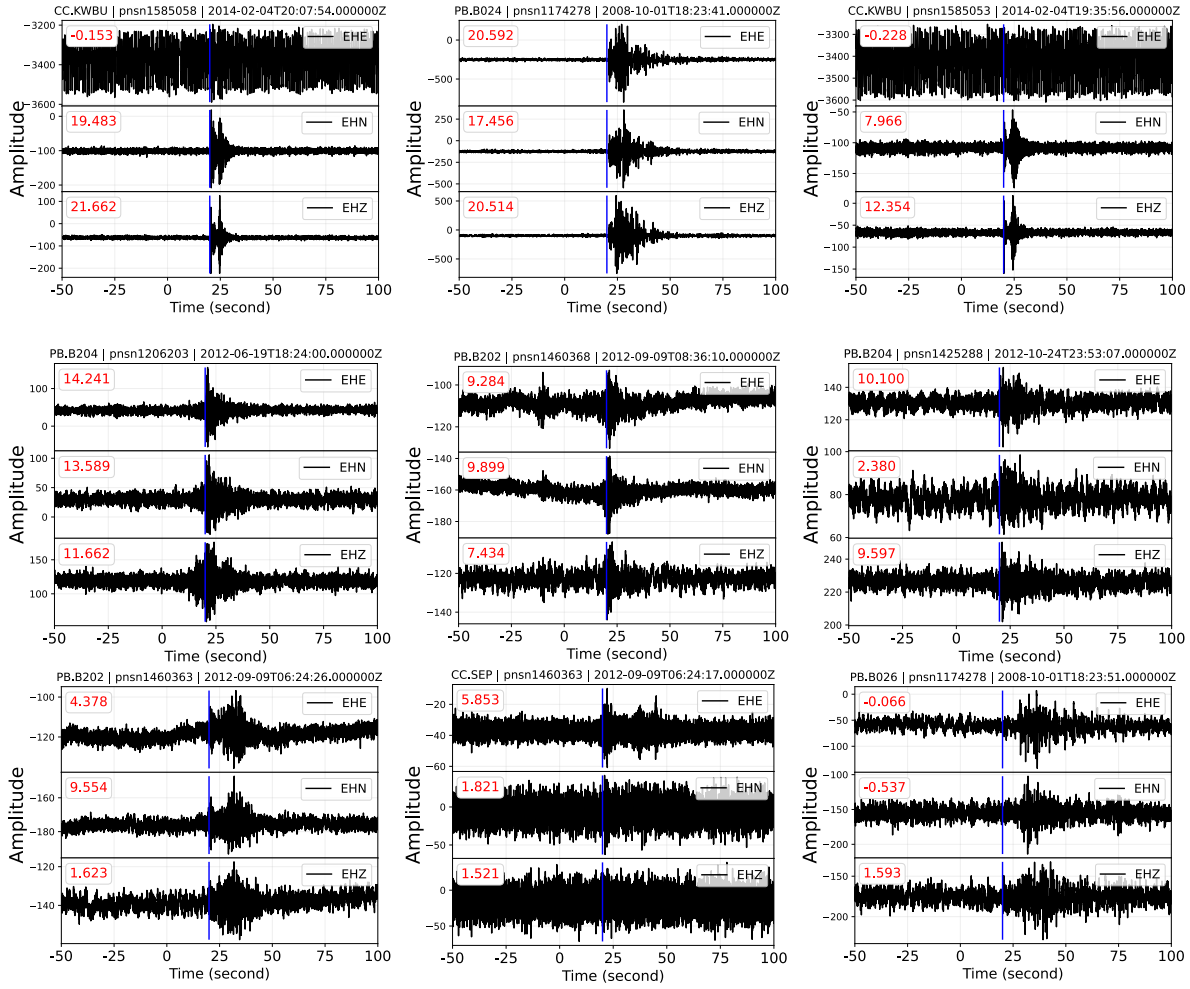

Figure S15 All waveform samples of exotic sonic boom events from short-period three-component EH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts.

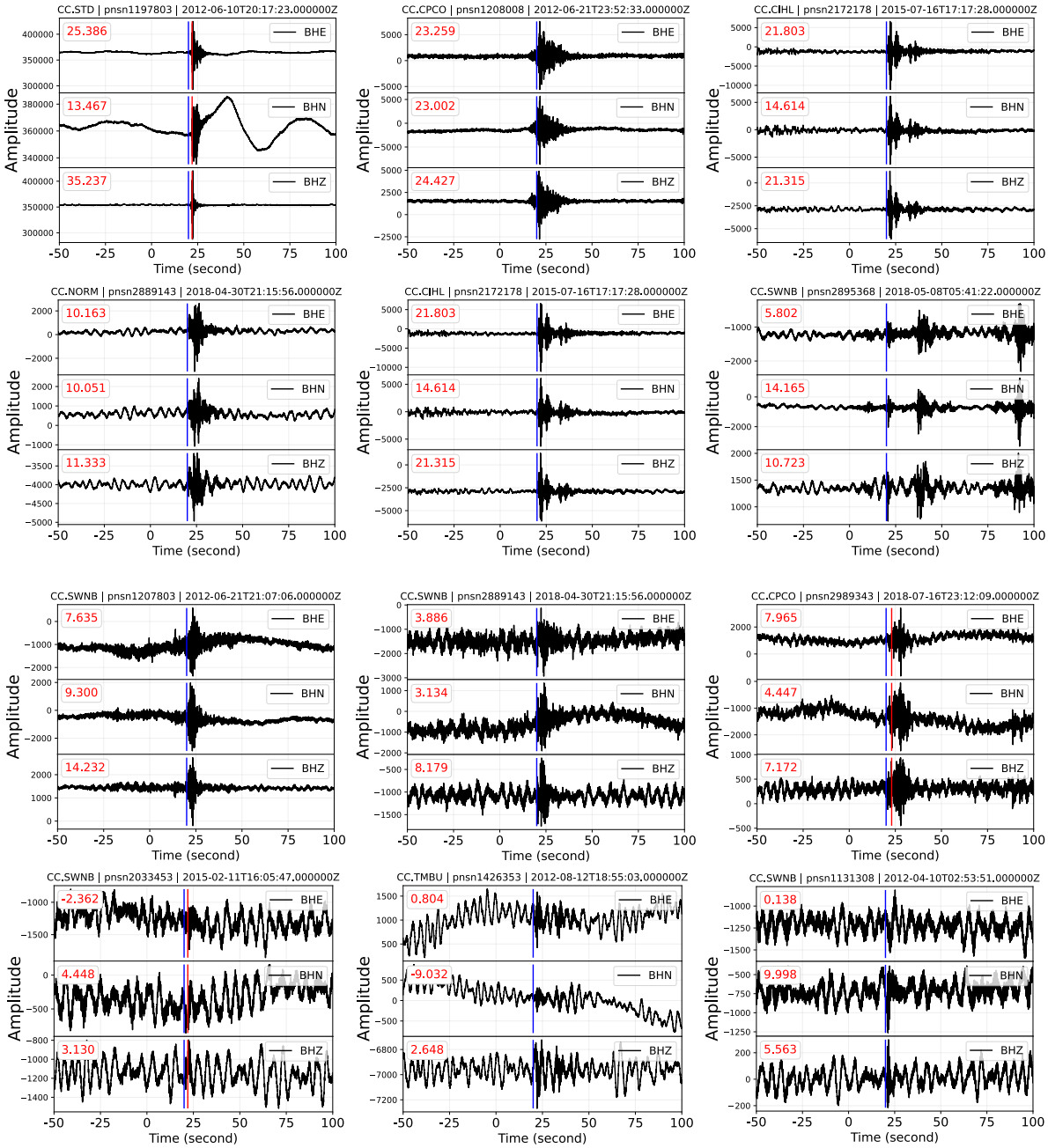

Figure S16 Randomly selected waveform samples of exotic sonic boom events from broad-band three-component BH? channels. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts.

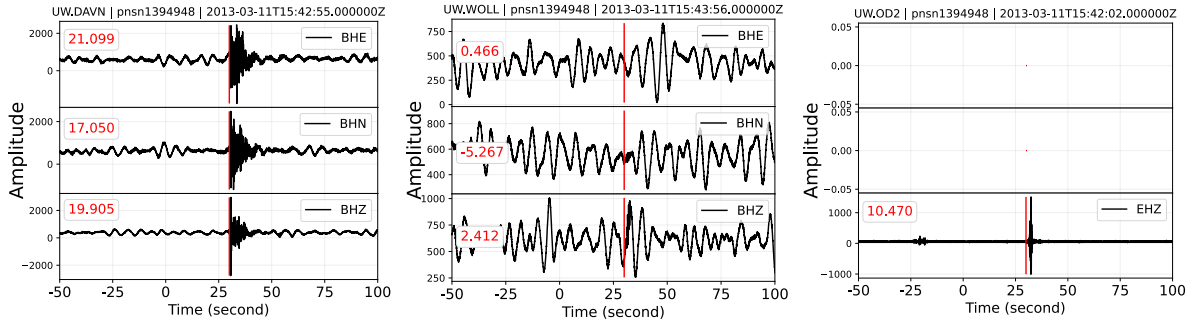

Figure S17 All waveform samples of exotic plane crash events. SNRs are marked on the upper left for each component. The blue line marks the P-wave arrival, and the red line (if any) marks the S-wave arrival. The amplitude units are in counts.

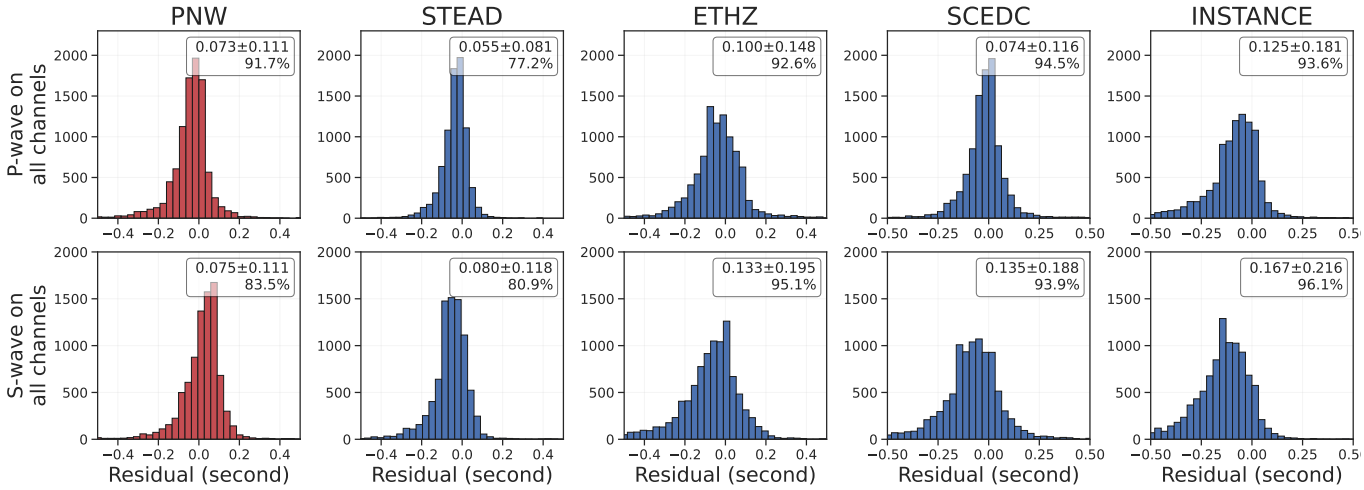

Figure S18 Histogram of P- and S-wave picking residuals ($t_{ML} - t_{PNSN}$) from the benchmark test data set of strong motion channels. The number in the upper right corner of each subplot shows the mean absolute error (MAE), the root-mean-square error (RMSE) as the uncertainties to the MAE, and the picking completeness in percentage with respect to the ground truth. The PNW-retrained Earthquake Transformer outperforms than other four pre-trained models from SeisBench (Woollam et al., 2022) in both picking accuracy and detecting completeness.

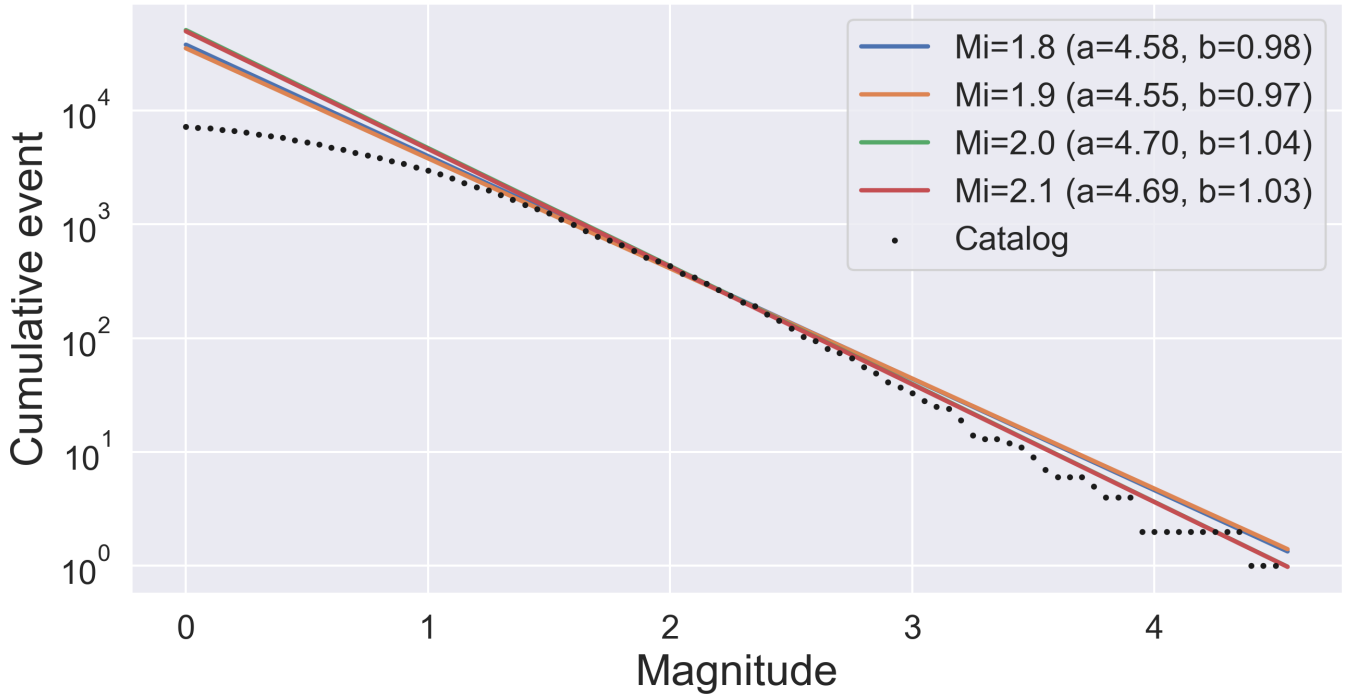

Figure S19 Fitting of Gutenberg-Richter (GR) power law distribution of magnitudes (solid lines) with 3 years of earthquake events cataloged by PNSN (black dots). The plot indicates that the minimum magnitude of completeness is around 2.

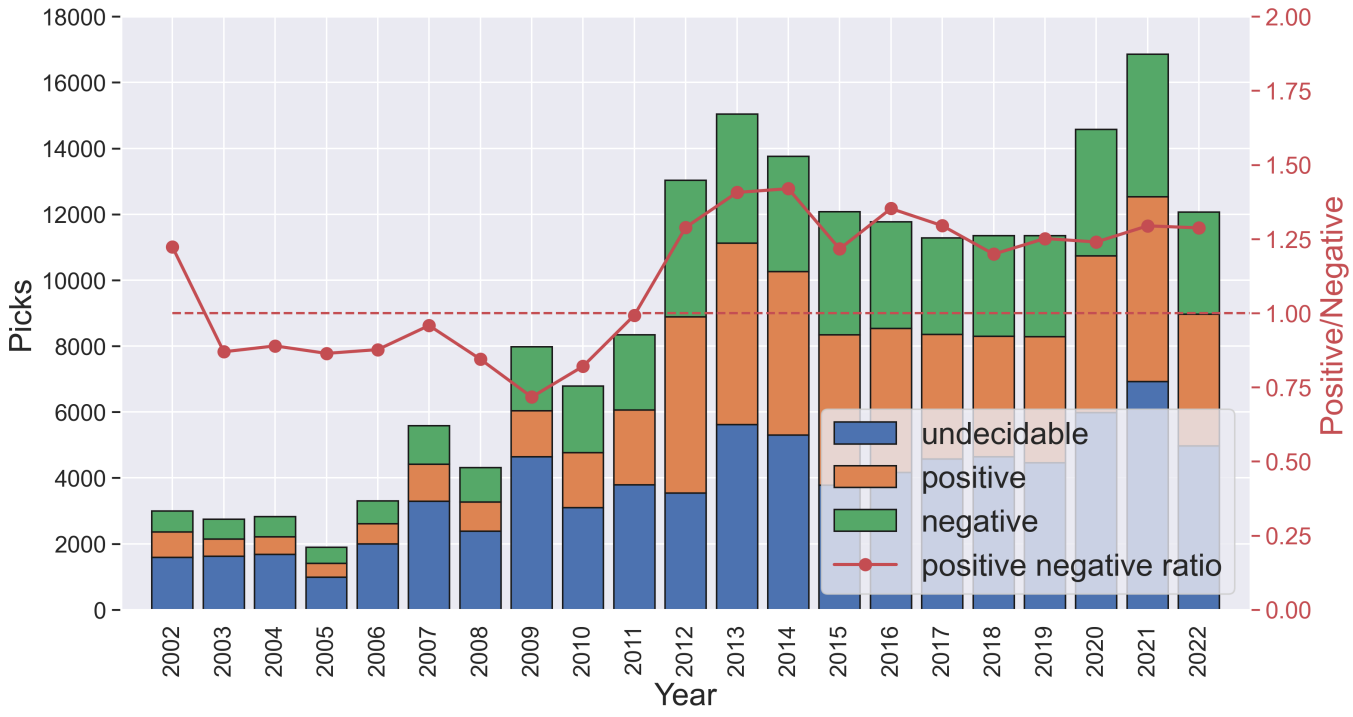

Figure S20 Number of picked P-wave polarity as a function of time. The red line marks the positive-negative ratio of the P-wave polarities.

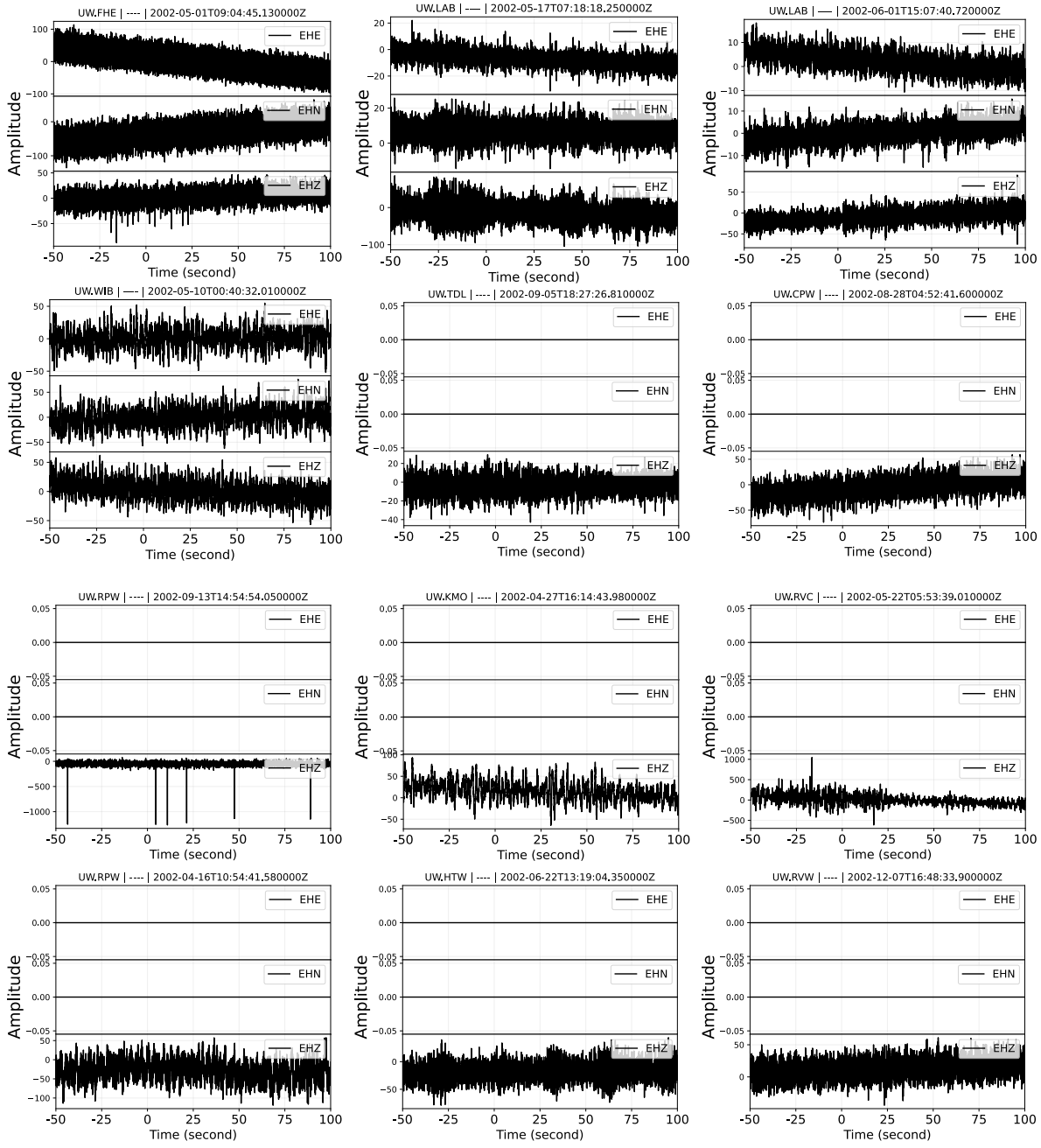

Figure S21 Randomly selected noise waveform samples from short-period EH? channels. The amplitude units are in counts but are detrended.

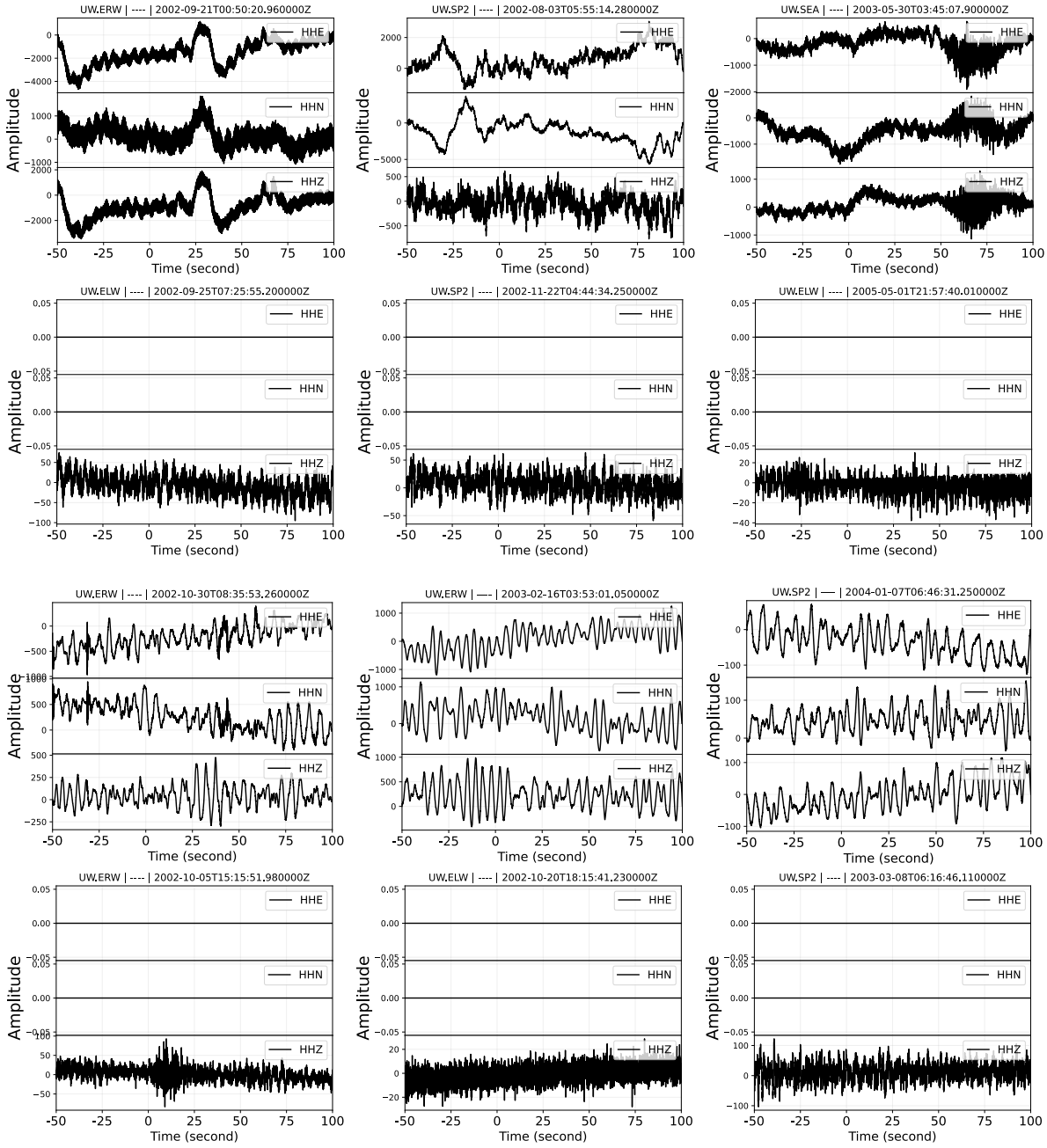

Figure S22 Randomly selected noise waveform samples from broad-band three-component HH? channels. The amplitude units are in counts but are detrended.

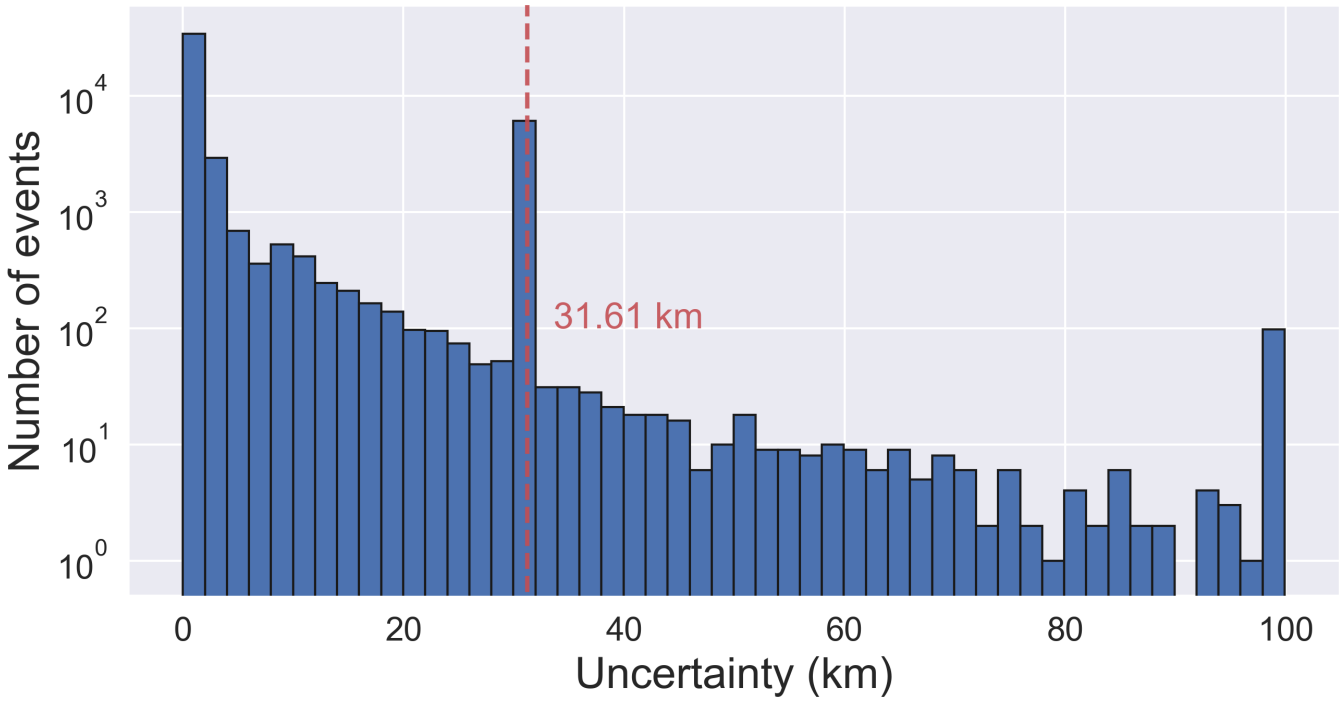

Figure S23 Histogram of depth uncertainties of ComCat events in log scale. A large number of events with 31.61 km depth uncertainty mostly come from locating a probable explosion event with a fixed depth.

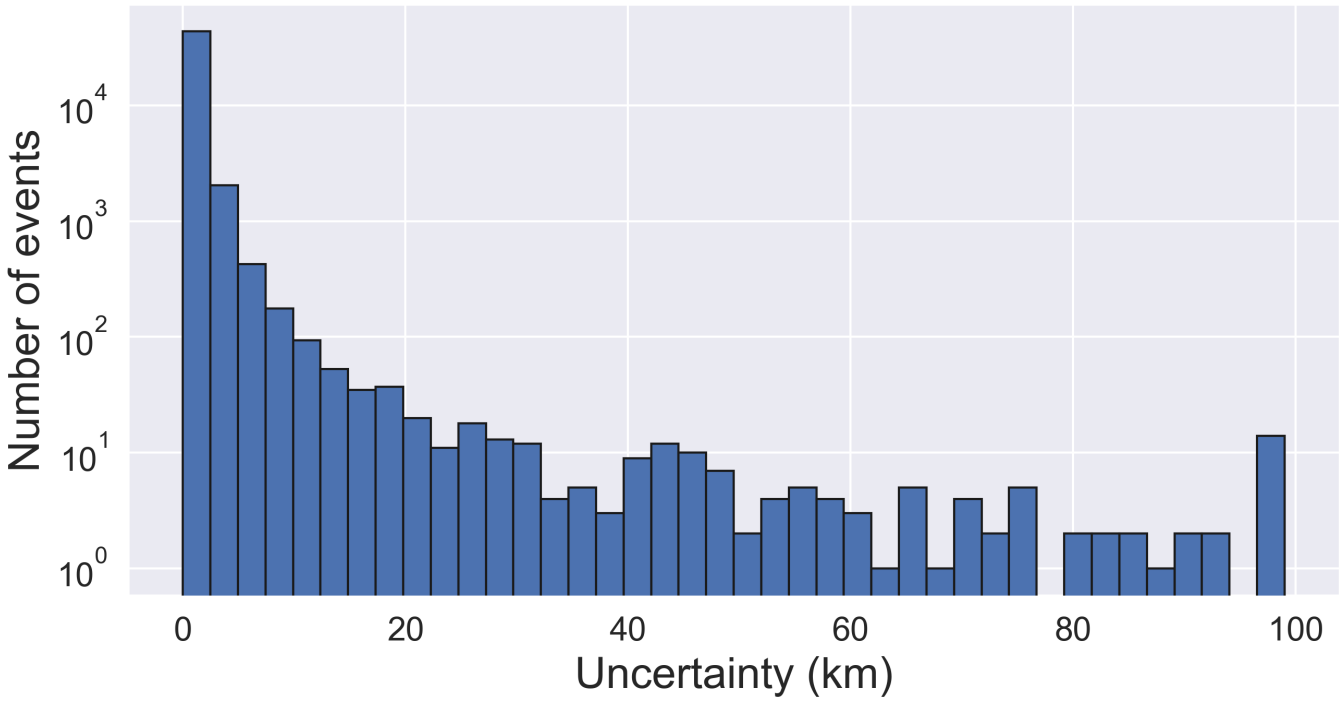

Figure S24 Histogram of horizontal uncertainties of ComCat events in log scale.

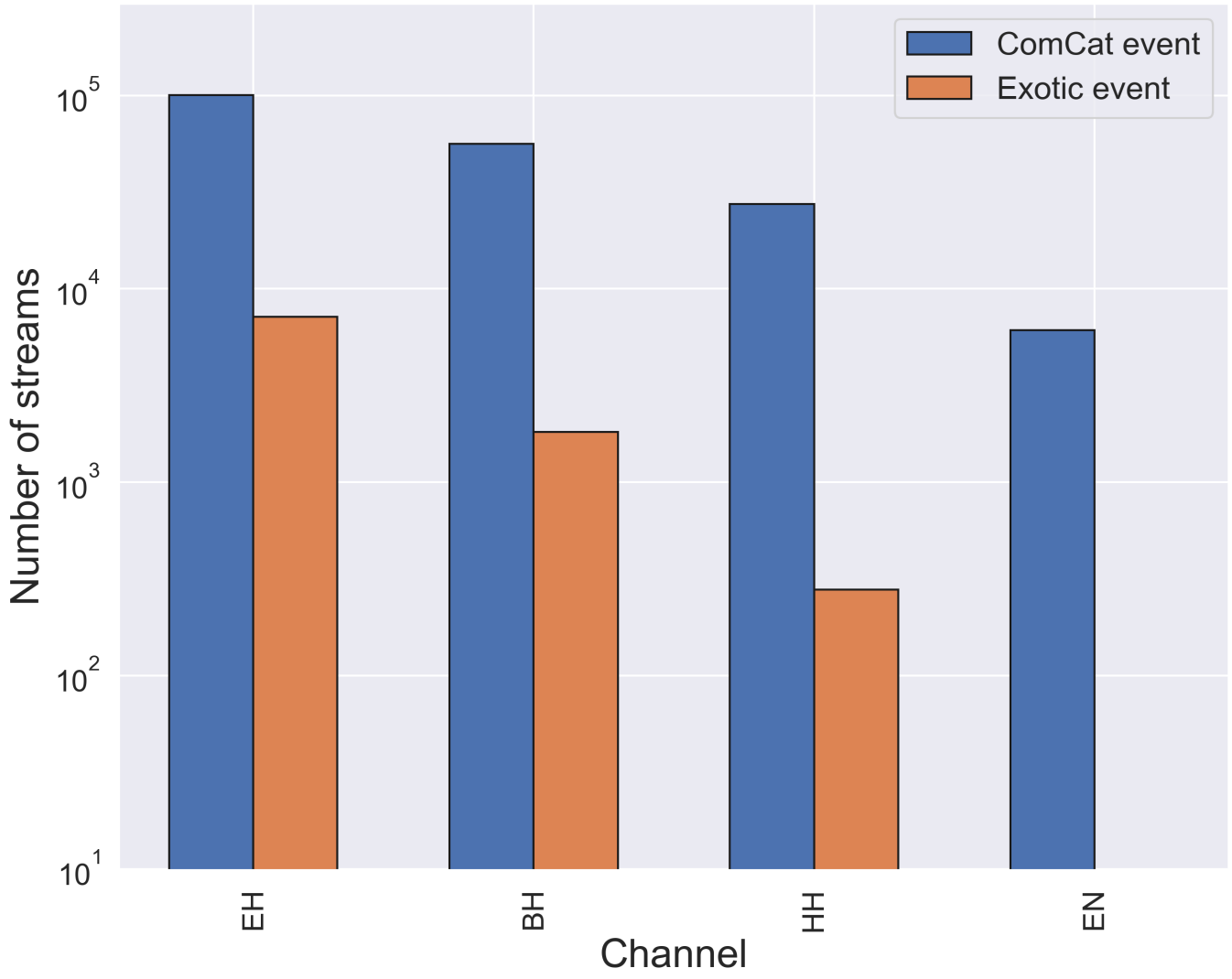

Figure S25 Number of streams arranged by the instrument type.

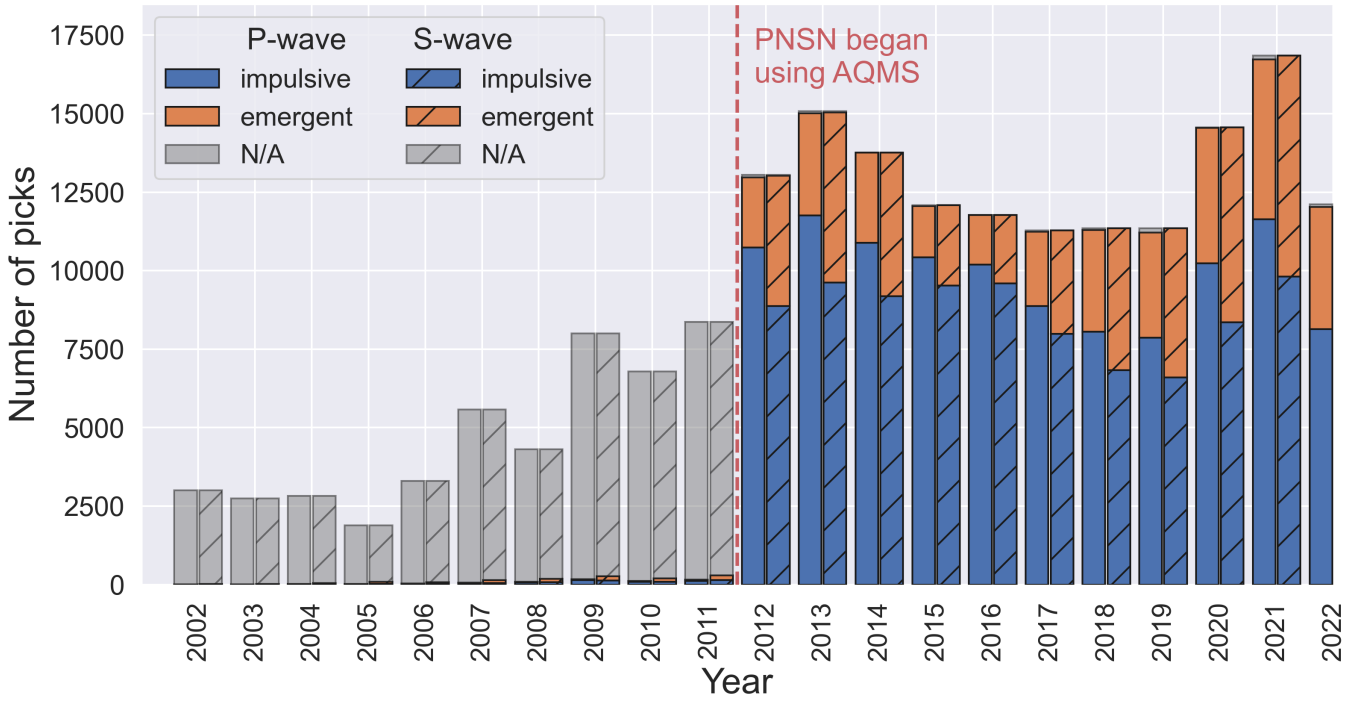

Figure S26 Number of picked P- and S-waves onsets as a function of time from ComCat events.

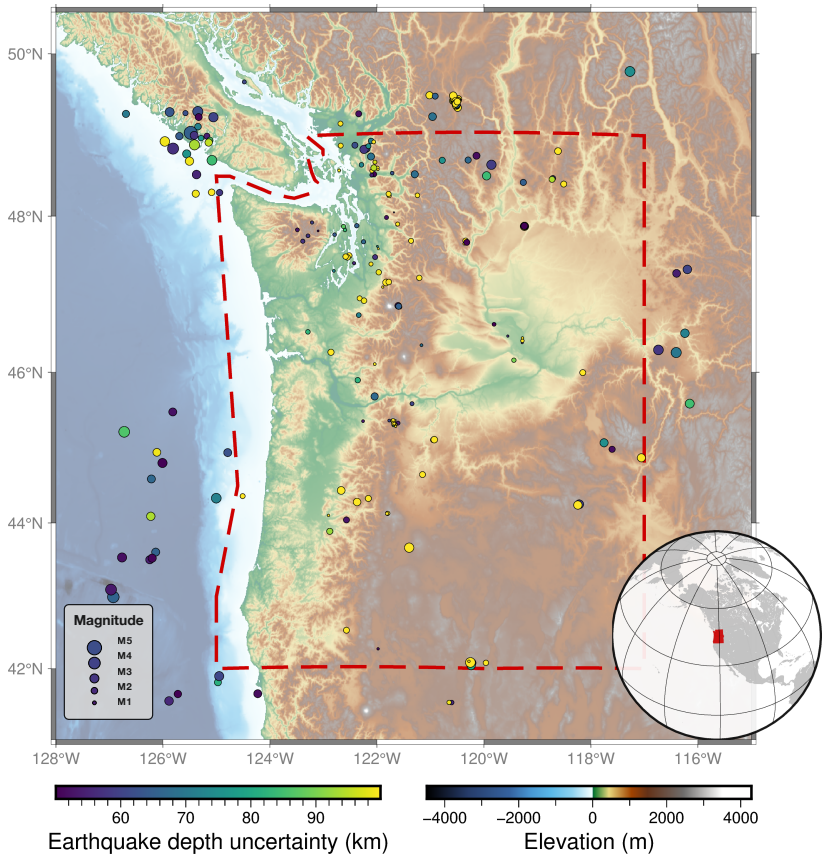

Figure S27 Location of the events with location horizontal uncertainty larger than 20 km. The red polygon denotes the authoritative region boundaries of PNSN.

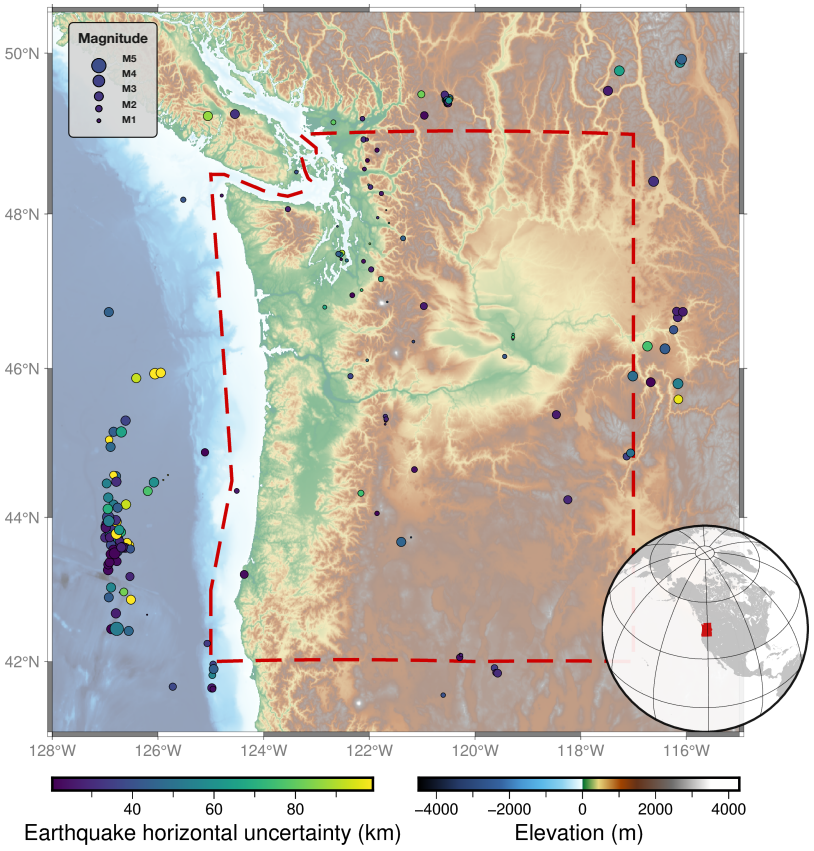

Figure S28 Location of the events with location horizontal uncertainty larger than 20 km. The red polygon denotes the authoritative region boundaries of PNSN.

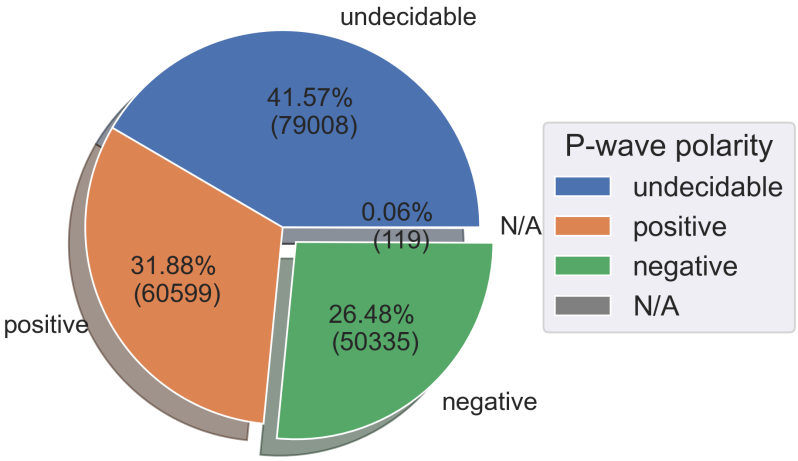

Figure S29 Percentages of the picked P-wave polarity of all streams from ComCat events.

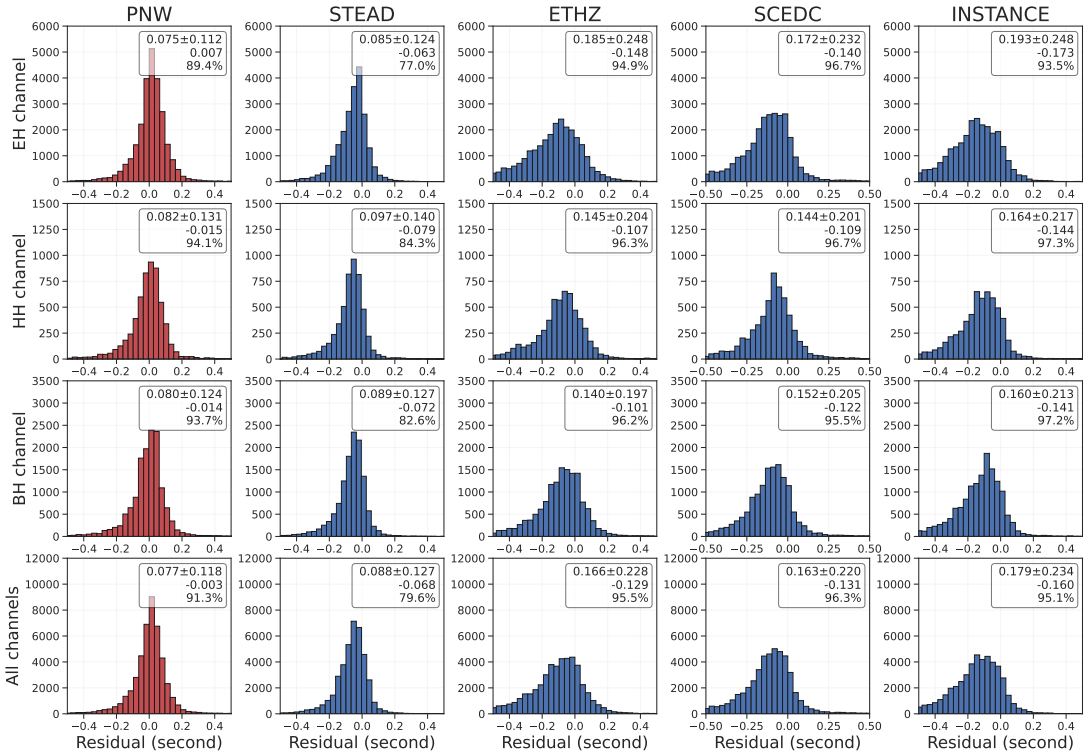

Figure S30 Histogram of P-wave picking residuals on velocity channels on the test data set. The number in the upper right corner of each subplot shows the mean absolute error (MAE), the root-mean-square-error (RMSE) as the uncertainties to the MAE, the mean value of the residual, and the picking completeness in percentage concerning the ground truth.

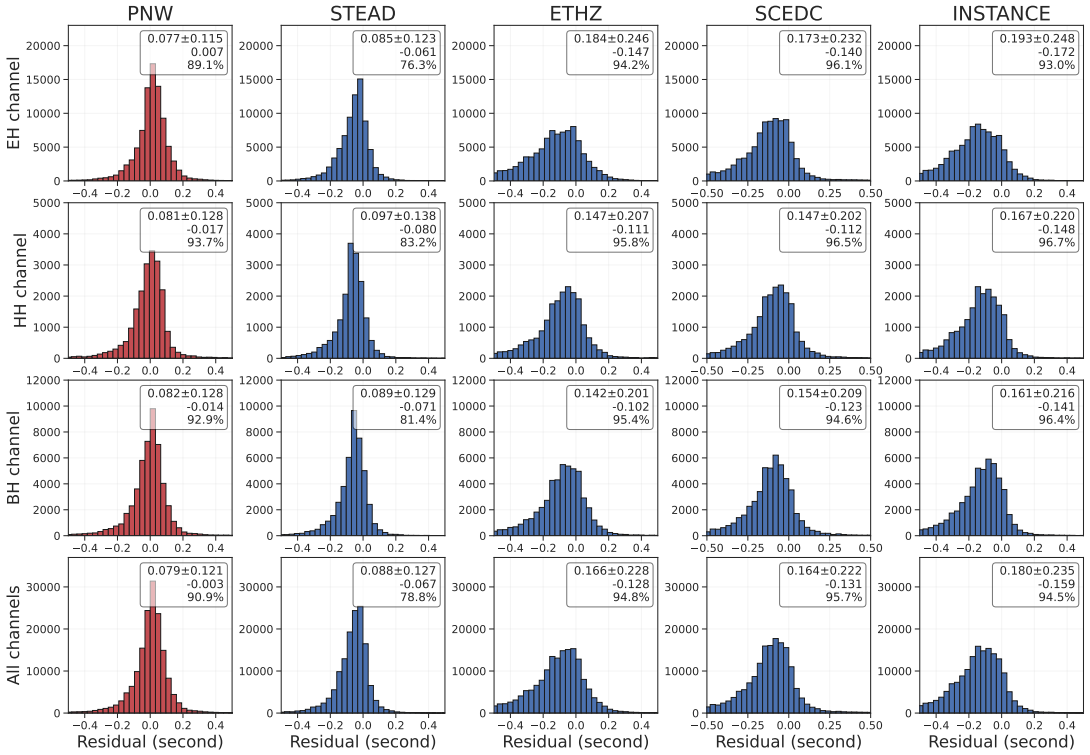

Figure S31 Histogram of S-wave picking residuals on velocity channels on the test data set. The number in the upper right corner of each subplot shows the mean absolute error (MAE), the root-mean-square-error (RMSE) as the uncertainties to the MAE, the mean value of the residual, and the picking completeness in percentage concerning the ground truth.

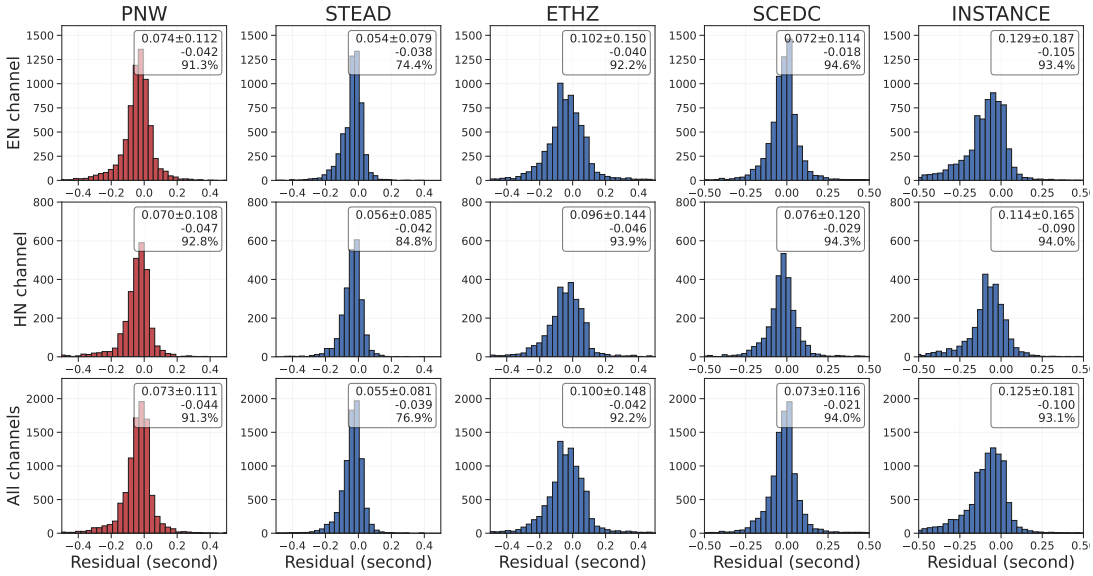

Figure S32 Histogram of P-wave picking residuals on strong motion channels. The number in the upper right corner of each subplot shows the mean absolute error (MAE), the root-mean-square-error (RMSE) as the uncertainties to the MAE, the mean value of the residual, and the picking completeness in percentage concerning the ground truth.

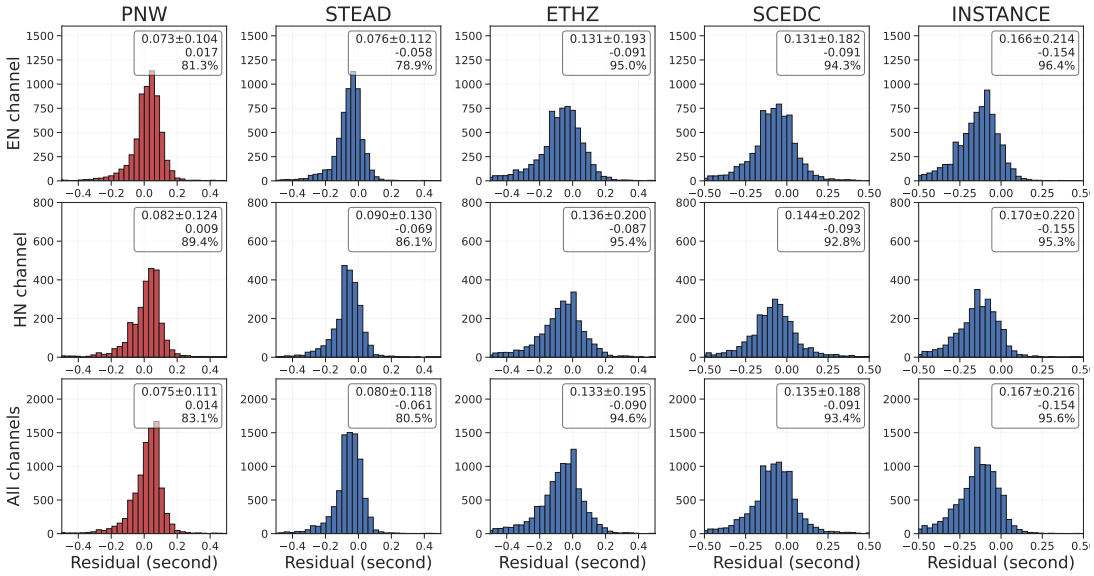

Figure S33 Histogram of S-wave picking residuals on strong-motion channels. The number in the upper right corner of each subplot shows the mean absolute error (MAE), the root-mean-square-error (RMSE) as the uncertainties to the MAE, the mean value of the residual, and the picking completeness in percentage concerning the ground truth.

References

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., et al. Seisbench—a toolbox for machine learning in seismology. *Seismological Society of America*, 93(3):1695–1709, 2022. doi: 10.1785/0220210324.