Supporting Information for "Earthquake Early Warning using 3 seconds of records on a single station"

Pablo Lara[1,2], Quentin Bletery[1], Jean-Paul Ampuero[1], Adolfo Inza[2]

[1]Université Côte d'Azur, IRD, CNRS, Observatoire de la Côte d'Azur, Géoazur, France

[2]Instituto Geofísico del Perú, Lima, Perú

Contents of this file

1. Introduction

This supporting information includes the attributes used in this work, 9 supplementary figures and 1 supplementary table.

2. Attributes

We detail here the attributes computed to train the Machine Learning algorithms, with their corresponding domain and signal.

2.1. Attributes from 3-component seismograms

1. Maximum eigenvalue λ_1 of covariance matrix from the 3C seismogram.

2. Eigenvalue factor: ratio of the maximum eigenvalue to the sum of the remaining eigenvalues:

$$\lambda_{factor} = \lambda_1/(\lambda_2 + \lambda_3). \tag{1}$$

3. The 3 components of the eigenvector ν_1 associated with the maximum eigenvalue λ_1.

2.2. Attributes from each channel

N denotes the number of samples per channel within the time window. F_s denotes the sampling rate, in Hz. The envelope of the signal s is defined as $e = |s + iH\{s\}|$, where H is the Hilbert transform.

2.2.1. Time-domain attributes

4. Maximum energy of the preprocessed signal:

$$E_{max} = \max(s^2). \tag{2}$$

5. Time at which the maximum energy is reached:

$$t_{E_{max}} = argmax(s^2). \tag{3}$$

6. Total energy:

$$E_{ne} = \sum_{n=1}^{N} s^2[n]. \tag{4}$$

7. Energy centroid time:

$$CT = \frac{1}{E_{ne}} \sum_{n=1}^{N} n\ s^2[n]. \tag{5}$$

8. Bandwidth, characteristic duration, variance around the energy centroid:

$$BW_t = \sqrt{\frac{\displaystyle\sum_{n=1}^{N} (n - CT)^2 s^2[n]}{E_{ne}}}. \tag{6}$$

9. Skewness around bandwidth:

$$Skewness_{BW_t} = \begin{cases} \sqrt{SK_{pre}}, & \text{if } SK_{pre} \geq 0 \\ -\sqrt{-SK_{pre}}, & \text{otherwise.} \end{cases}$$

where

$$SK_{pre} = \frac{\sum_{n=1}^{N} (n - CT)^3 s^2[n]}{E_{ne}\ BW_t^3}. \tag{7}$$

10. Kurtosis around bandwidth:

$$Kurtosis_{BW_t} = \sqrt{\frac{\sum_{n=1}^{N} (n - CT)^4 s^2[n]}{E_{ne}\ BW_t^4}}. \tag{8}$$

11. Mean envelope:

$$< env >= \frac{1}{N} \sum_{k=1}^{N} e[k]. \tag{9}$$

12. Ratio of maximum amplitude envelope to its mean:

$$RMM_t = \frac{\max(e)}{< env >}. \tag{10}$$

13. Standard deviation of the envelope:

$$STD_{env} = \sqrt{\frac{\sum_{k=1}^{N} (e[k] - < env >)^2}{N}}. \tag{11}$$

14. Skewness of the envelope:

$$Skewness_{env} = \frac{1}{N} \sum_{k=1}^{N} \left(\frac{e[k] - < env >}{STD_{env}} \right)^3. \tag{12}$$

15. Kurtosis of the envelope:

$$Kurtosis_{env} = \frac{1}{N} \sum_{k=1}^{N} \left(\frac{e[k] - < env >}{STD_{env}} \right)^4. \tag{13}$$

16. Threshold-crossing rate of the envelope signal: how many times per second the signal envelope crosses the threshold of 80% of its maximum amplitude:

$$TCR_t = \frac{count(r[n]r[n-1] < 0)}{N/F_s}, \tag{14}$$

where:

$$r = e/\max(e) - 0.8. \tag{15}$$

17. Fraction of envelope samples that exceed a threshold of 80% of the envelope maximum:

$$fract(TCR_{env}) = count(e \geq 0.8 \max(e))/N. \tag{16}$$

18. Shannon entropy of the envelope, with $N_{bins} = 200$.

$$Shannon_{env} = -\sum_{i=1}^{N_{bins}} Prob_e[i] \log_2(Prob_e[i]), \tag{17}$$

where:

$$Prob_e[i] = Histogram(e, N_{bins}). \tag{18}$$

19. Renyi entropy of the envelope, with $\alpha = 2$.

$$Renyi_{env} = \frac{\log_2 \sum_{i=1}^{N_{bins}} Prob_e^{\alpha}[i]}{1-\alpha}, \tag{19}$$

20. Zero crossing rate, how many times per second the signal s changes sign:

$$ZCR_t = \frac{count(s[n]s[n-1] < 0)}{N/F_s} \tag{20}$$

2.2.2. Spectral-domain attributes

Attributes extracted from $p = PSD(s)$, the Welch's Power Spectral Density of the signal s. Here N denotes the number of frequency samples in the spectrum up to the Nyquist frequency $F_s/2$.

21. Mean PSD:

$$< PSD >= \frac{1}{N} \sum_{k=1}^{N} p[k]. \tag{21}$$

22. Maximum spectral energy:

$$PSD_{max} = \max(p). \tag{22}$$

23. Frequency index of maximum spectral energy:

$$f_{PSD_{max}} = argmax(p). \tag{23}$$

24. Centroid frequency of the spectrum:

$$CF = \frac{\sum_{k=1}^{N} k\ p[k]}{\sum_{k=1}^{N} p[k]}. \tag{24}$$

25. Frequency bandwidth, variance around the spectral centroid:

$$BW_f = \sqrt{\frac{\sum_{k=1}^{N} (k - CF)^2\ p[k]}{\sum_{k=1}^{N} p[k]}}. \tag{25}$$

26. Skewness of the spectrum:

$$Skewness_{BW_f} = \begin{cases} \sqrt{SK_{pre}}, & \text{if } SK_{pre} \geq 0 \\ -\sqrt{-SK_{pre}}, & \text{otherwise.} \end{cases}$$

where

$$SK_{pre} = \frac{\sum_{k=1}^{N} (k - CF)^3\ p[k]}{BW_f^3 \sum_{k=1}^{N} p[k]}, \tag{26}$$

27. Kurtosis of the spectrum:

$$Kurtosis_{BW_f} = \sqrt{\frac{\sum_{k=1}^{N}(k - CF)^4 \; p[k]}{BW_f^4 \sum_{k=1}^{N} p[k]}}. \tag{27}$$

28. Standard deviation of the PSD:

$$STD_{PSD} = \sqrt{\frac{\sum_{k=1}^{N}(p[k] - < PSD >)^2}{N}}. \tag{28}$$

29. Skewness of PSD:

$$Skewness_{PSD} = \frac{\sum_{k=1}^{Na} \left(\frac{p[k] - < PSD >}{STD_{PSD}} \right)^3}{N}. \tag{29}$$

30. Kurtosis of PSD:

$$Kurtosis_{PSD} = \frac{\sum_{k=1}^{N} \left(\frac{p[k] - < PSD >}{STD_{PSD}} \right)^4}{N}. \tag{30}$$

31. Shannon entropy, with $N_{bins} = 50$:

$$Shannon_{PSD} = - \sum_{i=1}^{N_{bins}} Prob_p[i] \; \log_2(Prob_p[i]), \tag{31}$$

where:

$$Prob_p[i] = Histogram(p[k], N_{bins}). \tag{32}$$

32. Renyi entropy, with $\alpha = 2$:

$$Renyi_{PSD} = \frac{\log_2 \sum_{i=1}^{N_{bins}} Prob_p^{\alpha}[i]}{1 - \alpha}. \tag{33}$$

33. Ratio of maximum PSD amplitude to its mean.

$$RMM_f = \frac{\max(p)}{< PSD >}. \tag{34}$$

34. Threshold-crossing rate of the PSD, how many times the PSD crosses a threshold of 40% of its maximum amplitude:

$$TCR_f = \frac{count(r[k]r[k-1] < 0)}{N/F_s}, \tag{35}$$

where:

$$r = PSD/\max(PSD) - 0.4 \tag{36}$$

35. Relative number of samples that exceed a threshold of 40% of its maximum.

$$fract(TCR_{PSD}) = count(p \geq 0.4\max(p))/N. \tag{37}$$

2.2.3. Cepstral-domain attributes

36. The 13 first mel-frequency cepstrum coefficients (MFCC):

$$MFCC[m] = DCT\{log[\sum\{|F\{s\}|^2\Lambda_m\}]\}, \tag{38}$$

where DCT is the Discrete Cosine Transform, $F\{.\}$ is the Discrete Fourier Transform, and Λ is a triangular filter bank function linearly spaced from 1 to 45 Hz in a Mel scale. In this work, we use $m = 26$ filter banks, and are compute as in (Kopparapu & Laxminarayana, 2010).

3. Figures

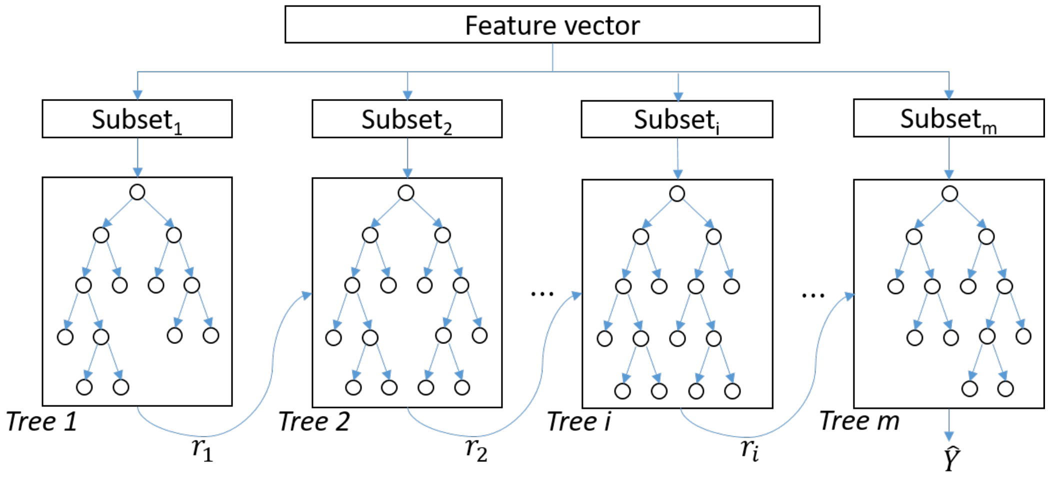

Figure S1. General architecture XGB.

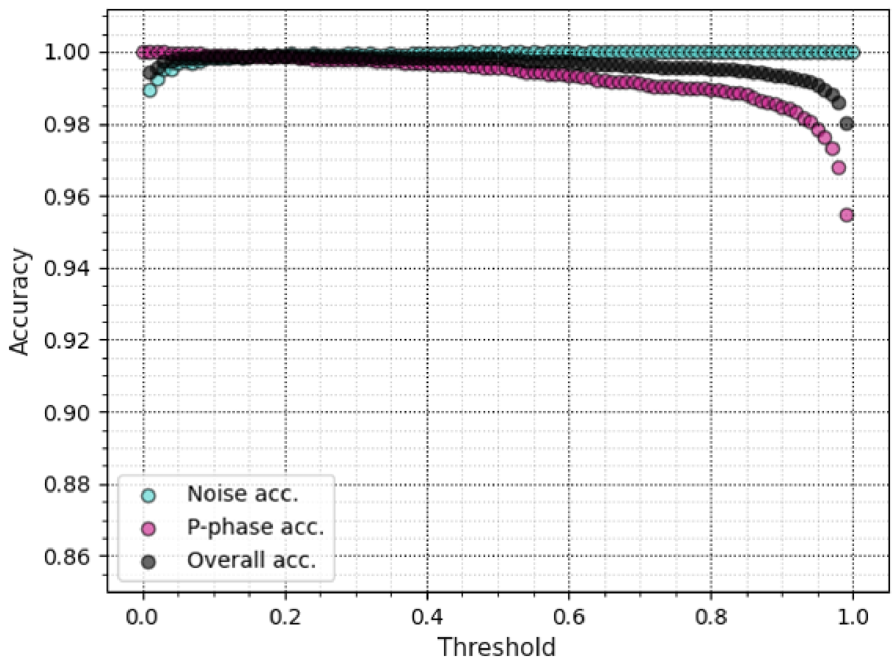

Figure S2. Accuracy of noise and earthquake classification, using different thresholds.

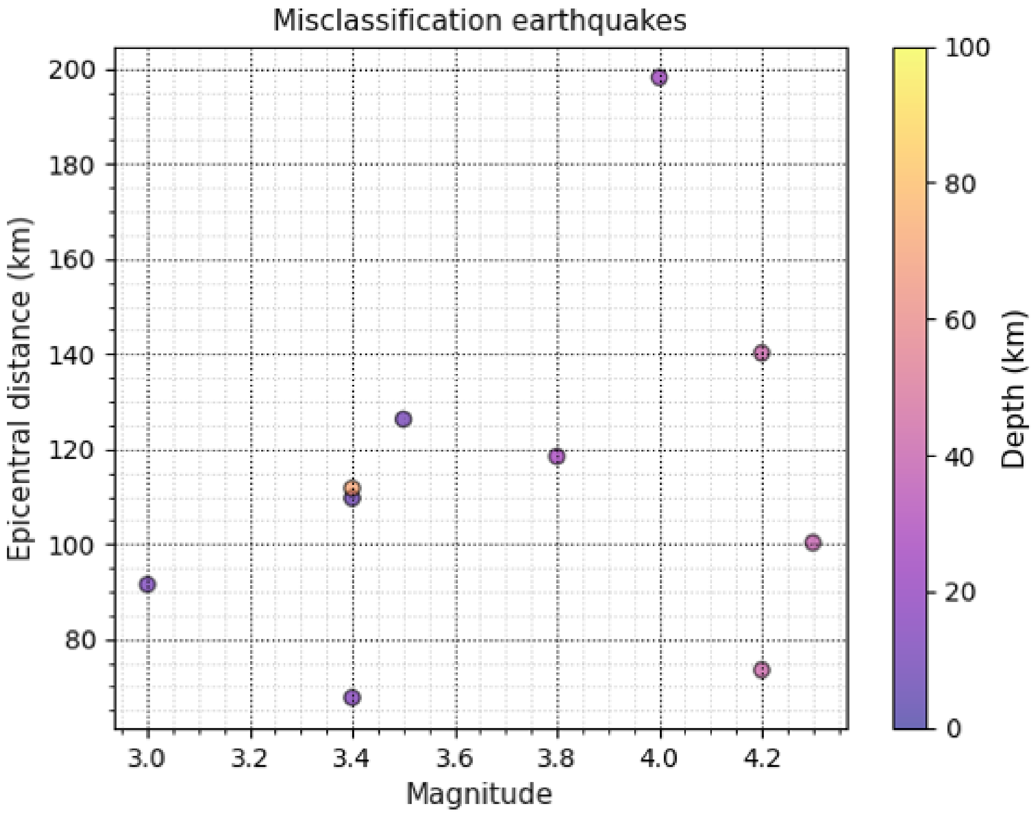

Figure S3. Magnitude, epicentral distance and depth of the misclassified signals shown in Fig. S4.

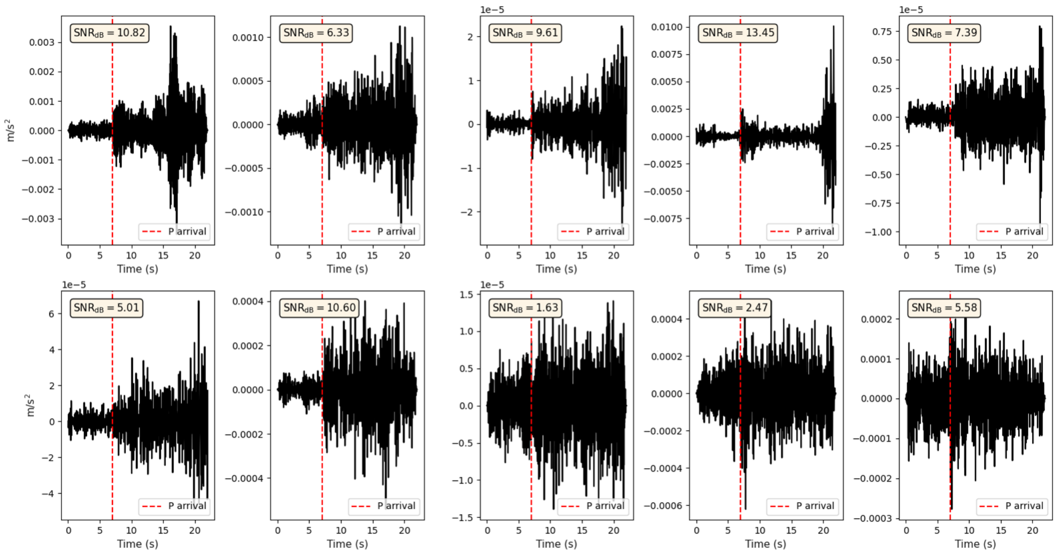

Figure S4. Earthquake signals misclassified as noise, and their signal-to-noise ratios (SNR).

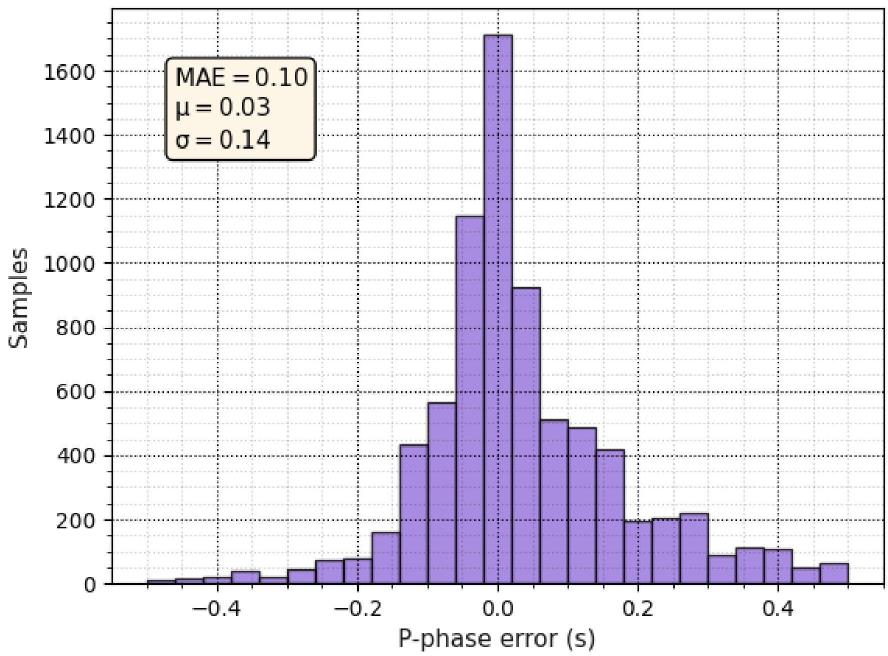

Figure S5. histogram of errors between the true and predicted P-phase arrival times.

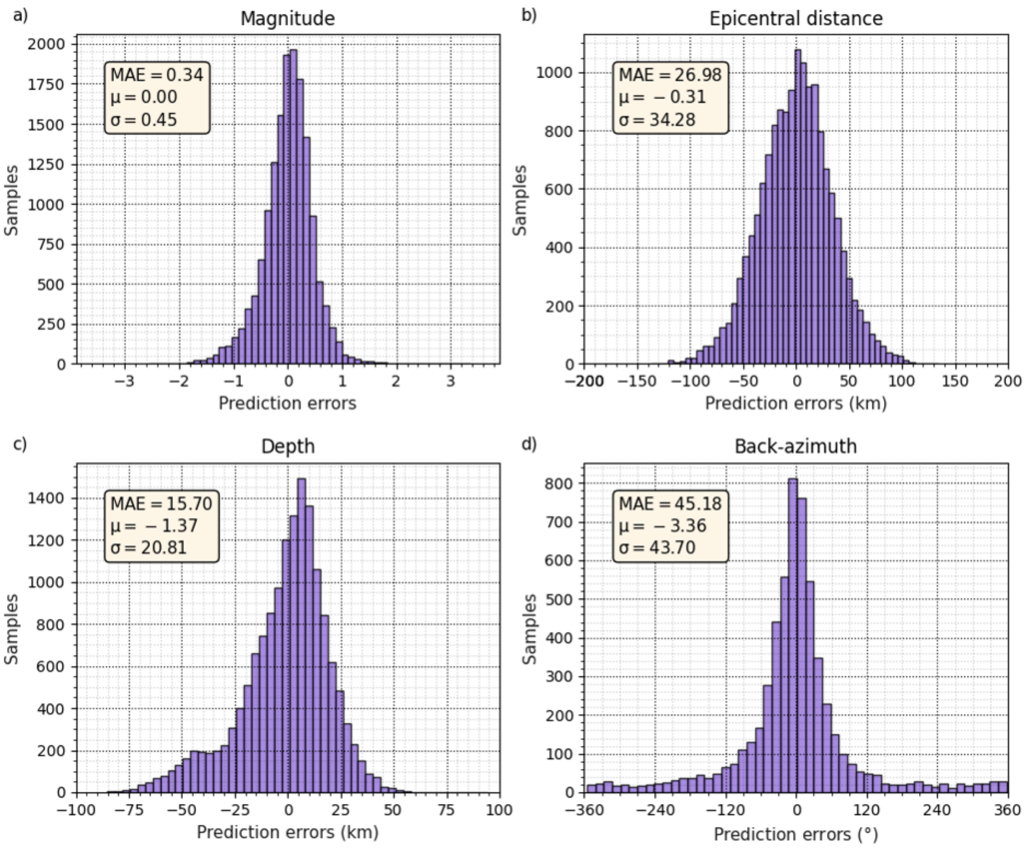

Figure S6. Histogram of the errors in the source characterization predictions using 3 s of P-wave.

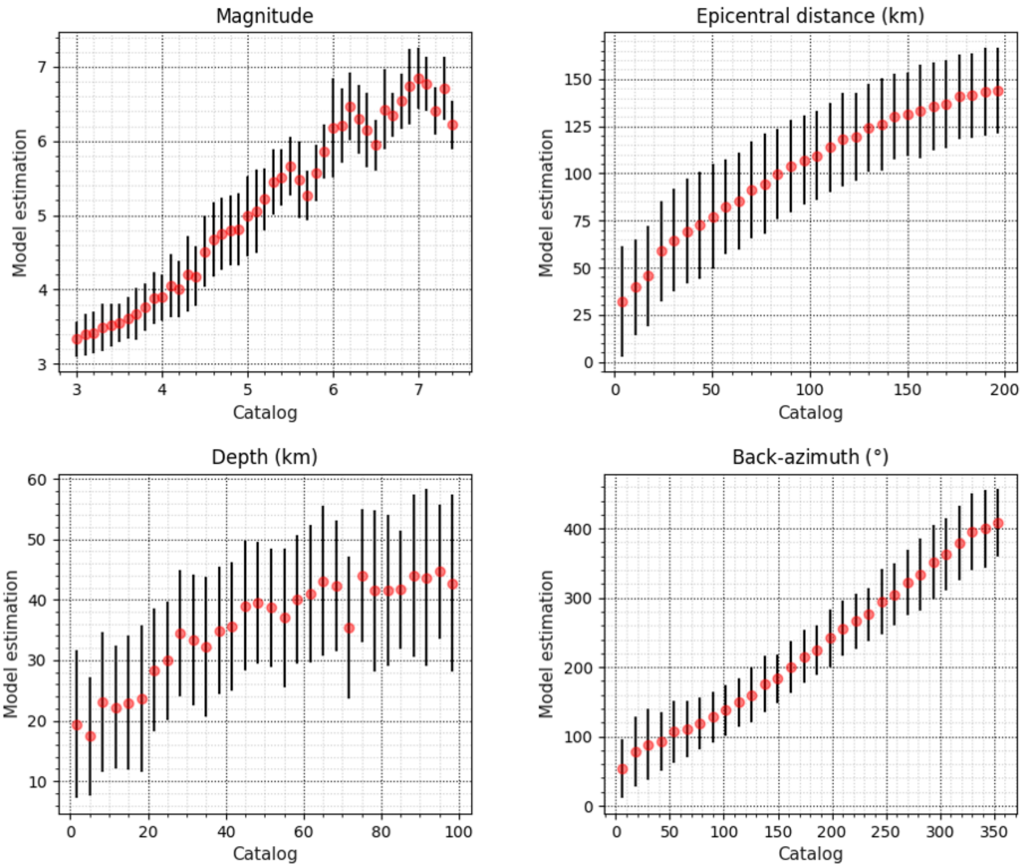

Figure S7. Mean (circle) and STD (bar) predictions per bin using 3 s of P-wave.

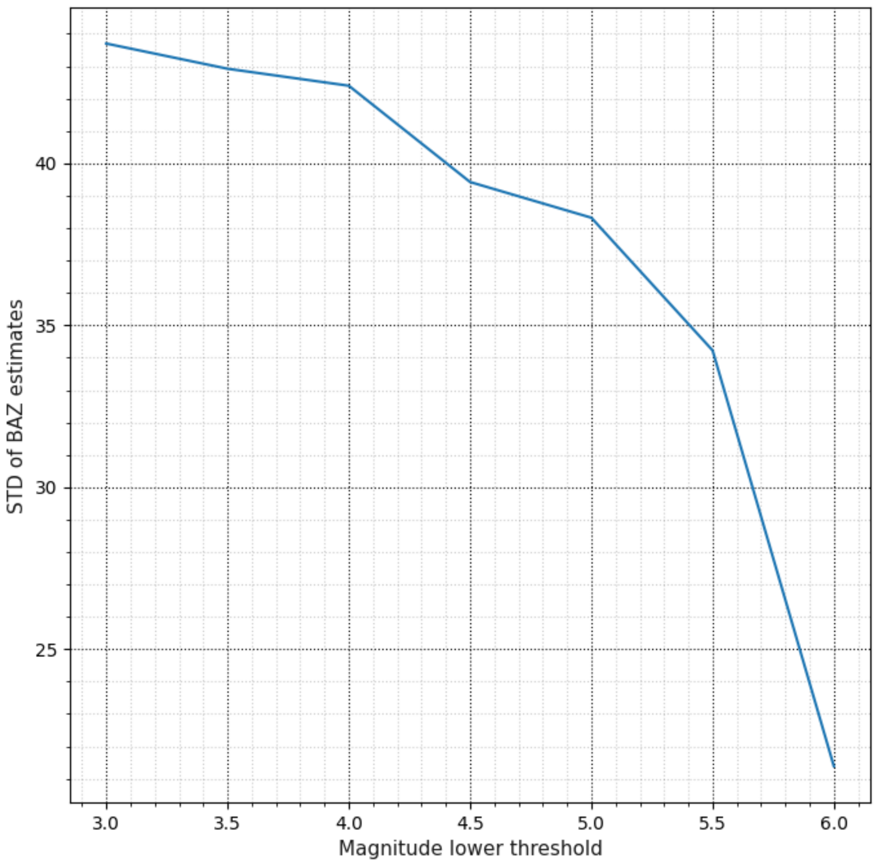

Figure S8. STD of the back-azimuth estimates, using different lower thresholds of magnitude.

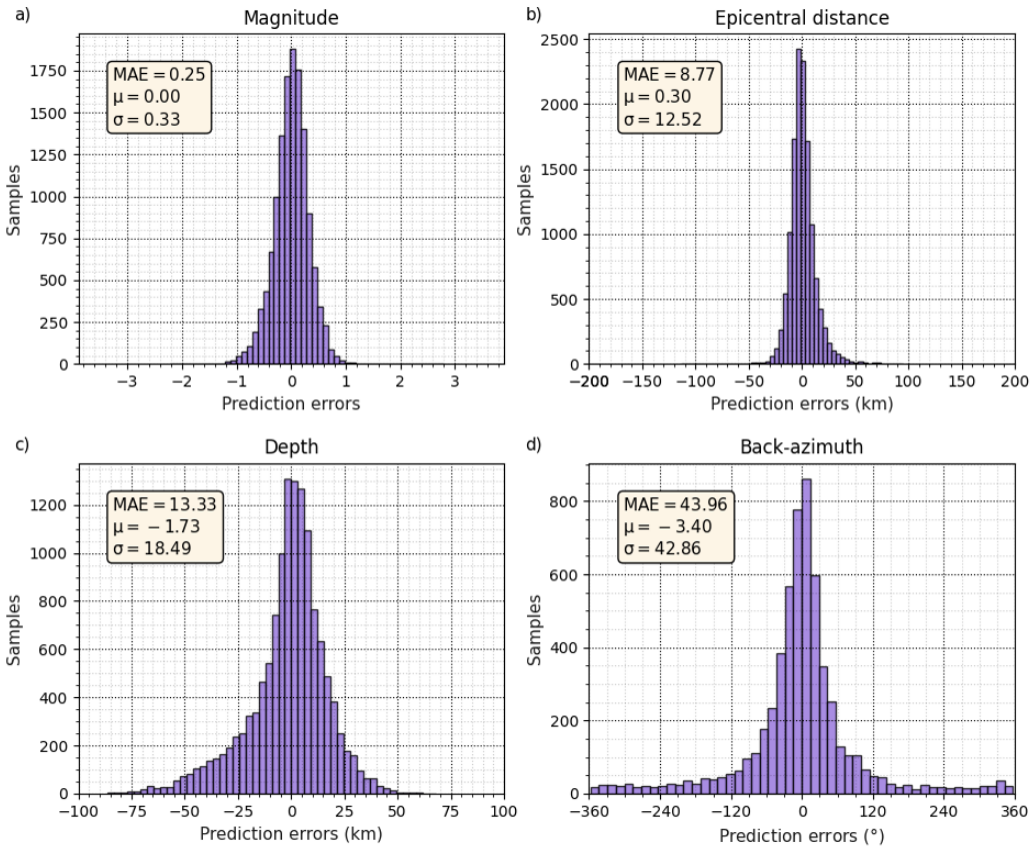

Figure S9. Histogram of the errors in the predictions using 46 s of P-wave.

Table S1. Real-time earthquake detection by E-EEWS using 3 s of P-wave in a continuous month.

Magnitude	Distance (km)	Depth (km)
3.1	175	13
3.5	179	56
3.3	195	85
3.2	145	78
3.2	173	54
3.3	89	50
3.3	98	86
3.2	163	17
3.7	162	19
3.3	159	84
3.8	155	10
3.0	47	35
3.4	189	88
3.5	138	47
3.5	97	49

References

Kopparapu, S. K., & Laxminarayana, M. (2010). Choice of mel filter bank in computing mfcc of a resampled speech. In *10th international conference on information science, signal processing and their applications (isspa 2010)* (pp. 121–124).