Focal mechanisms of small earthquakes beneath the whole Japan Islands based on first-motion polarities picked using deep learning

Takahiko Uchide[1,1,1,1,1]

[1]National Institute of Advanced Industrial Science and Technology

November 30, 2022

Abstract

Knowledge of crustal stress field is essential for understanding tectonics and earthquake generation. One way to estimate the crustal stress field is based on focal mechanisms of earthquakes. This study investigated focal mechanisms of ˜ 110 thousand microearthquakes in Japan Islands shallower than 20 km based on the first-motion polarities picked by a simple neural network model, which was trained using two data sets: moderate to large earthquakes all over Japan and microearthquakes in two regions in Japan. The threshold of the confidence score from the neural network model was so chosen as to maximize the overall quality of focal mechanism solutions. The P- and T-axes of the numerous focal mechanism solutions provide more detailed distributions of crustal stress field. For example, in Chugoku region, there exist slight differences in the trend of P-axes azimuths between the northern and southern areas are observed, corresponding to the geodetic observations in space.

1   # Focal mechanisms of small earthquakes beneath the

2   # Japanese islands based on first-motion polarities picked

3   # using deep learning

4   Takahiko Uchide[1,*]

5   [1] Geological Survey of Japan, National Institute of Advanced Industrial Science and

6   Technology (AIST)

7   Abbreviated Title: Focal Mechanisms in the entire Japanese islands

8   * Corresponding Author: Takahiko Uchide (t.uchide@aist.go.jp)

9 Summary

10 Knowledge of crustal stress fields is essential for understanding tectonics and

11 earthquake generation. One approach for estimating the crustal stress field is based on

12 the focal mechanisms of earthquakes. This study investigated the focal mechanisms of

13 approximately 110,000 microearthquakes in the area of the Japanese islands that

14 occurred at a depth shallower than 20 km, based on the first-motion polarities picked by

15 a simple neural network model. The model was first trained using a data set of mainly

16 moderate to large earthquakes throughout Japan. Following on, the model was

17 re-trained using a data set of microearthquakes in two regions of Japan. The threshold of

18 the confidence score from the neural network model was chosen to maximize the

19 overall quality of the focal mechanism solutions. The P- and T-axes of the numerous

20 focal mechanism solutions provided more detailed distributions of the crustal stress

21 field. For example, in the Chugoku region, small differences were observed in the trend

22 of P-axes azimuths between the northern and southern areas, spatially corresponding to

23 geodetic observations. The results of this study are useful for revealing the crustal stress

24 field, and, as such, for assessing past and current tectonic activities and potential future

25 earthquake generation.

26 Keywords

27 ● Earthquake source observations

28 ● Seismicity and tectonics

29 ● Neural networks, fuzzy logic

1. Introduction

Crustal stress field data is crucial for understanding tectonics and seismic activity; however, measuring the stress at depths over a large area is challenging. Direct measurements at specific boreholes (e.g., Wu *et al.* 2007; Huffman *et al.* 2016; Brodsky *et al.* 2017; Townend *et al.* 2017) offer detailed information but only as it concerns one spot. In contrast, seismology provides indirect measurements for a wider area but includes more uncertainty. The focal mechanisms, which indicate the fault plane and slip directions, specify the orientation of seismogenic stress. The World Stress Map (Heidbach *et al.* 2008; Heidbach *et al.* 2016; Heidbach *et al.* 2018) presents the global compilation of this information.

In the past, routinely determined moment tensor solutions were used for estimating regional stress fields (Terakawa and Matsu'ura 2010; Hardebeck 2015). However, non-estimated areas remain, even in seismically active areas such as the Japanese islands. More complete knowledge of the seismogenic stress field requires focal mechanisms for microearthquakes, particularly in low seismicity areas (e.g., Imanishi *et al.* 2011; Imanishi *et al.* 2012; Matsumoto *et al.* 2015). Comprehensive investigations of microearthquake focal mechanisms contributed to revealing the regional stress field (e.g., Iio *et al.* 2018; Imanishi *et al.* 2019).

The focal mechanisms of moderate or larger earthquakes can be automatically determined using the full waveform from the local (e.g., Dreger and Helmberger 1993; Fukuyama *et al.* 1998) or global (e.g., Ekström et al. 2012) seismic network. However, the mechanisms of small earthquakes cannot be similarly determined due to the difficulty of modeling high-frequency seismograms. Instead, the first-motion polarity is

53   typically used in this regard, where the vertical component initially moves either

54   upward or downward. Automatic polarity-picking methods include an approach based

55   on the sign of the first extremum after the P arrival (Nakamura 2004; Chen and Holland

56   2016). Pugh et al. (2016a) proposed a Bayesian approach using the first extremum and a

57   probability function of P arrival time. Recently, deep learning allowed for automatically

58   picking the polarity (Ross *et al.* 2018; Hara *et al.* 2019). Deep learning has the

59   approximation capabilities of a multilayer neural network to arbitrary functions (Hornik

60   1991). This presents an advantage in the polarity-picking problem, which involves the

61   polarity of the first pulse and the various forms of seismic noise.

62   This study aimed to obtain focal mechanism solutions for the Japanese islands, one of

63   the most seismically active regions in the world. The first-motion polarities were picked

64   using a neural network model and seismic data from nationwide seismic networks.

65   Finally, the focal mechanism solutions and spatial trends in P- and T-axes were

66   reviewed.

67   ## 2. Training the Neural Network Model

68   ### 2.1. Design of the Neural Network Model

69   Fig. 1 summarizes the neural network used in this study. The input of the neural

70   network model was a 256-sample-long seismogram in which the 156th sample

71   corresponded to the P-arrival time already picked either manually or automatically. The

72   output comprised two scores (confidence scores) corresponding to the upward and

73   downward polarities. Note that, in the case of Southern California, Ross et al. (2018)

74   classified the polarity as "up," "down," and "unknown"; however, in this study, the

75 "unknown" class was not set. The data set included many seismograms with impulsive

76 onset but no polarity information (Fig. 2), and the lack of polarity information did not

77 indicate "unknown" in this case. Thus, no training data set was available for the

78 "unknown" class and such a class could not be set for the present study. In the cases

79 with first-motion polarities being ambiguous, the confidence scores for both polarities

80 are expected to have similar values. A confidence threshold was set for this research at a

81 subsequent stage to accept suggested polarity information for determining the focal

82 mechanism.

83 A simple neural network model (Fig. 1) similar to those used in prior studies (Ross *et*

84 *al.* 2018; Hara *et al.* 2019) was designed. The neural network model started with two

85 convolution layers, followed by three units comprising convolution, batch normalization

86 (Ioffe and Szegedy 2015), and the maximum pooling layers. The model ended with two

87 fully connected layers. The kernel size of the convolutional layers was 11. For all but

88 the final layers, the activation function was the rectified linear units (known as "ReLU")

89 (Nair and Hinton 2010); the SoftMax function was chosen for the final layer:

93
$$\mathrm{softmax}(\mathbf{z})_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}, \tag{1}$$

90 where $\mathbf{z} = (z_1, z_2)$ is the output of the final layer corresponding to the upward and

91 downward polarities, respectively. The final outputs (confidence scores) are

92 non-negative and their summation is always 1.

94 ## 2.2. Training Procedure

95 Generally, the training of neural network models requires training, validation, and test

96 data sets. In this study, the data sets comprised the seismograms and the polarity. The

97 training data set was used for optimizing the parameters in the neural network model.

98 The initial values of the model parameters were randomly set. The negative

99 log-likelihood function was employed as a loss function to quantify the difference

100 between the output of the model and the polarity expectation from the training data set,

101 i.e., (1, 0) for the upward polarity and (0, 1) for the downward polarity. The value of the

102 loss function was simply called a loss. The parameters of the neural network model

103 were updated by back-propagating the loss (Rumelhart et al. 1986) optimized by the

104 adaptive moment estimation (known as "Adam") method (Kingma and Ba 2014). This

105 optimization was iteratively performed to decrease the loss.

106 The problem to be carefully addressed is the overfitting problem that the trained

107 model works well with the training data but poorly with unseen data. This study

108 addressed the overfitting problem in two ways. One was by using the dropout technique

109 (Srivastava et al. 2014), where 50% of randomly selected perceptions were muted

110 during the training. The second approach involved performing cross-validation using a

111 validation data set that was independent of the training set. If the loss for the validation

112 data set increased as the training progressed, while that for the training data set

113 decreased for 10 iterations (epochs), the training was stopped.

114 Following the completion of the training, the performance of the trained model was

115 examined using another data set (test data set) of small to microearthquakes, as the

116 present study focused on such small events.

117 Since the data set of microearthquakes available in this study was small, training of

118 the neural network model was performed in two stages. In the first stage, training was

119 carried out using a large data set of moderate to large earthquakes to learn the

120   polarity-picking problem in general. The second stage employed microearthquake data

121   for re-training the model. This reflected a fine-tuning approach (Girshick et al. 2014)

122   during which all the parameters were optimized in both stages. The merit of this

123   two-stage approach will be illustrated later by the results of the training.

124   2.3. Data

125   In the first stage, the Hi-net and the Japan Meteorological Agency (JMA) data of

126   19,341 earthquakes with P arrival and polarity data in the JMA catalog were used. Most

127   of these earthquakes were larger than M 3 (Fig. 3). All the data were then spatially

128   divided into the training and validation data sets (Fig. 4a; Table 1). Here, we should care

129   the data leakage problem that the validation data set includes similar or identical data to

130   one included in the training data set and cannot properly detect the overfitting problem.

131   The spatial division of the data set was adopted in order to reduce the risk that similar

132   data from collocated events at the same stations are included both in the training and

133   validation data sets.

134   In the second stage, the model was trained using the P arrival time and polarity of

135   microearthquakes in the Kanto and Chugoku regions, which were manually picked by

136   Geological Survey of Japan, National Institute of Advanced Industrial Science and

137   Technology (AIST). The Kanto data was also used by Imanishi et al. (2019). The

138   majority of earthquakes in these data sets were smaller than M 3 (Fig. 3). The

139   magnitude–frequency distributions were similar to that of the target data set (see Section

140   3), which will be introduced later. The spatial distributions of events in the Kanto and

141   Chugoku regions are shown in Figs. 4b and 4c, respectively. The number of seismogram

142   sets and earthquakes is summarized in Table 1.

143   In both stages, seismograms of the vertical component were used. The seismograms

144   had 256 samples, i.e., 156 samples before and 100 samples after P arrival. The samples

145   were 2.56 s long, as the data was sampled at 100 Hz. Low-frequency noise was

146   removed by applying a high-pass filter at 1 Hz. The initial portion of the P-wave was

147   emphasized by clipping seismograms at a particular threshold. Finally, each seismogram

148   was normalized by its maximum value. Seismograms with amplitudes larger than the

149   clipping threshold were normalized by the clipping threshold.

150   Furthermore, the data was augmented four times by flipping the seismograms upside

151   down and by using time-shifting. The flipping procedure equalized the number of

152   upward and downward polarity data. The time shift addressed the potential

153   misalignment of data due to uncertainties in the arrival-time picking. Various values of

154   the clipping threshold and the time-shift range were subsequently examined.

155   ## 2.4. Results

156   A hundred cases with randomly selected clipping thresholds in the range of 10^{-6} to

157   10^{-4} m/s and with the half-width of the time-shift ranging 0–30 samples were examined.

158   Note that, since the amplitudes of most of the seismograms were smaller than 10^{-4} m/s,

159   in the case of such a large clipping threshold, seismograms were not clipped and were

160   simply normalized by their maximum values. The results were evaluated based on the

161   loss value for the test data set. The results indicated that the shorter the half-width of the

162   time-shift range, the smaller the loss (Fig. 5a). The clipping threshold did not correlate

163   with the loss value (Fig. 5b).

164   Hereafter, time-shift was not applied and 10^{-5} m/s was used as the clipping threshold.

165   The neural network model was trained using these values.

166   The performance of the trained neural network model was quantified based on

167   precision and recall. The precision represented how reliable the model prediction was.

168   By focusing on one of the classes (upward or downward), the precision is defined as:

177
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

169   where TP represents a true positive, where the neural network model correctly predicts

170   the designated class, and FP is a fault positive, where the model incorrectly predicts the

171   designated class. Thus, TP and FP for the upward and downward polarities are:

178
$$\begin{aligned} \text{TP}_{\text{up}} &= N_{\text{uu}}, & \text{FP}_{\text{up}} &= N_{\text{du}}, \\ \text{TP}_{\text{down}} &= N_{\text{dd}}, & \text{FP}_{\text{down}} &= N_{\text{ud}}, \end{aligned} \tag{3}$$

172   where N_{ij} is of class i and predicted as class j, and "u" and "d" indicate the upward and

173   downward polarities, respectively. Here, this study acknowledges that the model

174   predicted the upward and downward polarity only when the confidence scores were

175   larger than the confidence threshold. Otherwise, the model prediction is considered to

176   have been ambiguous.

179   The recall represents how much data of the designated class is recovered by the

180   neural network model, and is defined as:

185
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

181   where FN is a false negative, where the model incorrectly predicts the designated class,

182   and

186
$$\text{FN}_{\text{up}} = N_{\text{ud}} + N_{\text{ua}}, \qquad \text{FN}_{\text{down}} = N_{\text{du}} + N_{\text{da}}, \tag{5}$$

183   where the subscript "a" indicates ambiguous, where the confidence score is lower than

184   the confidence threshold.

9

187   As the precision and the recall are functions of the confidence threshold, a curve of

188   precision and recall can be drawn as shown in Fig. 5c. The better the model

189   performance, the closer to the top right corner the precision-recall curve will be.

190   Comparing the curves for the model trained only by the JMA data set and for the model

191   trained also by the microearthquake data set, the latter showed observably better results.

192   The performances of the models for upward and downward polarities were almost the

193   same.

194 3. Application to Crustal Earthquakes in Japan

195   This study applied the trained model to the target data that is event data of

196   earthquakes that occurred in the period 2005–2019 at depths of less than 20 km within

197   the coastline in Japan islands, excluding the events for which polarity information was

198   already available in the JMA catalog (Table 1). The study employed 2,301,584

199   seismogram sets of 114,525 events from Hi-net and the JMA seismic network, with

200   P-wave arrival times taken from the JMA catalog. The magnitude–frequency

201   distribution of earthquakes analyzed is shown in Fig. 3. Preprocessing was done in the

202   same manner as in the model training. Good results were obtained for polarity picking

203   with high confidence scores (Fig. 6), even in noisy cases.

204   The focal mechanisms were determined using polarity information with confidence

205   scores larger than a confidence threshold and the HASH code (Hardebeck and Shearer

206   2002, 2008). The quality of focal mechanisms depended on the confidence threshold

207   (Fig. 5d). If the threshold was too high, the very small number of polarity picks could

208   not properly constrain focal mechanisms. If the threshold was lower than 0.7, the

209 fraction of A and B ranks given by the HASH code (Hardebeck and Shearer 2008) was

210 almost constant. This study adopted a confidence threshold of 0.7 and, accordingly,

211 1,800,937 polarity picks were used.

212 Fig. 7 shows the estimated focal mechanisms and their P- and T-axes, in addition to

213 the NIED F-net moment tensor solutions for reference. The focal mechanisms of

214 114,525 events were estimated, while those of 5,715 events were undetermined because

215 the number of seismic stations was smaller than eight. Ranks A, B, C, and D, given by

216 the HASH code (Hardebeck and Shearer 2008), were assigned to 1.582, 22.604, 43,486,

217 and 41,138 events, respectively. The focal mechanism solutions covered much more

218 space than those in a routine catalog.

219 ## 4. Discussion

220 ### 4.1. Quality of the First-motion Polarity Picking and Focal Mechanism Solutions

221 Somewhat surprisingly, the narrower the time-shift range of the data, the better the

222 model performed (Fig. 5a). This was because the time-shifting made the model more

223 flexible and robust to uncertainties in arrival-time picking. There were two potential

224 reasons for this; one was that the arrival times in the test data were accurate because of

225 careful review by an analyst, and, accordingly, the time-shift was not required. Another

226 possible reason was that the training was not sufficient due to the shortage of

227 microearthquake training data.

228 The stability of the performance of the trained model was measured by, in addition to

229 the test data set, a comparison of the confidence scores for similar seismic data. First,

230 using a repeating earthquake catalog (Igarashi 2020), 1,673 repeating earthquake pairs

231   were taken. For each event pair, seismogram pairs from identical stations were included.

232   In total, 26,228 pairs of seismograms were selected. Next, the confidence scores of the

233   pairs were compared using the product of the modified confidence scores as a measure

234   of similarity, where the modified confidence score was confidence score of the

235   downward polarity subtracted from that of the upward polarity, to establish their range

236   from −1 to 1. The histogram of the products is shown in Fig. 8. Most of the products

237   were close to 1, which meant a high similarity for the outcome of the trained model.

238   Some cases were close to −1. Examples of the seismograms in such cases are shown in

239   Fig. 9, where some had noticeable opposite polarities, while others may have been

240   affected by noise. This investigation suggested that the model worked well in most

241   cases, although room for improvement remains.

242   Determination of focal mechanisms from the first-motion polarities of P-waves

243   picked by the trained neural network model was also important for assessing the quality

244   of polarity picking. In this study, 40.9% of the focal mechanism solutions were ranked

245   D or "undetermined". In a study on the determination of the focal mechanisms of

246   earthquakes in Southern California using manually picked P-wave polarity and the

247   amplitude ratio of P and S waves (Yang et al. 2012), the results showed that 56.6%

248   (101,309 among 178,899) of events were ranked D, comparable to the result of the

249   present study. Thus, the present study yielded a reasonable quality of P-wave

250   first-motion polarity picking; however, based on several factors, this comparison is not

251   entirely straightforward. These factors include differences in observational conditions

252   such as the magnitude range and station density. Focal mechanism determination can be

253   improved in several ways, e.g., by the introduction of P-wave amplitude (e.g.,

254   Matsushita and Imanishi 2015; Pugh *et al.* 2016b) and the ratio of P- and S-wave

255   amplitudes (Hardebeck and Shearer 2003; Yang *et al.* 2012), alongside the application

256   of advances in P-wave polarity picking.

257   The quality of the focal mechanism solutions is shown by region in Fig. 5e. In

258   particular, the quality in the Hokkaido region was much worse than in other regions.

259   The reason for this was examined by focusing on the number of stations. First, the

260   quality of the focal mechanism solutions showed a good correlation with the number of

261   stations (Fig. 5f). Next, the number of usable stations was smaller in Hokkaido than in

262   other regions. This was likely because of the spatial density of seismic stations (Fig. 7e).

263   Hence, it was more difficult to determine the focal mechanisms in Hokkaido than

264   elsewhere.

265   4.2. Implications for the Stress Field

266   The spatial distribution of the P- and T-axis azimuths (Figs. 7c and 7d) suggested the

267   placement of crustal stress field in the Japanese islands. In this regard, first the

268   consistency of this study using the F-net catalog on a large scale is confirmed, before

269   briefly discussing the advantages of this study, particularly at small spatiotemporal

270   scales.

271   The overall features in the spatial distributions of the P- and T-axis azimuths from

272   this study and the F-net catalog were similar. Note that the events investigated by this

273   study and those included in the F-net catalog were largely different, as implied by the

274   different magnitude distributions (Fig. 3). Accordingly, an overview of the spatial trend

275   is provided by introducing two parameters from directional statistics, i.e., mean

276   direction and circular standard deviation (CSD) (e.g., Ley and Verdebout 2017). The

277 mean direction of directional data $(\theta_1, \theta_2, \ldots, \theta_N)$ is defined as $\bar{\theta}$, which satisfies the

278 following equations:

290
$$\cos \bar{\theta} = \bar{C} = \frac{1}{N} \sum_{i=1}^{N} \cos \theta_i, \qquad \sin \bar{\theta} = \bar{S} = \frac{1}{N} \sum_{i=1}^{N} \sin \theta_i. \tag{6}$$

279 The CSD, V, in the radian is defined as:

291
$$V = (-2 \ln \bar{R})^{1/2} \tag{7}$$

280 where the mean resultant length, $\bar{R}$, is:

292
$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} \tag{8}$$

281 The mean directions of P- and T-axis azimuths from this study and the F-net catalog at

282 0.5-degree grids are shown in Fig. 10. The differences in the mean directions from this

283 study and those in the F-net catalog were ~7°–8°, which were within the CSD (~24°) of

284 the present study and thus not significant. No spatial trends were observed in the

285 differences. Interestingly, the CSD of the P- and T-axes mean directions were

286 systematically larger in the northeast Japan than in the southwest Japan. This was likely

287 because of actual variation in the axes, rather than uncertainty in the focal mechanism

288 determination because the qualities of the focal mechanism solutions did not include

289 such a trend (Fig. 5e).

293 In the Chugoku region, the P-axes struck in the east–west direction in the northern

294 area (San-in area), whereas those that struck in the northwest–southeast direction did so

295 in the southern area (Sanyo area). The contrast between P-axes azimuths in western

296 Tottori was reported by Kawanishi et al. (2009). The present study also indicated a

297 similar trend for the entire Chugoku region. This contrast geographically corresponded

298 to the San-in shear zone (Meneses-Gutierrez and Nishimura 2020). A combination of

299   the current study alongside geodetic implications will further help to enhance an

300   understanding of seismotectonics.

301   Local abnormal stress fields can be detected by focal mechanisms. In the Kansai

302   region (Fig. 11), the majority of earthquakes exhibited east–west striking P-axes and

303   north–south T-axes, while groups of events with different trends were also observed, i.e.,

304   one group of normal faulting events with P-axes in the north–south direction, and

305   another group of events with northeast–southwest P-axes and northwest–southeast

306   T-axes. Knowing the origins of such local stresses will be helpful to tectonic studies.

307   Rich focal mechanism data will also assist the study of temporal changes in the stress

308   field. Fig. 12 shows the orientations of the T-axes before and after the 2016 Kumamoto

309   earthquake in central Kyushu. T-axes tend to be in the north–south direction, as

310   indicated by existing studies (e.g., Matsumoto *et al.* 2015; Savage *et al.* 2016). Small

311   changes in the T-axes azimuths at the 2016 Kumamoto earthquake may have been due

312   to, e.g., the stress rotation effected by large foreshocks and the mainshock (Yoshida et al.

313   2016), or as a result of different fault geometries. Additional investigation is needed in

314   this regard.

315   Following the 2011 Tohoku-oki earthquake, several events occurred with T-axes in

316   the east–west direction along the Pacific coast of the Kanto and Tohoku regions,

317   whereas the majority of events in the entire Kanto and Tohoku regions had P-axes in

318   this direction. This seismicity was activated following the 2011 Tohoku-oki earthquake.

319   The focal mechanisms of microearthquakes prior to the 2011 Tohoku-oki earthquake

320   were similar to those following the same earthquake in the northern Ibaraki prefecture,

321   the Fukushima Hamadori area (Imanishi *et al.* 2012; Otsubo *et al.* 2018), and the

322   Kitakami mountain range (Yoshida et al. 2015). In the present study, the small number

323   of focal mechanisms in the Boso peninsula before the Tohoku-oki earthquake were

324   similar to those after this earthquake (Fig. 13). The east–west extensional stress field in

325   these areas may have existed before the 2011 Tohoku-oki earthquake.

326   Despite the dramatic increase in focal mechanism solutions, there remain unstudied

327   areas among the Japanese islands, where seismicity is relatively low. Accordingly, this

328   type of study may be required for even smaller earthquakes, a more complex challenge

329   than that addressed in the present study. Additional campaign seismic observations may

330   improve the focal mechanism solutions for very small earthquakes. Furthermore,

331   combining these observations with research that includes geological, geographical, and

332   geodetical observations will improve our understanding of the crustal stress field and its

333   origin.

334 5. Conclusions

335   In this study, the focal mechanisms of small to microearthquakes were estimated for

336   gaining a better understanding of the crustal stress fields in the area of the Japanese

337   islands. The focal mechanisms were derived using P-wave first-motion polarities,

338   picked by a neural network model that takes the vertical component of seismograms

339   with P arrival times as input. The focal mechanisms of almost all microearthquakes in

340   the entire area of the Japanese islands were successfully determined. The focal

341   mechanism solutions were generally consistent with the stress regime on a large scale.

342   For example, the slight but observable differences in the P-axis azimuths in the northern

343   and southern parts of the Chugoku region were consistent with the geodetic

344   observations for this region. The results of this study will be useful for revealing the

345   crustal stress field, and, accordingly, for assessing past and current tectonic activities

346   and future earthquake generation.

Acknowledgements

367   AIST was employed.

368   References

369   Amante, C. & Eakins, B.W., 2009. ETOPO1 1 Arc-Minute Global Relief Model:

370   Procedures, Data Sources and Analysis. *in NOAA Technical Memorandum*

371   *NESDIS NGDC-24* National Geophysical Data Center, NOAA,

372   doi:10.7289/V5C8276M.

373   Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010.

374   ObsPy: A Python Toolbox for Seismology, *Seismol. Res. Lett.,* 81**,** 530-533,

375   doi:10.1785/gssrl.81.3.530.

376   Brodsky, E.E., Saffer, D., Fulton, P., Chester, F., Conin, M., Huffman, K., Moore, J.C. &

377   Wu, H.-Y., 2017. The postearthquake stress state on the Tohoku megathrust as

378   constrained by reanalysis of the JFAST breakout data, *Geophys. Res. Lett.,* 44**,**

379   8294-8302, doi:10.1002/2017GL074027.

380   Chen, C. & Holland, A.A., 2016. PhasePApy: A Robust Pure Python Package for

381   Automatic Identification of Seismic Phases, *Seismol. Res. Lett.,* 87**,** 1384-1396,

382   doi:10.1785/0220160019.

383   Dreger, D.S. & Helmberger, D.V., 1993. Determination of source parameters at regional

384   distances with three-component sparse network data, *J. Geophys. Res.,* 98**,**

385   8107-8125, doi:10.1029/93JB00023.

386   Ekström, G., Nettles, M. & Dziewoński, A.M., 2012. The global CMT project 2004–

387   2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Plant. Inter.,*

388   200–201**,** 1-9, doi:10.1016/j.pepi.2012.04.002.

389   Fukuyama, E., Ishida, M., Dreger, D.S. & Kawai, H., 1998. Automated seismic moment

390   tensor determination by using on-line broadband seismic waveforms, *Zisin 2nd*

391   *Ser.,* 51**,** 149-156, doi:10.4294/zisin1948.51.1_149. (*in Japanese with English*

392   *abstract*)

393   Girshick, R., Donahue, J., Darrell, T. & Malik, J., 2014. Rich Feature Hierarchies for

394   Accurate Object Detection and Semantic Segmentation. *in 2014 IEEE*

395   *Conference on Computer Vision and Pattern Recognition*, pp. 580-587,

396   doi:10.1109/CVPR.2014.81.

397   Hara, S., Fukahata, Y. & Iio, Y., 2019. P-wave first-motion polarity determination of

398   waveform data in western Japan using deep learning, *Earth Planets Space,* 71**,**

399   127, doi:10.1186/s40623-019-1111-x.

400   Hardebeck, J.L., 2015. Stress orientations in subduction zones and the strength of

401   subduction megathrust faults, *Science,* 349**,** 1213-1216,

402   doi:10.1126/science.aac5625.

403   Hardebeck, J.L. & Shearer, P.M., 2002. A new method for determining first-motion

404   focal mechanisms, *Bull. Seismol. Soc. Am.,* 92**,** 2264-2276,

405   doi:10.1785/0120010200.

406   Hardebeck, J.L. & Shearer, P.M., 2003. Using S/P amplitude ratios to constrain the focal

407   mechanisms of small earthquakes, *Bull. Seismol. Soc. Am.,* 93**,** 2434-2444,

408   doi:10.1785/0120020236.

409   Hardebeck, J.L. & Shearer, P.M., 2008. HASH: A FORTRAN program for computing

410   earthquake first-motion focal mechanisms – v1.2 –.

411   Heidbach, O., Rajabi, M., Cui, X., Fuchs, K., Müller, B., Reinecker, J., Reiter, K.,

412   Tingay, M., Wenzel, F., Xie, F., Ziegler, M.O., Zoback, M.-L. & Zoback, M.,

413   2018. The World Stress Map database release 2016: Crustal stress pattern across

414   scales, *Tectonophysics,* 744**,** 484-498, doi:10.1016/j.tecto.2018.07.007.

415   Heidbach, O., Rajabi, M., Reiter, K., Ziegler, M. & Team, W., 2016. World Stress Map

416   Database Release 2016. *in GFZ Data Services*, doi:10.5880/WSM.2016.001.

417   Heidbach, O., Tingay, M., Barth, A., Reinecker, J., Kurfeß, D. & Müller, B., 2008. The

418   World Stress Map database release 2008, doi:10.1594/GFZ.WSM.Rel2008.

419   Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks,

420   *Neural Networks,* 4**,** 251-257, doi:10.1016/0893-6080(91)90009-T.

421   Huffman, K.A., Saffer, D.M. & Dugan, B., 2016. In situ stress magnitude and rock

422   strength in the Nankai accretionary complex: a novel approach using paired

423   constraints from downhole data in two wells, *Earth Planets Space,* 68**,** 123,

424   doi:10.1186/s40623-016-0491-4.

425   Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment, *Comput. Sci. Eng.,* 9**,**

426   90-95, doi:10.1109/MCSE.2007.55.

427   Igarashi, T., 2020. Catalog of small repeating earthquakes for the Japanese Islands,

428   *Earth Planets Space,* 72**,** 73, doi:10.1186/s40623-020-01205-2.

429   Iio, Y., Kishimoto, S., Nakao, S., Miura, T., Yoneda, I., Sawada, M. & Katao, H., 2018.

430   Extremely weak fault planes: An estimate of focal mechanisms from stationary

431   seismic activity in the San'in district, Japan, *Tectonophysics,* 723**,** 136-148,

432   doi:10.1016/j.tecto.2017.12.007.

433   Imanishi, K., Ando, R. & Kuwahara, Y., 2012. Unusual shallow normal-faulting

434   earthquake sequence in compressional northeast Japan activated after the 2011

435   off the Pacific coast of Tohoku earthquake, *Geophys. Res. Lett.,* 39**,** L09306,

436   doi:10.1029/2012GL051491.

437   Imanishi, K., Kuwahara, Y., Takeda, T., Mizuno, T., Ito, H., Ito, K., Wada, H. & Haryu,

438   Y., 2011. Depth-dependent stress field in and around the Atotsugawa fault,

439   central Japan, deduced from microearthquake focal mechanisms: Evidence for

440   localized aseismic deformation in the downward extension of the fault, *J.*

441   *Geophys. Res.,* 116**,** B01305, doi:10.1029/2010JB007900.

442   Imanishi, K., Uchide, T., Ohtani, M., Matsushita, R. & Nakai, M., 2019. Construction of

443   the crustal stress map in the Kanto region, central Japan, *Bull. Geol. Surv. Jpn.,*

444   70**,** 273-298. (*in Japanese with English abstract*)

445   Ioffe, S. & Szegedy, C., 2015. Batch normalization: accelerating deep network training

446   by reducing internal covariate shift. *in ArXiv e-prints*.

447   Kawanishi, R., Iio, Y., Yukutake, Y., Shibutani, T. & Katao, H., 2009. Local stress

448   concentration in the seismic belt along the Japan Sea coast inferred from precise

449   focal mechanisms: Implications for the stress accumulation process on intraplate

450   earthquake faults, *J. Geophys. Res. Solid Earth,* 114, doi:10.1029/2008jb005765.

451   Kingma, D.P. & Ba, J., 2014. Adam: A method for stochastic optimization, *arXiv*

452   *preprint arXiv:1412.6980.*

453   Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. &

454   Wassermann, J., 2015. ObsPy: a bridge for seismology into the scientific Python

455   ecosystem, *Comput. Sci. Discov.,* 8, doi:10.1088/1749-4699/8/1/014003.

456   Ley, C. & Verdebout, T., 2017. *Modern directional statistics*, pp. 190, CRC Press, New

457   York, doi:10.1201/9781315119472.

458   Matsumoto, S., Nakao, S., Ohkura, T., Miyazaki, M., Shimizu, H., Abe, Y., Inoue, H.,

459   Nakamoto, M., Yoshikawa, S. & Yamashita, Y., 2015. Spatial heterogeneities in

460   tectonic stress in Kyushu, Japan and their relation to a major shear zone, *Earth*

461   *Planets Space,* 67**,** 172, doi:10.1186/s40623-015-0342-8.

462 Matsushita, R. & Imanishi, K., 2015. Stress fields in and around metropolitan Osaka,

463   Japan, deduced from microearthquake focal mechanisms, *Tectonophysics,* 642**,**

464   46-57, doi:10.1016/j.tecto.2014.12.011.

465 Megies, T., Beyreuther, M., Barsch, R., Krischer, L. & Wassermann, J., 2011. ObsPy –

466   What can it do for data centers and observatories?, *Ann. Geophys.,* 54**,** 12,

467   doi:10.4401/ag-4838.

468 Meneses-Gutierrez, A. & Nishimura, T., 2020. Inelastic deformation zone in the lower

469   crust for the San-in Shear Zone, Southwest Japan, as observed by a dense GNSS

470   network, *Earth Planets Space,* 72**,** 10, doi:10.1186/s40623-020-1138-z.

471 Nair, V. & Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann

472   machines. *in 27th International Conference on Machine Learning (ICML-10),*

473   pp. 807-814, Haifa, Israel.

474 Nakamura, M., 2004. Automatic determination of focal mechanism solutions using

475   initial motion polarities of P and S waves, *Phys. Earth Plant. Inter.,* 146**,**

476   531-549, doi:10.1016/j.pepi.2004.05.009.

477 National Research Institute for Earth Science and Disaster Resilience, 2020. NIED

478   Hi-net, National Research Institute for Earth Science and Disaster Resilience.

479   doi:10.17598/NIED.0003.

480 Otsubo, M., Miyakawa, A. & Imanishi, K., 2018. Normal-faulting stress state associated

481   with low differential stress in an overriding plate in northeast Japan prior to the

482   2011 Mw 9.0 Tohoku earthquake, *Earth Planets Space,* 70,

483   doi:10.1186/s40623-018-0813-9.

484 Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z.,

485   Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z.,

486   Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. & Chintala,

487   S., 2019. PyTorch: An imperative style, high-performance deep learning library.

488   *in Advances in Neural Information Processing Systems*, pp. 8024-8035, Curran

489   Associates, Inc.

490   Pugh, D.J., White, R.S. & Christie, P.A.F., 2016a. Automatic Bayesian polarity

491   determination, *Geophys. J. Int.,* 206**,** 275-291, doi:10.1093/gji/ggw146.

492   Pugh, D.J., White, R.S. & Christie, P.A.F., 2016b. A Bayesian method for microseismic

493   source inversion, *Geophys. J. Int.*, doi:10.1093/gji/ggw186.

494   Ross, Z.E., Meier, M.-A. & Hauksson, E., 2018. P wave arrival picking and first-motion

495   polarity determination with deep learning, *J. Geophys. Res. Solid Earth,* 123**,**

496   5120-5129, doi:10.1029/2017JB015251.

497   Rumelhart, D.E., Hinton, G.E. & Williams, R.J., 1986. Learning representations by

498   back-propagating errors, *Nature,* 323**,** 533-536, doi:10.1038/323533a0.

499   Savage, M.K., Aoki, Y., Unglert, K., Ohkura, T., Umakoshi, K., Shimizu, H., Iguchi, M.,

500   Tameguri, T., Ohminato, T. & Mori, J., 2016. Stress, strain rate and anisotropy in

501   Kyushu, Japan, *Earth Planet. Sci. Lett.,* 439**,** 129-142,

502   doi:10.1016/j.epsl.2016.01.005.

503   Shearer, P.M., Prieto, G.A. & Hauksson, E., 2006. Comprehensive analysis of

504   earthquake source spectra in Southern California, *J. Geophys. Res.,* 111**,** B06303,

505   doi:10.1029/2005JB003979.

506   Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R., 2014.

507   Dropout: a simple way to prevent neural networks from overfitting, *J. Mach.*

508   *Learn. Res.,* 15**,** 1929-1958.

509   Terakawa, T. & Matsu'ura, M., 2010. The 3-D tectonic stress fields in and around Japan

510   inverted from centroid moment tensor data of seismic events, *Tectonics,* 29**,**

511   TC6008, doi:10.1029/2009TC002626.

512 Townend, J., Sutherland, R., Toy, V.G., Doan, M.-L., Célérier, B., Massiot, C., Coussens,

513   J., Jeppson, T., Janku-Capova, L., Remaud, L., Upton, P., Schmitt, D.R., Pezard,

514   P., Williams, J., Allen, M.J., Baratin, L.-M., Barth, N., Becroft, L., Boese, C.M.,

515   Boulton, C., Broderick, N., Carpenter, B., Chamberlain, C.J., Cooper, A., Coutts,

516   A., Cox, S.C., Craw, L., Eccles, J.D., Faulkner, D., Grieve, J., Grochowski, J.,

517   Gulley, A., Hartog, A., Henry, G., Howarth, J., Jacobs, K., Kato, N., Keys, S.,

518   Kirilova, M., Kometani, Y., Langridge, R., Lin, W., Little, T., Lukacs, A.,

519   Mallyon, D., Mariani, E., Mathewson, L., Melosh, B., Menzies, C., Moore, J.,

520   Morales, L., Mori, H., Niemeijer, A., Nishikawa, O., Nitsch, O., Paris, J., Prior,

521   D.J., Sauer, K., Savage, M.K., Schleicher, A., Shigematsu, N., Taylor-Offord, S.,

522   Teagle, D., Tobin, H., Valdez, R., Weaver, K., Wiersberg, T. & Zimmer, M., 2017.

523   Petrophysical, Geochemical, and Hydrological Evidence for Extensive

524   Fracture-Mediated Fluid and Heat Transport in the Alpine Fault's Hanging-Wall

525   Damage Zone, *Geochem. Geophys. Geosyst.,* 18**,** 4709-4732,

526   doi:10.1002/2017GC007202.

527 Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J. & Wobbe, F., 2013. Generic Mapping

528   Tools: Improved Version Released, *EOS,* 94**,** 409-410,

529   doi:10.1002/2013eo450001.

530 Williams, M.C., 2014. HASHpy, doi:10.5281/zenodo.9808.

531 Wu, H.-Y., Ma, K.-F., Zoback, M., Boness, N., Ito, H., Hung, J.-H. & Hickman, S., 2007.

532   Stress orientations of Taiwan Chelungpu-Fault Drilling Project (TCDP) hole-A

533   as observed from geophysical logs, *Geophys. Res. Lett.,* 34,

534   doi:10.1029/2006GL028050.

535   Yang, W., Hauksson, E. & Shearer, P.M., 2012. Computing a large refined catalog of

536   focal mechanisms for southern California (1981–2010): Temporal stability of the

537   style of faulting, *Bull. Seismol. Soc. Am.,* 102**,** 1179-1194,

538   doi:10.1785/0120110311.

539   Yoshida, K., Hasegawa, A. & Okada, T., 2015. Spatial variation of stress orientations in

540   NE Japan revealed by dense seismic observations, *Tectonophysics,* 647–648**,**

541   63-72, doi:10.1016/j.tecto.2015.02.013.

542   Yoshida, K., Hasegawa, A., Saito, T., Asano, Y., Tanaka, S., Sawazaki, K., Urata, Y. &

543   Fukuyama, E., 2016. Stress rotations due to the M6.5 foreshock and M7.3 main

544   shock in the 2016 Kumamoto, SW Japan, earthquake sequence, *Geophys. Res.*

545   *Lett.,* 43**,** 10097-10104, doi:10.1002/2016GL070581.

546

547 Tables

548 Table 1

549   Numbers of seismograms and earthquakes included in the data sets.

Region	Type of Data Set	Seismogram Sets	Earthquakes
Entire Japan	Training	404,989	18,721
	Validation	8,636	620
Kanto	Training	12,803	1,262
	Validation	782	56
	Test	1,483	113
Chugoku	Training	63,327	2,259
	Validation	7,672	322
	Test	12,818	595
Entire Japan	Application	2,301,584	114,525

550

551 Figures

Input ——————∿⋎⋀∿ 1 ch × 256 samples

ConvBN

40 × 256

ConvBNP

60 × 128

ConvBNP

80 × 64

ConvBNP

100 × 32

Flatten

3200

FC

40

FC & LogSoftMax

Up ○   ○ Down
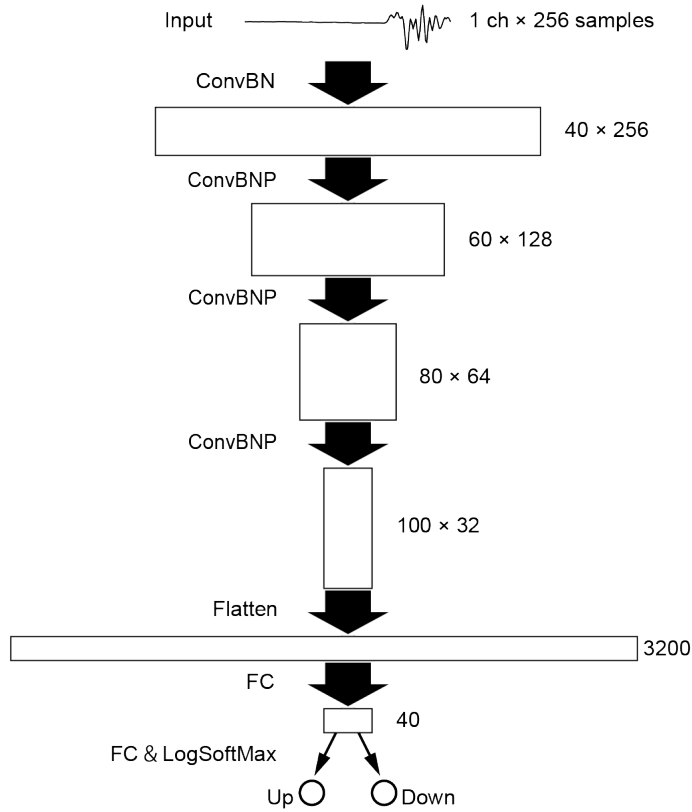
552

553 Fig. 1

554 The design of the neural network model. The numbers on the right indicate the

555 number of channels and samples, while "ConvBN," "Conv BNP," and "FC" denote

556 convolution and batch normalization layers, convolution, batch normalization, and

557 pooling layers, and fully connected layers, respectively.

558
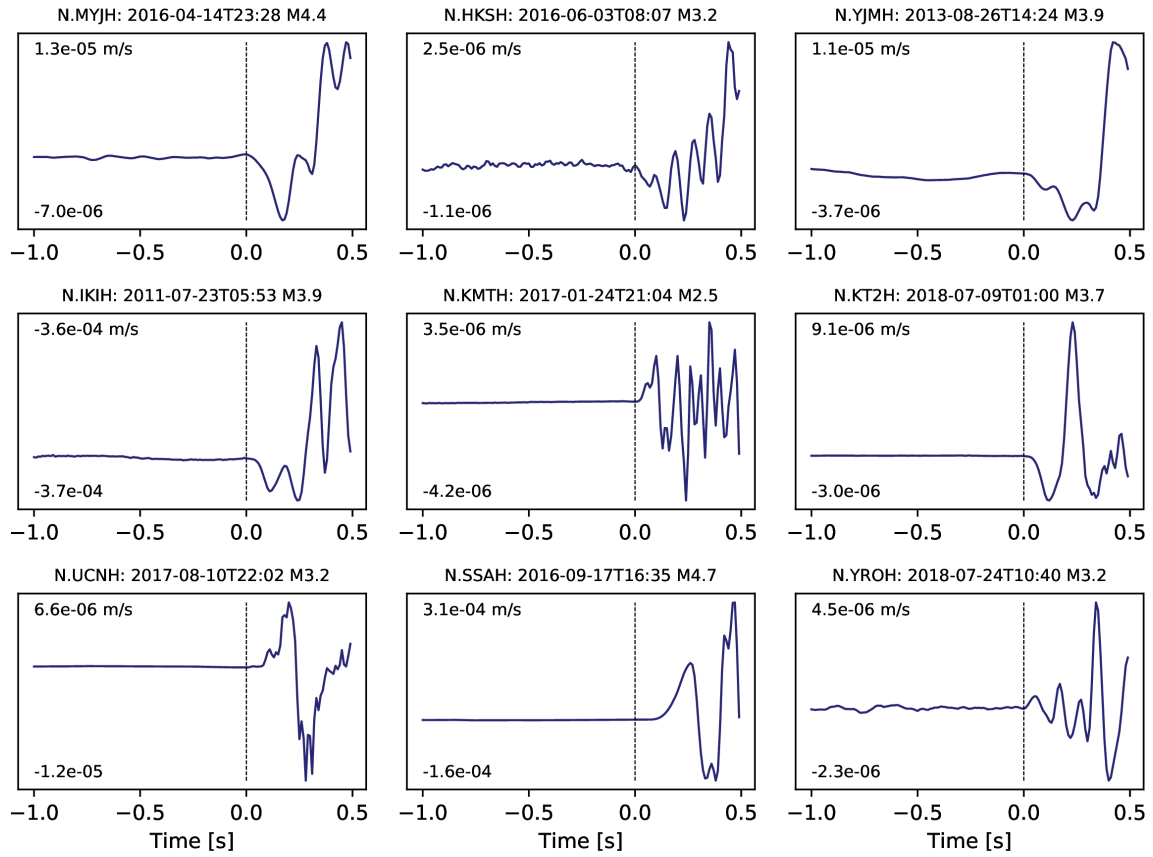
Fig. 2

560   Examples of seismograms with impulsive onsets but no accompanying polarity data

561   in the JMA Unified Earthquake Catalog. Station names and the origin time and

562   magnitude of the events are indicated above each panel. Note that this data was not used

563   in the study.

564

565 Fig. 3
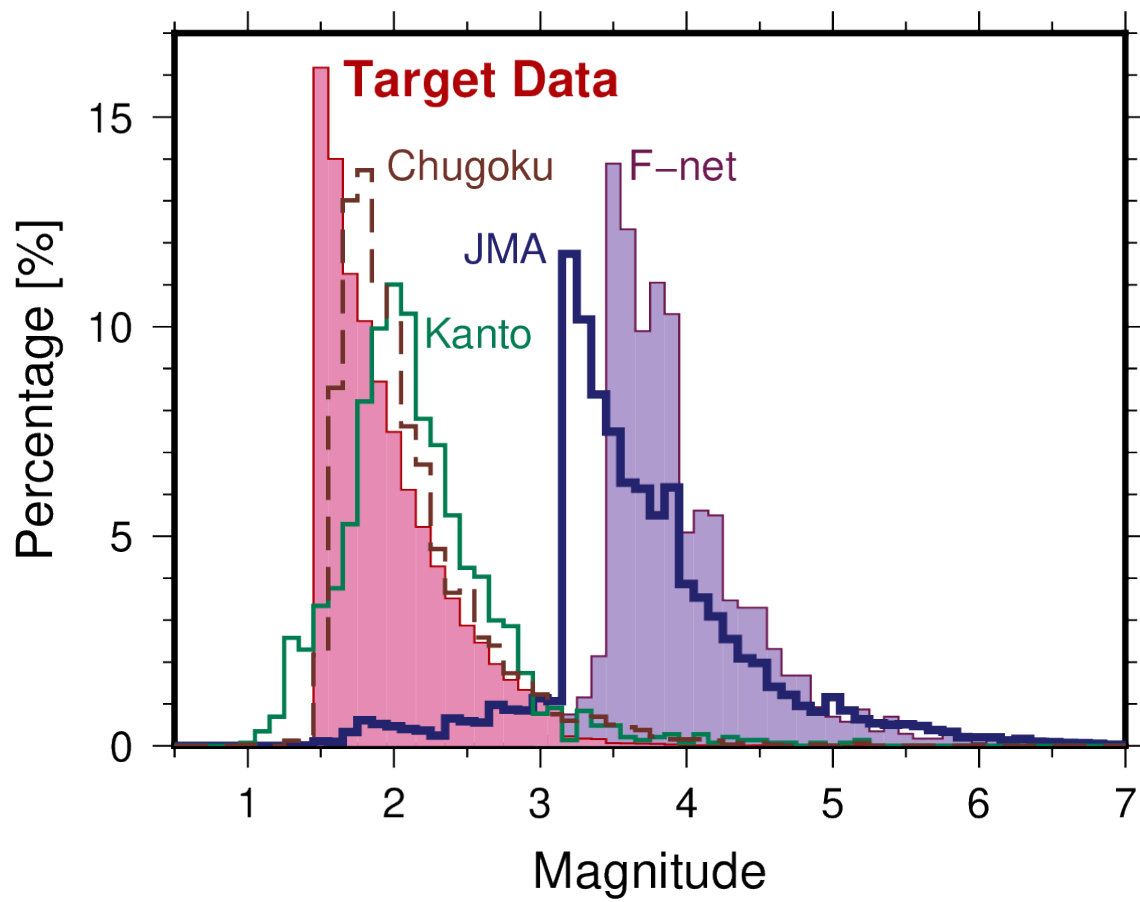
566 Magnitude–frequency distributions in the data for the target (pink shade), F-net

567 (purple shade), the JMA (thick navy), Kanto (thin green), and Chugoku (dashed brown)

568 data sets.

569
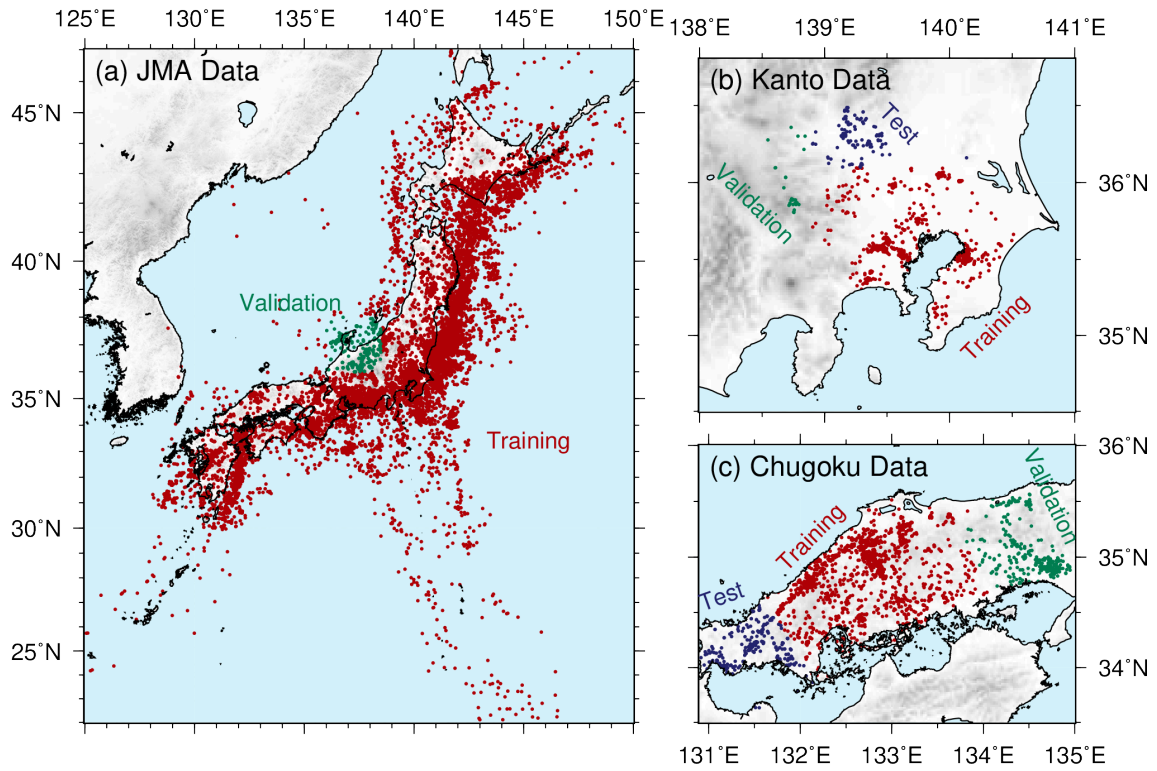
Fig. 4

Distribution of the epicentres of the earthquakes used for training (red), validation (green), and testing (blue) of the neural network model. The topography was taken from ETOPO1 (Amante and Eakins 2009).

574

575

576   Fig. 5
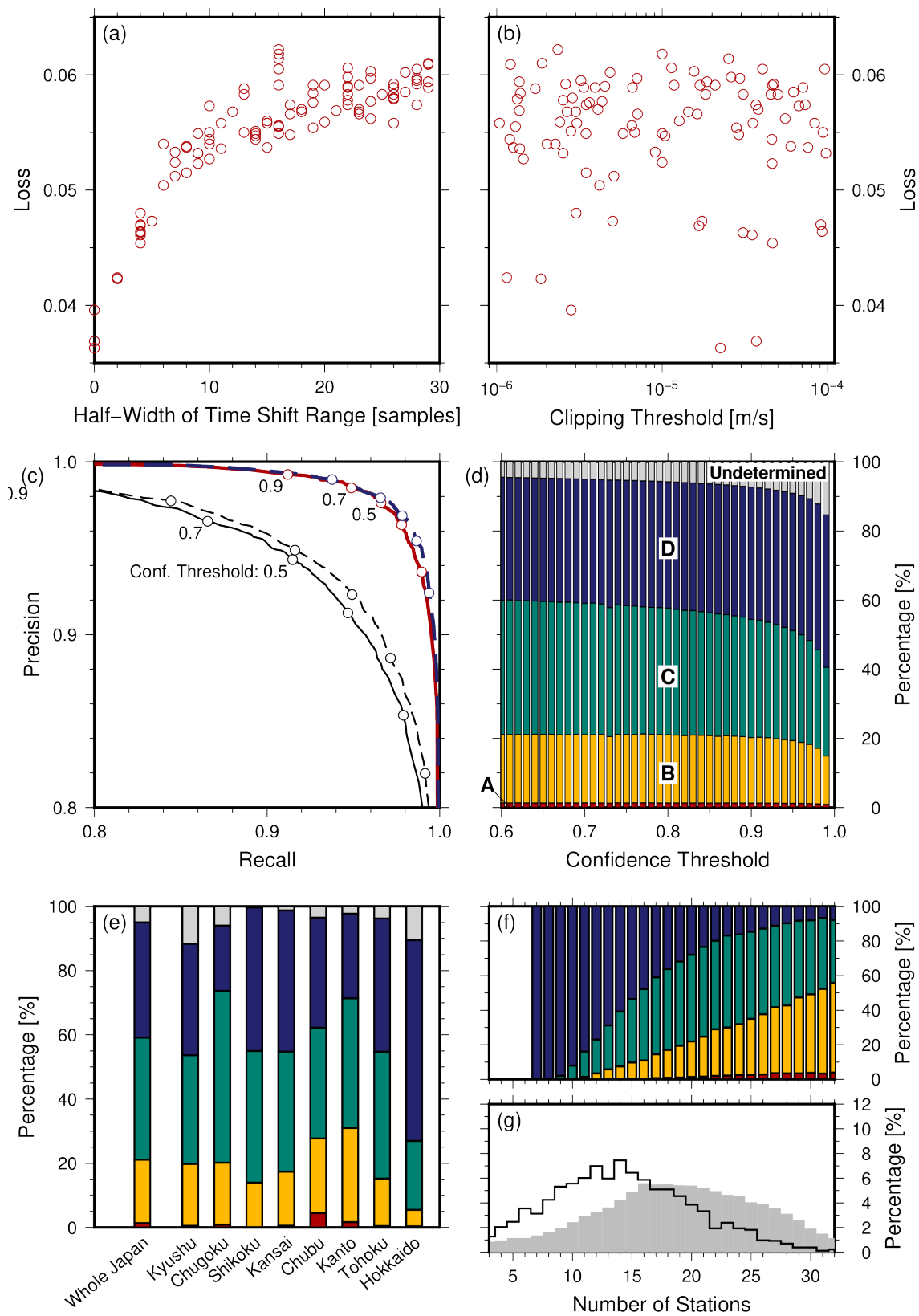
577   Summary of the results. Here, (a) and (b) show the loss function values as functions

578   of the time-shift range and the clipping threshold, respectively. (c) The precision-recall

579   curve of the trained model for the test data set. The thick red solid and blue dashed lines

580   refer to the upward and downward polarities, respectively, in the case of the two-stage

581   training. The thin black solid and dashed lines refer to the upward and downward

582   polarities, respectively, where the model was trained using only the JMA data set.

583   Circles correspond to every 0.1 units of the confidence thresholds. (d) Bar graphs of the

584   rank of the focal mechanism solutions as a function of the confidence threshold. Red,

585   yellow, green, and blue bars from the bottom to the top represent the A, B, C, and

586   D-ranks, respectively. Gray bars on the top indicate undefined focal mechanisms. (e)

587   Bar graphs of the rank of focal mechanism solutions for the whole of Japan and eight

588   specific regions (see Fig. 7e). Here, a model with a confidence threshold of 0.7 was

589   used. The color scheme is the same as in the case of (d). (f) Bar graphs as a function of

590   the number of seismic stations. The color scheme is the same as for (d). (g) Histograms

591   of events as a function of the number of stations. The black line and gray shaded region

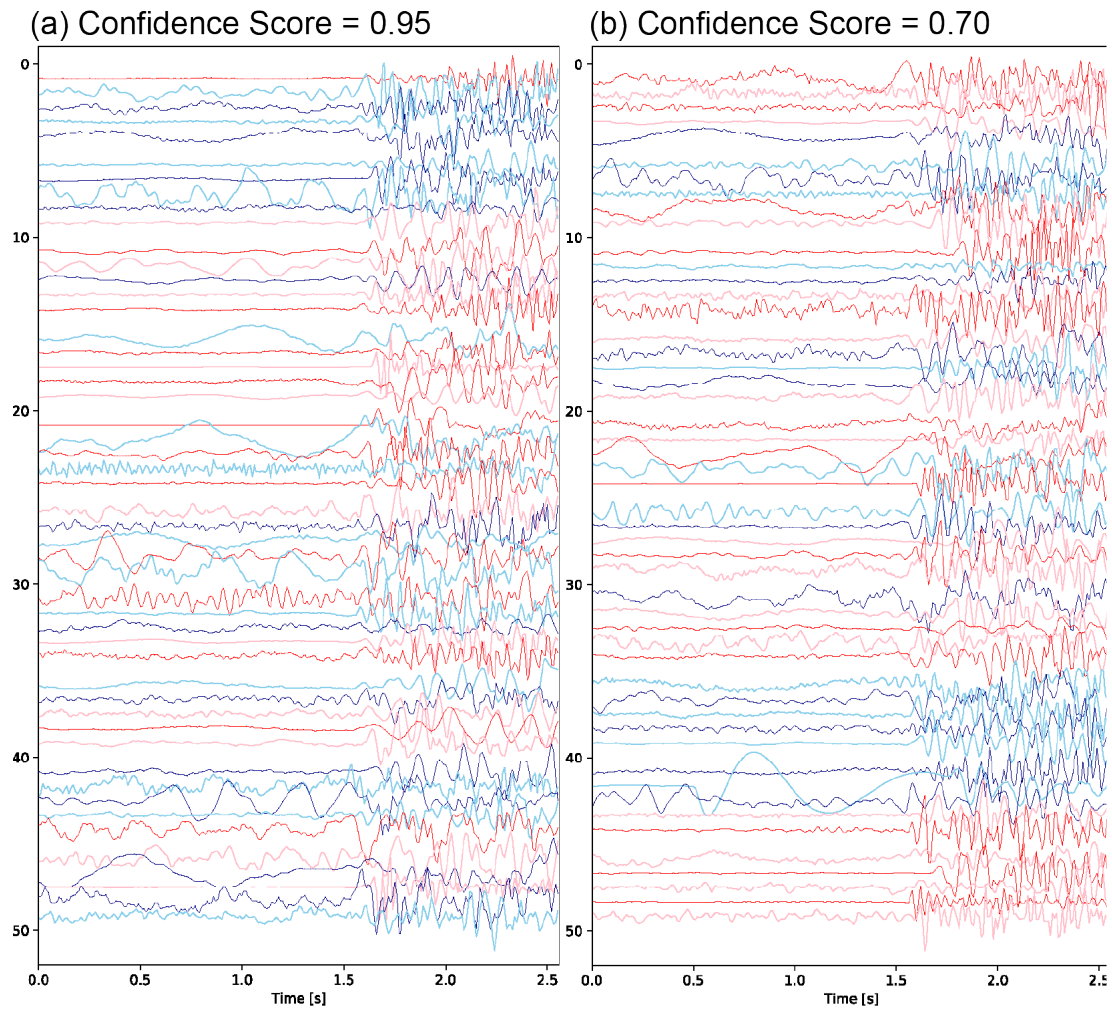592   indicate the values for Hokkaido and other regions, respectively.

593

594   Fig. 6

595   Examples of seismograms with polarities picked by the neural network model with

596   confidence scores of (a) 0.95 and (b) 0.70. Seismograms with downward polarities are

597   flipped. If correctly picked, the first motion appears upward in this figure. Light and

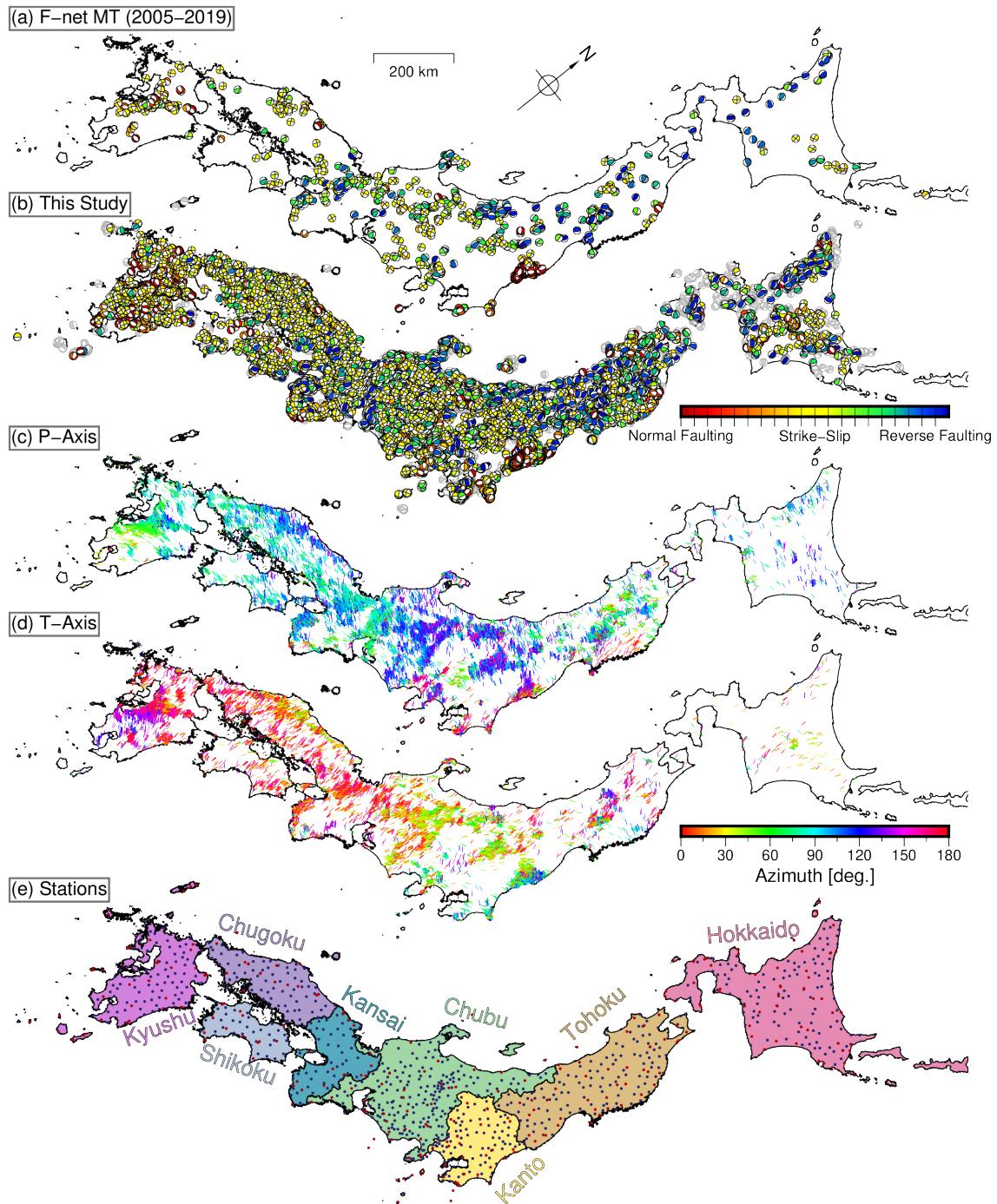598   dark colors are alternated for convenience.

599

Fig. 7

(a) Moment tensor solutions of earthquakes from 2006–2019 provided by the NIED
F-net project (Fukuyama et al. 1998) for reference. (b) Focal mechanism solutions
derived by this study. Solutions ranked A–C (Hardebeck and Shearer 2008) are colored
according to focal mechanism types (Shearer et al. 2006). The solutions ranked "D" is

34

605 shown by gray beach balls. (c) Azimuths of the P-axes of the estimated focal

606 mechanism solutions ranked A–C and with plunges less than 30°. Colors indicate the

607 azimuths. (d) Azimuths of the T-axes. (e) Seismic station distribution.

608

609 Fig. 8
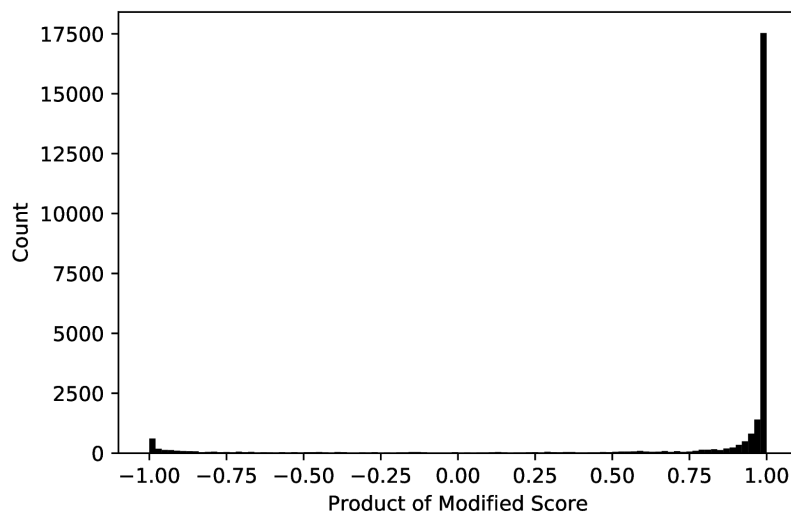
610 Histogram of the product of the modified scores of repeating earthquake pairs, where

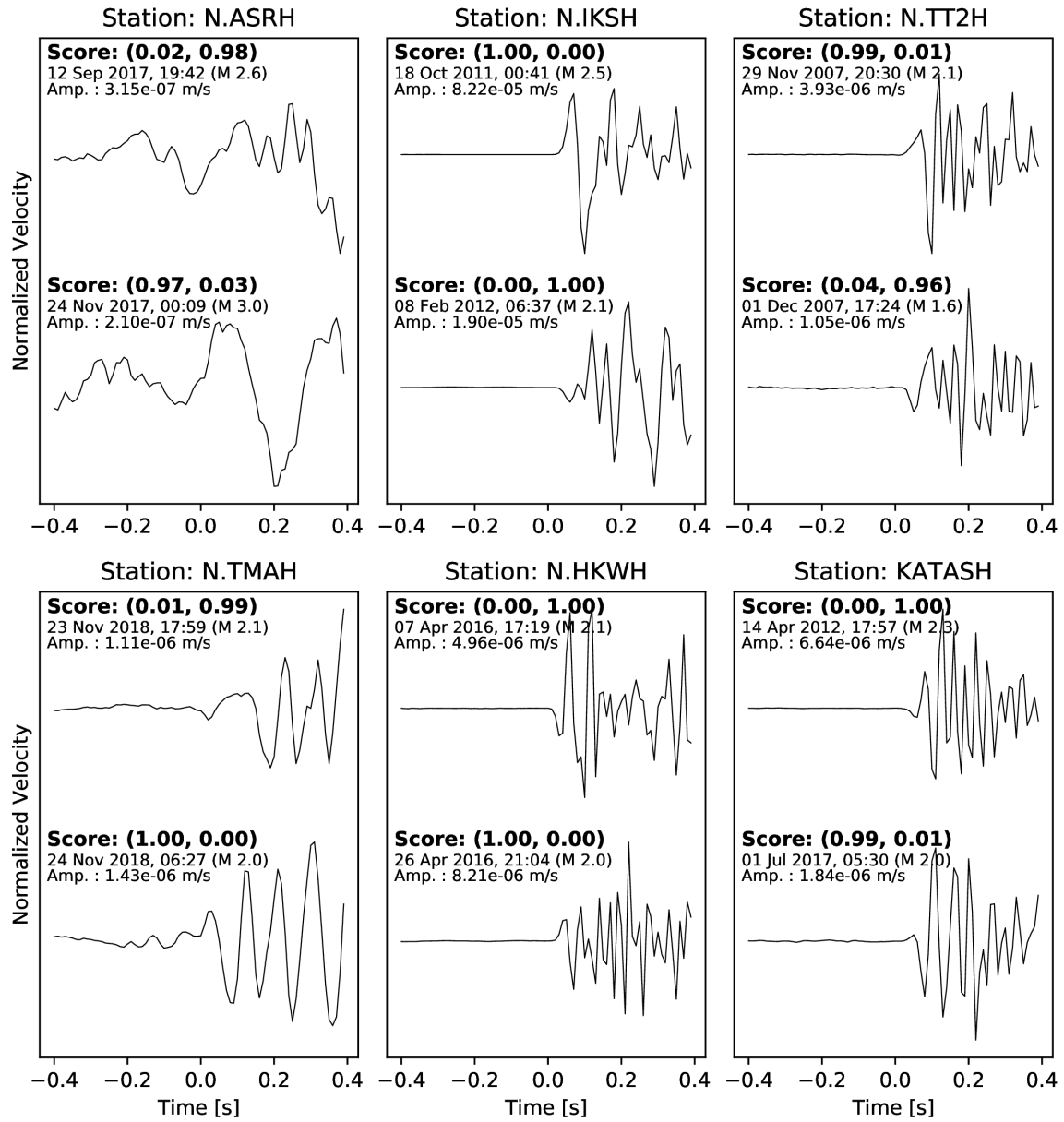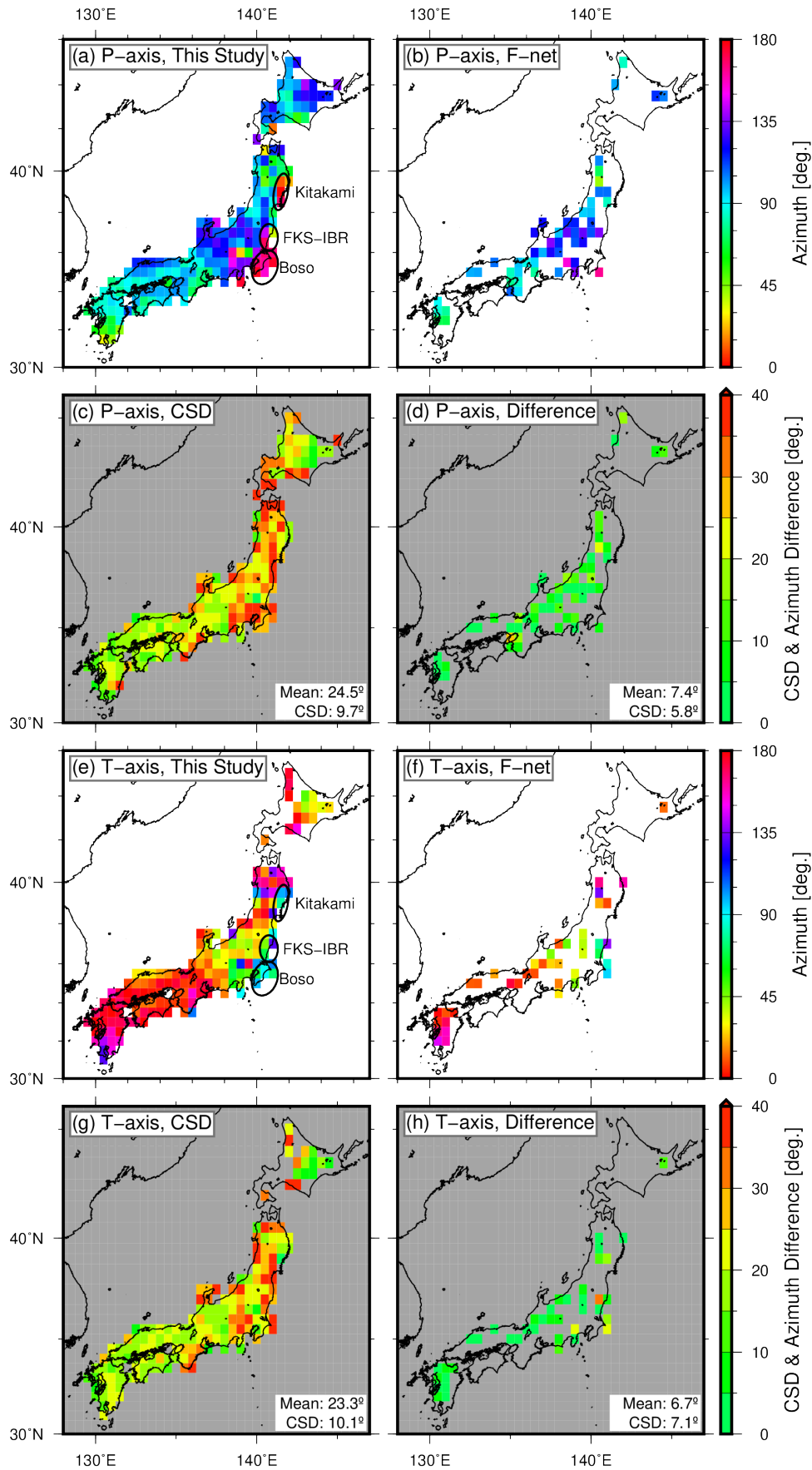611 the modified score is (confidence score for the upward polarity – that for the downward

612 polarity).

613 Fig. 9

615 Comparisons of seismograms of repeating earthquake pairs at the same stations. The

616 scores shown in the panels are the confidence scores for the upward polarity.

618 Fig. 10

619   Spatial distributions of (a) the mean directions of P-axes azimuths from this study and

620 (b) the F-net catalog, (c) the circular standard deviation (CSD) (eq. (7)) of P-axes

621 azimuths from this study, and (d) the absolute values of differences of (a) and (b); (e)–

622 (h) are the values for T-axes. The P- or T- axes with lunges less than 30° and focal

623 mechanism solutions ranked A–C were used. The values at 0.5-degree grids with four or

624 more events are shown. In (a) and (e), the areas in the Kitakami mountain range,

625 Fukushima Hamadori and northern Ibaraki prefecture (FKS-IBR), and Boso peninsula
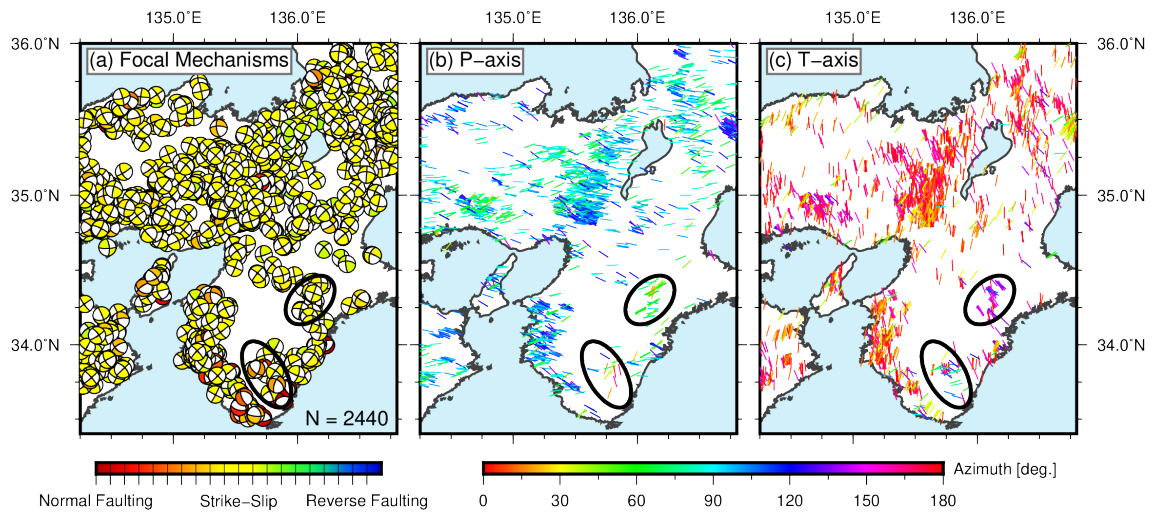
626 are marked by ellipses.

627

628   Fig. 11

629   (a) Focal mechanisms, (b) P-axes, and (c) T-axes in the Kansai area. In (b) and (c),

630   the axes with plunges less than 30° and the focal mechanism solutions ranked A–C were

631   plotted. The ellipses indicate areas with distinct focal mechanisms.
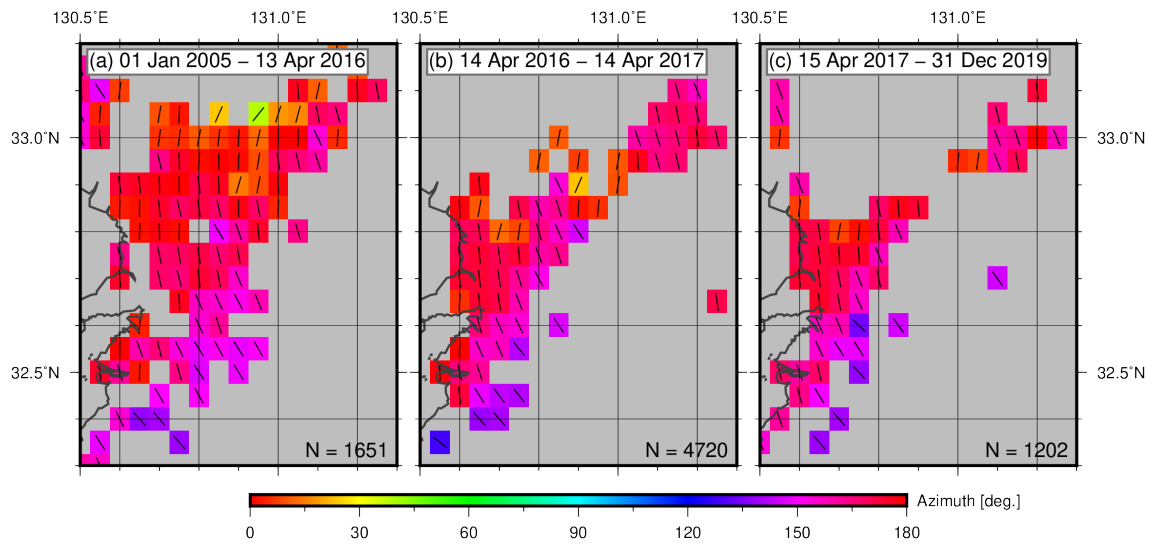
632

633   Fig. 12

634   The mean directions of the T-axes azimuths at 0.05-degree grids (a) before the 2016

635   Kumamoto earthquake, (b) within one year of the Kumamoto earthquake, and (c) after

636   this time, indicated by short bars and colors. T-axes with plunges less than 30° and the

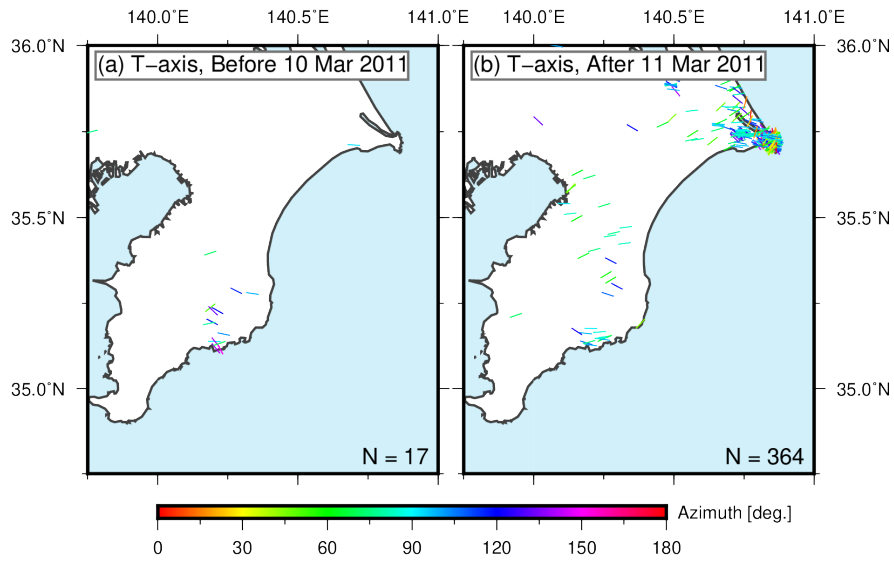637   focal mechanism solutions ranked A–C were used.

638

639   Fig. 13

640   The T-axis azimuths in the Boso peninsula (a) before and (b) after the 2011

641   Tohoku-oki earthquake. T-axes with plunges less than 30° and the focal mechanism

642   solutions ranked A–C were plotted.