

1 **Application of Two-way Fixed Effects Regression on the**
2 **Error Analysis of Seismic Magnitude Estimation:**
3 **Hypotheses, Theory and Testing**

4 Guo-quan Zhang¹

5 1 Wenxian Emergency Management Bureau, Jiaozuo, Henan, China

6 [https://orcid.org/0000-0002-9468-6132\(GZ\)](https://orcid.org/0000-0002-9468-6132(GZ))

7
8 Correspondence to: Guo-quan Zhang, zgq083516@163.com

9
10 Key Points:

- 11 1. A statistical method is introduced to numerically solve the relative values of
12 source, path and site effects
- 13 2. The magnitude uncertainty caused by source, path and site effects is determined
14 and compared quantitatively for Japan's KiK-net
- 15 3. A series of check tests are designed to examine the robustness of the decoupling
16 algorithm

17

18 **Abstract** Seismic magnitude has been an indistinct concept since modern seismology
19 was established. The author discusses it thoroughly , provides an general expression
20 of definition that is only relevant to the source characteristics and explains the
21 magnitude conversion problem physically. Based on the essential understanding of
22 magnitude estimation problems, the author introduces a method called two-way fixed
23 effects regression, which has been applied in seismology since Keiiti Aki studied the
24 properties of coda waves. A series of check tests are carefully designed to examine the
25 hypotheses of this method and then the author applies it to the error analysis of the
26 peak ground acceleration magnitude. The source, path and site errors are estimated by
27 this approach and the robustness of separation is checked. The acquired knowledge
28 about the errors in magnitude estimation may help us to improve the precision of
29 magnitude estimation during earthquake early warning.

30

31 **Plain Language Summary**

32 The purpose of seismology is to explain all the phenomena associated with
33 earthquakes by studying the laws behind them. To find these laws, the first effort
34 should be to describe the many earthquakes. However, in general, deep insight into
35 these laws is needed to adequately characterize earthquakes. Seismic magnitude is a
36 quantity introduced soon after modern seismology was established to categorize
37 earthquakes but is still a concept that troubles geophysicists, as there are many
38 fundamental seismological problems that have not yet been answered thoroughly. In
39 this paper, the author tries to give a strict definition of magnitude and analyze the
40 errors in it quantitatively from a statistical perspective. Two-way fixed effects
41 regression allows part of the relation to be nonlinear; thus, it is an appropriate tool to
42 explore the complicated laws behind earthquake phenomena. This paper explains the
43 theoretical assumptions and physical meanings of this statistical technique and applies
44 it to separate the fixed effects of different factors that affect the amplitude of seismic
45 record.

46

47 **1. Introduction**

48 The seismic magnitude is an estimated parameter of the relative size of an earthquake
49 (Bormann et al., 2013). The ‘estimated parameter’ here suggests that (1) this value
50 contains a considerable error and often not the direct measure of the earthquake size;
51 (2) this value is a parameter estimated from a certain sample (not the population), for
52 instance, the amplitudes recorded by a local, nationwide, or even worldwide
53 seismographic network but anyway from limited points on the ground; and (3) the
54 magnitude of an earthquake may be multivalued, whereas the size should be only, as
55 different observers can estimate the size from different perspectives, such as the
56 duration of the quake (duration magnitude M_d), the maximum amplitude of the body
57 waves (body wave magnitude M_b) or surface waves (surface wave magnitude M_s).
58 However, the word ‘size’ in this definition is ambiguous and consensus-less. Bormann
59 (2013) tends to regard the radiated energy (energy magnitude M_e) as the size of an
60 earthquake, as this value relates directly to the potential damage of an earthquake. On
61 the other hand, the United States Geology Survey (USGS) practically uses the work
62 done by the stress (moment magnitude M_w) as the preferred published magnitude
63 (USGS, 2023a). Moreover, the difference between these two fundamental physical
64 quantities is a scale factor known as the seismic efficiency.

65

66 Although the seismic magnitude is statistical and error-significant, we have little
67 knowledge about the statistical properties of the magnitude or its error. For a
68 theoretical point-shaped source, the amplitude of a certain earthquake at a certain
69 station is relevant to the size of the earthquake, the source spectrum and the dominant
70 frequency of the observed component, the angle between the earth’s surface and the
71 fault plane or the focal mechanism, the attenuation on the propagation path, the
72 radiation pattern and the direction of the station relevant to the source, and the site
73 amplification of the station, while the current magnitude formula usually estimates the
74 magnitude from the amplitude after nonexact and empirical corrections of attenuation.
75 In other words, the errors from the prevailing magnitude estimating techniques are
76 typically variable-omitting biases, as they take too few factors into account.

77

78 The bias from omitting source-related properties cannot be reduced by averaging the
79 measured values of multiple stations, while it is popular that some official agencies
80 and scholars treat the standard deviation as the uncertainty of the magnitude. On the
81 website of the USGS, one can easily find such instances (USGS, 2023b). In fact, the
82 uncertainty consists of fixed bias (systematic uncertainties) and stochastic error
83 (random uncertainties) (Taylor, 2022), and the estimation of magnitude from multiple
84 measurements is not definitely an unbiased estimation. Furthermore, different
85 magnitudes of the same earthquake may be inconsistent. ('Consistency' here is also a
86 statistic term. This means that the estimation converges to the true value as the
87 quantity of samples increases. Obviously, the true value should be only. Therefore, we
88 generally state that all the consistent estimations are consistent, as they have the same
89 limit.) Seggern (1970) reported that the focal mechanism exerts a different influence
90 on body and surface waves: the variation in the focal mechanism may enhance one
91 while suppressing the other. This will be incomprehensible if both the body wave
92 magnitude and the surface wave magnitude are consistent estimations of the
93 earthquake size.

94

95 The bias in magnitude estimation is annoying but nearly inevitable. Physically, this is
96 partly caused by the regular deployment of seismometers, which are all installed near
97 the surface of the earth and violate the principle of random sampling. Mathematically,
98 there are various methods to address fixed bias or endogeneity and obtain a consistent
99 estimation. In this paper, the author will apply two-way fixed effects regression
100 (2-FER) to the magnitude estimation to treat this endogeneity problem.

101

102 Numerous seismometers have acquired massive amounts of seismic data on
103 earthquakes. However, during the regression of the magnitude estimating formula,
104 geophysical researchers are attempting to determine a calibration function that is used
105 to correct the attenuation on the path by records over decades and even centuries
106 (Bormann, 2013). As discussed before, it will not help to obtain a fine estimation to
107 simply increase the data volume and wastes the massive acquired information to

108 finally obtain only one or two parameters in the equation. A large amount of data
109 (called panel data in statistics in our problem) provides the possibility for conducting
110 complex statistical methods and motivates us to mine the information inside the raw
111 data and interpret it.

112

113 Some researchers have realized that simple linear regression is not a proper tool to
114 analyze the magnitude problem, while they introduce orthogonal regressions to the
115 magnitude regression problem (Bormann, 2013; Das et al., 2018). However, a good
116 orthogonal regression requires knowledge of the error deviation ratio between the
117 magnitude and the earthquake size (Das et al., 2018), but this knowledge is not yet
118 quite clearly derived. The error analysis of magnitude is also important because
119 without this knowledge, it is difficult to distinguish errors and fallacies. For example,
120 the difference in the local magnitudes of earthquakes on the junction between Sichuan
121 and Yunnan (in Changning County in particular, where shale gas is extracted) given
122 by these two provincial networks is sometimes as large as 0.6. Is this difference a
123 fallacy? Or what extent of difference is acceptable? The heads of the Instrument
124 Operation and Maintenance Groups of these two provinces both denied that they had
125 set wrong instrument responses in the software and they did have incentives to hide
126 their faults if they actually existed.

127

128 Fixed effects regression (FER) is a popular statistical method (Allison, 2009) in many
129 applications, such as econometrics (Greene, 2011) and biostatistics (Diggle, 2002;
130 Gardiner, 2009), to treat some regression problems of panel data or longitudinal data,
131 as it is simple but effective. Seismic observations naturally generate panel data, as we
132 observe identical earthquakes by multiple stations, which constitute a matrix, a table
133 or a panel. Numerous geophysicists have adopted fixed effects models spontaneously
134 to address seismology problems but may not specifically consult related statistical
135 materials (Phillips & Aki, 1970; Joyner & Boore 1981; Brillinger & Preisler, 1984;
136 Andrews, 1986; Iwata & Irikura, 1986; Takemura et al., 1991; Moya et al.,
137 2003; Wang & Shearer, 2019; Torres-Sánchez & Castro, 2023 and so on).

138

139 The above papers are focused on the research of coda waves, P-waves or S-waves,
140 which are recorded by a strong motion network or broadband network. Certainly,
141 there is no essential distinction between the strong motion network and the
142 Earthquake Early Warning (EEW) network. Therefore, the relevant method and
143 conclusions are also applicable to EEW. For instance, the major revision of the EEW
144 magnitude estimation formula by JMA (Aketagawa et al., 2010) was heavily based on
145 the conclusions of Joyner & Boore (1981) from strong motion records.

146

147 Magnitude estimating techniques in EEW are mainly traditionally semiphysical
148 (amplitude-based [Wu & Zhao, 2006], period-based [Allen & Kanamori, 2003; Atefi
149 et al., 2017] and combined [Wang et al., 2022]) or purely mathematical or statistical
150 (artificially intelligent [AI] [Song et al., 2023; Zhu et al., 2023, Furumura & Oishi,
151 2023]). The result of AI methods looks much better than that based on traditional
152 seismology and shows great potential in EEW. On the other hand, it is debatable
153 whether one can determine the magnitude before rupture completion (Münchmeyer et
154 al., 2022). People hope that early rupture signals can predict the final earthquake size
155 because of its enormous social value and have obtained results as quickly as AI
156 methods (Colombelli et al., 2014, 2020), but others argue that this optimistic result is
157 an unreal vision due to the selection of a small data set and accidental errors
158 (Trugman et al., 2019; Meier et al., 2021) and oppose rupture determinism in a
159 probabilistic view. If the fallacy of Colombelli (2014, 2020) is caused by the careless
160 treatment of error, then is it possible that the evidence of earthquake evolution
161 similarity is caused by noisy observations and is not as reliable as the noise buries
162 precursors in return? Otherwise, why is the AI's result so astonishing? Did the AI
163 methods cheat? Or are there some empirical laws of the rupture process still
164 unrecognized? What is the upper limit of the performance of the EEW system? What's
165 the theoretical form of the optimal estimation of the earthquake size or the ground
166 motion with limited information? Can we obtain it in a comprehensive way? Or can
167 the traditional methods generate results as fast and as accurately as the current AI at

168 least? Would AI completely replace the weak-looking traditional seismological
169 methods?

170

171 The author will not answer the many debatable questions above in this paper. Instead,
172 the author will focus on the property of the magnitude error as a first step toward
173 these answers, try to reduce the error of the traditional magnitude estimating method,
174 and take the peak ground acceleration (PGA) recorded by the KiK-net in Japan to
175 illustrate it.

176

177 **2. Data**

178 The author selects 228 earthquakes with magnitudes larger than 4 (Figure 1a)
179 recorded by the KiK-net in Japan (National Research Institute for Earth Science and
180 Disaster Resilience, 2019). The KiK-net is composed of 699 stations (Figure 1b) with
181 paired seismometers (one in the borehole and another on the ground). Trugman (2019)
182 selected the population of observed earthquakes to avoid the suspicion of selecting a
183 particular data set or to ensure the objectivity of his conclusions. However, to cover
184 the entire of Japan as evenly as possible and avoid too large weights of small
185 earthquakes, the author selects 228 earthquakes manually preferably in the area where
186 earthquakes occur infrequently and selects fewer than 6 earthquakes in every 0.1
187 magnitude unit (m.u.). Unilaterally, the author thinks this subjective selection will not
188 harm the objectivity of this research. The program of this research is open (please
189 refer to the Data Availability Statement), and it is easy to check the result on the entire
190 data set if one has such a data set in hand. Moreover, the source parameters and peak
191 ground acceleration (PGA) in the following study are from the files that are directly
192 downloaded from the official K- & KiK-net website. The author adopts them without
193 roundly checking and believes the JMA completely.

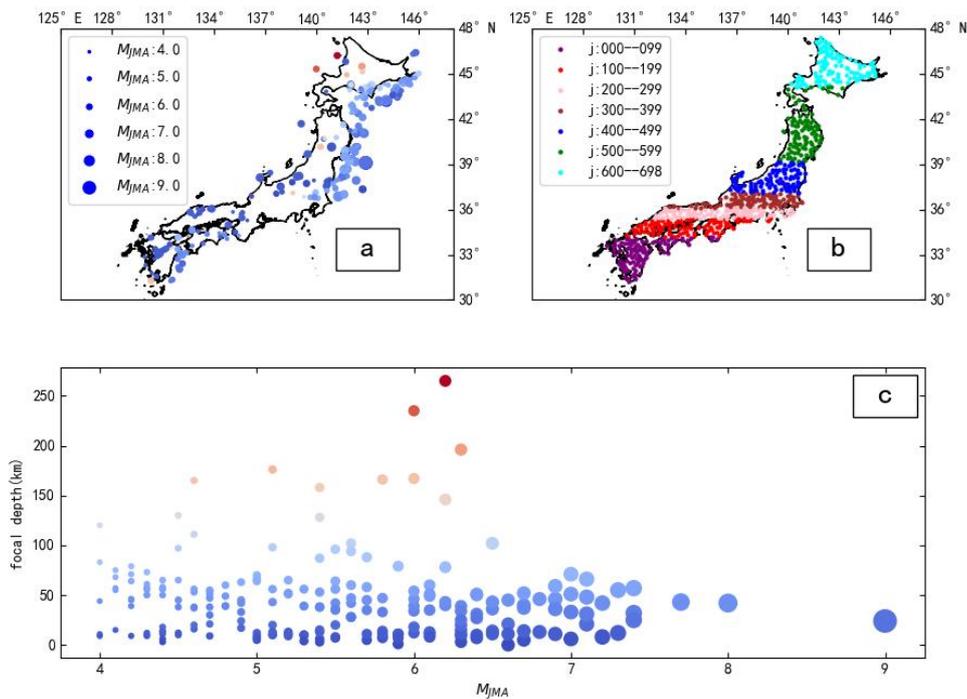
194

195 In addition, 6 channels at each station consist of KiK-net. Channels 1-3 constitute the
196 seismometer in the borehole, while Channels 4-6 constitute the seismometer on the
197 ground. Channels 1 and 4 record the NS component, Channels 2 and 5 record the EW

198 component, and Channels 3 and 6 record the UD component.

199

200 Why most researchers prefer to estimate the magnitude by peak ground displacement
201 (PGD) (for example, Wu & Zhao, 2006) but the author of this paper chooses the PGA?
202 PGD actually estimates the magnitude better than peak ground velocity (PGV), and
203 PGV does it better than PGA (Zhang et al., 2023). We know the PGD is the
204 integration of PGV and PGV the integration of PGA, so will it be better if we
205 continue to integrate PGD? What will happen if we integrate unlimited times? These
206 questions relate to a question we raised in the Introduction, namely, what is the
207 optimal estimation of the earthquake size. The author believes that the answer lies in
208 another domain, for example, the frequency domain maybe.



209

210 Figure 1. (a) Spatial distribution of the 228 selected earthquakes. The warmer color indicates a
211 deeper focal depth (see subplot [c]). (b) Spatial distribution of the 699 stations in KiK-net. They
212 are numbered from south to north. (c) The depths of the 228 earthquakes determined by JMA. The
213 warmer color indicates a deeper focal depth.

214

215 We know that an integration operator is transformed to a division operator by Fourier
216 transform, which is a characteristic that is often used to solve differential equations,
217 including the wave equation:

$$218 \quad g(t) = \int h(t) dt \Leftrightarrow G(f) = \frac{H(f)}{f}$$

219 where G is the Fourier transform of g and H is the Fourier transform of h .
220 Therefore, if h is monofrequent, integrations of h are meaningless and abundant as we
221 multiply a constant before it each time. However, the real seismic signal is
222 multifrequent, and its integration will suppress the high-frequency component due to a
223 larger divisor. Unlimited integrations of the seismic signal will make it converge to
224 the amplitude at 0 frequency (M_w in fact). Therefore, the author of this paper thinks
225 integration is a fancy operation in magnitude estimation and makes it more
226 complicated. If it is truly necessary, then filtering can replace it completely. The
227 integration estimates a more accurate magnitude because of the narrower signal
228 bandwidth or lower frequency component. It is not clear which of these two factors is
229 dominant. Here, we just shallowly analyzed this issue and will avoid touching the
230 essence. Below, we concentrate on the purpose of this paper.

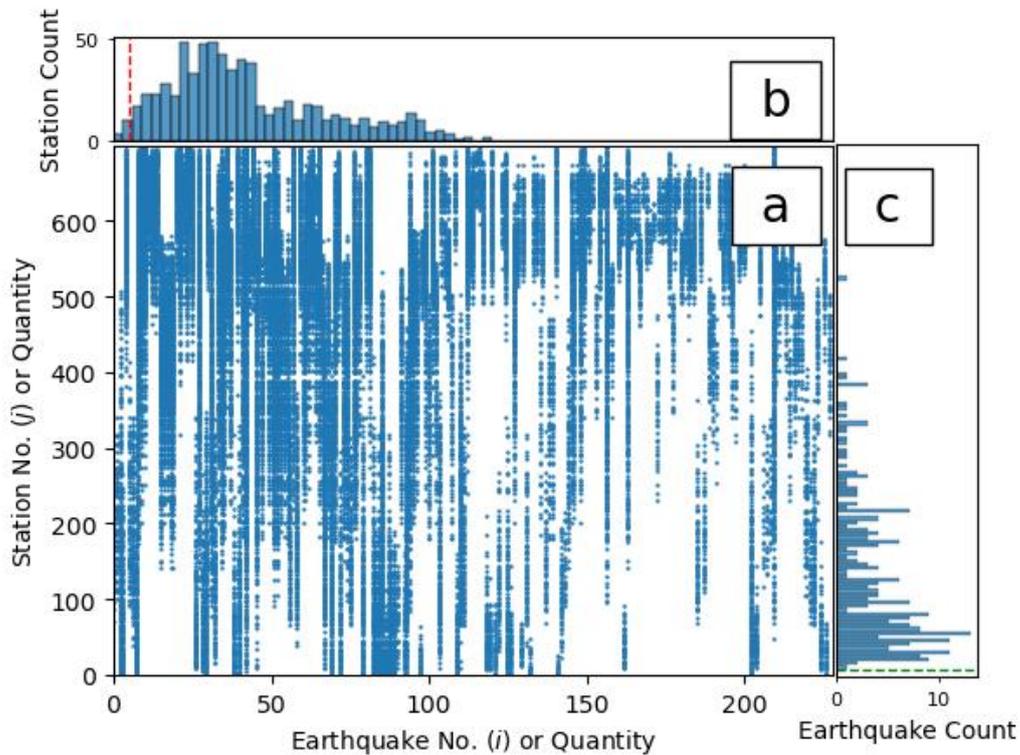
231

232 **3. Mathematical Modeling**

233 We assume that the amplitude of earthquake i at station j in the frequency domain
234 is:

$$235 \quad A_{ij}(f) = S_i(f)P_{ij}(f)L_j(f), \text{ for each } (i, j) \text{ in } I \times J \quad (1)$$

236 where $A_{ij}(f)$ can be the observed maximum amplitude of a certain phase, f is the
237 frequency, $S_i(f)$ is the source effect of earthquake i ($i = 1, 2, 3 \dots I$), $L_j(f)$ is the
238 site effect at station j ($j = 1, 2, 3 \dots J$), $P_{ij}(f)$ is the path effect determined by both
239 earthquake i and station j , and set $I \times J$ is the population of integer pairs (i, j)
240 (Figure 2a).



241

242 Figure 2. The joint distribution and marginal distribution of the 228 earthquakes and the 699

243 stations in Figure 1. (a) The scatter plot of integer pair (i, j) in which i is the earthquake number

244 and j is the station number. (b) The distribution describing how many earthquakes are recorded

245 by the stations. The x axis is the earthquake quantity, the y axis is the station count

246 corresponding to it, and point (x, y) means there are y stations that record exactly x

247 earthquakes. There are 13 stations that record fewer than 5 earthquakes (stations numbered 4, 94,

248 100, 105, 131, 165, 210, 273, 287, 321, 376, 430 and 515, namely, the bars on the right of the red

249 dotted line). (c) The distribution describing how many stations that one earthquake triggered. The

250 y axis is the station quantity, the x axis is the earthquake count corresponding to it, and point

251 (x, y) means there are x earthquakes that are recorded by exactly y stations. All earthquakes

252 are recorded by more than 10 stations (it is blank below the green dotted line).

253

254 In this paper, the author will try not to talk about frequency. We assume that the

255 frequency is a constant during the following derivation.

256

257 Taking the logarithm of both sides of Equation (1) and ignoring the frequency, we

258 obtain:

$$259 \quad \lg A_{ij} = s_i + p_{ij} + l_j, \text{ for each } (i, j) \text{ in } I \times J \quad (2)$$

260 where $s_i = \lg S_i$, $r_i = \lg R_i$ and $p_{ij} = \lg P_{ij}$.

261

262 The real site amplification is also related to the direction of incidence of the waves

263 (Papageorgiou, 1991), so it relates both to i and j . Denote the mean of site

264 amplification or site effect as l_j , and we assume l_j is a fixed value or does not

265 change much with i :

$$266 \quad l_j = \bar{l}_{ij} = l_j - \varepsilon_{ij}^l \quad (3)$$

267

268 We denote:

$$269 \quad p_{ij} = g_{ij} + h_{ij} + \varepsilon_{ij}^p + \varepsilon_{ij}^l \quad (4)$$

270 where g_{ij} is the geometrical spreading, h_{ij} is the anelastic attenuation and ε_{ij}^p is an

271 error term caused by the omission of the radiation pattern, as this factor is determined

272 both by i and j . Certainly, the 2-D radiation pattern error is related to the orientation

273 of the path vector, so it is comprehensible to classify this omitted variable bias as part

274 of the path effect.

275

276 The geometrical spreading g_{ij} is supposed to be dominated by the hypocentral

277 distance r_{ij} if the source is simply point-shaped:

$$278 \quad g_{ij} = g(r_{ij}) + \varepsilon_{ij}^g \quad (5)$$

279 where ε_{ij}^g is the uncertainty caused by the irregular shape of the earth.

280

281 Similarly, the anelastic attenuation h_{ij} is also supposed to be dominated by the

282 hypocentral distance r_{ij} if the source is simply point-shaped:

283
$$h_{ij} = h(r_{ij}) + \varepsilon_{ij}^h \quad (6)$$

284 where ε_{ij}^h is the uncertainty caused by the heterogeneity and anisotropy of the earth.

285

286 We denote the attenuation function as:

287
$$p(r_{ij}) = g(r_{ij}) + h(r_{ij}) \quad (7)$$

288 and define the magnitude corresponding to the selected phase as:

289
$$M_i = s_i - p(0) \quad (8)$$

290 We rigorousize the seismic magnitude definition in consideration of (a) Richter (1935)
 291 used the amplitude recorded by the stations at an epicentral distance of 100 kilometers,
 292 and this value is influenced not only by the earthquake size but also by the local
 293 attenuation within 100 km, so we use s_i , which is the logarithmic amplitude at the
 294 source, to characterize the earthquake size instead. (b) The limit of the amplitude at
 295 the source is infinite, so we subtract it by another infinitely large quantity $p(0)$ to
 296 obtain a finite quantity.

297

298 As previously mentioned, magnitude is an estimation of the earthquake size S_i :

299
$$M_i = d(S_i) \quad (9)$$

300 where $d(*)$ is a function to be determined. Certainly, S_i can be regarded as another
 301 kind of magnitude scale, and we usually presume that the relation between different
 302 magnitude scales is linear (Das, 2018):

303
$$d(S_i) = \alpha S_i + \beta + \varepsilon_i \quad (10)$$

304 where α and β are constants and ε_i is the uncertainty caused by the source
 305 spectrum and focal mechanism for a point-shaped source, as discussed in the
 306 Introduction. However, if one of the two is saturated, the linearity may be broken.
 307 Furthermore, please note that if the fault exhibits a certain scale, the rupture process is
 308 more appropriate to describe the earthquake than the source spectrum theory, so
 309 Equation (10) should be reconsidered.

310

311 As discussed in the Introduction, there is currently no strict definition of earthquake

312 size. In this paper, regarding the convenience of our study, we adopt the magnitude
313 published by JMA as the earthquake size:

$$314 \quad S_i = M_{\text{JMA}_i} \quad (11)$$

315

316 Equation (2) is clearly undetermined because there are more variables to be
317 determined than observed variables. The key is that the amount of p_{ij} is as large as
318 $\lg A_{ij}$. However, we know that not all the solutions of Equation (2) are physically
319 reasonable. At least, $g(r_{ij})$ and $h(r_{ij})$ should look well regulated. Castro adds a
320 smooth condition on p_{ij} to solve Equation (2) (for the latest example, Torres-Sánchez
321 & Castro, 2023). However, because p_{ij} contains error and is not smooth at all, Castro
322 plays a good trick in that he selects a portion of p_{ij} and requires the distance between
323 two closest p_{ij} larger than 10 km to ensure that p_{ij} is almost smooth. As Castro
324 represents the attenuation or calibration function by a sequence rather than a function
325 that contains only one or two parameters, this method is named the nonparametric
326 method. The disadvantage of the nonparametric method is that the acquired
327 calibration function is applicable only to the distance within which stations exist and
328 cannot be extrapolated, and as the limitation of distance between p_{ij} , it seems
329 difficult to reduce random error by increasing the data volume, at least it has not fully
330 utilized the data. On the other hand, the parametric method is much more popular and
331 presupposes the functional forms of the calibration function from physical models or
332 intuitions. By replacing p_{ij} with only one or two unknown parameters, Equation (2)
333 will be much easier to solve. However, this method is subjective and not strict most of
334 the time, which makes the extrapolation poor. In this paper, we adopt a parametric
335 method and will show that our residual sequence is stationary to the distance, and then
336 the fitting function is supposed to be extrapolated well.

337

338 Equation (2) is also undetermined because its coefficient matrix is rank deficient. This

339 feature is well specified in most FER and dummy variable regression materials. We
340 just briefly explain it: if s_i, l_j, p_{ij} is a set of solutions of Equation (2), then $s_i - \gamma, l_j +$
341 $\gamma - \delta, p_{ij} + \delta$ also satisfies Equation (2). Certainly, $h(0)$ should be 0, and it seems
342 that $p_{ij}|_{r_{ij}=0} - g(0)$ should be small, but there is still one degree of freedom in it.
343 That is, the solution space of Equation (2) is at least one-dimensional rather than a
344 point. As only the relative values of s_i and l_j can be determined by Equation (2), if
345 one would like to obtain the absolute value, one should achieve it by other means, for
346 example, part of feathers of the source spectrum (Moya et al., 2003; Wang & Shearer,
347 2019) or the site amplification of a certain station (Andrews, 1986) is known. As the
348 frequency is not discussed in this paper and the magnitude is a relative value, we
349 arbitrarily set $\gamma = \bar{l}_j$, as in Torres-Sánchez & Castro (2023).

350

351 We denote the path error as:

$$352 \quad \varepsilon_{ij} = \varepsilon_{ij}^p + \varepsilon_{ij}^g + \varepsilon_{ij}^h + \varepsilon_{ij}^r \quad (12)$$

353 Assume:

$$354 \quad g(r_{ij}) = -algr_{ij} \quad (13)$$

355 and:

$$356 \quad h(r_{ij}) = -br_{ij} \quad (14)$$

357 Rewrite Equation (2):

$$358 \quad \lg A_{ij} = s_i + l_j - algr_{ij} - br_{ij} + \varepsilon_{ij} \quad (15)$$

359 Expand each term in Equation (15):

379 theoretical hypotheses that are almost impossible to satisfy strictly in practice. It is
380 essential to design tests to check the influence of these dissatisfactions and to what
381 extent these hypotheses are approximately satisfied. The basic idea is very simple: no
382 observations, no testing. As we observe identical physical quantities much more times
383 than necessary, we can design check tests and infer the influence of the unmeasured
384 variables from these tests. In addition, without analyzing the properties of the
385 coefficient matrix and simply from artificial experiments, we can infer the properties
386 of the above model.

387

388 4.1. Test of Source Effect

389 In the Introduction, we have pointed out that the traditional earthquake magnitude
390 estimation method takes too few variables into account. However, its results have
391 been used for a very long time, and they seem reliable to some extent. How large is
392 the difference between them? Is improvement necessary? In this part, we try to
393 quantitatively analyze this issue. As there is no fixed effect in the traditional method,
394 we denote it as the 0-FER method. The method that considers only the source effect
395 and ignores the station effect (1-FER) is detailed in our former paper (Zhang et al.,
396 2023). Their statistical hypotheses are detailed in Appendix A.

397

398 Figure 3 a-c give the results of the a posteriori estimate of magnitudes in the three
399 models, namely, the normalized source term $(s_i - \beta)/\alpha$. As the distribution of the
400 source error is nonnormal and even nonsymmetric (Figure 3 e-f), we adopt median
401 regression in the 1-FER and 2-FER models instead of LSM regression. The LSM
402 regression will produce the minimum deviation $D\varepsilon_i$, but the author is afraid that the
403 result is biased in such cases. As more effects are taken into account in the models, the
404 regression seems more reliable with less $D\varepsilon_i$ and a good-looking distribution of ε_i .
405 In other words, the discrete points are more concentrated. However, the shapes of
406 discrete points are similar in Figure 3 a-c. This result indicates that the energy of the
407 site effect will leak to the source effect if not corrected; however, this leak is
408 stochastic but not i -uncorrelated, which means that the dotted line will be deflected

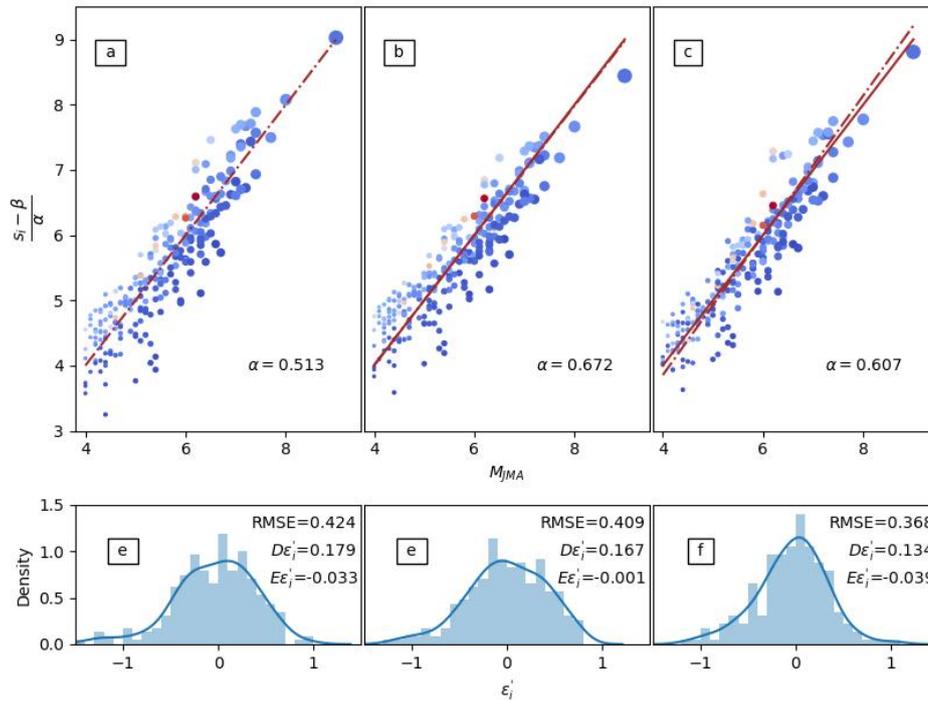
409 with an unknown but possibly small angle (i.e., the varying slope α). In addition, this
410 means that the additional conditions $\sum_j l_j = 0$ and $\sum_i s_i = 0$ in condition (AC3) are
411 not strictly met, so $D\varepsilon_i$ differs slightly. Please note that the custom deviation is
412 generated by the difference between the measured data and its mean, and it assumes
413 that the true value is unmeasurable, while the root mean square error (RMSE) is
414 generated by the difference between the measured data and the true value (namely,
415 M_{JMA}). The relation between deviation $D\varepsilon$ and RMSE $\sqrt{E\varepsilon^2}$ is:

$$416 \quad E\varepsilon^2 = D\varepsilon + (E\varepsilon)^2$$

417 The above equation shows that the error consists of two parts: a random error $D\varepsilon$ and
418 a fixed bias $E\varepsilon$.

419

420 The linearity of discrete points in Figure 3 is obvious. We have found that in Yunnan,
421 the P-wave EEW magnitude is not saturated even at a surface wave magnitude of 6.4
422 (Zhang et al., 2023). The author has been expecting to observe a plateau level and
423 preparing to adopt a nonlinear regression and extremely worried how to estimate the
424 earthquake size in Yunnan as he supposed the P-wave EEW magnitude would saturate
425 at approximately 5 m.u. If so, there will be no P-wave maximum amplitude difference
426 between an earthquake of 6 m.u. and 7 m.u. However, now, it seems an unnecessary
427 worry. In fact, the results of JMA also violate the saturation theory (Aketagawa et al.,
428 2010; Kiyomoto et al., 2010). Melgar (2015) noted that the PGD from GPS will not
429 be saturated. The further discussion of magnitude saturation is beyond the purpose of
430 this paper. We may discuss this issue in detail at a future date.



431
 432 Figure 3. (a-c) The regression of the source effect of the 0-FER, 1-FER and 2-FER models. The
 433 dotted lines are LSM regression models, while the solid lines are median regression models. (e-f)
 434 The source term error of the 0-FER, 1-FER and 2-FER models, i.e., the distance between the
 435 discrete points and the dotted line in the corresponding subplot above. Except for special
 436 instructions, all the results below are from Channel 3, i.e., the UD direction of the seismometers in
 437 the borehole.

438

439 If the source error, site error (we regard the whole site term as error here as generally
 440 it is not corrected in conventional magnitude estimating) and the path error are
 441 randomly combined, in statistics, we have the conservation of energy:

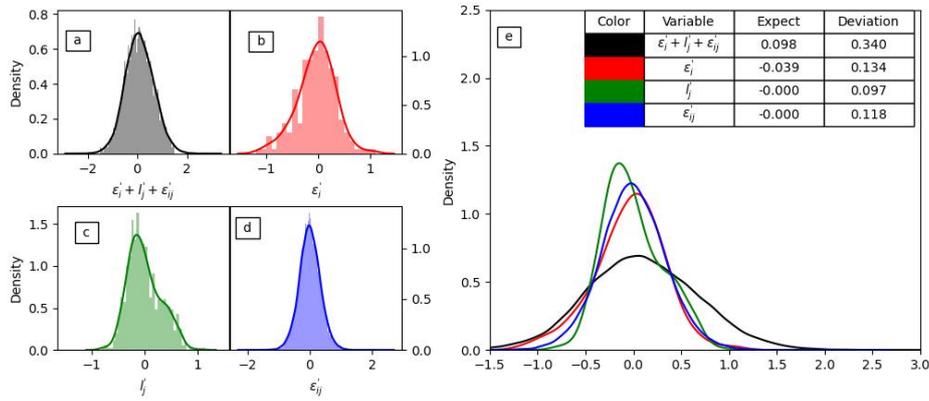
$$442 \quad D(\epsilon_i + r_j + \epsilon_{ij}) = D\epsilon_i + Dr_j + D\epsilon_{ij}$$

443 In our 2-FER model, the above equation is approximately held (Figure 4e):

$$0.340 \approx 1.03 * 0.340 = 0.349 = 0.134 + 0.097 + 0.118$$

444 The coefficient 1.03 indicates that the endogeneity has not been eliminated clean;
 445 otherwise, it will be 1. Energy leakage still exists. The author thinks it is caused
 446 primarily by the nonrandom pairing of i and j (Figure 1). The total error of magnitude

447 is not normal (black line in Figure 5), and we have pointed out that it is not quite
 448 proper to represent the error by only its standard error as if it was a normal
 449 distribution (Zhang et al., 2023). By the 2-FER technique, the author thinks that the
 450 three error components are successfully separated, and the path error looks normal, as
 451 expected (Figure 4d). Therefore, he estimates that the three errors contribute
 452 comparably to the uncertainty of magnitude, and it is necessary to handle any of them
 453 cautiously. It is quite imprecise to evaluate the magnitude uncertainty of an
 454 earthquake by its deviation, but RMSE will be a much better evaluation. Although it is
 455 difficult to obtain the RMSE just at the moment the earthquake occurs, their
 456 confounding will lead to an overoptimistic estimation of error.



457
 458 Figure 4. (a-d) The distributions of the 4 variables listed in the table of subplot (e) and their
 459 parameters. (e) Comparison and parameters of the 4 distributions.

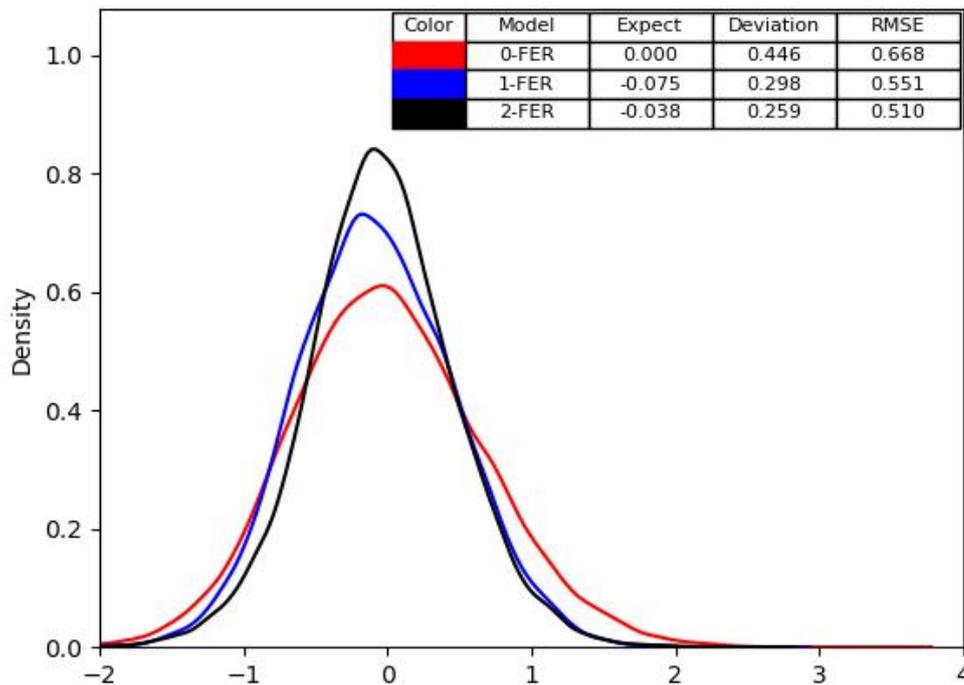
460

461 The source and path errors cannot be corrected when an earthquake just occurs. If one
 462 applies the above three models to the magnitude calculation, only the site effect and
 463 path attenuation can be corrected. We denote the left error as ε . Please note that the
 464 relation between the parameters of the 2-FER model in Figure 4 and of Figure 5 is:

$$D\varepsilon = D(\varepsilon_i + \varepsilon_{ij}) \approx D\varepsilon_i + D\varepsilon_{ij}$$

465 The magnitude error energy (deviation) drops considerably after the source effect is
 466 introduced. Surprisingly, the LSM (0-FER) model generates a result with a larger error.
 467 This is typically due to the regression's objective, and the author has explained it in
 468 Appendix A. Another explanation is that the raw data are endogenous and the basic

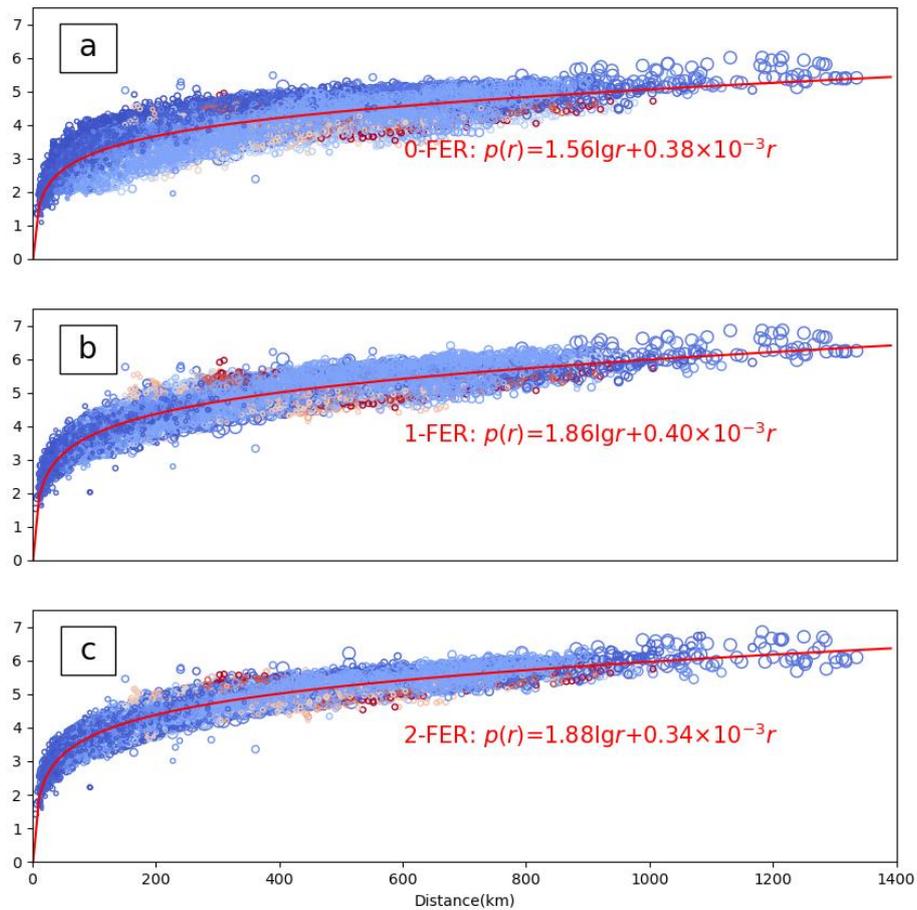
469 assumption of LSM is not satisfied. For a long time, we used LSM to address
 470 magnitude problems, pinning our hope on increasing the data volume and long-term
 471 observations, but without seriously checking whether the hypotheses of the model are
 472 met on earth.
 473



474
 475 Figure 5. The distributions and single-station magnitude errors and their parameters of the three
 476 models.

477
 478 The plot of errors versus distance (Figure 6) explains why LSM (0-FER) cannot
 479 generate a result with the minimum error from another perspective. The distance
 480 between the discrete points and the attenuation curve is the error terms in the brackets
 481 of Models (AM0), (AM1) and (AM2), i.e., the hypothetical path error. In the 0-FER
 482 model (Figure 6a) and 1-FER model (Figure 6b), obviously, the attenuation curve
 483 does not pass through the center of the discrete points. This means that the path effect
 484 is not corrected appropriately, and the close station will underestimate the magnitude,
 485 while the far station will overestimate it. To achieve a steady result, we will be forced

486 to calculate the station at various distances to average the errors. The distortion is due
 487 to the leakage of energy of source and site terms and makes the total error in Figure 5
 488 enlarged. The 2-FER model overcomes this shortcoming best, and as the curve passes
 489 through the center of the points (Figure 6c), we can deduce that this form of
 490 attenuation function is suitable for solving our physical problem.



491
 492 Figure 6. The regression of attenuation functions (solid red curve) of the three models. The colors
 493 of the points are the same as those in Figure 1c, indicating the depth of the source.

494
 495 Regarding Figure 6 as evidence, the author asserts that with suitable modeling, the
 496 effects of source, path and site can be reliably separated. As the distance increases, the
 497 variance becomes homogeneous, and the stochastic process becomes stationary. After
 498 attenuation is compensated, it is not necessary to reject far stations for their precision.

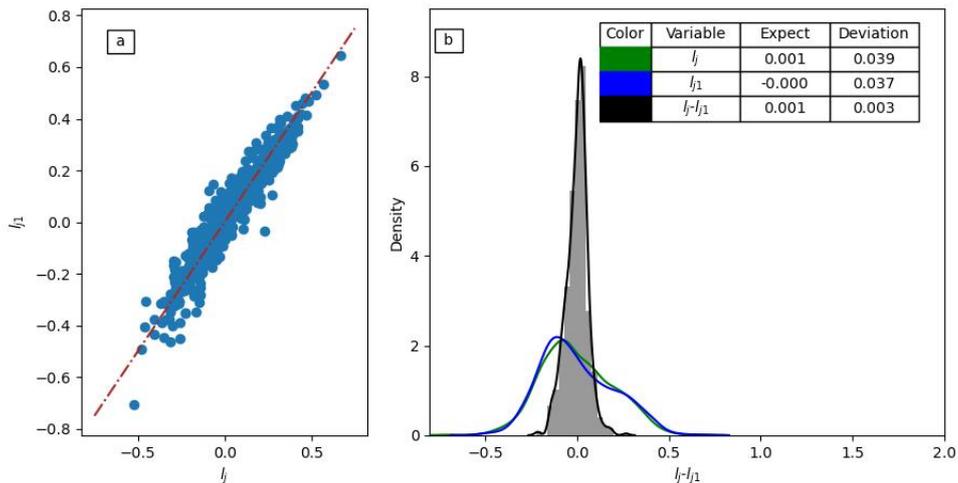
499

500 4.2. Test of Site Effect

501 We know that the model with more parameters tends to produce more accurate results
502 naturally mathematically, and a good result does not definitely indicate that a model is
503 more physically reasonable. We call this phenomenon overfitting. In this part, we
504 focus on the stability of l_j or the change in l_j over time and evaluate the overfitting by
505 the model's generalization ability.

506

507 One of our important hypotheses is that l_j is steady and does not vary with i too much.
508 We check this hypothesis by removing the data after 2015. Then, 13759 of 29290
509 records are removed. As mentioned in Appendix A, more station effects will be
510 unreliable as the data lessen, and the ridge regression of unreliable stations will
511 generate values close to 0. After these stations are removed, 661 stations remain. The
512 difference between the site effects in the two sets is approximately
513 $2 \times 0.003 / (0.037 + 0.039) = 7.9\%$ (Figure 7b). There is a strong linear correlation between
514 them (Figure 7a), which indicates that the result before 2015 is applicable up to 2023.
515 Thus, we prove that the stability assumption holds approximately.



516

517 Figure 7. (a) The relation between the subset site effect l_{j1} and complete set site effect l_j . (b) The
518 difference between l_j and l_{j1} .

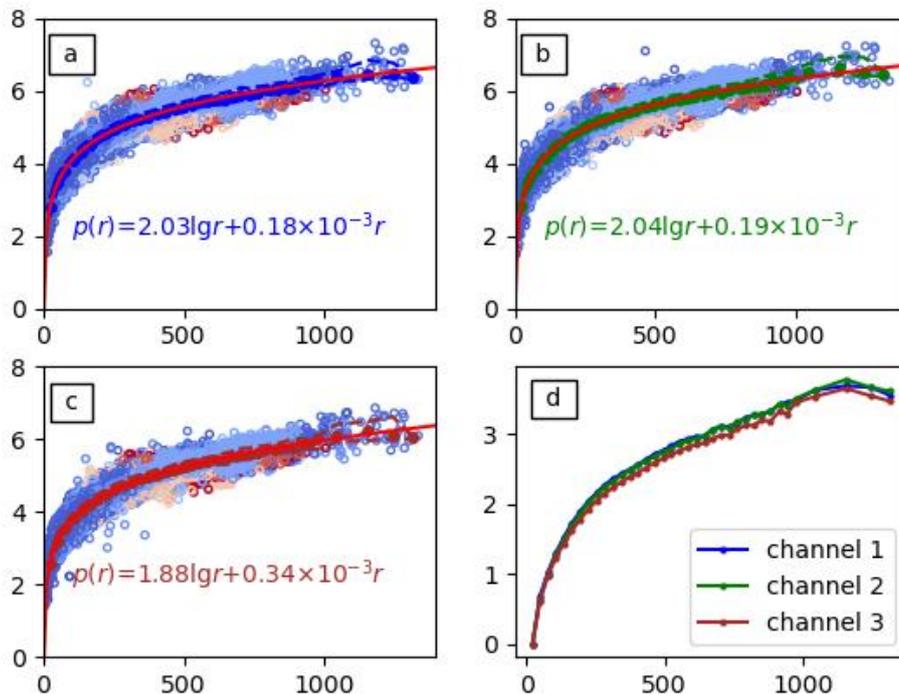
519

520 4.3. Test of Path Effect

521 As previously mentioned, we use only two parameters to characterize the attenuation.
522 Conversely, the nonparameter method introduces a large number of parameters. The
523 real attenuation may be piecewise (Motaghi & Ghods, 2012), and it may not be
524 enough to describe the attenuation by two parameters. This will result in another
525 problem - underfitting. In addition, the effects of source, path and site are coupled,
526 which means inappropriate assumptions on one of them will influence the others, and
527 results in that the attenuation function cannot approximate the path effect well.

528

529 The attenuation of three components of one seismometer should be the same if the
530 effect separating makes sense. We have checked this for the seismometers on the
531 ground, and the coefficients in $p(r_{ij})$ change slightly. An accident occurred for the
532 seismometers in the borehole: the coefficients changed slightly significantly (Figure 8
533 a-c). However, the relative trends of the curves are comparable (Figure 8d). It reminds
534 us that only the relative value of three effects can be determined as mentioned in
535 Mathematical Modeling. The individual parameter in $p(r_{ij})$ does not make sense
536 physically, and it is only an approximation for the real attenuation. One should not
537 regard the anelastic attenuation coefficient in this paper as the real property of the
538 Earth's crust. Due to the collinearity of $g(r_{ij})$ and $h(r_{ij})$, we cannot separate them
539 numerically. In the three channels, the regressed attenuation curves pass through the
540 center of the points. The author has tried to fix the coefficient before $\lg r_{ij}$ at one, and
541 then the attenuation curve does not coincide with the curve of piecewise medians, i.e.,
542 it does not pass through the center. Therefore, not all assumptions of the form of the
543 attenuation function can ensure that the regressed curve performs well with distance
544 r_{ij} . Judged by our naked eyes, it appears that the underfitting in Figure 8 is slight and
545 acceptable. Thus, subjectively, the author believes that two parameters are sufficient
546 to characterize the attenuation. The parameter method in this paper does not work
547 worse than the nonparameter method.



548

549 Figure 8. (a-c) The regression of the attenuation of Channels 1, 2 and 3. The red line represents the
 550 regressed attenuation, the line with circle markers represents the piecewise median regression, and
 551 the dotted lines represent the upper and lower quartiles. (d) The comparison of three piecewise
 552 median regressions that have been moved down so start at 0.

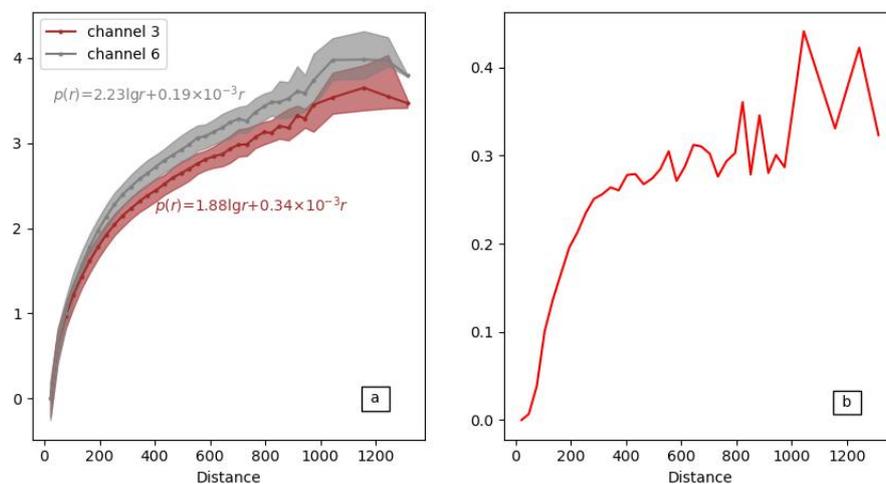
553

554 5. Extension and Discussion

555 (1) In this paper, we adopt the 2-FER model to obtain an estimation of the earthquake
 556 size, i.e., the magnitude of JMA. The author used the complete set and a subset to
 557 show the stability of the site effect, because the author found the complementary set is
 558 of low quality by plotting a figure like Figure 2. One should ensure that an earthquake
 559 contains enough stations and that a station is covered enough times before adopting
 560 the 2-FER model so that condition (AC3) is satisfied. Otherwise, one should choose a
 561 less precise model such as 1-FER or 0-FER, especially when the data volume is finite.
 562

563 (2) The author has avoided talking all the issues about frequency in this paper, but it
 564 does not mean it is not important. Instead, perhaps it is closer to the essence of the

565 magnitude estimation issue as we analyzed in Data. Initially, the author supposed that
 566 only the site effects of the two seismometers are different, and those of the source and
 567 path are identical. However, the separated path effects are different (Figure 9a). The
 568 author hypothesized that this should be due to the different dominant frequency
 569 components for seismometers in the borehole and on the ground. If so, the difference
 570 in the attenuation should mainly be due to the anelastic attenuation and should be
 571 linear. However, it seems to be logarithmic (Figure 9b). We leave this issue to the
 572 future.



573
 574 Figure 9. (a) The piecewise attenuation function of Channels 3 and 6. The shaded area is between
 575 the upper and lower quartiles, i.e., 50% of the points fall in this area. (b) The difference between
 576 the attenuation of channel 6 and channel 3.

577
 578 (3) We know that the site effect is caused by coherence emphasis and is strongly
 579 related to sediment thickness and wave frequency. In this paper, the site effect seems
 580 not very prominent, but that is possibly due to the instrument characteristics. When
 581 estimating the magnitude, we prefer a frequency at which the site effect is weak;
 582 however, when estimating the ground motion, we are more concerned about the
 583 frequency at which the site effect is strong. On the other hand, for large earthquakes, it
 584 is not enough to simplify the source to a point. Böse et al. (2012, 2018) applied an
 585 image identification method to estimate the fault length in real time. The author

586 especially stresses which equations are based on the point source assumption in
587 Mathematical Modeling. It is everyone's aim who specializes in EEW to predict
588 ground motion in real time. The author personally hopes these careful assumptions
589 will be conducive to his future research. JMA has applied the site effect correction of
590 PGD to P-wave attenuation and found virtually no use (Kiyomoto et al., 2010). This is
591 expected as the dominant frequency is different. However, it is essential to design
592 tests to study whether the site effects of P-waves and S-waves at the same frequency
593 are identical.

594

595 (4) We have mentioned the collinearity of $g(r_{ij})$ and $h(r_{ij})$ in our former works
596 (Zhang et al., 2023) and wishfully thought the coefficient in $g(r_{ij})$ should be 1 from
597 a point source model in the homogeneous medium of a semi-infinite half-space. The
598 author apologizes that this is a mistake, and it is proven to be wrong by the mid-field
599 data of Japan's KiK-net. Please note that the coefficients do not make sense in physics
600 in his papers actually; they are actually a mathematical approximation.

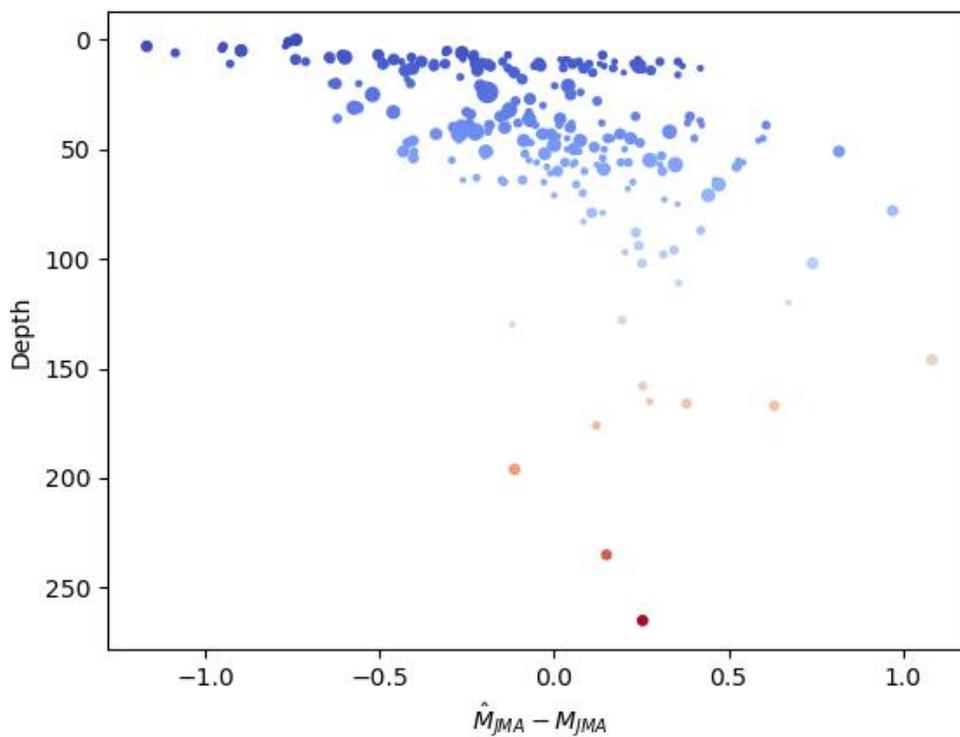
601

602 (5) The RMSE of the multistation-average magnitude of the PGD magnitude formula
603 by JMA is 0.309 after correction of the site effect and 0.319 before (Kiyomoto et al.,
604 2010). The corresponding error to this paper is approximately 0.368. The author is not
605 sure if JMA calculates the magnitude after simulation, but according to the published
606 literature, its initial equation is from the manual earthquake catalog (Aketagawa,
607 2010). Regardless, the author infers that the estimation precision will be higher in the
608 frequency domain, both because the original physical quantities are frequency related
609 and α of long-period waves is larger. This paper focuses on the basic properties of
610 errors in magnitude estimation, and it is still far from the optimal estimation of
611 magnitude. Different magnitudes are measured with different components in the
612 waves; thus, the magnitude estimation technique actually estimates one component
613 according to the measurement of another component. In addition to using a specific
614 frequency in the waveform to estimate the magnitude, it seems better to use the whole

615 observed bandwidth to predict an unobserved component. This idea is called spectral
616 magnitude (Duda & Yanovskaya, 1993), and it may be useful to reduce the source
617 error ε_i . Before looking for the optimal map from the source spectral to magnitude
618 (namely, the field of functional analysis), the author thinks the work of this paper is
619 essential and fundamental.

620

621 (6) One may notice that in Figure 3c, the colors of points above the line are mostly
622 light cool or warm, whereas those below are heavily cool. However, in Figure 6c, the
623 warm points are on both sides of the line. Therefore, it looks the magnitude relates to
624 depth. Gutenberg and Richter (1956) thought that attenuation also relates to depth. As
625 the source and path effects are coupled, we are not sure that the depth dependency is
626 numerically caused by attenuation or source. In addition, the data set we used is too
627 small to determine this dependency (Figure 10). JMA has used a linear term to correct
628 the effect of focal depth during EEW (Aketagawa et al., 2010; Kiyomoto et al., 2010).



629

630 Figure 10. Single-earthquake magnitude estimation error versus focal depth. The size of the circle

631 is directly proportional to the magnitude, and a warmer color indicates a deeper depth.

632

633 **6. Conclusion**

634 In this paper, the author introduces a statistical technique named 2-FER. It is difficult
635 and not enough to analyze the mathematical property of this technique in theory. For
636 example, its coefficient matrix is not full rank; therefore, its condition number cannot
637 tell us if the relative value of the solution is robust or not. The author designs a series
638 of check tests to carefully examine hypotheses of this mathematical approach and
639 conclude that the relative value of the source, path and site effects is reliable. Thus,
640 2-FER is a proper tool to analyze the errors in magnitude estimation. After checking
641 the deviation additivity and stationarity of the error process with distance r_{ij} , the
642 author asserts that the endogeneity of the original issue is finely addressed, the energy
643 leaks slightly and the different errors in the estimated magnitude are decoupled
644 correctly. In addition, the author compares them quantitatively and explains in detail
645 which physical factors impact each of the three errors.

646

647 The focal mechanism affects both the source and path effects. The dip affects the
648 source effect, and the trend affects the path effect. The author has noted that the path
649 error constituents are complicated by Equation (12). The author is not sure if it is
650 mainly from the focal mechanism or radiation pattern and if its accuracy is sufficient
651 to determine the radiation pattern in real time.

652

653 The simple point source assumption is not a suitable assumption to address near-field
654 ground motion prediction, which is an important issue in EEW. The station density of
655 the EEW Network is sufficient to determine a complex source in real time (Böse et al.,
656 2012, 2018). A large earthquake usually consists of a few main ruptures and thus can
657 be regarded as a few moderate earthquakes. If the waves from different moderate
658 sources can be separated, then we can replace a large earthquake with a few points.
659 Then, the analysis of large earthquakes during EEW will be largely simplified and
660 closer to the physical matter. We have proven that 2-FER is a powerful tool to handle

661 point sources. However, it is nearly impossible to separate waves from different
662 orientations in only one channel, but it may be possible for three-component seismic
663 data (Lei, 2005). AI methods may also help in this area.

664

665 The author makes an effort to analyze the magnitude error as he hopes to determine
666 the rupture as quickly as possible from minor differences during EEW in the future.
667 The regression of Equation (9) is not necessarily linear. Thus, it does not require us to
668 measure the maximum amplitude, and we may research the time evolution of PGD
669 more carefully with 2-FER. As this evolution is nonlinear and error-sensitive,
670 mainstream geophysicists disagree with rupture determinism in a probabilistic view
671 (Trugman et al., 2019; Meier et al., 2021). Thus, methods with a higher precision are
672 quite crucial no matter whether to support it or against it in a physical way.

673

674 Although the author is introducing 2-FER for EEW purposes, the analysis method in
675 this paper is not limited. To decouple the errors successfully, the key is a proper
676 choice of attenuation function when applying it to surface waves or P-waves in other
677 regions. The error process stationarity can help us to determine if the attenuation is
678 appropriate. However, the author thinks that current tests of the application of 2-FER
679 in seismology are not sufficient. Before being adopted, careful and further inspection
680 is still essential.

681

682 **Appendix A: Preconditions and Objectives of Different Regressions**

683 The statistical model of 0-FER is:

$$684 \quad \lg A_{ij} = \alpha S_i - a \lg r_{ij} - b r_{ij} + \beta + (\varepsilon_{ij} + l_j + \varepsilon_i) \quad (\text{AM0})$$

685 The objective of Regressing Model (AM0) is to find the minimum of:

$$686 \quad \min_{a,b} \sum_{i,j} (\lg A_{ij} - \alpha S_i + a \lg r_{ij} + b r_{ij} - \beta)^2 \quad (\text{AO0})$$

687 Additionally, to achieve reliable $a \lg r_{ij} + b r_{ij}$ by the least minimum square method

688 (LSM), we should constrain:

689
$$\sum_{i,j} (\varepsilon_{ij} + l_j + s_i) \rightarrow 0 \quad (\text{AC0})$$

690

691 The statistical model of 1-FER is:

692
$$\lg A_{ij} = \sum_{k=1}^N s_i \delta_{ik} - a \lg r_{ij} - b r_{ij} + (\varepsilon_{ij} + l_j) \quad (\text{AM1})$$

693 The objective of Regressing Model (AM1) is to find the minimum of:

694
$$\min_{s,a,b} \sum_{i,j} (\lg A_{ij} - \sum_{k=1}^N s_i \delta_{ik} + a \lg r_{ij} + b r_{ij})^2 \quad (\text{AO1})$$

695 where $s = (s_1, s_2 \dots s_N)$ and δ_{*k} is the Kronecker function, which is also referred to
696 as a dummy variable in statistics.

697 Additionally, to achieve reliable s_i by LSM, we should constrain:

698
$$\sum_j (\varepsilon_{ij} + l_j) \rightarrow 0, \text{ for each } i \quad (\text{AC1})$$

699

700 The statistical model of 2-FER is:

701
$$\lg A_{ij} = \sum_{k=1}^N s_i \delta_{ik} + \sum_{k=1}^M l_j \delta_{jk} - a \lg r_{ij} - b r_{ij} + (\varepsilon_{ij}) \quad (\text{AM2})$$

702 The objective of Regressing Model (AM2) is to find the minimum of:

703
$$\min_{s,l,a,b} \sum_{i,j} (\lg A_{ij} - \sum_{k=1}^N s_i \delta_{ik} - \sum_{k=1}^M l_j \delta_{jk} + a \lg r_{ij} + b r_{ij})^2 \quad (\text{AO2})$$

704 where $\mathbf{l} = (l_1, l_2 \dots l_M)$.

705 Additionally, to achieve reliable s_i and r_j by LSM, we should constrain:

706
$$\begin{cases} \sum_i \varepsilon_{ij} \rightarrow 0, \text{ for each } j \\ \sum_j \varepsilon_{ij} \rightarrow 0, \text{ for each } i \end{cases} \quad (\text{AC2})$$

707 The objective of ridge regression is to find the minimum of:

708
$$\min_{s,l,a,b} \sum_{i,j} (\lg A_{ij} - \sum_{k=1}^N s_i \delta_{ik} - \sum_{k=1}^M l_j \delta_{jk} + a \lg r_{ij} + b r_{ij})^2 + \lambda (s^2 + \mathbf{l}^2 + a^2 + b^2)$$

709 Therefore, the ridge regression will find an estimator that not only produces a result
710 with less error but also smaller coefficients. In this paper, the author has not rejected

711 the data that may harm condition (AC2) (namely, has not removed stations covered
712 less than 5 times, as mentioned in Figure 2c). The author chooses ridge regression to
713 make the result look not that strange, as there will be some outliers for unreliable r_j .
714 This operation covers up the bad results but please note this is not fabrication or
715 falsification and the author just does not want the few abnormally large values mask
716 the true trends in the following figure.

717

718 The logical relationships between the three preconditions are as follows:

$$719 \quad \left. \begin{array}{l} AC2 \\ \sum_j l_j = 0 \end{array} \right\} \Rightarrow AC1 \left. \begin{array}{l} \\ \sum_i s_i = 0 \end{array} \right\} \Rightarrow AC0 \quad (AC3)$$

720

721 Please note that as we concern only the relative value in this paper, $l_j = 0$ here
722 means l_j is a constant that is uncorrelated to j , as is s_i .

723

724 In the 0-FER model, if we denote

$$725 \quad \alpha \widehat{S}_{ij} = \lg A_{ij} + a \lg r_{ij} + b r_{ij} - \beta$$

726 where \widehat{S}_{ij} is the estimated size of earthquake i at station j . Then, the objective
727 function can be rewritten as:

$$728 \quad \min_{a,b} \sum_{i,j} \alpha (\widehat{S}_{ij} - S_i)^2$$

729 This objective prefers a small α . If one wants an optimal magnitude estimation, i.e.,
730 the desired objective is

$$731 \quad \min_{a',b'} \sum_{i,j} (\widehat{S}'_{ij} - S_i)^2 \quad (AC0')$$

732 the Model (AM0) should be rewritten as

$$733 \quad S_i = \alpha' \lg A_{ij} + a' \lg r_{ij} + b' r_{ij} - \beta' - (\varepsilon'_{ij} + l'_j + \varepsilon'_i) \quad (AM0')$$

734 The Model (AM0') minimizes the error of the single-station magnitude instead of the
735 amplitude, while Model (AM1) minimizes the error of the single-earthquake or

736 multistation-average magnitude. The hyperplane defined by (AM0') performs even
737 slightly better than (AM2), although the author did not plot it in Figure 5. The
738 objective (AC0') requires (AM0') to be the best estimation of single-station magnitude
739 mathematically, but the author does not think it is physical, as it overestimates the
740 magnitude of far stations even more seriously than (AM0). On the other hand, the
741 model (AM0') performs slightly worse than (AM0) in the view of single-earthquake
742 magnitude error in the data set of this paper, but the coefficients differ greatly, which
743 means that the ordinary linear regressions are not robust.

744

745 JMA explained that it is their preference to fix the coefficient before $\lg A_{ij}$ at one
746 (Aketagawa et al., 2010). Wu and Zhao (2006) also adopt the model form of (AM0).
747 However, this form makes the error term appear small as it multiplies the real error
748 term ε'_i by $\alpha < 1$, especially for a short time window such as 3 seconds and then the
749 α is quite small (Wu & Zhao, 2006). JMA found that the regression model calculated
750 by stations in 200 km has a better RMSE in 500 km than the model in 500 km. This
751 result is comprehensible, as the α in the model of 200 km is larger. JMA thought the
752 model with the larger error is better, as it is from a larger data volume (Aketagawa et
753 al., 2010).

754

755 To remain consistent with the former works, the author also fixes the coefficient
756 before $\lg A_{ij}$ at one. However, please note that the author does not think it is a
757 definitely good tradition and it brings an overoptimistic looking result. To reduce its
758 influence, the author did not regress on Equation (10) directly in the 1-FER and
759 2-FER models. Instead, in this paper, the author regressed on the following model:

760

$$S_i = \alpha' M_i + \beta' + \varepsilon'_i$$

761

762 **Acknowledgments**

763 The regressions in this paper were completed with statsmodels and sklearn, the
764 seismic data were preprocessed with obspy, and the drawings in this paper were

765 plotted with matplotlib and seaborn. The author would like to thank the developers of
766 these open-source third-party packages of Python.

767

768 **Data Availability Statement**

769 The intermediate data and programs used for processing and plotting are available in
770 Mendeley Data, V1 (<https://data.mendeley.com/datasets/y4sjgg7vxs/1>,doi:
771 10.17632/y4sjgg7vxs.1). The original waveform data can be downloaded from the
772 official site of K- & KiK-net (<https://www.kyoshin.bosai.go.jp>).

773

774 **References**

775 Allen, R. M., & Kanamori, H. (2003). The potential for earthquake early warning in
776 southern California. *Science*, 300(5620), 786-789.

777 <https://doi.org/10.1126/science.1080912>

778

779 Allison, P. D. (2009). *Fixed Effects Regression Models*. America: SAGE Publications.

780

781 Atefi, S., Heidari, R., Mirzaei, N., & Siahkoochi, H. R. (2017). Rapid Estimation of
782 Earthquake Magnitude by a New Wavelet-Based Proxy. *Seismological Research*

783 *Letters*, 88(6), 1527-1533. <https://doi.org/10.1785/0220170146>

784

785 Aketagawa Tamotsu, et al (2010). Improvement of P-wave magnitude estimation for
786 earthquake early warnings from JMA. *Quarterly Journal of Seismology* 73. (In

787 Japanese)

788

789 Andrews, D. J. (1986). Objective determination of source parameters and similarity of
790 earthquakes of different size. *Earthquake source mechanics*, 37, 259-267.

791

792 Böse M, Smith D E, Felizardo C, Meier M-A, Heaton T H, Clinton J F (2018), FinDer
793 v.2: Improved real-time ground-motion predictions for M2–M9 with seismic

794 finite-source characterization, *Geophysical Journal International*, Volume 212, Issue 1,

795 January 2018, Pages 725–742, <https://doi.org/10.1093/gji/ggx430>

796

797 Böse M, Thomas H. Heaton, Egill Hauksson (2012), Real-time Finite Fault Rupture
798 Detector (FinDer) for large earthquakes, Geophysical Journal International, Volume
799 191, Issue 2, November 2012, Pages 803–812,
800 <https://doi.org/10.1111/j.1365-246X.2012.05657.x>

801

802 Brillinger, D. R., & Preisler, H. K. (1984). An exploratory analysis of the
803 Joyner-Boore attenuation data. Bulletin of the Seismological Society of
804 America, 74(4), 1441-1450.

805

806 Bormann, P., Wendt, S., DiGiacomo, D. (2013): Seismic Sources and Source
807 Parameters. - In: Bormann, P. (Ed.), New Manual of Seismological Observatory
808 Practice 2 (NMSOP2), Potsdam : Deutsches GeoForschungs Zentrum GFZ, 1-259.
809 https://doi.org/10.2312/GFZ.NMSOP-2_ch3

810

811 Colombelli, Simona; Festa, Gaetano; Zollo, Aldo (2020). Early rupture signals predict
812 the final earthquake size. Geophysical Journal International, (), ggaa343–.
813 doi:10.1093/gji/ggaa343

814

815 Colombelli, S., Zollo, A., Festa, G., & Picozzi, M. (2014). Evidence for a difference
816 in rupture initiation between small and large earthquakes.
817 Nature Communications, 5, 3958. <https://doi.org/10.1038/ncomms4958>

818

819 Das, Ranjit; Wason, H. R.; Gonzalez, Gabriel; Sharma, M. L.; Choudhury, Deepankar;
820 Lindholm, Conrad; Roy, Narayan; Salazar, Pablo (2018). Earthquake Magnitude
821 Conversion Problem. Bulletin of the Seismological Society of America, 108(4),
822 1995–2007. doi:10.1785/0120170157

823

824 Diggle, Peter J.; Heagerty, Patrick; Liang, Kung-Yee; Zeger, Scott L. (2002). Analysis

825 of Longitudinal Data (2nd ed.). Oxford University Press.
826 pp. 169–171. ISBN 0-19-852484-6.
827
828 Duda, S. J., and Yanovskaya, T. B. (1993). Spectral amplitude-distance curves for
829 P-waves: effects of velocity and Q-distribution. *Tectonophysics*, 217, 255-265.
830
831 Furumura, T., & Oishi, Y. (2023). An early forecast of long-period ground motions of
832 large earthquakes based on deep learning. *Geophysical Research Letters*, 50,
833 e2022GL101774. <https://doi.org/10.1029/2022GL>
834
835 Gardiner, Joseph C.; Luo, Zhehui; Roman, Lee Anne (2009). "Fixed effects, random
836 effects and GEE: What are the differences?". *Statistics in Medicine*. 28 (2):
837 221–239. doi:10.1002/sim.3478. PMID 19012297. S2CID 16277040.
838
839 Greene, W.H., 2011. *Econometric Analysis*, 7th ed., Prentice Hall
840
841 Gutenberg B, Richter CF (1956) Magnitude and energy of earthquakes. *Ann Geofis*
842 9:1–15
843
844 Iwata, Tomotaka; Irikura, Kojiro (1986). Separation of Source, Propagation and Site
845 Effects from Observed S-Waves. *Zisin (Journal of the Seismological Society of Japan*.
846 2nd ser.), 39(4), 579–593. doi:10.4294/zisin1948.39.4_579 (in Japanese)
847
848 Lei Jun (2005). A method for non-orthogonal seismic polarization-vector separation. ,
849 162(3), 965–974. doi:10.1111/j.1365-246x.2005.02709.x
850
851 Joyner, W. B. and D. M. Boore (1981), Peak horizontal acceleration and velocity from
852 strong-motion records including records from the 1979 Imperial Valley, California,
853 earthquake, *Bull. Seism. Soc. Am.* 71, 2011-2.
854

855 Kiyomoto Masashi, et al (2010). Investigation of Technical Issues for Earthquake
856 Early Warning. Quarterly Journal of Seismology 73. (In Japanese)
857

858 Meier, M. A., Ampuero, J. P., Cochran, E., & Page, M. (2021). Apparent earthquake
859 rupture predictability. Geophysical Journal International, 225(1), 657-663.
860

861 Melgar, D., B. W. Crowell, J. Geng, R. M. Allen, Y. Bock, S. Riquelme, E. M. Hill, M.
862 Protti, and A. Ganas (2015), Earthquake magnitude calculation without saturation
863 from the scaling of peak ground displacement, Geophys. Res. Lett., 42, 5197–5205,
864 doi:10.1002/2015GL064278.
865

866 Moya, A. (2000). Inversion of Source Parameters and Site Effects from Strong
867 Ground Motion Records using Genetic Algorithms. Bulletin of the Seismological
868 Society of America, 90(4), 977–992. doi:10.1785/0119990007
869

870 Motaghi, K., & Ghods, A. (2012). Attenuation of Ground-Motion Spectral Amplitudes
871 and Its Variations across the Central Alborz Mountains. Bulletin of the Seismological
872 Society of America, 102(4), 1417–1428.
873

874 Münchmeyer, J., Leser, U., & Tilmann, F. (2022). A probabilistic view on rupture
875 predictability: All earthquakes evolve similarly. Geophysical Research Letters, 49,
876 e2022GL098344. <https://doi.org/10.1029/2022GL098344>
877

878 National Research Institute for Earth Science and Disaster Resilience (2019), NIED
879 K-NET, KiK-net, doi:10.17598/NIED.0004
880

881 Nuttli, O. W.(1973), Seismic wave attenuation and magnitude relations for eastern
882 North America, J. Geophys. Res., 78, 876-885,.
883

884 Papageorgiou, A. S., & Kim, J. (1991). Study of the propagation and amplification of

885 seismic waves in Caracas Valley with reference to the 29 July 1967 earthquake: SH
886 waves. *Bulletin of the Seismological Society of America*, 81(6), 2214-2233.
887
888 Phillips, W. Scott, Keiiti Aki; Site amplification of coda waves from local earthquakes
889 in central California. *Bulletin of the Seismological Society of America* 1986;; 76 (3):
890 627–648. doi: <https://doi.org/10.1785/BSSA0760030627>
891
892 Parolai, S. (2004). Comparison of Different Site Response Estimation Techniques
893 Using Aftershocks of the 1999 Izmit Earthquake. *Bulletin of the Seismological*
894 *Society of America*, 94(3), 1096–1108. doi:10.1785/0120030086
895 Richter, C. F. (1935). An instrumental earthquake magnitude scale, *Bull. Seism. Soc.*
896 *Am.* 25, 1-32.
897
898 Seggern , David von (1970); The effects of radiation patterns on magnitude estimates.
899 *Bulletin of the Seismological Society of America*; 60 (2): 503–516. doi:
900 <https://doi.org/10.1785/BSSA0600020503>
901
902 Song, Jindong; Zhu, Jingbao; Li, Shanyou. MEANet: Magnitude Estimation Via
903 Physics-based Features Time Series, an Attention Mechanism, and Neural Networks.
904 *Geophysics*. 2023. 88(1): V33-V43.
905
906 Takemura, Masayuki; Kato, Kenichi; Ikeura, Tomonori; Shima, Etsuzo (1991). Site
907 Amplification of S-Waves from Strong Motion Records in Special Relation to Surface
908 Geology. *Journal of Physics of the Earth*, 39(3),
909 537–552. doi:10.4294/jpe1952.39.537
910
911 Taylor, John R. (2022), *An Introduction to Error Analysis: The Study of Uncertainties*
912 *in Physical Measurements (Third Edition)*, USA: University Science Books
913
914 Torres-Sánchez, E. M., & Castro, R. R. (2023). P-and S-wave attenuation in the

915 northern region of the Gulf of California, Mexico. *Journal of Seismology*, 1-17.
916

917 Trugman, D. T., Page, M. T., Minson, S. E., & Cochran, E. S. (2019). Peak ground
918 displacement saturates exactly when expected: Implications for earthquake early
919 warning. *Journal of Geophysical Research: Solid Earth*, 124, 4642–4653.
920 <https://doi.org/10.1029/2018JB017093>
921

922 United States Geology Survey (2023a): Magnitude Types,
923 <https://www.usgs.gov/programs/earthquake-hazards/magnitude-types>
924

925 United States Geology Survey (2023b): Origin of M 4.9 - 115 km SE of Honiara,
926 Solomon Islands,
927 <https://earthquake.usgs.gov/earthquakes/eventpage/us7000kwk7/origin/detail>
928

929 Wang, W., & Shearer, P. M. (2019). An improved method to determine coda-Q,
930 earthquake magnitude, and site amplification: Theory and application to southern
931 California. *Journal of Geophysical Research: Solid Earth*, 124, 578–598.
932 <https://doi.org/10.1029/2018JB015961>
933

934 Wang, Y., Li, X., Li, L., Wang, Z., & Lan, J. (2022). New magnitude proxy for
935 earthquake early warning based on initial time series and frequency. *Seismological*
936 *Research Letters*, 93(1), 216-225.
937

938 Wu, Y.M. & Zhao, L., 2006. Magnitude estimation using the first three seconds
939 P-wave amplitude in earthquake early warning, *Geophys. Res. Lett.*, 33, L16312,
940 doi:10.1029/2006GL026871
941

942 Zhang, G., D. Li, Y. Gao, and J. Yang (2022). A Magnitude Estimation Approach and
943 Application to the Yunnan Earthquake Early Warning Network, *Seismol. Res. Lett.* 94,
944 234–242, doi: 10.1785/0220220093.

945

946 Zhang, G. (2023). Programs of Two-way Fixed Effects Regression on the Error
947 Analysis of PGA Magnitude (Version 1) [Program]. Mendeley Data.

948 <https://doi.org/10.17632/y4sjgg7vxs.1>

949

950 Zhu, Jingbao; Li, Shanyou; Song Jindong. Hybrid Deep-Learning Network for Rapid
951 On-Site Peak Ground Velocity Prediction. 2022. IEEE Transactions on Geoscience
952 and Remote Sensing. vol. 60, pp. 1-12, 2022, Art no. 5925712