

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Comment on "Two Foreshock Sequences Post Gulia and Wiemer (2019)" by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky Comments and Replies

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Comment on "Two Foreshock Sequences Post Gulia and Wiemer (2019)" by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky Comments and Replies / Gulia, L; Wiemer, S. - In: SEISMOLOGICAL RESEARCH LETTERS. - ISSN 0895-0695. - STAMPA. - 92:5(2021), pp. 3251-3258. [10.1785/0220200428]

Availability:

This version is available at: <https://hdl.handle.net/11585/854900> since: 2022-02-10

Published:

DOI: <http://doi.org/10.1785/0220200428>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Laura Gulia; Stefan Wiemer, Comment on “Two Foreshock Sequences Post Gulia and Wiemer (2019)” by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky, *Seismological Research Letters* (2021) 92 (5): 3251–3258.

The final published version is available online at:
<https://doi.org/10.1785/0220200428>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

1 **Comment on “Two foreshock sequences post Gulia and Wiemer (2019)”**

2 **by Kelian Dascher-Cousineau, Thorne Lay, Emily E. Brodsky**

3
4 Laura Gulia^{1*} and Stefan Wiemer²

5
6 ¹ University of Bologna, Department of Physics and Astronomy, Bologna.

7 ² Swiss Seismological Service, ETH Zurich, Switzerland.

8
9 L. Gulia: Laura Gulia (laura.gulia@unibo.it)

10 S. Wiemer: Stefan Wiemer (stefan.wiemer@sed.ethz.ch)

11
12 **corresponding author: Laura Gulia, laura.gulia@unibo.it*

13
14 **Abstract**

15 Kelian Dascher-Cousineau et al. (2020) apply the so-called Foreshock Traffic Light
16 System (FTLS) model proposed by Gulia and Wiemer (2019) to two earthquake
17 sequences that occurred after the submission of the model: the 2019 Ridgecrest (M7.1)
18 and the 2020 Puerto Rico (M6.4) earthquakes. We show in this comment that the
19 method applied by Kelian Dascher-Cousineau (2020) deviates in at least six substantial
20 and not well documented aspects from the original FTLS method. As a consequence,
21 they used for example in the Ridgecrest case only 1% of the data available to estimate b-
22 values and from a small sub volume of the relevant mainshock fault. In the Puerto Rico
23 case, we document here substantial issues with the homogeneity of the magnitude scale
24 that in our assessment make a meaningful analysis of b-values impossible. We conclude
25 that the evaluation by Kelian Dascher-Cousineau et al. (2020) is misrepresentative and

26 a not a fair test of the FTLS hypothesis.

27

28 ***Introduction and context***

29 Kelian Dascher-Cousineau (2020, from now on DC2020) apply the so-called Foreshock
30 Traffic Light System (FTLS) model proposed by Gulia and Wiemer (2019, from now on
31 GW2019) to two earthquake sequences that occurred after the submission of the model:
32 the 2019 Ridgecrest (M7.1) and the 2020 Puerto Rico (M6.4) earthquakes. We
33 appreciate that DC2020 decided to evaluate our model and hypothesis pseudo-
34 prospectively on independent data, and partially with their own code implementation.
35 This is exactly how science needs to work: hypotheses proposed by one group need to
36 be evaluated independently by others. For this reason, we provided as part of GW2019
37 also the source code used for the analysis. However, in our assessment documented
38 here, the study by DC2020 contains substantial deviations from the originally proposed
39 method, including demonstratable errors, which then lead the authors to partially
40 incorrect conclusions.

41

42 Since DC2020 did not provide their source code nor their datasets as part of the
43 publication, we requested them directly from the authors, who kindly supplied them for
44 the Ridgecrest case. This comment addresses the deviations introduced by DC2020 in
45 their study for each of the two mainshocks individually and draws some common
46 conclusions.

47

48

49

50

51 **Ridgecrest case study**

52 For the first sequence (Ridgecrest), the analysis by DC2020 resulted in a red FTLS alert
53 after the M6.4 event and in an orange alert after the M7.1 event. Meanwhile in the same
54 SRL issue, Gulia et al. (2020) published their own pseudo-prospective assessment of
55 this sequence, reporting also a red FTLS alert after the M6.4, but a green alert following
56 the M7.1 two days later. The observed differences between these two papers in the
57 FTLS setting and in the underlying b-value time series are a direct consequence of the
58 substantial deviations from the GW2019 approach as implemented by DC2020. Below
59 we document these deviations in methodology introduced by DC2020 step by step.

60

61

62 **1. Correctly establish the reference b-value for the first mainshock**

63

64 A critically important parameter to be established in the FTLS model is the local
65 reference b-value, because the FTLS decisions are based on the difference in percent
66 between the sequence-specific b-values and the reference b-value. According to the
67 GW2019 hypothesis, it is important to establish the reference b-value such that: 1) it is
68 only based on earthquake immediately near the initiating mainshock fault (i.e., within 3
69 km of the fault), since b-values vary substantially with space; 2) uses a long time series,
70 to have the statically most robust estimate that averages over temporal variations. For
71 the background of Californian sequences in Gulia et al. (2018), we start our analysis
72 from 1981, when the network was greatly improved (e.g., Tormann et al., 2014).
73 Therefore, for our Ridgecrest analysis presented in Gulia et al. (2020), we use a 39-year-
74 long background catalog. DC2020 started their analysis only in the year 2000, resulting
75 in a factor of two reduction of the data used. This choice was made to avoid the

76 influence of the aftershocks of Landers (M7.3 in 1992) and Hector Mine (M 7.1 in 1999)
77 (*Kelian Dascher-Cousineau*, personal communication). However, these two sequences
78 occurred, about 200 and 170 km from the Ridgecrest mainshock, respectively, distances
79 well beyond the Gardner and Knopoff (1974) radii of influence for both magnitudes and
80 neither of these mainshocks had a noticeable impact on earthquake rates in the
81 Ridgecrest area. We thus consider 1981 the better justified starting date, but this choice
82 is indeed a subjective one and not fully automate in the approach yet.

83

84 **Deviation 1.1:** DC2020 use a catalog from the year 2000, while we would advise (and
85 do so in Gulia et al. (2018, 2020)) to use data from 1981.

86

87 Estimating reliable b-values also requires a robust, automated estimation of the
88 magnitude of completeness. In GW2019, we use the so-called maximum curvature
89 method by Woessner and Wiemer (2005) and apply it as suggested in their paper: we
90 first cut for robustness the catalog close to the overall catalog completeness and then
91 re-estimate M_c for each time-step. We differentiate in purpose between the background
92 b-values estimation, where we apply as an overall M_c cut ($M_{c_maxCurve_overall} - 0.2$),
93 and aftershocks sequence that are both data rich and have strongly varying M_c with
94 time, where we apply as on overall M_c cut of ($M_{c_maxCurve_overall}$, thus 0.2 higher. In
95 both cases, we then re-estimate M_c in each time bin using $M_{c_MaxCurv} + 0.2$. This
96 procedure has been documented in the paper and in detail in the source code.

97

98 DC2020 argued that the approach outlined above is actually an error in our code that
99 they detected (which it is not) and modified it such that they added an additional M_c
100 increment I of +0.2 for estimating the background b-values also.

101

102 **Deviation 1.2:** DC2020 apply erroneously a 'safety' M_c increment of +0.4 rather than
103 +0.2 for the background b-value calculations.

104

105 These two deviations from our published method decrease the number of events
106 available to establish the reference b-value with 3 km of the fault plane by 93% percent,
107 from 1154 to 89, which then is well below the critical threshold of 250 events defined as
108 a quality criterion in GW2019. Therefore, DC2020 select events in a circle around the
109 M6.4 epicenter (the alternative method used by GW2019 for inferior datasets), instead
110 of along the actual fault plane.

111

112 **Deviation 1.3:** To establish the reference b-value, DC2020 sample events in circular
113 region of about 10 km around the epicenter, while Gulia et al. (2020) use events in a box
114 within 3 km of the rupture plane.

115

116 The combined impact of these three deviations is illustrated in Figure 1. Figure 1 A-B
117 shows the fault plane projection of the M6.4 event (black grid), superimposed is the
118 catalog used by DC2020 to establish the background b-value (red dots). It is composed
119 of the 250 events nearest the mainshock since 2000, events up to about 10 km from the
120 epicenter. Shown in comparison is the dataset used by Gulia et al. (2020, blue dots),
121 composed of events with a maximum distance of 3 km from the fault plane. Note that
122 DC2020 also used shallow events that are more than 3 km from the fault plane and thus
123 not included in the GW2019 approach. As a consequence of these differences, the
124 background b-value in DC2020 is 0.90, based on about 7 events above completeness per

125 year and averaged over 19 years. Using the GW2019 approach, we compute $b=0.97$,
126 based on about 22 events per year, averaged over 39 years.

127

128 **2. Correctly selecting the mainshock fault plane and events between the** 129 **mainshock**

130

131 Among the two nodal planes defined by the focal mechanism, GW2019 proposed to use
132 the one with the highest number of immediate aftershocks within 3 km of the fault,
133 since the method needs to run fully automatically and in near real-time. For the 31
134 sequences analyzed in Gulia et al. (2018) as well as for the three sequences analyzed in
135 GW2019, we determined the mainshock plane based on the first 24 hours of aftershock
136 data. This is a commonly used time interval sufficiently long to allow for stable
137 detection of the active fault in most cases (see also Kanamori, 1977): however, it is true
138 that we did not document this choice in GW2019 explicitly. DC2020 decided to use a
139 much shorter time interval of only one hour to establish the mainshock fault, resulting
140 as explained below in the choice of the alternative fault plane.

141

142 The initial M6.4 Ridgecrest mainshock was a complex rupture and it took several days
143 before geodetic, seismic, and relocated seismicity data provided a reliable view of this
144 complex sequence. Ross et al. (2019) identified three simultaneous subevents and
145 hypothesized that the rupture had been a cascading phenomenon. The purely statistical
146 method used in GW2019 based on the first 24 hours of aftershocks selects the
147 northwest-trending fault plane that represented the initial rupture (Figure 1D, blue
148 symbols). DC2020, on the contrary, select the orthogonal plane (Figure 1D, red dots).
149 Given the complex rupture pattern, both choices are actually defensible. Note that

150 deviation 1.1 and 1.2 apply on top for this part of the analysis. In addition, DC2020 did
151 not limit the depth of selected events.

152

153 **Deviation 2.1:** DC2020 selected aftershock of the first hour, rather than the first 24
154 hours to define the active fault. They thus selected the alternative fault plane for
155 estimating the b-values of the aftershocks following the first mainshock.

156

157 **Deviation 2.2:** DC2020 do not limit the analysis to events with 3 km depth below and
158 above the fault plane, but extend the sampling down to 20 km.

159

160 As a consequence of these deviations, DC2020 compute on the alternative nodal plane a
161 b-value for all the in-between events of $b=0.83$, while Gulia et al. (2020) compute
162 $b=0.74$, based on a much larger data sets due to the lower M_c (Figure 1E). Note that
163 despite these 5 deviations, the overall result of the FTLS assessment given by DC2020
164 remains unchanged: a red FTLS setting.

165

166

167 **3. Correctly selecting the second mainshock fault plane, the new reference**
168 **background and aftershocks**

169

170 According to the FTLS model, once a second and larger mainshock occurs as part of a
171 sequence, the FTLS assessment process restarts: first, the new fault plane is determined
172 based on the seismicity within 24 hours of this mainshock. Next, the background b-
173 value is redetermined based on events within 3 km of this – longer – fault plane and
174 then compared to the b-values of the aftershocks near the new fault to estimate the new

175 FTLS status. This is typically the most data-rich part of the analysis, since it involves
176 larger fault areas and numerous aftershocks. Here, DC2020 also apply the deviations
177 D1.1 (start date 2000), D1.2 (Mc double counted for the background) and D1.3 (circular
178 sampling instead of along the fault plane), but the resulting impact is much bigger since
179 the M7.1 fault is considerably longer.

180

181 **Deviation 3.1:** To establish the reference b-value for the M7.1 fault, DC2020 sample
182 events in circular region of about 3 km around the epicenter, while Gulia et al. (2020)
183 use events in a box within 3 km of the about 60 km long rupture plane.

184

185 As shown in Figure 2A, DC2020 select events that only cover a small subset of the fault,
186 about 10%; added to this is the higher Mc and shorter duration catalog duration. The
187 background b-value estimation of DC2020 thus is based only on about 1% of the data
188 used by the GW2019 approach (Figure 3C). As a consequence, the background b-value
189 for the second event established by DC2020 is not unexpectedly very different from the
190 one the GW2019 approach will compute (Figure 2C): DC2020 estimate $b = 1.10$, Gulia et
191 al. (2020) estimate $b = 0.87$. Note also that the frequency magnitude distribution of
192 DC2020, being based on a small data set, shows a substantial break in slope around
193 magnitude 3 (Figure 2C, red symbols). This difference in background b-value then will
194 results in very different changes in percent when compared to the aftershock b-values,
195 and ultimately results in the difference in the FTLS setting observed between DC2020
196 and Gulia et al. (2020).

197

198 For the computation of the b-values of the aftershocks, DC2020 then correctly use
199 events within 3 km of the mainshock fault (Figure 2D, E), although deviations 1.2 and

200 2.2 still apply. However, while the absolute aftershock b-values are quite similar
201 between the two papers, the all-important changes in percent normalized to the
202 background b-values are very different (-10% for DC2020 → orange alert; +26% for
203 Gulia et al. (2020) → green alert), largely because of the different background b-values
204 that they are normalized to ($b = 1.10$ versus $b = 0.87$).

205

206 Because there are at least six substantial deviations from the GW2019 approach, it is no
207 surprise that Gulia et al. (2020) report quite different results from DC2020 for the
208 Ridgecrest sequence. We will discuss the appropriateness of these deviations and the
209 meaningfulness of the comparison given these deviations in the discussions and
210 conclusion section.

211

212

213 **Puerto Rico Case study**

214

215 The second case study discussed by DC2020 is the January 7, 2020, Puerto Rico event:
216 DC2020 reported a red alert after the mainshocks, indicating an upcoming larger event,
217 which has not yet occurred at 24 February 2020 (and not until December 19 2020),
218 thus suggesting a false positive for the FTLS evaluation. As DC2020 themselves state,
219 this case is not an actual test of the GW2019 hypothesis:

220

221 *“For the source region surrounding this event used for computing a b-value, we relax the nominal spatial*
222 *window of 3 km from the source to 10 km to determine stable b-values. For this reason, the time series pro-*
223 *duced for the M_w 5.0 foreshock is not a strict test of the method proposed by [Gulia and Wiemer \(2019\)](#) but is*
224 *nonetheless inter- interesting to consider.”*

225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249

We would add to this statement that:

- 1) in GW2019, we explicitly exclude from the test offshore sequences, because hypocenter accuracy but also completeness are inevitably much inferior. In our assessment the quality of offshore catalogs is typically too low to allow to select enough earthquakes near the rupture plane and with sufficient confidence.
- 2) DC2020 performed a time-series on an M5, a much lower magnitude compared to the minimum one (M6) required for the model of GW2019. Because stress changes scale with magnitude, we have argued in Gulia et al (2018) that in order to apply the method to smaller magnitudes, only events close by should be considered, for example within 1 km of an M5.

Even though DC2020 in the Puerto Rico study did not test the GW2019 hypothesis in the first place, we also like to point out that their analysis in our opinion flawed, or biased. Data quality issues related to the homogeneity of the estimate magnitudes across the magnitude scale were not considered, leading to arbitrary estimates of b-values, as explained below.

In a first step, we evaluated the FTLS method on the M6.4 mainshock, using the original published and unchanged method and selection criteria by GW2019 and the same catalog of the Puerto Rico National Seismic Network used also by DC2020 (although we could not check if it had been updated in between downloads). We select events within a 3-km distance from the fault plane of the M6.4 event, applying a preliminary Magnitude cut-off at a minimum level of completeness (here $2.3+0.2$ correction factor). The results obtained without any modifications in the released code are shown in

250 Figure 3. In our analysis of the M6.4 event, the b-value increases by 30% after the
251 mainshock, resulting in a green alert. The Puerto Rico sequence would thus represent
252 an additional and further positive test of the GW2019 hypothesis; however, as
253 explained below, the quality check applied in GW2019 estimates the FMDs not reliable
254 enough to consider this a successful case study.

255

256 The challenge with magnitude-scale reporting homogeneity of the Puerto Rico catalog is
257 illustrated in Figure 4, where we show the overall b-value of earthquakes within about
258 50 km from the island of Puerto Rico for the period 2003-2019, plotted as a function of
259 cut-off magnitude (red curve). This kind of plot is a simple check for both M_c and the
260 homogeneity of reporting (Woessner and Wiemer, 2005; Wiemer and Wyss, 2000). The
261 expected behavior is that the b-value is strongly underestimated as long as the catalog
262 is incomplete, and, once M_c is approached, the b-value levels off and a plateau emerges.
263 The plots for the Puerto Rico catalog reveal no such plateau (the ones for Ridgecrest, for
264 example, do). Instead, it signals a very high sensitivity of b-values to the choice of M_c ,
265 with b-values ranging from below 1.0 to 1.6, depending on the choice of M_c . Similar
266 behavior is found for the M6.4 mainshock region analyzing the 2020 data only (blue line
267 in Figure 4). Such a peak rather than a plateau is indicative of an upwards-bend of the
268 frequency magnitude distribution, typical for example if different procedures are used
269 to estimate magnitudes in different magnitude bins.

270

271 The impact of this magnitude-scale compression on the frequency-magnitude
272 distribution near the mainshocks is shown in Figure 4B. We selected events within 3 km
273 distance to of the M6.4 mainshock fault. The resulting FMD does not only look non-
274 linear to the eye, but it also does not pass the non-linearity filter (Tormann et al., 2014)

275 that we apply as a quality check in GW2019 to ensure compliance with a linear power-
276 law model. The substantial 'kink' in the distribution around magnitude 3.0 - 3.5 leads to
277 the aforementioned strong sensitivity of the background b-value on the choice of M_c .
278 Figure 4 implies that a stable b-value analysis may only be possible from about
279 magnitude 4.0, but then almost no data would be left for analysis.

280

281 The main difference between our analysis and the one by DC2020 lies in the b-values of
282 the aftershock sequence and it is ultimately related to the aforementioned data quality
283 issue. For the background, DC2020 compute a rather high b-value ($b = 1.2$) when
284 compared to our analysis ($b = 0.87$). This is a consequence of different sampling
285 volumes (large circles versus fault plane) but also a result of the 'upwards' bend of the
286 FMD for events below magnitude 3. DC2020 use a much lower M_c here (about 2.0), we
287 would use $M_c = 2.5$. DC2020 then compute an aftershock b-value of about 0.5 - 0.6
288 (their Figure 3). Our analysis, shown in Figures 3, results in a $b = 1.1$. We cannot fully
289 explain how DC2020 obtain such an unusual low b-value, and we note that their FMD
290 does not fit the data for most of the range - too low for small magnitudes, too high for
291 larger ones (Figure 4b).

292

293 We recognize that the dependence on the two free parameters of our analysis, the no-
294 alert time and the magnitude of completeness, is potentially creating an arbitrariness in
295 the analysis. To address this limitation, we introduced in Gulia et al. (2020) a systematic
296 scan of the free parameter space, to assess the robustness of the analysis. We repeated
297 the analysis for the Puerto Rico case. If the M_c of the aftershocks is below completeness,
298 then b-values are much too low, and an erroneous red alert is found. Once M_c is high

299 enough, and for all possible constellations of Mc and no-alert time, a green alert after
300 the M6.4 results.

301

302 **Discussions and Conclusions**

303 Testing earthquake forecasts in rigorous ways is highly important and the past 40 years
304 of research have seen a rather spotty record of the seismology community on testing
305 (Jordan 2006; Jackson, 1996; Kagan, 1999; Zechar et al., 2016). One of the challenges is
306 that often the models are a moving target. There is a broad consensus in the community
307 (e.g., Strader et al., 2017, Jordan, 2006; Marzocchi et al., 2015; Schorlemmer et al., 2018;
308 Zechar et al., 2011) that prospective and pseudo-prospective testing in earthquake
309 sciences (no different from medicine or other sciences) must follow strict rules, and
310 community efforts such as CSEP have been created for this purpose (e.g., Gerstenberger
311 and Rhoades, 2010; Werner et al., 2010; Zechar et al., 2010; Zechar et al., 2013; Tsuruoka et
312 al., 2012).

313

314 One of the most fundamental rules for evaluating hypotheses in science is that that the
315 hypothesis to be tested cannot be changed arbitrarily, otherwise, biases (in favor or
316 against a hypothesis) are likely to influence the test and endless discussion may occur.
317 Another basic rule of science is the fact that quality limitation of the data must be
318 accepted and respected, even if we do not like them. Otherwise, the garbage in –
319 garbage out criteria will almost inevitably apply.

320

321 The study by DC2020 has violated these two basic rules of hypothesis testing in several
322 respects, biasing their analysis against the GW2019 hypothesis. We demonstrate here
323 and in Gulia et al. (2020) that, when applied correctly, the Ridgecrest cases study would

324 be fully in line with the FTLS hypothesis. DC2020 has deviated in at least six steps from
325 the analysis; these are in parts major deviations, changing by 99% the data to be
326 analyzed. As much as we appreciate that DC2020 evaluated our hypothesis, in our
327 opinion this test is meaningless, or actually misleading because the method and data
328 processing of DC2020 are substantially different. DC2020, therefore, test their own
329 hypothesis, not ours, but they do not state so in their paper. We consider this confusing
330 for the community and we even had pointed out some of these shortcomings to the
331 authors in a review before publication.

332

333 One might argue that a method or hypothesis should be robust enough to work also
334 with somewhat modified parameters, as a measure of robustness, a point raised by
335 DC2020. We respond first of all that even before such a useful sensitivity analysis, one
336 obviously needs to test the actual unmodified hypothesis and also document the
337 changes transparently. However, much more important is in our view that the
338 deviations applied by DC2020 are unjustified in several ways:

- 339 • **The deviations violate the *physical framework*** of GW2019: We consider it
340 critically important and physically plausible to sample events in the immediate
341 vicinity of the actual fault plane, because stress changes due to the mainshock
342 are strongest here.
- 343 • **The deviations violate the *statical framework*** of GW2019 that aims to
344 maximize the amount of data and hence robustness of the analysis. Instead,
345 DC2020 use only a small fraction of the data available for no apparent reason
346 (Figures 1 and 2).

- 347 • **The deviations violate the *principle of reproducibility*** since they are not
348 documented and possibly not intended modifications (e.g., all depth selected, Mc
349 add-on double-counted).

350

351 The Puerto Rico case is more complex to interpret. Both groups agree that this case
352 study does not represent a test of the GW2019 hypothesis in the first place. But in
353 addition, also here DC2020 introduced inconsistencies in the analysis, in parts possibly
354 due to the same deviations stated here for Ridgecrest, but even more so by ignoring the
355 limitations of the data as well as the minimum required magnitude (M6) to implement
356 GW2019. Re-doing the analysis using the original GW2019 approach with no
357 modifications, we find also that this case would support the FTLS hypothesis (Figure 3).
358 Nevertheless, we argue that the offshore data quality is too poor, and the magnitude
359 scale shows unexplained bends (Figure 4) to allow for robust analysis. The automatic
360 procedures for quality control in GW2019 would reject this case also.

361

362 Every forecast model has several free parameters. Some are obvious, first order free
363 parameter, such as the sample sizes used or the width of the volume sampled, and these
364 can be readily analyzed in a sensitivity analysis. Some are related to the automated
365 quality analysis, such as the determination of Mc, and here the uncertainty in Mc
366 determination can be used to estimate sensitivity. A third set of 'free' parameters are
367 resulting from expert choices, based for example on data quality, such as the start time
368 of the catalog or the fault plane used. In Gulia et al. (2020), we explore some of the free
369 parameter space, confirming the robustness of the FTLS model to first-order free
370 parameters, but a complete search of the free parameter space is difficult. It would
371 require a logic tree approach such as the ones used in probabilistic seismic hazard

372 assessment, capturing aleatory and epistemic uncertainties. In forecast, the preferred
373 method instead is to perform fully prospective test of models under controlled
374 conditions and against predefined, authoritative data sources (e.g. Schorlemmer et al.,
375 2018). The GW2019 hypothesis may well fail such a test, but it deserves to be tested
376 fairly. DC2020 did in our assessment - unfortunately - not conduct such a fair test of the
377 actual hypothesis, nor did it perform a systematic sensitivity test,

378

379 **Final comment after reading the Reply by the authors**

380 We carefully read the reply by Dascher-Cousineau et al. to our comment and thank the
381 authors for the detailed discussion as well as the clarifications and corrections applied
382 to their analysis. We still believe that all deviations we listed in our comment are
383 correctly identified and justified. The aim of GW2020 was to implement the published
384 FTLS *without any modifications* and this is what we did.

385

386 As stated before in our comment, we welcome the independent evaluation of the FTLS
387 by DC2020 and welcome also their response to our criticisms raised. Details matter in
388 science, and we are struck again how difficult it is in earthquake forecasting to not only
389 ensure full reproducibility but to write down a 'recipe' that other qualified scientist can
390 apply to new cases and reach the same conclusions. Cooking is a good analogue: Even a
391 detailed recipe will not ensure the same outcome. For evaluating earthquake
392 forecasting related hypothesis our experience documented in the paper and replies also
393 highlights the need for a collaborative and fully prospective testing environment such as
394 the one provided by CSEP, with community-agreed rules and decoupling between
395 modellers and evaluators.

396

397 **Data and Resources**

398 For Ridgecrest: events from the Advanced National Seismic System (ANSS)
399 Comprehensive Earthquake Catalog (ComCat) and Shelly (2020, SRL); for Puerto Rico:
400 events from the Puerto Rico Seismic Network.

401 Data about European Real-time earthquake rIsk reduction for a reSilient Europe project
402 are available at www.rise-eu.org. Both figures and calculations were performed by
403 MATLAB, available at www.mathworks.com/products/matlab.

404

405 **Declaration of Competing Interests**

406 The authors declare no competing interests

407

408 **Acknowledgments**

409 The authors thank the Editor in chief, Allison Bent, the Associated Editor, Anastasia
410 Pratt, and the anonymous reviewer. This study was supported by the Real-time
411 earthquake rIsk reduction for a reSilient Europe (RISE) project, funded by the European
412 Union's Horizon 2020 research and innovation program under Grant Agreement
413 Number 821115.

414

415

416 **References**

417

418 *Dascher-Cousineau, K., T. Lay, and E. E. Brodsky (2020). Two Foreshock Sequences Post*
419 *Gulia and Wiemer (2019), Seismol. Res. Lett. doi: 10.1785/0220200082*

420

421 *Gardner, J. K. and Knopoff, L. (1974). Is the sequence of earthquake in Southern California,*
422 *with aftershocks removed Poissonian? Bull. Seism. Soc. Am., 66, 1271–1302.*
423

424 *Gerstenberger, M. C., and D. A. Rhoades (2010), New Zealand Earthquake Forecast Testing*
425 *Centre, Pageoph, 167(8-9), 877-892, doi:10.1007/s00024-010-0082-4.*
426

427 *Gulia, L., S. Wiemer, and G. Vannucci (2020). Pseudoprospective Evaluation of the*
428 *Foreshock Traffic-Light System in Ridgecrest and Implications for Aftershock Hazard*
429 *Assessment, Seismol. Res. Lett., doi: 10.1785/0220190307*
430

431 *Gulia, L., and S. Wiemer (2019). Real-time discrimination of earthquake foreshocks and*
432 *aftershocks, Nature, 574, 193–199.*
433

434 *Gulia, L., A. P. Rinaldi, T. Tormann, G. Vannucci, B. Enescu, and S. Wiemer (2018). The*
435 *effect of a mainshock on the size distribution of the aftershocks, Geophys. Res. Lett. 45, no.*
436 *B1, doi: 10.1029/2018GL080619.*
437

438 *Jackson, D. D. (1996), Hypothesis testing and earthquake prediction, Proceedings of the*
439 *National Academy of Science United States of America, 93, 3772-3775.*
440

441 *Jordan, T. H. (2006), Earthquake predictability, brick by brick, Seism. Res. Lett., 77(1), 3-6.*
442

443 *Kagan, Y. Y. (1999), Is earthquake seismology a hard, quantitative science?, Pageoph,*
444 *155(2-4), 233-258.*
445

446 Marzocchi, W., T. H. Jordan, and G. Woo (2015), *Operational earthquake forecasting and*
447 *decision making*, *Ann. of Geophys.*, 58(4), doi:ARTN RW0434
448 10.4401/ag-6756.

449

450 Schorlemmer, D., M. J. Werner, W. Marzocchi, T. H. Jordan, Y. Ogata, D. D. Jackson, S. Mak,
451 D. A. Rhoades, M. C. Gerstenberger, N. Hirata, M. Liukis, P. J. Maechling, A. Strader, M.
452 Taroni, S. Wiemer, J. D. Zechar, and J. C. Zhuang (2018), *The Collaboratory for the Study of*
453 *Earthquake Predictability: Achievements and Priorities*, *Seism. Res. Lett.*, 89(4), 1305-
454 1313, doi:10.1785/0220180053.

455

456 Strader, A., Schneider M., Schorlemmer D. (2017). *Prospective and retrospective evaluation*
457 *of five-year earthquake forecast models for California*, *Geophys. J. Int.*, 211 (1), 239-
458 251, <https://doi.org/10.1093/gji/ggx268>

459

460 Tormann, T., Wiemer, S., & Mignan, A. (2014). *Systematic survey of high-resolution b value*
461 *imaging along Californian faults: Inference on asperities*. *Journal of Geophysical Research:*
462 *Solid Earth*, 119, 2029–2054. <https://doi.org/10.1002/2013JB010867>

463

464 Tsuruoka, H., Hirata, N., Schorlemmer, D., Euchner F., Nanjo K. Z. and Jordan T. H.
465 (2012), *CSEP Testing Center and the first results of the earthquake forecast testing*
466 *experiment in Japan*. *Earth Planet Sp* 64, 661–671 (2012).
467 <https://doi.org/10.5047/eps.2012.06.007>

468

469 Werner, M. J., J. D. Zechar, W. Marzocchi, S. Wiemer, and C. S.-I. W. Grp (2010),
470 Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts,
471 *Annals of Geophysics*, 53(3), 11-30, doi:10.4401/ag-4840

472

473 Woessner, J., and S. Wiemer (2005). Assessing the quality of earthquake catalogues:
474 Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* 95,
475 684–698.

476

477 Zechar, J. D., J. L. Hardebeck, A. J. Michael, M. Naylor, S. Steacy, S. Wiemer, J. C. Zhuang, and
478 C. W. Grp (2011), *Community Online Resource for Statistical Seismicity Analysis*, *Seism.*
479 *Res. Lett.*, 82(5), 686-690, doi:10.1785/gssrl.82.5.686.

480

481 Zechar, J. D., W. Marzocchi, and S. Wiemer (2016), *Operational earthquake forecasting in*
482 *Europe: progress, despite challenges*, *Bulletin of Earthquake Engineering*, 14(9), 2459-
483 2469, doi:10.1007/s10518-016-9930-7.

484

485 Zechar, J. D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, and T. H. Jordan
486 (2010), *The Collaboratory for the Study of Earthquake Predictability perspective on*
487 *computational earthquake science*, *Concurr Comp-Pract E*, 22(12), 1836-1847,
488 doi:10.1002/cpe.1519.

489

490 Zechar, J. D., D. Schorlemmer, M. J. Werner, M. C. Gerstenberger, D. A. Rhoades, and T. H.
491 Jordan (2013), *Regional Earthquake Likelihood Models I: First-Order Results*, *Bull. Seismol.*
492 *Soc. Am.*, 103(2a), 787-798, doi:10.1785/0120120186.

493

494

495 **Authors'adress**

496 Laura Gulia, University of Bologna, Department of Physics and Astronomy, Viale Berti

497 Pichat, 8 - 40127 Bologna (Italy)

498

499 Stefan Wiemer, Swiss Seismological Service, ETH, NO H61, Sonneggstrasse 5, CH-8092

500 Zurich

501

502

503 **Figure captions**

504

505 **Figure 1 A-F.** *A-B) Map of Ridgecrest region, shown are the selected mainshock plane of*

506 *the M6.4 mainshock on 4 July 2019 (black grid) and the selected background seismicity by*

507 *DC2020 (red dots) and by GW2020 (blue dots). C) Annualized frequency-magnitude*

508 *distribution for the two datasets show in A-B. D-E) Map of Ridgecrest region, shown are*

509 *the selected mainshock plane of the M6.4 mainshock on 4 July 2019 (black grid) and the*

510 *selected 'in-between' events by DC2020 (red dots) and by GW2020 (blue dots). F)*

511 *Annualized frequency-magnitude distribution for the two datasets show in D-E.*

512

513 **Figure 2 A-D.** *A-B) Seismicity maps showing the fault plane (in black) and the events*

514 *preceding the M7.1 event on 6 July 2019 selected by DC2020 (red dots) and by GW2020*

515 *(blue dots). C) the relative frequency-magnitude distribution for the two datasets in A-B.*

516 *D-E) Seismicity maps showing the fault plane (in black) and the events following the M7.1*

517 *event 2019 selected by DC2020 (red dots) and by GW2020 (blue dots).*

518

519 **Figure 3 A-B:** left Performance of the foreshock traffic-light system (FTLS) for the M6.4
520 event in Puerto Rico. A) Frequency-magnitude distributions (FMDs) for the source of the
521 Mw 6.4 event for two time periods: background in blue and maximum b-value reached in
522 the first weeks of aftershocks. B) b-value time series for the M 6.4; blue dashed line is the
523 reference b-value; red dashed vertical line indicates the time of the M6.4 event. All the
524 estimates are above the reference value.

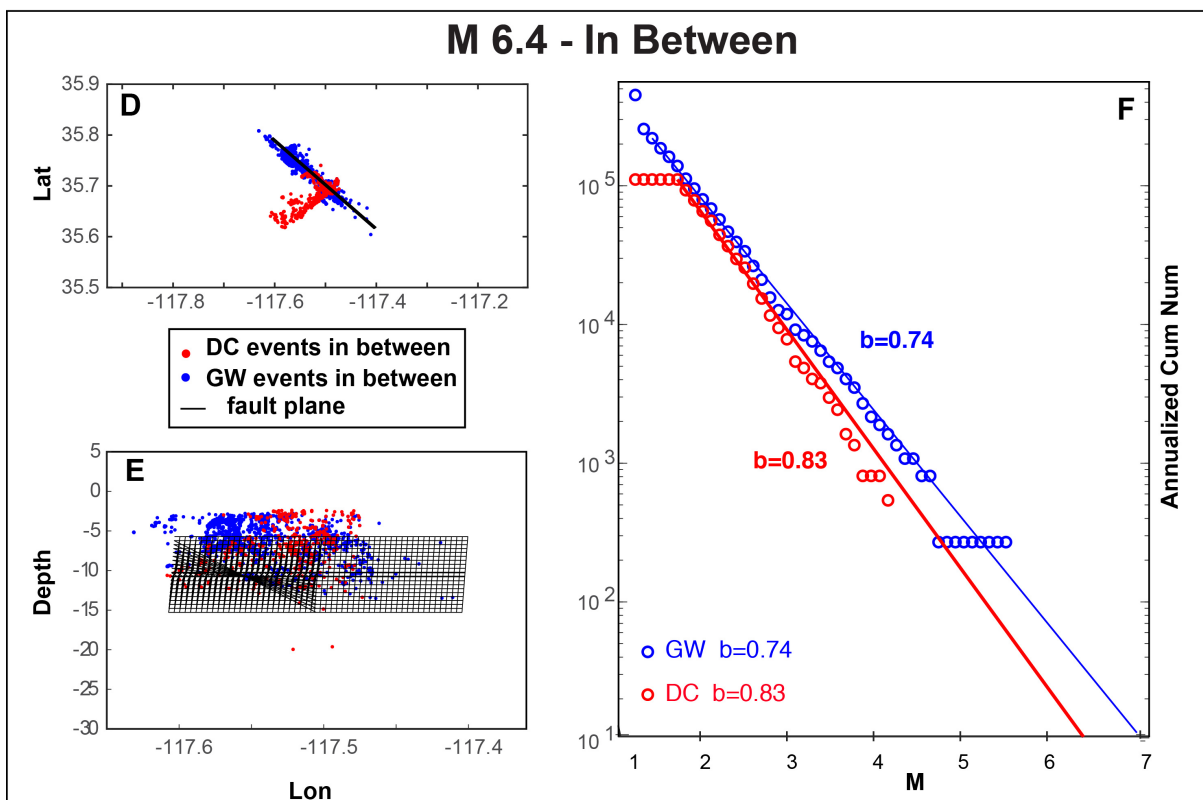
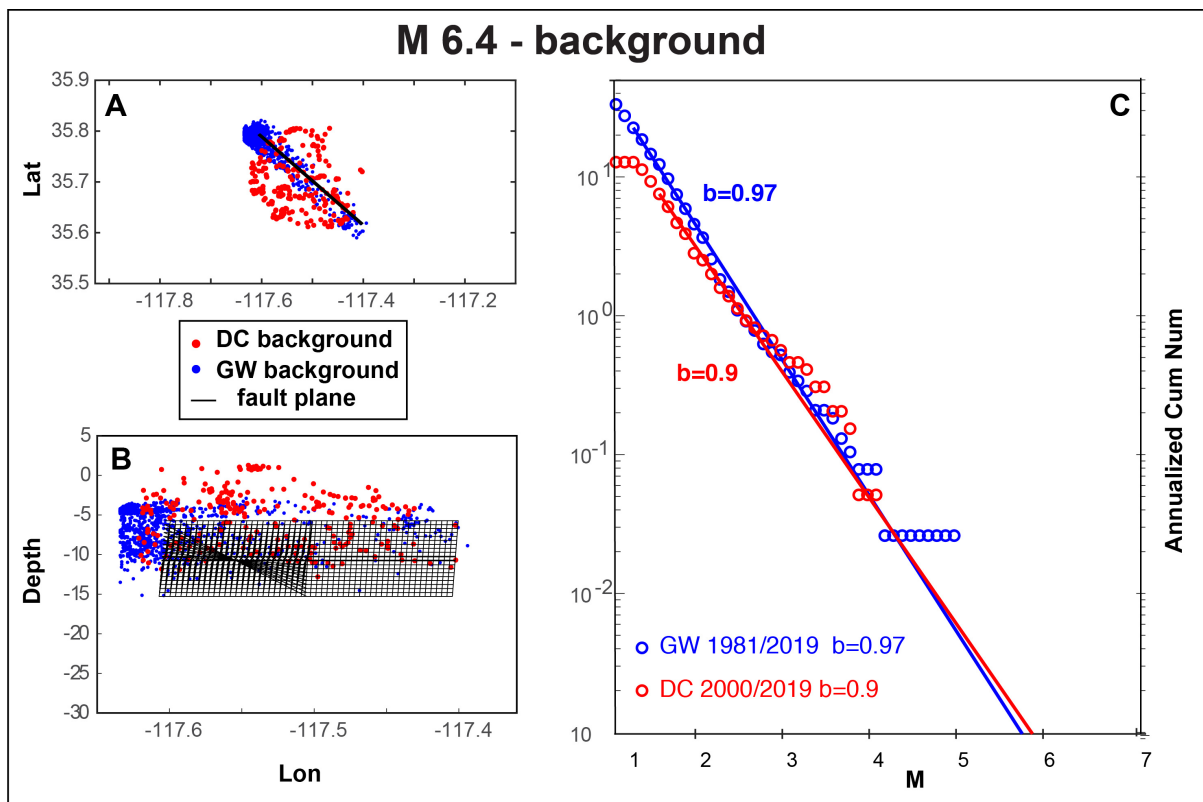
525

526 **Figure 4 A-B –** A) b-value as a function of magnitude of completeness for the Puerto Rico
527 catalog, for the periods 2003-2019 (red) and 2020 (blue). B) Annualized Frequency
528 Magnitude Distribution of the background (blue circles) at two different magnitude of
529 completeness and relative b-values for the M6.4 event dataset.

530

531 Figure 1

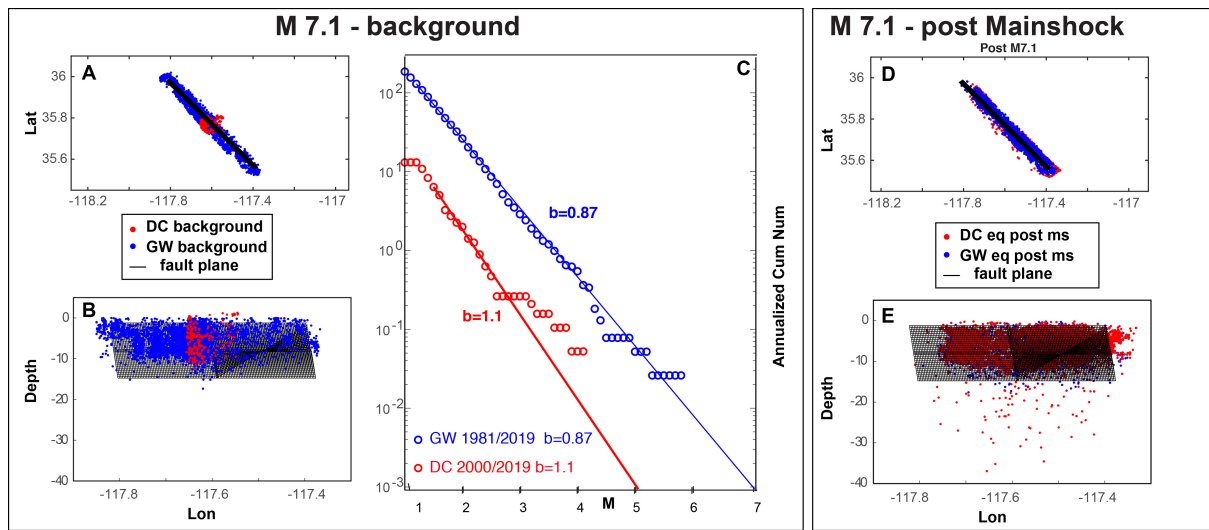
532



533

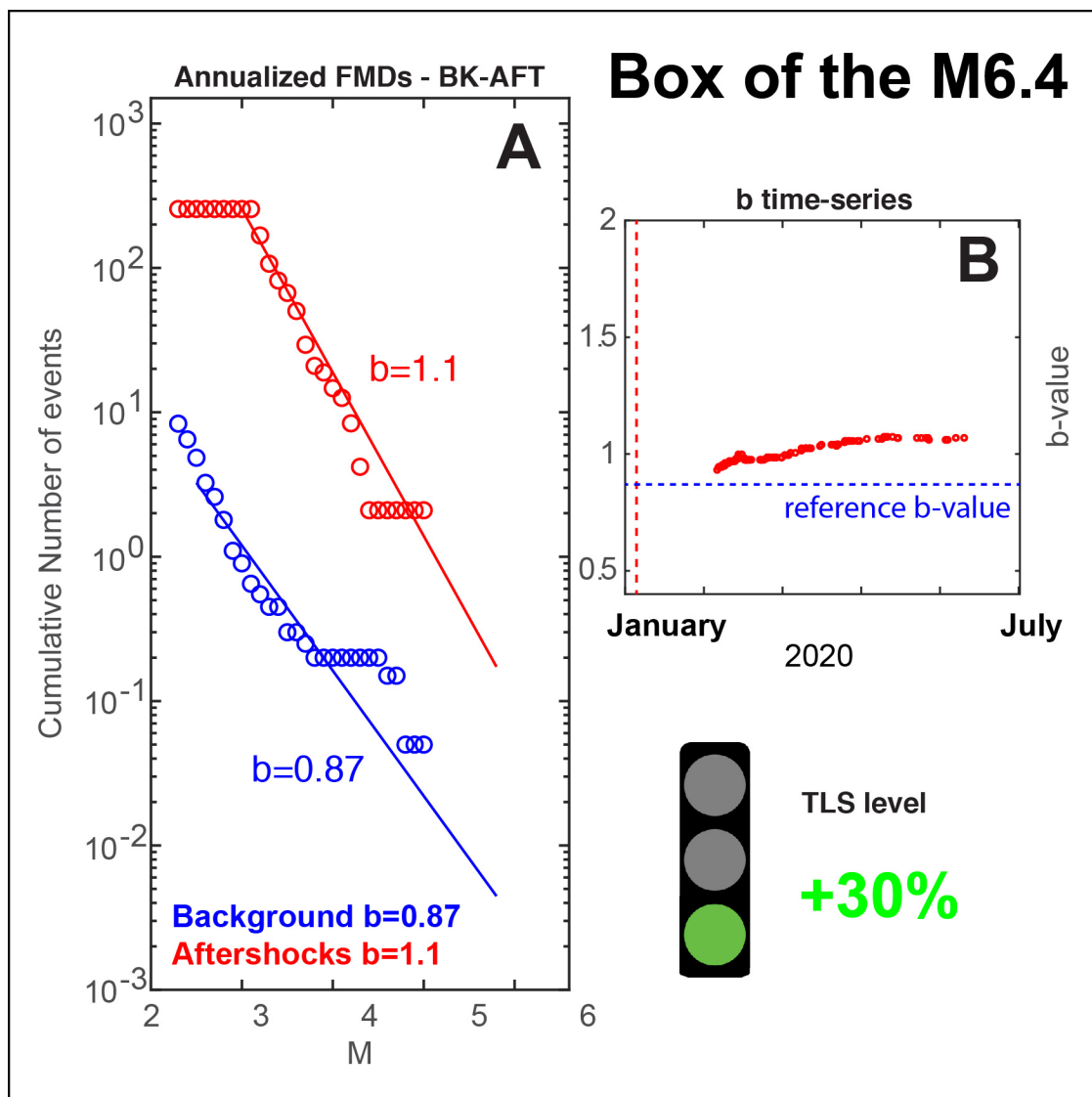
534

535 Figure 2



536

537



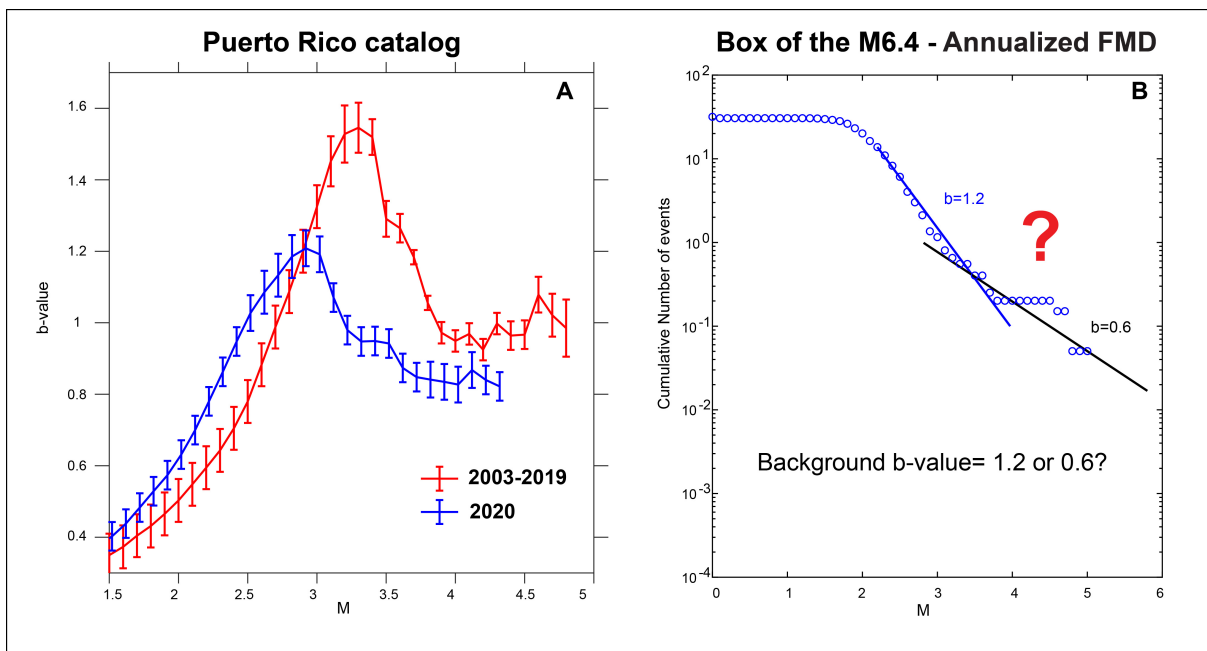
539

540

541

542 Figure 4

543



544

545