Comment on "Two foreshock sequences post Gulia and Wiemer (2019)"

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

Laura Gulia[1*] and Stefan Wiemer[2]

[1] University of Bologna, Department of Physics and Astronomy, Bologna.

[2] Swiss Seismological Service, ETH Zurich, Switzerland.

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

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

[*]corresponding author: Laura Gulia, laura.gulia@unibo.it

Abstract

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

a not a fair test of the FTLS hypothesis.

Introduction and context

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

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

Ridgecrest case study

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

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

A critically important parameter to be established in the FTLS model is the local reference b-value, because the FTLS decisions are based on the difference in percent between the sequence-specific b-values and the reference b-value. According to the GW2019 hypothesis, it is important to establish the reference b-value such that: 1) it is only based on earthquake immediately near the initiating mainshock fault (i.e., within 3 km of the fault), since b-values vary substantially with space; 2) uses a long time series, to have the statically most robust estimate that averages over temporal variations. For the background of Californian sequences in Gulia et al. (2018), we start our analysis from 1981, when the network was greatly improved (e.g., Tormann et al., 2014). Therefore, for our Ridgecrest analysis presented in Gulia et al. (2020), we use a 39-year-long background catalog. DC2020 started their analysis only in the year 2000, resulting in a factor of two reduction of the data used. This choice was made to avoid the influence of the aftershocks of Landers (M7.3 in 1992) and Hector Mine (M 7.1 in 1999) (*Kelian Dascher-Cousineau*, personal communication). However, these two sequences occurred, about 200 and 170 km from the Ridgecrest mainshock, respectively, distances well beyond the Gardner and Knopoff (1974) radii of influence for both magnitudes and neither of these mainshocks had a noticeable impact on earthquake rates in the Ridgecrest area. We thus consider 1981 the better justified starting date, but this choice is indeed a subjective one and not fully automate in the approach yet.

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

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

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

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

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

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

The combined impact of these three deviations is illustrated in Figure 1. Figure 1 A-B shows the fault plane projection of the M6.4 event (black grid), superimposed is the catalog used by DC2020 to establish the background b-value (red dots). It is composed of the 250 events nearest the mainshock since 2000, events up to about 10 km from the epicenter. Shown in comparison is the dataset used by Gulia et al. (2020, blue dots), composed of events with a maximum distance of 3 km form the fault plane. Note that DC2020 also used shallow events that are more than 3 km from the fault plane and thus not included in the GW2019 approach. As a consequence of these differences, the background b-value in DC2020 is 0.90, based on about 7 events above completeness per year and averaged over 19 years. Using the GW2019 approach, we compute b=0.97, based on about 22 events per year, averaged over 39 years.

2. Correctly selecting the mainshock fault plane and events between the mainshock

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

The initial M6.4 Ridgecrest mainshock was a complex rupture and it took several days before geodetic, seismic, and relocated seismicity data provided a reliable view of this complex sequence. Ross et al. (2019) identified three simultaneous subevents and hypothesized that the rupture had been a cascading phenomenon. The purely statistical method used in GW2019 based on the first 24 hours of aftershocks selects the northwest-trending fault plane that represented the initial rupture (Figure 1D, blue symbols). DC2020, on the contrary, select the orthogonal plane (Figure 1D, red dots). Given the complex rupture pattern, both choices are actually defendable. Note that deviation 1.1 and 1.2 apply on top for this part of the analysis. In addition, DC2020 did not limit the depth of selected events.

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

<u>Deviation 2.2:</u> DC2020 do not limit the analysis to events with 3 km depth below and above the fault plane, but extend the sampling down to 20 km.

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

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

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

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

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

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

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

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

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

Puerto Rico Case study

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

"*For the source region surrounding this event used for computing a b-value, we relax the nominal spatial window of 3 km from the source to 10 km to determine stable b-values. For this reason, the time series produced for the M_w 5.0 foreshock is not a strict test of the method proposed by Gulia and Wiemer (2019) but is nonetheless inter- interesting to consider.*"

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

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

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

The impact of this magnitude-scale compression on the frequency-magnitude distribution near the mainshocks is shown in Figure 4B. We selected events within 3 km distance to of the M6.4 mainshock fault. The resulting FMD does not only look non-linear to the eye, but it also does not pass the non-linearity filter (Tormann et al., 2014) that we apply as a quality check in GW2019 to ensure compliance with a linear power-law model. The substantial 'kink' in the distribution around magnitude 3.0 - 3.5 leads to the aforementioned strong sensitivity of the background b-value on the choice of Mc. Figure 4 implies that a stable b-value analysis may only be possible from about magnitude 4.0, but then almost no data would be left for analysis.

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

We recognize that the dependence on the two free parameters of our analysis, the no-alert time and the magnitude of completeness, is potentially creating an arbitrariness in the analysis. To address this limitation, we introduced in Gulia et al. (2020) a systematic scan of the free parameter space, to assess the robustness of the analysis. We repeated the analysis for the Puerto Rico case. If the Mc of the aftershocks is below completeness, then b-values are much too low, and an erroneous red alert is found. Once Mc is high enough, and for all possible constellations of Mc and no-alert time, a green alert after the M6.4 results.

Discussions and Conclusions

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

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

The study by DC2020 has violated these two basic rules of hypothesis testing in several respects, biasing their analysis against the GW2019 hypothesis. We demonstrate here and in Gulia et al. (2020) that, when applied correctly, the Ridgecrest cases study would be fully in line with the FTLS hypothesis. DC2020 has deviated in at least six steps from the analysis; these are in parts major deviations, changing by 99% the data to be analyzed. As much as we appreciate that DC2020 evaluated our hypothesis, in our opinion this test is meaningless, or actually misleading because the method and data processing of DC2020 are substantially different. DC2020, therefore, test their own hypothesis, not ours, but they do not state so in their paper. We consider this confusing for the community and we even had pointed out some of these shortcomings to the authors in a review before publication.

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

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

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

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

Every forecast model has several free parameters. Some are obvious, first order free parameter, such as the sample sizes used or the width of the volume sampled, and these can be readily analyzed in a sensitivity analysis. Some are related to the automated quality analysis, such as the determination of Mc, and here the uncertainty in Mc determination can be used to estimate sensitivity. A third set of 'free' parameters are resulting from expert choices, based for example on data quality, such as the start time of the catalog or the fault plane used. In Gulia et al. (2020), we explore some of the free parameter space, confirming the robustness of the FTLS model to first-order free parameters, but a complete search of the free parameter space is difficult. It would require a logic tree approach such as the ones used in probabilistic seismic hazard assessment, capturing aleatory and epistemic uncertainties. In forecast, the preferred method instead is to perform fully prospective test of models under controlled conditions and against predefined, authoritative data sources (e.g. Schorlemmer et al., 2018). The GW2019 hypothesis may well fail such a test, but it deserves to be tested fairly. DC2020 did in our assessment - unfortunately - not conduct such a fair test of the actual hypothesis, nor did it perform a systematic sensitivity test,

Final comment after reading the Reply by the authors

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

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

Data and Resources

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

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

Declaration of Competing Interests

The authors declare no competing interests

Acknowledgments

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Authors'adress

Laura Gulia, University of Bologna, Department of Physics and Astronomy, Viale Berti Pichat, 8 - 40127 Bologna (Italy)

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

Figure captions

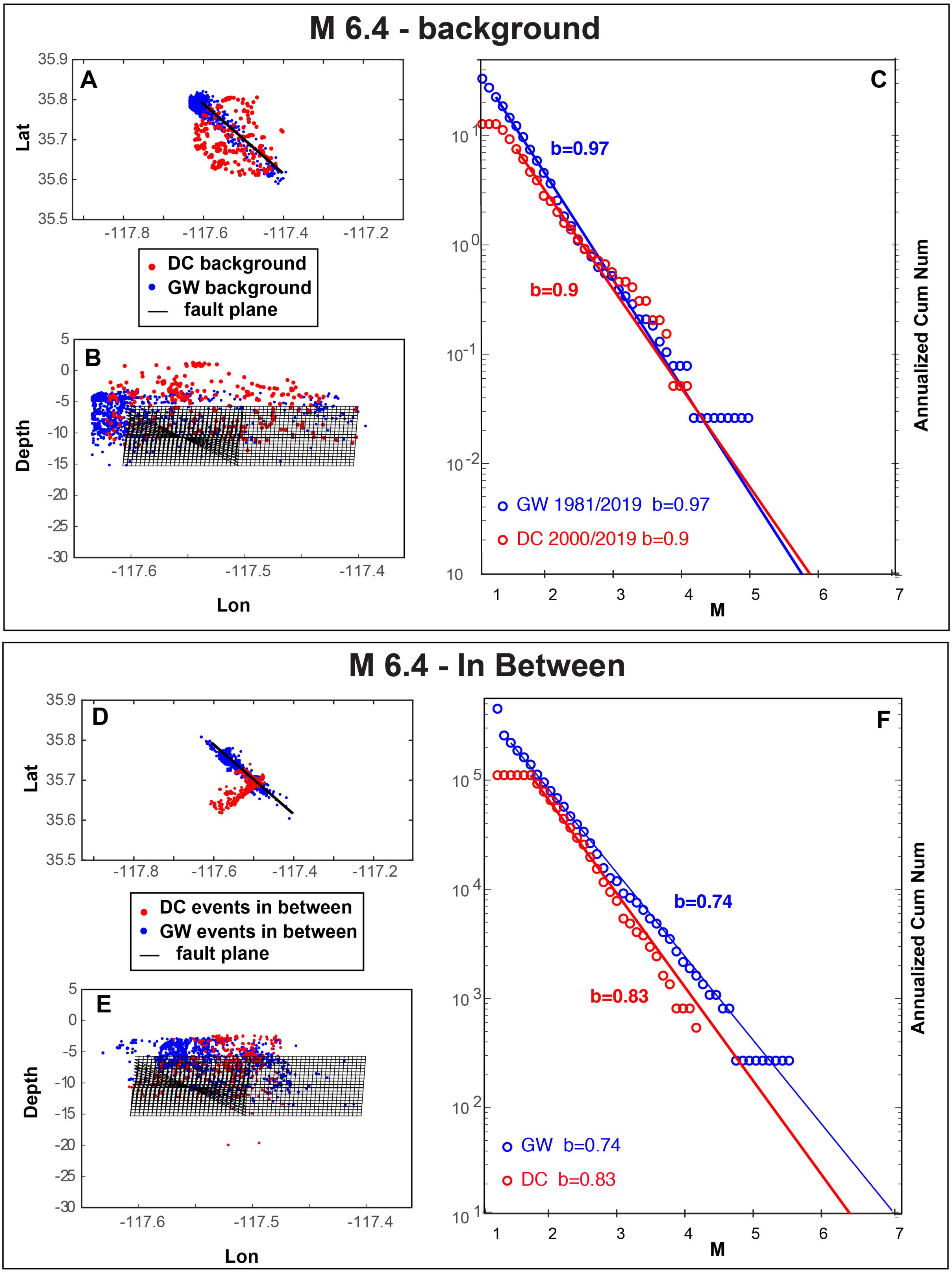
Figure 1 A-F. *A-B) Map of Ridgecrest region, shown are the selected mainshock plane of the M6.4 mainshock on 4 July 2019 (black grid) and the selected background seismicity by DC2020 (red dots) and by GW2020 (blue dots). C) Annualized frequency-magnitude distribution for the two datasets show in A-B. D-E) Map of Ridgecrest region, shown are the selected mainshock plane of the M6.4 mainshock on 4 July 2019 (black grid) and the selected 'in-between' events by DC2020 (red dots) and by GW2020 (blue dots). F) Annualized frequency-magnitude distribution for the two datasets show in D-E.*

Figure 2 A-D. *A-B) Seismicity maps showing the fault plane (in black) and the events preceding the M7.1 event on 6 July 2019 selected by DC2020 (red dots) and by GW2020 (blue dots). C) the relative frequency-magnitude distribution for the two datasets in A-B. D-E) Seismicity maps showing the fault plane (in black) and the events following the M7.1 event 2019 selected by DC2020 (red dots) and by GW2020 (blue dots).*

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

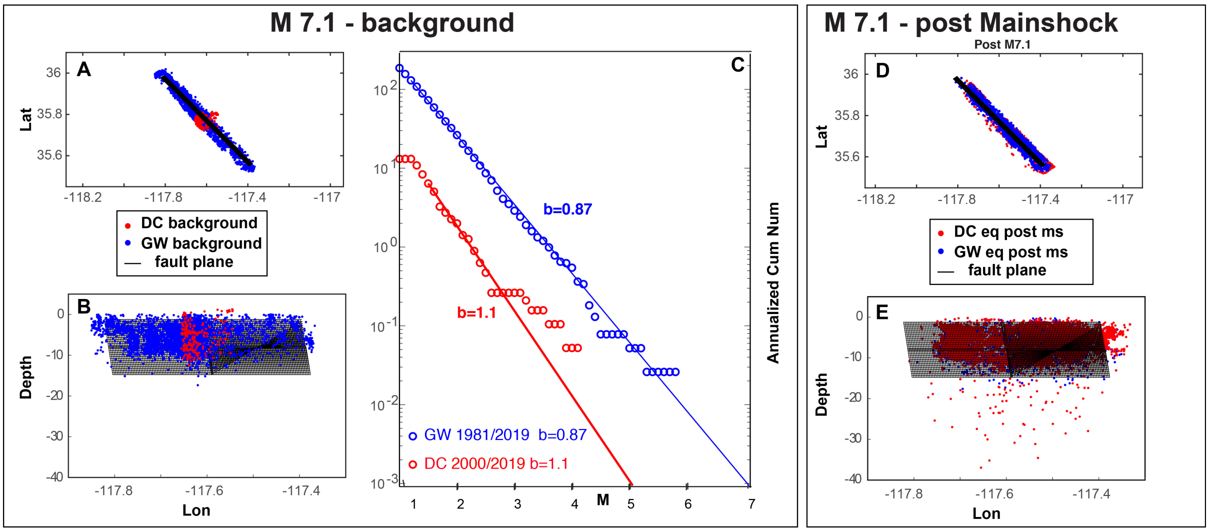
Figure 4 A-B *– A) b-value as a function of magnitude of completeness for the Puerto Rico catalog, for the periods 2003-2019 (red) and 2020 (blue). B) Annualized Frequency Magnitude Distribution of the background (blue circles) at two different magnitude of completeness and relative b-values for the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

Figure 1



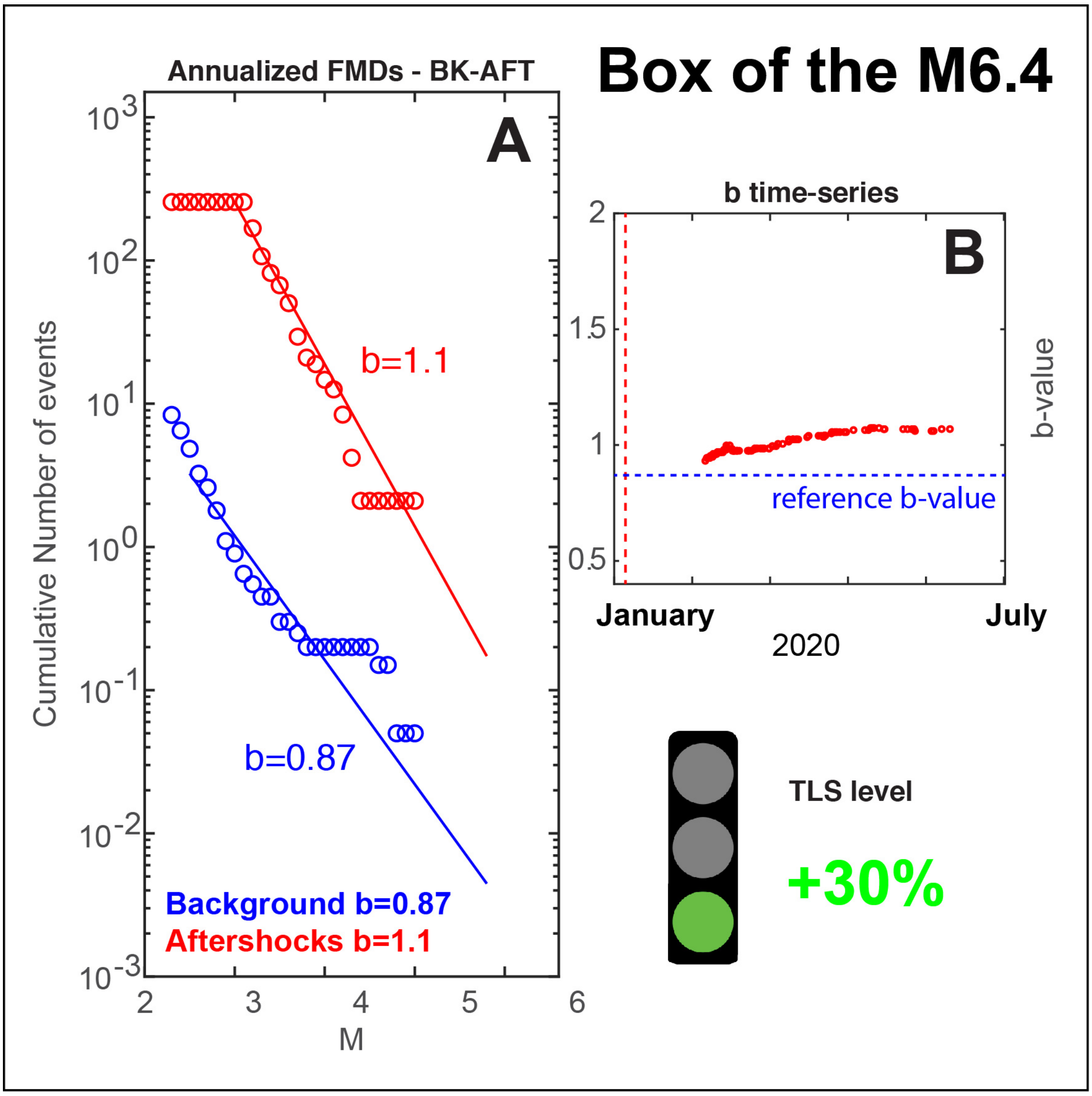
M 6.4 - background
A
Lat
Lon
DC background
GW background
fault plane
B
Depth
C
b=0.97
b=0.9
GW 1981/2019 b=0.97
DC 2000/2019 b=0.9
Annualized Cum Num
M
M 6.4 - In Between
D
DC events in between
GW events in between
fault plane
E
F
b=0.74
b=0.83
GW b=0.74
DC b=0.83

Figure 2



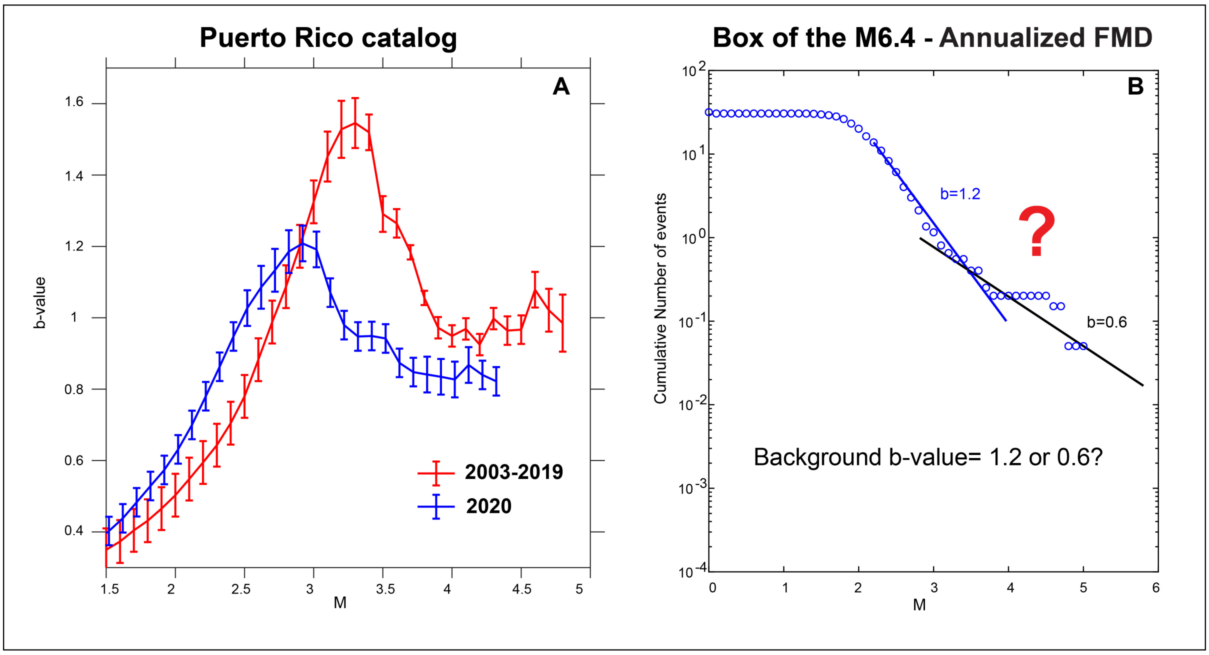
M 7.1 - background
A
B
C
Lat
Depth
Lon
DC background
GW background
fault plane
b=0.87
b=1.1
GW 1981/2019 b=0.87
DC 2000/2019 b=1.1
M
Annualized Cum Num
M 7.1 - post Mainshock
Post M7.1
D
E
DC eq post ms
GW eq post ms
fault plane

Figure 3



Box of the M6.4
Annualized FMDs - BK-AFT
A
Cumulative Number of events
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
b=1.1
b=0.87
Background b=0.87
Aftershocks b=1.1
2
3
4
5
6
M
b time-series
B
2
1.5
1
0.5
b-value
reference b-value
January
July
2020
TLS level
+30%

Figure 4


Puerto Rico catalog
A
b-value
M
2003-2019
2020
Box of the M6.4 - Annualized FMD
B
Cumulative Number of events
b=1.2
?
b=0.6
Background b-value= 1.2 or 0.6?
M