Summary of the key points raised by the reviewers:

Reviewer #1 agrees the study is important and well conducted, but highlights that your current conclusion may overstate the inconsistency with static Coulomb stress triggering. They suggest that a complete evaluation should include true negatives (locations where no aftershock occurred) and propose a modified approach based on the CRS model and integration over the full stress shadow region. They also request clarifications on slip model distances, ETAS parameter choices, and a few technical and editorial improvements to figures and phrasing.

We thank Reviewer #1 for their insightful comments. We have added a new test to address their comments about true negatives, which is not exactly the test they suggested, but does incorporate CRS modeling over the full stress shadow region. The results further support our interpretation that the number of aftershocks observed in the stress shadows exceeds what would be expected. We have also addressed their clarification questions, please see responses to individual comments below.

Following his/her comments:

The authors analyze the possible explanations for aftershocks occurring in so-called stress shadows where the estimated static Coulomb stress has decreased. These aftershocks seem to contradict the static stress-triggering hypothesis. However, it has previously been hypothesized that they might have actually experienced positive static stress changes due to (1) inaccurate stress calculations, (2) unusual fault orientations, (3) different frictional conditions, and (4) secondary stress changes due to early aftershocks or postseismic deformation. The authors analyze these potential causes in detail for the rich aftershock data sets of the Kumamoto and Ridgecrest sequences. In particular, they use the available slip models to evaluate the (epistemic) uncertainties of the stress calculations and the focal mechanisms of background and aftershocks to characterize the receiver mechanisms. By performing 2000 Monte Carlo calculations of the stress changes at each aftershock location, they define stress shadow events as earthquakes where the stress change is positive in less than 1/3 of the cases. To avoid uncertain near-field stress calculations, the analysis focuses only on those events that are more than 4 km away from the mainshock rupture. For these events, the authors test the four different hypotheses and find that none of them can sufficiently explain the occurrence.

The topic is very interesting, the sophisticated analysis is done with care, and the paper is well structured and written. I therefore recommend that the results of this analysis should be published. However, my main criticism is that the conclusion that "aftershocks in stress shadows are inconsistent with static Coulomb stress triggering" (e.g., title) cannot be easily drawn (see below). Therefore, the conclusions should be partially revised before publication.

Main point:

The authors claim that while there is a chance of experiencing positive stress for any earthquake in the defined stress shadow area, the total of 355 and 285 events in the Kumamoto and Ridgecrest cases, respectively, cannot be explained. In fact, the resulting probability of positive stress at all locations would be close to 0 if the individual probabilities of positive stress (i.e., values $p < 1/3$) were simply multiplied.

However, it is not enough to consider only those locations in the stress shadow where an earthquake occurred ("false negatives"). One must also consider the locations where no earthquake was triggered ("true negatives").

This is a good point, thank you. We have endeavored to respond to this comment by demonstrating that the number of aftershocks in the stress shadow as a whole is larger than what is predicted by the CRS model. Considering the aftershock rate accounts for both the true and false negatives contained within the shadow.

Let's say there are 1000 locations in the stress shadow region with (uncorrelated) stress changes taken from a distribution with $p=1/3$ for a positive stress change. Then, on average, a positive stress change (earthquake) would be expected at 333 locations, which would roughly match the observations.

This is an interesting thought problem, but is unworkable as an actual test because it is completely dependent on the grid size that defines the locations. On one extreme, one could make the grids so large that there are only 3 locations and we expect 1 aftershock. One could make the 1000 grid points the reviewer suggests and expect 333 aftershocks. Then one could make 10,000 grid points and expect 3333 aftershocks. And so on to the other extreme where we expect an arbitrarily large number of aftershocks.

One way to consider the entire stress shadow region (including "true negatives") would be to apply the CRS model for the calculated variability of the stress changes. Specifically, the stress shadowing areas can be divided into a grid, and the empirical stress change distribution for a slip model (with foreshocks) can be calculated at each location using 2000 random selections of the aleatory uncertainties (receiver mechanism and friction coefficient). In each grid (index i), the total number of expected aftershocks N_i can then be calculated by averaging the 2000 values calculated by equation (3). Summing over all grid points in the stress shadow (each associated with a volume V_i) gives the total expected number N of aftershocks. N depends on Asigma, ta, and the background rate. While the first two parameters can be set to literature values, the latter can be approximated by the background rate of earthquakes with a magnitude larger the the completeness magnitude (Mc) in a larger region (volume V) multiplied by the factor V_i/V. If N is significantly less than the observed number of m>Mc aftershocks in the stress shadow, then the conclusion that "the aftershocks in stress shadows are inconsistent with static Coulomb stress triggering" is indeed justified.

Thank you for suggesting this approach, which should be less grid-size dependent. We tried implementing this test and ran into a few problems, however.

The first problem is that the average number of aftershocks from the suite of realizations is unstable because it is largely controlled by the largest rate increases from just a few realizations with the largest positive Coulomb stress changes. Because the CRS event rate is exponential with Coulomb stress change, these larger N/Nback values can span several orders of magnitude and these high tails are poorly sampled. An example for a grid cell in the stress shadow is shown below. The three points at the high tail (out of 2000 realization) have an outsized impact on the average N/Nback: including/excluding the highest point alone can change the average by ~15% and all three by ~30%. Using the median is more stable, but is unsatisfactory because it does not sample the positive Coulomb stress changes at all for locations in the shadows.

The second problem is that, because of the very large N/Nback values in the high tail, the expected value of N/Nback ends up being >1 almost everywhere, including in the identified stress shadow where most realizations have N/Nback~0. This essentially erases the stress shadows from the CRS model, similar to the Woessner et al. (2011) implementation of the mainshock uncertainty. The problem with this is that while the stress shadows might disappear when you average over all of the considered mainshock slip models, there is one actual mainshock slip distribution and it must produce a stress shadow somewhere.

Therefore, we chose to address this comment by considering the whole stress shadow for each realization, and finding the distribution of the shadow area and predicted number of aftershocks. This is shown in new Figure 2. The left panels show that the size of the observed stress shadow (P(DCS>0)<(1/3)) is smaller than most of the modeled stress shadows, meaning that it appears to be a conservative estimate of the stress shadow extent. The right panel shows the distribution of the number of aftershocks in the modeled shadows. The number of events in the observed stress shadow is much larger (red vertical line). We also looked for an "optimized" stress shadow in the observed data by finding the area the size of the median modeled stress shadow that has the minimum number of aftershocks. This optimized stress shadow also has more aftershocks than any of the realizations (green vertical line). This means that there isn't a shadow in the observed data with the same size and lack of aftershocks as forecast by the CRS model. The rate of aftershocks in the stress shadow (however the shadow is identified) is therefore higher than predicted by CRS modeling.

Minor points:

- Only aftershocks larger than 4 km away from the rupture are considered. However, the alternative slip models have at least partially different geometries. Therefore, it needs to be clarified whether the 4 km is related to a particular slip model or it is meant that those events are at least 4 km away for all analyzed slip models.

We use the median distance to the multiple slip models. This is now noted in the text.

- Figure 1(a): It would be nice to add known fault traces.

The fault traces closely follow the planes of the mainshock source model, so this does not add much to the figure and makes the mainshock source model less clear.

- Line 242, caption of Figure 1: Be more precise: "... >= 4 km from the mainshock rupture" instead of "... >= 4 km from the mainshock"

Done.

- Line 262: "M = 0 - 3.3" instead of "M0-3.3"

Done.

- Line 263: "M = 0.3 - 3.7" instead of "M0.3-3.7"

Done.

- Declustering with the ETAS model: The results depend on the parameter choice of the ETAS model. Therefore, the chosen ETAS parameters should be provided and their choice justified.

The ETAS parameters were fit to each catalog. The ETAS equation and parameter values are now given.

Reviewer #2 is supportive of the manuscript but offers numerous insightful comments to clarify language and strengthen interpretation, including whether both nodal planes are considered, explanations of figure captions, robustness of the 4 km threshold, and additional references that could enrich the context.

We thank Reviewer #2 (Debi Kilb) for her many helpful suggestions to improve the clarity and presentation of the manuscript. We have adopted most of her suggestions, please see response to individual comments below.

This work explores why aftershocks occur in the regions of mainshock generated static coulomb stress decreases (i.e., shadows), locations where aftershocks theoretically should be suppressed. The authors investigate five different hypotheses. Of these five, none of them can adequately explain these shadow aftershock's occurrence. They conclude that other triggering models must be at play, suggesting one possibility is delayed triggering from dynamic stress changes that results from transient passage of large amplitude seismic waves from the mainshock. This paper is easy to read, I love the title, short, simple and to the point, the figures are clear and informative. This manuscript is top-notch. Typically, when I review manuscripts, I suggest three primary aspects to consider modifying to strengthen the paper, but for this paper I struggled to come up with even two aspects! Here they are:

1) Consider adding a few sentences about the treatment of the two nodal planes. Do you
consider both planes in your work?

Yes, we do consider both nodal planes. We have clarified in the Methods section that we randomly choose one of the two nodal planes for each realization. We are unable to confidently select a preferred fault plane due to the complexity of faulting (e.g. the sets of orthogonal right- and left-lateral faults at Ridgecrest.)

2) I like to include a short summary statement to figure captions to solidify what the main
point is of the figure. You might consider this in your figure captions. If that is not your style, no worries, leave as is.

We've added short summaries, or edited the existing text to better start with a short summary, for some of the figure captions. Other figures captions already start with a short summary.

Additional minor comments, listed by line number, are included below.
If the authors have any questions about my suggestions, they should feel free to contact me directly. I look forward to seeing this manuscript in print.

Dr. Debi Kilb (OK to release my name to the authors)

Minor suggestions listed by line number.

40-41: OLD: This implies that other physical triggering models are needed, for example delayed triggering by dynamic stress changes from the passing seismic waves. NEW: This implies that other physical triggering models are needed, for example transient processes such as delayed triggering by dynamic stress changes from the passing seismic
waves. {unless that makes your abstract go over the word count limit}

Added.

L45: “Earthquakes usually trigger more earthquakes, referred to as aftershocks.” Should ‘earthquakes’ be changed to “relatively large” earthquakes? Do M1 earthquakes have aftershocks?

Yes, we think M1 earthquakes cause aftershocks, they are just usually too small to make it into earthquake catalogs. So we haven’t changed this.

L47: Perhaps introduce static and dynamic here, noting that static can be positive and negative, but dynamic only positive or zero.

This seems like too much detail for a non-technical summary. We think it’s easier to follow if we focus on the tests of the static stress change model, which is the main point of the paper, and introduce dynamic stress triggering later as an alternative model. We’ve added the word “permanent” to indicate that we’re focusing on static stress changes.

L49: Just to be clear OLD: ‘stress shadows’ NEW: ‘stress shadows’, where no aftershocks are expected.

Added.

L56: Traversing the region

Added.

L62: Perhaps introduce what a ‘receiver fault’ is here, in relation to the mainshock.

Definition added.

L68: Consider moving the explanation about receiver faults above Equation 1.

We think it's helpful to have the Coulomb stress change equation earlier in the paragraph, followed by the details about how we determine each of the parameter values.

L72: Is 'a few km' true for all mainshock sizes? Maybe list your assumed mainshock size?

This shouldn't be mainshock size dependent, as stress change should decrease similarly away from mainshocks of all sizes. The main controlling factor is what is assumed about the background stress magnitude in the calculation of the OOPs, and a few km is typical for the OOP calculations that appear in the literature.

L97: Maybe define the magnitude range for moderate to large mainshocks

Done.

L101-101: Consider expanding on high background rate a bit here – does this mean a lower
magnitude of completeness, and so a decrease is easier to detect?

It's very hard to see a rate decrease if the rate is already very low. We've added a reference to Marsan and Nalbant (2005) who discuss the difficulty of observing rate decreases.

L107: You might consider including information about this paper and this reference: https://www.nature.com/articles/s41586-021-03601-4

This is an interesting paper, but it addresses the opposite problem of why aftershocks didn't occur in a stress increase region. So we have chosen not to discuss it.

L107: A summary sentence might be nice here – these findings suggest the process is not
straightforward, and multiple additional explanations might be required to fully explain these observations.

Done.

L126: To me, 4km seems kind of small. Maybe test a couple of ranges to make sure the results remain robust. Or as an initial check, see how far the ‘unexplained’ fellows to the fault. Or maybe there is a reference that can back this up?

Citation added. Marsan (2006) shows that small-scale features of the slip model can dominate out to about 4 km.

L130: A few references here might be useful.

Added

L134-135: Instead of ‘several’ say “have at least N …”

Done.

L144-148: Not sure, maybe move this information to a table and simply list the number of slip models for each region.

We think a table would end up taking up too much additional space in the paper, compared to a relatively short list of references.

L177: OLD: letter quality, NEW: focal mechanisms quality metric ranking

Done.

L189-190: For belts and suspenders add your 4km from the mainshock info here.

Done.

L195: By receiver fault orientation do you mean strike only?

No, this is the uncertainty found from the mechanism variability in 3D. We’ve added “3D” to clarify this.

L204: To me, 45° seems like a lot – is there any correlation between the ‘unexplained’ and the Letter grades (Ridgecrest) or cataloged uncertainties (Japan)?

No, the events with larger mechanism uncertainty don't tend to be disproportionately unexplained by CRS. In fact, the distribution of uncertainty is similar for the stress shadow, stress increase, and ambiguous mechanisms, as shown in the figure below. We've added this information to the results for the test of Hypothesis 2, where this is most relevant.

L209: You might provide a reference for your friction choice of 0.2-0.8, the range is very reasonable, but a reference might be nice if easy to find.

Added

L206-214: Nicely explained!

Thanks!

L240: The foreshock for the Japan data looks like a single line, was it a 90-degree dipping fault? You might mention that in the caption or otherwise explain why it is only a straight line.

This is correct, it's a vertical fault. Similarly, the shown model for the Ridgecrest mainshock is a vertical fault. This explanation has been added to the figure capture.

L241-242: Fig caption 1, consider adding number of aftershocks in the maps (N=XXX), (N=YYY)

The insets show the cumulative number of events, so this information is already there.

L256: Instead of “many,” state the exact number, is it 2000?

Done.

L258:260: This Landers info might be better suited for the discussion section?

Done.

L262-263: Consider adding 355 out of how many and 285 out of how many? Or give percentages.

We don’t think that this is the most relevant information, as it would just give the relative rate of aftershocks in the shadows versus stress-increase and stress-ambiguous regions. The more relevant information is that this is much larger than the number of aftershocks we would expect to be in the shadows. We now compute this expected number (see response to Reviewer 1), and show that ~300 is a relatively large number of aftershocks.

L274-279: Interesting!

Thanks!

L287-289: Revisit – consider better explaining why there would be a rate increase followed by a rate decrease. Perhaps add another sentence of explanation?

This comes out of the CRS modeling, which is done in the referenced papers. Basically all of the loaded faults fail before all of the unloaded faults are reloaded. We’ve added a very brief statement to this effect.

L289-290: Swap order of references Toda / Marsan.

Done.

L322: Consider listing here what data you are using for the background quakes? Time range and spatial extent?

Added the time and distance ranges for the background events.

L321-330: Very nice figure! Consider adding 'N=XX' to either the subplots or the figure caption. How come the dots in (c) and (f) are larger than in the other panels. Is it because there are fewer points? Add a sentence explaining why there is this difference.

Thanks, we've added the number of events for each panel. We also note that the dots in those two panels are larger to make up for there being fewer events.

L335: Which nodal plane do you use? Both?

Yes, both. Noted.

L342: Figure 3. Consider N=X in shadow; N=y not in shadow to both titles

Added.

L372: Consider adding a summary statement to this figure.

Added.

L384-389: For this paragraph, for clarity note which mainshock you are discussing.

This refers to both sequences, which is already noted in this paragraph.

L406: For clarity, note if both nodal planes are explored, or maybe add that information to your methods section.

Thanks, we've expanded on this information in the methods section.

L411: It looks like you have some space in the subplots, to help the reader consider adding
"Shear failure only' and 'Normal failure only" etc. descriptive text within the sub-plots

Added

L414: OLD: Shown for aftershocks NEW: Shown for the NNN aftershocks (list total number of shadow aftershocks)

Changed

L438: Consider adding a summary sentence to the figure caption.

Added.

L449-451: Nice!

Thanks!

L455: OLD We include in the declustering NEW: We first decluster the catalog by removing ….

This is a list of all events that were input in the declustering algorithm, not the events that were removed. We have clarified this.

L465: Consider limiting the values to have only two digits of precision?

We have left Table 1 as-is. Limiting to two digits would leave some values with a precision of +/-0.1 and some +/-1, which makes this table awkward.

L475: Consider adding why these two models were selected over the other available models.

These were the only two models that we were able to find where the model itself was publicly available. If there are others, please let us know!

L502: Consider adding the number of aftershocks within 4 km, and what percent of the total aftershocks they represent.

We added these percentages.

L541: Consider adding how many aftershocks. Also add the spatial extent of the background earthquakes. Is this the full study area?

Added. Yes the background is the full study area.

Table 2: Add what percentage 285-355 represents. I wonder if there is an easy way to show
some kind of Venn diagram that shows the overlap of these results for the different datum. Would you expect that would be enlightening or not?

As we noted above, we don't think that this is the most relevant information, as it would just give the relative rate of aftershocks in the shadows versus stress-increase and stress-ambiguous regions. The more relevant information is that this is much larger than the number of aftershocks we would expect to be in the shadows. We now compute this expected number (see response to Reviewer 1), and show that ~300 is a relatively large number of aftershocks.

L587: OLD: None of these hypotheses. NEW: None of the individual hypothesis …

Done.

L592: Consider adding a reference or two

Reference added.

L604-605: Any correlation with depth?

This is a good question in general, about all of the observations. The aftershocks don't cover much depth range, but there is no observed systematic trends with depth for any of the tests.

L657: To be picky, should 'positive everywhere' be changed to 'never negative', because they can be zero.

Done.

L671-674: All references here are more than 10 years old, add some newer references to
indicate these hypotheses are still active.

We want to cite original literature. If there's a more recent paper the reviewer thinks we should cite, let us know!

L677: 15-60 is a pretty large range, if it makes sense to do so, consider listing the range for the two studies individually: … estimate XX-YY % and ZZ-AA% for the van der Elst and
Brodsky 2010 and Hardebeck and Harris 2022 studies, respectively. Or maybe just refer to van der Elst and Brodsky here as the Hardebeck and Harris findings are nicely listed later in this paragraph.

Both studies find this full range, when considering uncertainty. So it doesn't make much sense to separate them.

L697: OLD: stress shadows NEW: stress shadow regions

Done.

L702-704: Nice conclusion and very nicely backed-up / supported.

Thanks!

L706-709: Consider rewording this sentence for clarity. Or perhaps break into two sentences.

Done.

L712: Do you check both nodal planes? Maybe add that to the methods section.

The consideration of both nodal planes is now more clear in the methods section.

L728: Does it make sense to add information about what percentage are explained by two or more of the hypotheses?

We don't see that this would be particularly useful information. It's also implied in Table 2 for anyone who's interested, just by adding up the percentage explained by each model and comparing that to 50%.

L747-751: More clearly indicate which of the two mainshocks you are discussing.

Done.

L748: Consider for clarity: OLD: majority (58-71%) ; NEW: majority of the shadow aftershocks (58-71%).

Done.

Reviewer #3 encourages publication with minor revisions. They propose that your results might best be interpreted not as evidence against static stress triggering per se, but as a demonstration that Coulomb static stress estimates may not accurately capture the complexity of real static stress variations in the crust, especially in the presence of rough faults, damage zones, or fluid circulation. They suggest clarifying this interpretation in the conclusion, and potentially framing your results more conservatively, focusing on the epistemic limits of Coulomb stress calculations rather than ruling out static stress altogether. They also encourage you to cite recent relevant literature (e.g., Meade et al., 2017).

We thank Reviewer #3 (Davide Zaccagnino) for his supportive comments on the manuscript. While we seem to agree that Coulomb static stress as it is currently calculated is a poor predictor of aftershock locations, we do have a difference of opinion about how to interpret that result. The Reviewer appears to favor the Coulomb static stress model as the cause of aftershocks, and interprets the results as showing that the current modeling is inadequate to capture the true stress changes. We are not so attached to static Coulomb stress, and are more open to considering other models such as dynamic stress changes. The Reviewer makes some good points about the difficulty of modeling localized stress concentrations due to material heterogeneity. We have acknowledged throughout the paper that the calculated Coulomb stress changes are only an approximation of the true stress changes, and have added a section to the Discussion that considers unmodeled material heterogeneity as a possible explanation of our results. However, we decided not to reframe our results with this as the primary explanation.

We have added the Meade et al. (2017) reference. We would appreciate it if the Reviewer could give us some citations to his preferred model of localized stress concentrations due to material heterogeneity, which could help us better develop this new section of the Discussion.

Following his comments:

Dear authors and editor,

the manuscript "Aftershocks in Stress Shadows are Inconsistent with Static Coulomb Stress Triggering" is a relevant contribution to a crucial topic in earthquake physics concerning our understanding of the triggering mechanisms of seismicity. Authors analyse various features of seismic events belonging to two earthquake sequences reported in high quality seismic catalogs in Japan (Kumamoto 2016) and in California (Ridgecrest 2019) using different techniques. The paper is well-written, organized and enjoyable to read; the quality and statistical significance of results seem reliable (I did not reproduce them, but I cannot detect issues in methods and the displayed values are perfectly reasonable). The introduction is clear and provides a fair bird's-eye view on the current state of knowledge. Discussions and conclusions are mostly in agreement with authors' results.

Therefore, I endorse the publication of this article pending minor revisions.

Hereafter, I outline my major suggestion, while specific and minor comments are reported throughout the revised pdf attached below.

Aftershock triggering is usually attributed to increases of static stress on fault. Coulomb stress is an oversimplified, but usually considered effective method to assess the static stress under the hypotheses that the slip behavior of faults is completely controlled by their frictional properties (i.e., they are well modeled as individual planar surfaces with friction surrounded by elastic crustal volumes).

In my opinion, the paper would benefit from a statement clarifying the difference between the "real" static stress (which is unknown) and "estimations" of static stress (e.g., static stress).

Thank you, this is a good point. It is now explicitly stated in Methods, and other places, that the calculated Coulomb stress is only an approximation of the "real" stress.

Then, if we assume that an increase of static stress is the reason why aftershocks occur, the take-home message of this manuscript should be that Coulomb static stress estimations are not reliable at all when applied to high quality seismic catalogs available nowadays. In this framework, the following paper may be of interest and could be added among the references of the article:

Meade, B. J., DeVries, P. M., Faller, J., Viegas, F., & Wattenberg, M. (2017). What is better than Coulomb failure stress? A ranking of scalar static stress triggering mechanisms from 105 mainshock-aftershock pairs. Geophysical Research Letters, 44(22), 11-409.

as well the already cited paper by Sharma et al., JGR, 2020.

Thank you, we have added the Meade reference. Meade, Sharma, and others find that a simple isotropic spatial kernel explains aftershocks better than the Coulomb spatial kernel. This is consistent with aftershocks in the stress shadows degrading the performance of the Coulomb spatial kernel.

Possible explanations may be the following:
1) the brittle crust behaves as porous granular matter (with fluids circulation within) producing complex temporal and spatial arrangement of additional stress provided by seismicity.
2) fault damage zones and fractality may sustantially affect the stress field both in its amplitude and direction (self-amplification and stress focusing due to strongly nonlinear mechanical behavior of crustal volumes due to heterogeneity of rheological parameters and structural complexity, e.g, asperities) and the spatial resolution to assess real stress variations is not available.
3) other explanations already investigated by authors.

Thank you for sharing your ideas about what processes could be causing the shadow aftershocks. We do already address pore pressure changes as a possible triggering mechanism. We have added a new section to the Discussion to address the second point as a possible explanation (new section 4.2). However, we think that this section could be stronger if you could provide some specific literature to cite that develops these ideas and perhaps provides observational evidence. Thank you.

On the other hand, the authors may question the first assumption "If we assume that an increase of static stress is the reason why aftershocks occur, ...", but it would require a clear demonstration that observations, beyond any reasonable doubt, are not compatible with any possible physically-justified "real" static stress values. This is very hard to prove because the spatial variation of the "real" static stress can be extreme, expecially around rough fault interfaces (fault stress is a self-affine quantity - it follows a power law, and static stress is expected to dynamically converge to a power-law distibution).

So, I would not advocate new physical mechanisms as a first step, especially if authors are not sure that they are feasible (they suggest a possible role for "delayed" dynamic stress changes). I would rather focus on the take-home message which is more parsimonious (and, at the same time, with capital importance for earthquake seismology) and certainly supported by authors' results which may be summarised as follows: "The Coulomb static stress is a poor method for the estimation of the real static stress variations in the brittle crust, whose spatial and temporal variations may be severely underestimated, being possibly dominated by strong nonlinear amplification (and shadowing/suppression) mechanisms requiring unprecedented resolutions and techniques to be detected".

It is impossible to prove beyond any reasonable doubt that there aren't localized stress heterogeneities, due to unobservable material heterogeneities, that are located perfectly to load the faults of each of the shadow aftershocks. So this does not seem like a reasonable criterion to require before considering other physical mechanisms. Dynamic triggering is hardly a "new" physical mechanism, it is well-known to trigger far-field earthquakes. We have provided multiple citations supporting dynamic triggering in the near-field as well.

The hidden assumption in my comments above is that uncertainties usually associated with Coulomb stress values are underestimated.

Yes. However, the alternative model (localized stress concentrations due to material heterogeneity) doesn't produce any quantitative uncertainties that we could use in hypothesis testing.

I think there are chances that my hypothesis is true and I hope that authors will consider it in their manuscript.
I list a couple of motivations that may support my hypothesis:

1) Stress can focus around asperities and other structural heterogeneities on "rough" faults. Faults are fractals, so that at high spatial resolution this effect can become dominant producing isolated or clustered peaks and shodows of stress within the fault systems far beyond the "gaussian predictions" of the Coulomb static stress (and related uncertainties).

2) Static stress may have a dependence on time, although weak, due to the mechanical (porous, granular) behavior of the fault system during aftershocks, likely marked by fluid circulation.

We have added a new section to the Discussion to address these possible explanations (new section 4.2).

Thank you for considering my comments,

Davide Zaccagnino

Annotated manuscript:

L28: Aftershock triggering is usually attributed to increases of static stress. Coulomb stress is an oversimplified, but usually considered effective method to assess the state of stress on faults under the hypotheses that
they are completely controlled by frictional properties
-> they are well modeled as individual planar surfaces with friction

I think that it is important that authors to clarify in this paper the difference between the "real static stress" and "estimators of static stress" variations.

CS =! real static stress

It is now stated in Methods, and other places, that the calculated Coulomb stress is only an approximation of the real stress.

If we assume that an increase of static stress is the reason why aftershocks occur, the take-home message of this manuscript is that Coulomb static stress estimations are not reliable at all when applied to high quality seismic catalogs available nowadays.
Possible explanations may be the following:
1) the brittle crust behaves as a porous granular matter (with fluids circulation within) producing complex temporal and spatial arrangement of additional stress provided by seismicity.
2) fault damage zones and fractality may sustantially affect the stress field both in its amplitude and direction (self-amplification and stress focusing due to strongly nonlinear mechanical behavior of crustal volumes due to heterogeneity of rheological parameters and structural complexity, e.g, asperities) and the spatial resolution to assess real stress variations is not available.
3) ...other explanations such as those listed by authors.

We have added a new section to the Discussion to address these possible explanations (new section 4.2).

On the other hand, the authors may question the first assumption "If we assume that an increase of static stress is the reason why aftershocks occur, ...", but it would require a clear demonstration that static stress perturbations, beyond any reasonable doubt, are not compatible with any possible physically-justified "real static stress", which is very hard to prove because the spatial variation of the "real" static stress can be extreme, expecially around rough fault interfaces (fault stress is a self-affine quantity - it follows a power law, and static stress dynamically converges to a power-law distibution).

It is impossible to prove beyond any reasonable doubt that there aren't unobservable stress heterogeneities located perfectly to load the faults of each of the shadow aftershocks. So this does not seem like a reasonable criterion to require before considering other physical mechanisms.

L40: ... for the correct assessment of the "real value of static stress".
Also delayed (almost static) stress contributions may be considered of course, I agree.

But I would not advocate new physical mechanisms as a first step, especially if authors are not sure that they are feasible. I would rather focus on the take-home message which is more parsimonious (and, at the same time, with capital importance for earthquake seismology) and certainly supported by authors' results: "The Coulomb static stress is a poor method for the estimation of the real static stress variations in the brittle crust, whose spatial and temporal variations may be severely underestimated, being likely dominated by strong nonlinear amplification (and shadowing/suppression) mechanisms requiring unprecedented resolutions and techniques to be detected".

We've added to the Abstract a statement that the results apply to the currently-possible calculations of Coulomb stress change, which don't include the small-scale heterogeneity that the reviewer refers to.

L50: My assumption in my comments above is that uncertainties usually associated with Coulomb stress values are underestimations.
I think there is a good chance that my hypothesis is true and I hope that authors will consider it in their manuscript.

We've added to the Non-technical summary a statement that the results apply to the currently-possible calculations of Coulomb stress change.

I list some reasons that may support my hypothesis:

1) Stress can focus around asperities and other structural heterogeneity of "rough" faults. Faults are fractals, so that at high spatial resolution this affect can become dominant producing isolated or clustered peaks and shodows of stress within the fault systems far beyond the "gaussian predictions" of the Coulomb static stress and its estimated uncertainties.

2) Static stress may have a dependence on time, although weak, due to the mechanical (porous, granular) behavior of the fault system during aftershocks, likely marked by fluid circulation.

We have added a new section to the Discussion to address these possible explanations (new section 4.2).

3) Static stress does not consider the state of stress of receiver faults: if they are far from failure, even a relatively large static stress may be not sufficient to trigger earthquakes.

True, although this is more relevant to the opposite question of why there are some areas in stress increase regions that don't produce aftershocks.

4) ...

L206: This uncertainty is likely an excellent estimation of the variability of CS, but, in my view, it may be a rough underestimation of the "real static stress" variability within the fault system.

We do write there that this is the uncertainty of the "calculated" Coulomb stress change, not the real change.

L397: I think it is always important to stress that acoustic emissions in the lab, although providing useful information about the physical processes that may be at work during rupture processes and material physics, should be painstakingly scrutinized before upscaling them to natural fault systems. A short "warning" message should be added to recall this important concept to the readers.

We use the words "suggesting" and "may" in this sentence, so we think this is adequately careful.

Table 1: H1: what about fault roughness, fractality, structural heterogeneity, nonlocal focusing of stress etc.? (My previous comments)

We now clearly state that this hypothesis only encompasses the modeling choices currently available for calculating Coulomb stress change, and does not include possible effects from unobservable small-scale Earth structure or varying rheology.

L753: I would like authors add the hypothesis that the static stress variability may be underestimated.

We've added to the Conclusions a statement that the results apply to the currently-possible calculations of Coulomb stress change, which don't include the small-scale heterogeneity that the reviewer refers to.

Dear Dr. Petrillo,

Thank you for the reviews of the revised manuscript. We are pleased that the reviewers are in general satisfied with our responses to the first round of review. We have responded to the few remaining issues below.

Best wishes,
Jeanne Hardebeck

REVIEWER #1 -

The authors have done an excellent job addressing nearly all the points raised by the three reviewers. Therefore, I believe the paper is close to being acceptable.

Thank you.

However, I still have two questions/comments regarding the new test (lines 300 - 330, including the new Figure 2):

1. I agree that averaging the RS response for the various possible receiver mechanisms and friction coefficients may be unrealistic, as it assumes that all those fault orientations coexist within each volume. Instead, the authors appear to have performed a test where, if I understand correctly— please correct me if I'm wrong—they first randomly selected a focal mechanism and friction coefficient, and then applied these values uniformly across the entire region. If this is accurate, it indicates an assumption of uniform/single fault orientation throughout the region, which seems unrealistic given the natural heterogeneity.

This is not quite correct. For each realization, we chose a single mainshock model. Then at each location on the spatial grid, we chose a single focal mechanism from the distribution of background events at that location. This allows for the fault orientations to vary across the region, in accordance with the variation of background events. This is now explained at line 305.

As an alternative, I suggest running each simulation in the following way: At each location, randomly select a single receiver mechanism and friction coefficient independently from the other locations. Then calculate the size of the stress shadow region and the expected number of earthquakes within it for the corresponding simulation. This process could be repeated 2,000 times to get the distribution of the expected earthquakes in the stress shadow and the stress shadow size.

We actually did exactly this. Each simulation assumes a single mainshock source model, while the receiver mechanism at each location is chosen at random from the local mechanism distribution, and the friction coefficient is

randomly chosen as well. We have added text better describing what we did, starting at line 305.

2. The size of the observed stress shadow is smaller than most of the modeled stress shadows. This discrepancy suggests an inconsistency between the test calculations and the original analysis. Could this difference be due to the authors using the same receiver mechanism for the entire region in each simulation, as mentioned in my previous point?

We think that the observed stress shadow is smaller because we use a conservative definition of the stress shadow. It is roughly the intersection of the stress shadows computed with multiple different mainshock source models, which makes it smaller than the average size of the individual modeled shadows. We have added text with this explanation, starting at line 322. As we explained above, we do not actually use the same receiver mechanism for the whole region in each simulation.

REVIEWER #2 -
This is my second review of this work based on the updated manuscript. The authors have carefully responded to the reviewers' comments and suggestions and updated the manuscript accordingly. They were very thorough in their response, including additional in-depth tests and figures. Their careful response was very impressive. There were some reviewer suggestions that the authors did not find helpful, and they adequately stated the reasons why no changes were made.

Thank you.

The authors added a new figure (Figure 2) to the work, which I found helpful. They should, however, refer to (a), (b), (c) and (d) in the Figure 2 caption, and I would suggest they also add a summary sentence to the figure caption if they think that might be helpful.
I look forward to seeing this paper published.
Debi Kilb (OK to give my name to the authors)

We have revised the caption of Figure 2 to include a summary sentence and individual explanations of panels (a), (b), (c) and (d).

REVIEWER #3 -
Dear editor and authors,
the manuscript has been greatly improved following the suggestions of all the reviewers and it is now ready for publication.
I would like to thank the authors for considering my comments.

Thank you.

Following the authors' request

"Thank you for sharing your ideas about what processes could be causing the shadow aftershocks. We do already address pore pressure changes as a possible triggering mechanism. We have added a new section to the Discussion to address the second point as a possible explanation (new section 4.2). However, we think that this section could be stronger if you could provide some specific literature to cite that develops these ideas and perhaps provides observational evidence. Thank you."
Hereafter I list some research papers supporting the idea of "anomalous" stress concentration around asperities and weak points in disordered and heterogenous materials in different research fields. I agree with authors that very limited studies have been realized so far in earthquake science on this important topic. I hope they can be useful:
1) Wiese, K. J. (2022). Theory and experiments for disordered elastic manifolds, depinning, avalanches, and sandpiles. *Reports on Progress in Physics*, *85*(8), 086502.
2) Bonamy, D., & Bouchaud, E. (2011). Failure of heterogeneous materials: A dynamic phase transition?. *Physics Reports*, *498*(1), 1-44.
3) Hainzl, S., & Zöller, G. (2001). The role of disorder and stress concentration in nonconservative fault systems. *Physica A: Statistical Mechanics and its Applications*, *294*(1-2), 67-84.
4) Sornette, D. (1992). z– 3/2 powerlaw decay of Laplacian fields induced by disorder: consequences for the inverse problem. *Geophysical research letters*, *19*(24), 2377-2380.
5) Laubie, H., Radjai, F., Pellenq, R., & Ulm, F. J. (2017). Stress transmission and failure in disordered porous media. *Physical review letters*, *119*(7), 075501.
6) Ben-Zion, Y., & Rice, J. R. (1993). Earthquake failure sequences along a cellular fault zone in a three-dimensional elastic solid containing asperity and nonasperity regions. *Journal of Geophysical Research: Solid Earth*, *98*(B8), 14109-14131.
7) Shu, W., Lengliné, O., & Schmittbuhl, J. (2023). Collective behavior of asperities before large stick-slip events. *Journal of Geophysical Research: Solid Earth*, *128*(9), e2023JB026696.
8) Duan, H. L., Wang, J., Huang, Z. P., & Luo, Z. Y. (2005). Stress concentration tensors of inhomogeneities with interface effects. *Mechanics of Materials*, *37*(7), 723-736.
9) Nattermann, T., Shapir, Y., & Vilfan, I. (1990). Interface pinning and dynamics in random systems. *Physical review B*, *42*(13), 8577.

Thank you for sending these references. The papers that consider earthquakes (Hainzl & Zöller, 2001; Ben-Zion & Rice, 1993; Shu et al., 2023; as well as the earthquake section of Wiese, 2022) are mainly focused on explaining the Gutenberg-Richter b-value with models or lab experiments of a single fault with asperities. This is interesting, but not directly applicable to our problem of Coulomb stress changes at distances >4 km from the mainshock fault. The other papers are less directly related: Bonamy & Bouchaud (2011) consider crack propagation, Sornette (1992) explains the problems of representing heterogeneous fields with harmonic decomposition, Laubie et al (2017) study tensile failure in a porous material, Duan et al (2005) model a spherical inhomogeneity in a composite material, and Nattermann et al (1990) study the effects of defects on magnetic systems.

Therefore, we lack a good reference from the literature that presents a model of localized stress change concentrations due to material heterogeneity that we could apply in the context of Coulomb static stress change calculations. This means that while we can consider unmodeled material heterogeneity as one possible explanation of our results, developing a detailed model of the effects of material heterogeneity on Coulomb static stress change calculations, perhaps based in part on the suggested references, would be a substantial project that is outside of the scope of this paper.

To address this theme briefly, however, we have added a new reference, and a new sentence, starting at line 700: "Material heterogeneity has been linked to stress variability (e.g. Martínez-Garzón et al., 2025)."