Reviewer # 1 Comments

Recommendation: Revisions required

For Author and Editor

Review of "Fragility Modeling of Precariously Balanced Rocks: Calibration, Benchmarking, and Sensitivity" by Z. Chen et al.

Devin McPhillips
May 14, 2026

Precariously balanced rocks (PBRs) can provide unique information about the intensity of past earthquake shaking. This manuscript presents a body of work to better quantify PBRs constraints.

First, the manuscript presents a detailed description of the software VSR2, which is an update to VSR (as presented in Chen 2024). VSR2 is a tool for implementing physics engine simulations of PBRs, and the documentation, especially in the supplement, provides a useful baseline. I found S2 Contact Modeling in PyBullet particularly useful from the perspective of a potential user who wants to understand more deeply how this implementation of pybullet works.

Next, the manuscript calibrates the physics engine's macro-physical parameters using >500 real shake table experiments (from Saifullah and Wittich, 2023). Such a calibration is essential for applying the VSR2 with confidence.

Finally, the manuscript compares the VSR2 to a discrete element model (DEM) for the same set of shake table experiments. Results show that the VSR2 achieves similar accuracy to the DEM with orders of magnitude less computational cost.

I found the manuscript clearly written and illustrated. I believe it will contribute to the wider adoption of PBR data for improving hazard estimates. My concerns (below) are either about presentation and clarity or so broad as to be reasonably beyond the scope of this work. I suggest publication with minor revisions.

Comments

DEM comparison
1. I suggest describing the DEM work in Saifullah and Wittich 2023 in more detail. For example, is it fair to say they explored/calibrated the relevant DEM parameters already, and the best values were used for the comparison here? (Edit - I see that sentence now in the results paragraph, consider moving to methods)
2. I don't know much about comparing computational efficiency. You seem to be (a) solving the same problem, (b) using a similar hardware environment, (c) exploring a range of ground motions,

(d) presenting metrics for a range of scenarios, and (e) showing scaling behavior. A quick look at best practice suggests that this is good.

Calibration

1. In Figs 4 and 5, consider using open symbols for the FP and FN results--seems like there are a lot of points stacked on top of each other. Also consider printing the accuracy score in each subplot box.

2. In the text related to Fig. 6, the words monotonic and non-monotonic are weird words for describing the results. I get it, I think: you mean that increasing the parameter doesn't consistently increase or decrease the fragility. But I found it confusing because requires viewing the plots in an added dimension.

3. Of course, the work is limited by the use of just one PBR specimen. More is better, but I appreciate that shake table experiments are not easy. As it is, I'm glad to have so many ground motions for a real (complex) boulder.

4. Interesting to learn that pybullet allows for local crushing via state-dependent contact model.

Reviewer # 2 Comments

Recommendation: Accept Submission

For Author and Editor

This manuscript addresses an important need to have an open-source virtual tool that can simulate the overturning response of precariously balanced rocks (PBRs) and rapidly develop fragility curves for seismic hazard analyses. The authors compare their physics-engine results against discrete element method (DEM) simulation results produced through previous studies by this project team to assess the relative accuracy of their simulation tool and assess which specific model parameters have the most influence on overturning fragility. The authors emphasize the speed of their calculations relative to DEM simulations and carry out this study to effectively address whether speed still provides accuracy. I recommend this manuscript be published after addressing the comments below.

Major Comments

- Significant portions of this paper are devoted to comparisons with the discrete element method. I would like to see some more discussion of aspects that are also relevant to DEM simulations. The first is contact overlap. In DEM simulations, contact overlaps are expected to be very small relative to the object. In DEM simulation with spherical particles, Hanley et al. (2015) in *Soils and Foundations* suggest the timestep and contact stiffness should be selected to ensure the overlap between two spheres does not exceed 5% of their radii. Are overlaps between objects relevant to your simulations, and are they similarly affected by contact stiffness and timestep?

- What are the criteria for numerical stability in VSR2? DEM simulations are stable if their timestep is less than the critical timestep, and the critical timestep is proportional to the square root of mass over stiffness. Typically timesteps are selected as ~10% more or less of the critical timestep in DEM simulations. It makes sense that VSR2 would be significantly faster than DEM simulations if using a timestep that is orders of magnitude greater than DEM simulations. How does the timestep in your simulations compare with that used by Saifullah & Wittich?
- Did the authors compare the modes of failure between simulation and experiment, and is there any indication that these are consistent? For example, if the rock overturns in both experiment and simulation, do they typically overturn in the same direction, or are there many cases where the simulated rock instead slid or overturned in a different direction?

Minor Comments

- Is there an upper-bound frequency that your model can resolve accurately? Would it just be the inverse of the timestep?
- Figure 6: include units for contact damping and contact stiffness in the figure or the caption.
- Line 185 “Therefore, the VSR2 contact parameters should be interpreted as numerical controls…”: I’m moving this comment down to minor comments because Section 6 addresses most of this. Can the contact parameters be calibrated from physical measurements? For example, friction coefficient is often set to 0.5 in DEM simulations. This value is consistent with the sliding friction coefficient for blocks of quartz. Density is addressed, but do your other model parameters similarly reflect real physical values? I recognize that this question is addressed in an ad-hoc manner in the first paragraph of Section 6.1, but I’m wondering if such parameters can be determined without going through the optimization process of your study, particularly for contact stiffness and friction. Maybe the answer is that given uncertainty in physical properties, fragility models can be rapidly developed for a range of different parameters and incorporated as epistemic uncertainty in seismic hazard analyses. This latter point is mostly addressed in Section 6.2.

For editor only

For transparency in case this is a conflict of interest, I do know 5 of the authors personally. I participated in a workshop recently with 4 of them and will likely be a co-author with them in the near future. At the time of this review, though, I have no formalized collaboration with any of them beyond discussions of a formal collaboration.

Dear Editor,

We sincerely thank you and the reviewers for the careful evaluation of our manuscript, *“Fragility Modeling of Precariously Balanced Rocks: Calibration, Benchmarking, and Sensitivity.”* We appreciate the constructive comments and the reviewers’ positive assessment of the manuscript’s clarity and potential contribution to the wider use of PBR constraints in seismic hazard analysis.

We have revised the manuscript in response to the reviewers’ comments. A detailed point-by-point response to each reviewer comment is provided in the accompanying response document, and changes in the manuscript are marked using tracked revisions.

Thank you again for your time and consideration. We hope that the revised manuscript is now suitable for publication in *Seismica*.

Sincerely,

Zhiang Chen
on behalf of all co-authors

Reviewer # 1 Comments

Review of "Fragility Modeling of Precariously Balanced Rocks: Calibration, Benchmarking, and Sensitivity" by Z. Chen et al.

Devin McPhillips
May 14, 2026

Precariously balanced rocks (PBRs) can provide unique information about the intensity of past earthquake shaking. This manuscript presents a body of work to better quantify PBRs constraints.

First, the manuscript presents a detailed description of the software VSR2, which is an update to VSR (as presented in Chen 2024). VSR2 is a tool for implementing physics engine simulations of PBRs, and the documentation, especially in the supplement, provides a useful baseline. I found S2 Contact Modeling in PyBullet particularly useful from the perspective of a potential user who wants to understand more deeply how this implementation of pybullet works.

Next, the manuscript calibrates the physics engine's macro-physical parameters using >500 real shake table experiments (from Saifullah and Wittich, 2023). Such a calibration is essential for applying the VSR2 with confidence.

Finally, the manuscript compares the VSR2 to a discrete element model (DEM) for the same set of shake table experiments. Results show that the VSR2 achieves similar accuracy to the DEM with orders of magnitude less computational cost.

I found the manuscript clearly written and illustrated. I believe it will contribute to the wider adoption of PBR data for improving hazard estimates. My concerns (below) are either about presentation and clarity or so broad as to be reasonably beyond the scope of this work. I suggest publication with minor revisions.

Comments
DEM comparison
1. I suggest describing the DEM work in Saifullah and Wittich 2023 in more detail. For example, is it fair to say they explored/calibrated the relevant DEM parameters already, and the best values were used for the comparison here? (Edit - I see that sentence now in the results paragraph, consider moving to methods)
We appreciate the suggestion. We have added the description of the DEM comparison in Section 4.4 Benchmarking. Specifically, we clarify that the predictive reliability benchmark compares the VSR2 performance results with the DEM performance outcomes reported by Saifullah and Wittich (2023), including accuracy, precision, recall, and F1 score.

For the computational efficiency benchmark, we reproduced DEM simulations using the DEM parameter set reported by Saifullah and Wittich (2023). These values correspond to the DEM parameter set that produced the highest F1 score for Specimen 1 in Saifullah and Wittich (2023).

2. I don't know much about comparing computational efficiency. You seem to be (a) solving the same problem, (b) using a similar hardware environment, (c) exploring a range of ground motions, (d) presenting metrics for a range of scenarios, and (e) showing scaling behavior. A quick look at best practice suggests that this is good.
Thank you for the feedback. We have clarified that "This setup allowed the runtime comparison to evaluate the same prescribed overturning problems across various ground-motion durations, shaking intensities, and geometric complexities."

Calibration
1. In Figs 4 and 5, consider using open symbols for the FP and FN results--seems like there are a lot of points stacked on top of each other. Also consider printing the accuracy score in each subplot box.
Thank you for bringing this up. We have modified the figures as suggested.

2. In the text related to Fig. 6, the words monotonic and non-monotonic are weird words for describing the results. I get it, I think: you mean that increasing the parameter doesn't consistently increase or decrease the fragility. But I found it confusing because requires viewing the plots in an added dimension.
We agree on and have replaced monotonic and non-monotonic with more descriptive words.

3. Of course, the work is limited by the use of just one PBR specimen. More is better, but I appreciate that shake table experiments are not easy. As it is, I'm glad to have so many ground motions for a real (complex) boulder.
Thank you for the feedback.

4. Interesting to learn that pybullet allows for local crushing via state-dependent contact model.
Thank you for the feedback.

Reviewer # 2 Comments
This manuscript addresses an important need to have an open-source virtual tool that can simulate the overturning response of precariously balanced rocks (PBRs) and rapidly develop fragility curves for seismic hazard analyses. The authors compare their physics-engine results against discrete element method (DEM) simulation results produced through previous studies by this project team to assess the relative accuracy of their simulation tool and assess which specific model parameters have the most influence on overturning fragility. The authors emphasize the speed of their calculations relative to DEM simulations and carry out this study to effectively address whether speed still provides accuracy. I recommend this manuscript be published after addressing the comments below.

Major Comments
1. Significant portions of this paper are devoted to comparisons with the discrete element method. I would like to see some more discussion of aspects that are also relevant to DEM simulations. The first is contact overlap. In DEM simulations, contact overlaps are expected to be very small relative to the object. In DEM simulation with spherical particles, Hanley et al. (2015) in Soils and Foundations suggest the timestep and contact stiffness should be selected to ensure the overlap between two spheres does not exceed 5% of their radii. Are overlaps between objects relevant to your simulations, and are they similarly affected by contact stiffness and timestep?
We thank the reviewer for raising this important question. Contact overlap, contact stiffness, and timestep selection are closely connected to the reviewer's second comment regarding numerical stability. Therefore, we address these two comments together in the response below.

2. What are the criteria for numerical stability in VSR2? DEM simulations are stable if their timestep is less than the critical timestep, and the critical timestep is proportional to the square root of mass over stiffness. Typically timesteps are selected as ~10% more or less of the critical timestep in DEM simulations. It makes sense that VSR2 would be significantly faster than DEM simulations if using a timestep that is orders of magnitude greater than DEM simulations. How does the timestep in your simulations compare with that used by Saifullah & Wittich?
DEM and PyBullet/VSR2 use fundamentally different approaches for resolving contact. We have clarified these differences in terms of contact overlap, timestep selection, and the number of solver iterations in Section 6.3 of the revised manuscript, as follows:

"Contact overlap and timestep selection have different interpretations in VSR2 than in DEM, because the two methods use different contact formulations. In DEM, contact forces are computed from the overlap between bodies, then advanced with explicit time integration. The timestep must therefore be small enough to resolve the rapid contact response; otherwise, the explicit integration can produce unstable forces, velocities, or overlaps. In PyBullet, contacts are handled as constraints. When two bodies are detected in contact, PyBullet estimates contact points, contact normals, and penetration depth. The penetration depth is treated as a positional constraint error and enters the normal-contact impulse through a correction term scaled by the error reduction parameter, while contact stiffness and damping are mapped into the constraint force mixing and error reduction terms (Supplementary Material S2). Thus, contact overlap in PyBullet is used to correct constraint violation during the impulse-based contact solve, rather than being interpreted directly as physical deformation that generates a DEM-style contact force. DEM overlap criteria, such as limiting overlap to a fixed fraction of particle radius, and the DEM critical-timestep criterion are not directly transferable to VSR2.}

In VSR2, the timestep primarily controls the temporal resolution of the prescribed ground motion and impact events, whereas the number of solver iterations affects convergence of the iterative constraint-solution process. Increasing the solver iteration count allows PyBullet to further reduce velocity and positional constraint errors within each timestep. In this study, VSR2 used a fixed timestep of 0.001 s and 150 solver iterations per timestep. The reproduced DEM *simulations used a timestep calculated following the 3DEC procedure for explicit integration, in which the timestep is selected from the critical stability condition associated with the model's mass and stiffness."*

3. Did the authors compare the modes of failure between simulation and experiment, and is there any indication that these are consistent? For example, if the rock overturns in both experiment and simulation, do they typically overturn in the same direction, or are there many cases where the simulated rock instead slid or overturned in a different direction?
We thank the reviewer for this comment. In the present study, the primary benchmark was based on binary overturning or stability outcomes, consistent with the available DEM and shake-table comparison data from Saifullah and Wittich (2023) and with the requirements of PBR fragility modeling. We did not perform a systematic mode-by-mode comparison of overturning direction, sliding-dominated response, twisting, or other intermediate response modes across all 582 motions. However, VSR2 records the full translational and rotational time histories of the rock, so such a comparison is technically feasible. Our Limitations and Future Work has clarified this point.

Minor Comments
1. Is there an upper-bound frequency that your model can resolve accurately? Would it just be the inverse of the timestep?
Thank you for this comment. The timestep sets an upper limit on the frequencies that can be represented in the simulation, but this limit is not simply the inverse of the timestep. In this study, VSR2 used a timestep of 0.001 s, corresponding to a sampling rate of 1000 Hz and a Nyquist frequency of 500 Hz. The practical upper frequency that can be resolved accurately is lower than the Nyquist frequency and depends on solver convergence and contact-impact behavior. Quantifying this practical limit would require a dedicated timestep-convergence study and frequency-domain analysis of the input motions and simulated contact response, which are beyond the scope of the present work. We have clarified this limitation in the revised manuscript. "Future work should further examine how physics-engine simulation settings, such as timestep and solver iterations, affect PBR fragility estimates through timestep-convergence, solver-sensitivity, and frequency-domain analyses."

2. Figure 6: include units for contact damping and contact stiffness in the figure or the caption.
Thank you for the suggestion. We have added the units in the caption.

3. Line 185 "Therefore, the VSR2 contact parameters should be interpreted as numerical controls…": I'm moving this comment down to minor comments because Section 6 addresses most of this. Can the contact parameters be calibrated from physical measurements? For example, friction coefficient is often set to 0.5 in DEM simulations. This value is consistent with the sliding friction coefficient for blocks of quartz. Density is addressed, but do your other model parameters similarly reflect real physical values? I recognize that this question is addressed in an ad-hoc manner in the first paragraph of Section 6.1, but I'm wondering if such parameters can be determined without going through the optimization process of your study, particularly for

contact stiffness and friction. Maybe the answer is that given uncertainty in physical properties, fragility models can be rapidly developed for a range of different parameters and incorporated as epistemic uncertainty in seismic hazard analyses. This latter point is mostly addressed in Section 6.2.

The reviewer is right about these ideas. Some VSR2 contact parameters can be physically informed. For example, measured or literature-based rock–rock friction values can provide useful prior ranges for the friction coefficients in VSR2. Restitution is more difficult to constrain directly because it depends on local contact geometry, material properties, and impact conditions. However, for the single-axis motions considered here, our sensitivity results indicate that restitution has a smaller effect on fragility than lateral friction. Contact stiffness and damping require a different interpretation because they are modeled through the PyBullet contact solver and do not directly correspond to unique macroscopic physical properties. Therefore, they cannot be uniquely determined from direct physical measurements alone. Future shake-table experiments with additional PBR geometries and contact conditions will help determine whether these calibrated effects are specimen-specific or more generally applicable to PBR fragility modeling.