Response to reviewers

Preprint egusphere-2025-947 (https://doi.org/10.5194/egusphere-2025-947)

Constraining landslide frequency across the United States to inform county level risk reduction

Lisa V. Luna, Jacob B. Woodard, Janice L. Bytheway, Gina M. Belair, and Benjamin B. Mirus

Dear reviewers, dear editors,

We thank Maria Teresa Brunetti and an anonymous reviewer for their positive assessments of our study and their constructive feedback, which has resulted in an improved and clearer manuscript. Below we respond to the reviewer comments (marked as ***paragraphs in bold italics beginning with A***) and show how our revised manuscript has addressed these remarks. All figure, section, and line numbers refer to the **revised** manuscript.

Best wishes,

Lisa Luna on behalf of all co-authors

RC1: Maria Teresa Brunetti

The work aims to estimate the frequency of landslides in the US using available landslide inventories. Landslides are those triggered by earthquakes and precipitation.

The manuscript overall is well written and the objectives of the work are clear. The approach is also promising. Nevertheless, the part describing the models and techniques used (section 2.4) is somewhat cryptic and not easy to read for those with non-advanced skills on statistical distributions, such as negative binomial, applied to overdispersed data. I strongly suggest expanding the part on the models used, giving a more accessible explanation related to the purpose of the work.

A: We thank the reviewer for the positive assessment of our manuscript and approach. We appreciate the feedback that our methods description was not easy to follow for readers with less background in statistics. To address this comment, we have expanded sections 2.2 and 2.4 to include a more accessible explanation of our methods with more background information that better clarifies their purpose for our research objectives.

To keep this response document concise, we refer to the tracked changes manuscript version for the complete overview of changes and highlight key updates here:

In Section 2.2, we included more information to motivate our choice of the negative binomial distribution in lines 177 – 184: "**Negative binomial distributions are suitable for modelling counts, which in our case is the reported number of landslides in a year in a given area**

(White and Bennetts, 1996). Distributions of annual landslide counts per county in our dataset were typically heavily right-skewed, with few years showing many reported landslides and many years showing few reported landslides, and over-dispersed, with variances that exceeded means (Fig. 2). The negative binomial distribution can capture over-dispersion with its two parameters: a rate parameter (μ), which indicates the expected, or average, frequency and a shape parameter (ϕ), which together control the variance. We therefore preferred it to the Poisson distribution, an alternative count distribution, which requires the mean and variance to be equal (White and Bennetts, 1996).*"*

In Section 2.4, we included more background information to provide a more accessible explanation of our model set up, our choice of priors, and the fitting algorithms used. In particular:

***Lines 306 – 310 now further introduce negative binomial regression:* "Negative binomial regression is a generalized linear model that estimates landslide frequency as a function of predictors (Eq. 2). Other examples of generalized linear models include logistic regression, which relies on the binomial distribution to model probabilities, and Poisson regression, which uses the Poisson distribution to model frequencies or rates (McElreath, 2020)."**

***Lines 334 – 344 now provide more background information on multi-level models:* "We chose to include ecoregion as a grouping variable that served as a proxy for the many factors that may influence landslide frequency that we do not explicitly include in our models, for example, climate, land-cover, and geology. In contrast to the pooled models that estimated parameters for the whole domain (Eq. 2), these multi-level models explicitly modelled regional variation in landslide frequency by learning a different intercept for each ecoregion ($b_{0,r}$) while simultaneously learning the mean ($b_{0,p}$) and standard deviation (σ_r) of intercepts among ecoregions (Eq. 3). This means that for a fixed set of predictor values, the estimated landslide frequency is allowed to vary by ecoregion if the data supports this. Nevertheless, because each ecoregion's intercept must belong to the population-level distribution, the model is guarded from overfitting regions with many counties with reported landslides and estimates for areas with less available data are informed by data rich regions, which generally improves predictions (McElreath, 2020)."**

***Lines 360 – 363 introduce the concept of weakly informative priors:* "In Bayesian inference, priors can encode previous knowledge or beliefs about parameter values. Whereas uninformative priors consider all possible parameter values equally probable, weakly informative priors assign a probability to possible parameters, but do not exclude any values that might be learned from the data (Kruschke, 2014; McElreath, 2020)." *Lines 376 – 379 clarify the implications of these choices for our study:* "For datasets with many observations, like ours, these priors primarily serve as a starting point for the fitting algorithm (refer to next paragraph) and the posterior parameter estimates are generally insensitive to the exact choice of prior parameter values (Kruschke, 2014; McElreath, 2020)."**

Lines 382 – 391 clarify how the fitting algorithms work and motivate why this approach is useful for characterizing parameter uncertainty in our study: **"This is an advantage of Bayesian inference: we obtain a distribution of estimates for each parameter rather than, for example, a single maximum likelihood estimate. Bayesian statistical models thus inherently provide transparent estimates of parameter uncertainty (Kruschke, 2014; McElreath, 2020; van de Schoot et al., 2021), but require advanced algorithms to estimate the posterior distributions. To do so, we used Markov Chain Monte Carlo (MCMC) implemented via the R package brms v2.21.0 (Bürkner, 2017), which calls STAN v2.32.6, a statistical programming language that uses the No U-Turn Sampler (NUTS) Hamiltonian Monte Carlo fitting algorithm (Stan Development Team, 2023). MCMC is a stochastic process that samples from the posterior distribution and the NUTS Hamiltonian Monte Carlo algorithm is an MCMC method that generates efficient transitions that span the posterior (McElreath, 2020; Stan Development Team, 2023). We ran four independent chains, or sequences of samples, for 4000 iterations, discarding the first 1000 iterations as warm up, for a total of 12,000 post-warmup draws, or samples from the posterior. The Gelman-Rubin coefficient (R-hat) was 1.00 for all parameters, indicating that the four chains converged around the same distribution."**

Figures (especially 1, 2 and 6) are too dense of information and not easily readable. As a consequence, figure captions are also too long. Please consider moving some figures to a Supplementary Material section.

A: We appreciate this feedback. U.S. Geological Survey publications are required to specify data sources and cartographic projection information in figure captions. Although this leads to lengthier figure captions, it results in increased transparency and allows individual figures to stand on their own. To address this comment, we have made the following changes to figures, which also allowed us to reduce the length of their captions:

- ***Divided former Figure 1 into two figures, one focusing on maps (revised Figure 1) and one focusing on time series (revised Figure 2)***
- ***Moved former Figure 5 showing parameter estimates to the Appendix (revised Figure A1)***
- ***Divided former Figure 6 into two figures, one showing absolute error on a log-scale (revised Figure 6) and one showing error direction and the cross validation results, which we moved to the Appendix (revised Figure A2)***

We agree that former Figure 2 (revised Figure 3), which shows the predictors used in our models, contains dense information. However, we argue that this information is essential for readers to understand the results, particularly as the international audience of NHESS may not be familiar with the topographic, climatic, ecological, and seismic characteristics of the U.S. that underpin our study. We therefore advocate for leaving this figure in the main text, noting that its information density is not uncommon in NHESS publications.

In addition, as a general rule, figures must be cited in advance in the text. For multiple maps/graphs in the same figure, they must also be cited in the text in the order given.

A: We have ensured that all figures are properly cited in order in the revised manuscript. We have also rearranged panels in revised Figure 3 to make sure they are in the same order that they are introduced in the text.

Line by line comments from PDF document. Lines in the review comments refer to the preprint. Lines in the responses refer to the revised manuscript.

A: We are grateful for the specific comments.

Line 40: 2023a

A: Line 41: Added reference to Federal Emergency Management Agency, 2023a

Line 79: I would suggest to add "catalogue data" to "inventory data", since generally a landslide catalogue should contain temporal information on landslide occurrences and not necessarily include geometrical data.

A: The datasets we used here included location data (see Lines 86 – 87), so we refrained from making this change.

Line 117: This statement seems too generic to me. I would support here with examples or references.

A: We clarified that "**The obtained posterior parameter distributions, which show the probability of possible parameter estimates, allow us to transparently report model uncertainty given the available landslide inventory data*****"** **(Lines 117 – 119)**

Table 1, column "record length": If you include the earliest and the latest years, the numbers in this column must be increased by one

A: Thanks for catching this. We updated the values by 1 (Table A1)

Figure 1: Figures must be cited in advance in the text. In addition, each individual map/histogram must also be given in the text in the order given.

A: We have made the appropriate changes (see description above)

Line 229: Please, add a reference here

A: Added reference to Juang et al., 2019 (Line 228)

Line 319: This notation is unfair. I suggest using for beta the subscript i with (i= 1,..., 3).

A: Thanks for this suggestion. We have updated the notation to $\beta_{j=1,2,3}$ ***throughout the manuscript.***

Equation 3: This notation is not clear. Do you mean that this coefficient is extracted from a normalized Gaussian distribution?

Please be more descriptive for both equations and parameters.

A: We clarified that "$\beta_{0,r}$* is a group-level intercept for each ecoregion that belongs to the overarching distribution of intercepts across all ecoregions, which we modelled as a normal (Gaussian) distribution with a mean of zero and standard deviation σ_r." *(Lines 349 – 351) and made additional descriptive changes throughout (see response above and tracked changes document for further details)

Equation 4. The notation N here is different from that of Eq. 3. Please, be coherent.

A: Thanks for catching this. We updated the notation to consistently use "Normal" throughout.

Line 351. This part deserves a more in-depth description

A: We introduced our priors in more depth.* "In Bayesian inference, priors can encode previous knowledge or beliefs about parameter values. Whereas uninformative priors consider all possible parameter values equally probable, weakly informative priors assign a probability to possible parameters, but do not exclude any values that might be learned from the data (Kruschke, 2014; McElreath, 2020)." *(Lines 360 – 364)

Line 355. How did you assume exactly these values (-4.5,3)?

***A: We included additional justification and explanation:* "Our choice of prior for b_0 encodes our belief that landslide frequencies will be well below one landslide km^{-2} y^{-1} in areas with average predictor values. Through the log-link function that relates b_0 to m_c (Eq. 2, Eq. 3), the mean prior for b_0 of -4.5 corresponds to 0.01 landslides km^{-2} y^{-1} when all other predictors are at their mean." (Lines 374 – 377)**

Lines 360 – 365: This part is very cryptic and requires skills that not all readers of the journal may have. For greater clarity, it deserves a longer explanation of the techniques and algorithms used.

***A: We added a longer explanation of the methods and techniques. We also reference several textbooks and publications that interested readers can check for further details:* "Posterior distributions are probability distributions of all parameters that are consistent with the data, prior, and model. This is an advantage of Bayesian inference: we obtain a distribution of estimates for each parameter rather than, for example, a single maximum likelihood estimate. Bayesian statistical models thus inherently provide transparent estimates of parameter uncertainty (Kruschke, 2014; McElreath, 2020; van de Schoot et al., 2021), but require advanced algorithms to estimate the posterior distributions. To do so, we used Markov Chain Monte Carlo (MCMC) implemented via the R package brms v2.21.0**

(Bürkner, 2017), which calls STAN v2.32.6, a statistical programming language that uses the No U-Turn Sampler (NUTS) Hamiltonian Monte Carlo fitting algorithm (Stan Development Team, 2023). MCMC is a stochastic process that samples from the posterior distribution and the NUTS Hamiltonian Monte Carlo algorithm is an MCMC method that generates efficient transitions that span the posterior (McElreath, 2020; Stan Development Team, 2023). We ran four independent chains, or sequences of samples, for 4000 iterations, discarding the first 1000 iterations as warm up, for a total of 12,000 post-warmup draws, or samples from the posterior. The Gelman-Rubin coefficient (R-hat) was 1.00 for all parameters, indicating that the four chains converged around the same distribution. All model diagnostics indicated acceptable fitting algorithm performance (Kruschke, 2014; McElreath, 2020).” ***Lines 380 – 392***

Line 519. Could it result from biased information sources?

A: Although any landslide study relies on incomplete reporting and we understand the reviewer’s concern, we consider it exceedingly unlikely that the observed right skew in landslide counts results from biased information sources. If, hypothetically, landslide counts followed a Gaussian distribution in which most years had approximately the same number of landslide counts, the mean count would need to be very high to produce the highest values observed in these inventories. For example, if we disregard the years with no reported landslides in Marin County, CA, the mean count in years with reported landslides is 41 (Figure 1i). If biased reporting were to produce the observed right skewed distribution from a hypothetical underlying Gaussian distribution, it would imply that ~41 landslides per year would have gone unreported in most years. However, missing dozens of landslides per year is inconsistent with our assessment of monitoring in this area, as this area is a population center that is monitored by both the California Geological Survey and the U.S. Geological Survey. Furthermore, this issue would need to occur repeatedly across the inventories we included this study, requiring a consistent reporting bias across the various state and local actors that produced these inventories, which we would find extremely surprising. We added this point to the discussion: **“Although we acknowledge that these records are likely incomplete, we consider it unlikely that the observed right-skewed distributions result from reporting bias, given the consistent occurrence of such distributions across counties covered by different inventories.” (Lines 382 – 384)**

Line 558: Fig 7d – f

A: Included reference to figure panels.

Line 577: Fig 8d – f

A: We intended to reference the entire figure here.

RC2: Anonymous Referee #2

The study addresses a significant gap in current national-scale landslide susceptibility research by prioritizing the frequency of landslides, which is essential for effective hazard and risk management planning. The manuscript is well-structured and states its objectives clearly, emphasizing the importance of informative landslide hazard estimates for mitigation planning and risk reduction at a national level. The authors employ Bayesian negative binomial regression to model landslide frequencies at the county level using predictor variables that comprise susceptible area, earthquake potential, precipitation frequency, and ecological region. Authors highlight the extreme diversity of landslide susceptibility across the United States, with frequencies described as ranging from a few incidents in some areas to high concentrations in others. There are some minor issues to be modified in the manuscript before publication. The methodology should be presented more concisely and provide the reader with an explanation for further advancing the proposed methods. The proposed manuscript is a significant addition to landslide hazard assessment in that it provides more complete and spatially resolved frequency data than have been feasible with national-scale studies. It is a publication standard, providing valuable information that can be utilized to direct targeted risk reduction and mitigation efforts across the United States.

A: We thank the reviewer for the encouraging appraisal of our work and its implications for landslide risk reduction efforts. We appreciate the feedback that our methods should be presented more concisely. To address this comment and the related comments of Reviewer #1, we have clarified our approach in a concise manner in the revised methods section (see also response to Reviewer #1 above). We also provided a bullet point overview of our approach at the beginning of the methods section to orient readers (Lines 131 – 145).

Concerning the suggestion that we provide the reader with an explanation for further advancing the proposed methods, we direct readers to lines 635 to 670 of the discussion, where we discuss how linking weather-related landslide activity to the magnitude and frequency of precipitation events and earthquake-related landslide activity to seismological parameters may improve future estimates of landslide frequency.