March 31, 2022

Editor
Natural Hazards and Earth System Sciences (NHESS)

Subject: Resubmission of revised article "Equivalent Hazard Magnitude Scale" with manuscript number nhess-2021-87.

Dear editor,

We thank you and the referees for careful review of our manuscript (nhess-2021-87) entitled "Equivalent Hazard Magnitude Scale". To address the comments from the review team, we have made a major revision to the manuscript.

A detailed account of how we addressed the comments from the referees is attached below this response letter in a point-by-point style. The major changes to the manuscript are summarized as follows:

1) We have modified the order of introduction to hazard types in the Methodology section.

2) We have significantly revised the paragraph on the rationale for using logarithmic transformation on some hazard magnitude indicators.

3) We have added material in the Discussion section regarding the limitation of the tsunami magnitude indicator currently adopted in the study.

The revised manuscript is now 9 645 words long and contains seventy-one references, eight figures, two tables in the main text, two tables in the appendix, six supplementary data files, and one supplementary video.

We look forward to hearing back from you regarding our revision.

Sincerely,

Yi Victor Wang, Ph.D.
Postdoctoral Fellow
Institute for Earth, Computing, Human and Observing
Chapman University
ywang2@chapman.edu

Antonia Sebastian, Ph.D.
Assistant Professor
Department of Earth, Marine and Environmental Sciences
University of North Carolina at Chapel Hill
asebastian@unc.edu

Anonymous Referee #2

We thank you very much for your comments and questions. In the following, we copy your comments in *italics* and follow with our response. The major changes to the manuscript are summarized as follows:

1) We have modified the order of introduction to hazard types in the Methodology section.

2) We have significantly revised the paragraph on the rationale for using logarithmic transformation on some hazard magnitude indicators.

3) We have added material in the Discussion section regarding the limitation of the tsunami magnitude indicator currently adopted in the study.

Comment: *I have already reviewed this paper in its original version. I do appreciate the attempt made by the authors in trying to answer to my queries. However, I still have the same doubts*.

Response: Thank you very much for your time and consideration regarding our manuscript in its original version and the revised version after the first round of review. In the response below, we make another attempt to clear your same doubts.

Comment: *Here I try to summarize the most critical comments in a concise way to clarify my view; I hope that this may be of some usefulness for the authors*.

Response: Thank you very much for your summary. We appreciate and value your critical comments.

Comment: *Hazard analysis tries to forecast what the Nature is going to do; and the risk quantifies the impact of those events on humans and the environment*.

Response: Thank you very much for your comment. In the field of science, engineering, and management of hazards and disaster risks, the words "hazard" and "risk" are often used interchangeably to refer to different or the same conceptual domains. Etymologically, "hazard" and "risk" have few differences. According to Möller 2012 (https://doi.org//10.1007/978-94-007-1433-5_3), "risk" can be mainly used to refer to five conceptual domains, which we can further summarize into three: 1) the source/cause of unwanted event or situation (which is also called the hazard agent), 2) a risk quantity referring to the expected loss and/or associated probability distributions, and 3) the unwanted event or situation itself. If we look at literature and academic communications, it is not difficult to find the cases where the word "hazard" is used to describe the same three conceptual domains of "risk", i.e., risk source (hazard agent), risk quantity, and risk situation. Therefore, the terms "hazard analysis" and "risk analysis" can literally refer to exactly the same thing. Likewise, "hazard" and "risk" can literally quantify exactly the same thing. The differences in the use of these terms in practice are to a large extent subjectively, arbitrarily, and authoritatively (instead of being objectively, logically, and scientifically) determined and recognized by different groups of scholars.

Comment: *The hazards cannot be compared (how can we compare the ash fall loading of a volcanic eruption, ground shaking caused by an earthquake, or the run up of a tsunami?); the risk can be compared because they may refer to the same loss metric (number of human lives, economic losses, or else).*

Response: Thank you very much for your comment and question. Since "hazard" and "risk" are often used interchangeably in the public and academic domains, we need to be careful when we are using these words to describe quantitative measures. Here, we use the term "hazard strength" to refer to the measure of the force of a hazard agent associated with an event potentially resulting in adverse impacts. We agree that the impact and the expected impact (or "risk") can be compared across hazard types, as they may refer to the same loss metric. However, we disagree with the statement that the forces of hazard agents cannot be compared (e.g., ash fall loading of a volcanic eruption versus ground shaking of an earthquake). In fact, in this manuscript, we have demonstrated that it is possible to compare the strength of a hazard associated with events across different hazard types by using the expected impact given average exposed entities. It is worth noting that the expected impact given average exposed entities is not the same thing as merely the impact. We are not using the recorded impact to directly compare forces of hazard agents.

Comment: *It is not clear to me what is the "hazard strength". The authors write "... equivalent hazard strength, that is derived from impacts given average exposure and vulnerability based on a robust record of historical impacts.". So, it is not a hazard because it is calculated from the "average" impact of the event, but the authors claim that is neither the impact, even though it is linearly estimated by it.*

Response: Thank you very much for your comment. As mentioned previously, hazard strength is the force associated with the source of an event that may result in adverse impact prior to incorporating vulnerable entities exposed to the force. Hazard strength could therefore be a measure or function of ash fall loading of a volcanic eruption, ground shaking caused by an earthquake, or the run up of a tsunami, as the reviewer has suggested. Based on a systematic review of existing hazard strength metrics, such as earthquake moment magnitude, Saffir-Simpson hurricane wind scale, and tornado Fujita scale, included in another submission, we have noticed that these hazard strength metrics can be further categorized into four groups along two conceptual dimensions, i.e., the spatial dimension and the temporal dimension. Accordingly, the four groups include agential-durational metrics, location-durational metrics, agential-momental metrics, and locational-momental metrics. All hazard strengths can be categorized into one of these four hazard types. However, there are two major challenges to the comparison of hazard strengths across different hazard types.

First, it is only meaningful to compare the hazard strengths of the same spatial-temporal type. This can even be demonstrated by the different types of hazard strengths for a same hazard type. For example, the comparison between a moment magnitude and modified Mercalli intensity for earthquakes is meaningless, as the former is an agential-durational metric while the latter is a locational-durational one. However, it is possible and common to compare the Richter magnitude and the moment magnitude because both are agential-durational metrics.

Second, there is no existing method to quantify the equivalency of a hazard strength when comparing hazard strengths across hazard types. In this manuscript, we propose a methodology to compare the agential-durational hazard strengths across different hazard types by deriving the equivalency of these hazard strengths. To do so, we can build quantitative models to map considered hazard strengths to a common scale, which we define as the equivalent hazard strength scale. Such models can be calibrated by targeting the expected impact given average vulnerable entities exposed to the hazard strength. Even though the target can be considered as the average impact as pointed out by the reviewer, it is not the actual impact. The expectation of a variable and the variable itself are quantitatively and qualitatively different. Thus, it is problematic to assume that the average/expectation/mean of the variable and the variable itself are interchangeable since they refer to different conceptual domains. Here, by using the expectation of adverse impacts, we have demonstrated that it is possible to derive the equivalency of hazard strengths of events across different hazard types.

Comment: 2. *So, if it is not clear what the hazard strength means, how can it be used in practice?*

Response: Thank you very much for your comment. For clarity of the meaning of hazard strength, please see our response to the previous comment. The idea and utility of hazard strength is not new. However, what is novel in this paper is the summarization regarding the types of hazard strengths and how to derive the equivalency of hazard strengths across different hazard types. Regarding the utility of the concept of hazard strength, we are seeing it everywhere. When news media are reporting on an M8 earthquake, that "M8" refers to the hazard strength (an agential-durational one) of the earthquake event. When news media are reporting on a hurricane evolving from Category 1 to Category 5, that "Category 1" and "Category 5" are the measures of hazard strength (an agential-momental one) of the hurricane. However, it is difficult for the general public, or even emergency managers to understand how to prepare for hazard events that have not yet been experienced. Particularly in a new and more violent world (e.g., due to climate change), the ability to compare hazard strengths across multiple hazard types has immense value (see, e.g., the tornado experienced in New Orleans recently).

Comment: *In other words, what is the additional information carried out by the hazard strength with respect to hazard and risk?*

Response: Thank you very much for your comment. As mentioned previously, "hazard" and "risk" are often used interchangeably. To avoid confusion and to facilitate scientific communication, we try to find and use a term to refer to specifically the force associated with the hazard agent of an event that could result in adverse impacts when vulnerable entities are exposed to such a force. There are candidates such as "magnitude" and "intensity". Based on our systematic literature review for another publication, however, we have noticed that these two words, "magnitude" and "intensity" are used to refer to different types of hazard strengths along the conceptual spatial and temporal dimensions across hazard types, even though for each individual hazard type, the use of "magnitude" and "intensity" may follow certain rules (e.g., in earthquake research). Therefore, for our purposes, we find it necessary to use another term to refer to things such as "magnitude" and "intensity" and we settle down with the term "hazard strength".

Comment: *The authors write "Our proposed scale has applicational significance in providing benchmark measures of hazard strength for vulnerability and resilience analyses". That's exactly* *what the risk is meant to do. Why may the use of hazard strength provide additional relevant information on this point?*

Response: Thank you very much for your comment and question. First, "vulnerability" and "resilience" can be used to refer to multiple conceptual domains. Here, for demonstration as an example, let's use the word "vulnerability" to refer to the tendency to suffer adverse impacts given a certain level of hazard strength. Such a vulnerability can be manifested as the casualty ratio, building damage ratio, or economic loss ratio given a certain level of hazard strength. Suppose there are two communities. community A has just experienced an M5 earthquake while community B has just experienced a C5 hurricane. Suppose both communities A and B lost 10% of their population due to the earthquake and the hurricane, respectively. Can we say that these communities seem to have the same vulnerability to adverse impacts because they experienced the same loss ratio? Probably not because we still need to know how to compare the "M5" and "C5". Can we use the existing risk modeling efforts to compare "M5" and "C5"? Probably not because they are not capable of doing so. In this case, then, we can use our proposed concept of equivalent hazard strength, which is what we mean by the benchmark measures to support vulnerability and resilience analyses.

Comment: *In my understanding, the authors claim that the hazard strength is a spatially averaged quantity whereas the risk is more localized. This is not true, because we can calculate the risk at any spatial scale, for example we can calculate the probability that one specific loss (e.g., 1 billion of euros or more) will be caused in the next 10 years by earthquakes, tsunamis, floods, or else, in Europe or at any spatial region.*

Response: Thank you very much for your comment. However, the reviewer's understanding is incomplete. Hazard strength is not necessarily a spatially averaged quantity. As mentioned previously, there are four types of hazard strength, i.e., the agential-durational metric, the locational-durational metric, the agential-momental metric, and the locational-momental metric. Only the agential-durational and agential-momental metrics are spatially aggregated quantities. The other two types are specific to locations. In this manuscript, we only present the agential-durational type of equivalent hazard strength, which is a spatially aggregated metric. Our typology on the 4 types of hazard strength can actually also be applied to impact metrics and risk metrics.

Comment: *Again, not very clear what the hazard strength tells us more than what we already know.*

Response: Thank you very much for your comment. The contribution of our study is on the equivalent hazard strength, not merely on the hazard strength. In our study, we have demonstrated that it is possible to compare hazard strengths of events across different hazard types. We have also proposed a quantitative modeling framework for derivation of equivalency of hazard strength across hazard types.

Comment: *The authors also claim that "In addition, the derived equivalency of hazard strengths can be used to create multi-hazard hazard maps to show the distribution of exceedance probability of hazard strength across different hazard types." That's a bold statement, but to me it is not clear at all how this measure can be used for multi-hazard purposes (cascading events, etc…)*

Response: Thank you very much for your comment. First, before we reach cascading events, it's important to consider the aggregation of expected frequency distributions of hazard strengths of different hazard types for hazard mapping. Existing practices regarding such an aggregation are simply summation without any justification for the weights used for each hazard type. With the derivation of equivalent hazard strength, we can convert different hazard strengths onto a same scale and aggregate the influences of these hazard strengths by using the equivalency of hazard strengths to create a multi-hazard hazard map involving multiple types of hazards. This will be a significant improvement to the current modeling effort for a multi-hazard hazard map. Then, regarding cascading/compound events, our next step would be to apply the typology to a minimum of four types of hazard strength to systematically develop methods to define the hazard strength metrics and to measure these hazard strengths of cascading/compound events before we can develop and train models to compute the equivalency of hazard strengths for these cascading/compound events in a way that is consistent with what we have proposed. Once this is achieved, cascading/compound events can also be included in a multi-hazard hazard map.

Comment: *The authors agree that the database used is not long enough to have a complete view of the losses that can be caused by all natural events (most of them having a long return period), but they claim that "… as the purpose of the paper is to propose an empirical method to derive the equivalent hazard strength across hazard types, the EM-DAT data base is sufficiently robust for demonstration of the proposed method."*

Response: Thank you very much for your comment. This manuscript presents a pioneering study to propose the concept of equivalency of hazard strength and how to compute it. Therefore, the EM-DAT database is sufficient for our purposes herein. However, to systematically compute the equivalent hazard strengths across all four different types along the conceptual spatial and temporal dimensions and across major hazard types including cascading/compound events, we need a database with much higher quality in terms of records of indicators of hazard strengths of events than any existing databases of disaster events. Such an ideal database will also need to include events with zero or almost zero losses and cover a long period of time to account for events with a long return period as much as possible.

Comment: *Moreover, they calculate the hazard strength using a regression analysis on parameters that are not statistically correlated. The authors write "However, the purpose of the regression modeling is not to provide statistical inference between hazard magnitude indicators and impact metric, nor is it to make predictions of impact metric with hazard magnitude indicators. Instead, the purpose is to provide a computational tool to map the hazard magnitude indicators to the equivalent hazard magnitude, which is correlated with the expected value of impact metric.". In my opinion, this view is quite debatable. What is the scientific reliability of a measure that is obtained by a non statistically significant regression? And using a too short catalog?*

Response: Thank you very much for your comment and questions. Among our regression models for equivalent hazard magnitude, many are of statistical significance, e.g., for earthquake and tropical cyclone. Some are not due to a lack of data points, e.g., for cold wave and heat wave. However, the purpose of the study is not to make statistical inference nor predictions of adverse impacts, but rather to provide a quantitative tool to convert hazard strengths to an equivalent hazard strength across different hazard types. The scientific significance of a study such as this presented one is relative to its purpose. Indeed, we agree that, to improve the scientific reliability, we need better data to support our modeling so that we may have a sufficiently large amount of data points to obtain statistically significant results for all considered hazard types and we have proposed this as future work.

Moreover, our study demonstrates the need for database of disaster events with a much higher quality in terms of indicators of hazard strength metrics and impact metrics, which would result in more data points for improvement of the modeling in terms of scientific reliability not only for equivalency of hazard strength but also further for empirical disaster vulnerability and many other metrics that can be used for disaster risk reduction.

Regarding the short catalog issue. Then the question also goes to "how short is too short?" and "how long is long enough?" With a database with data of 1,000 years, we would still miss data points for events once in 1,000 or 2,000 years. The tail distribution of the frequency of extremely large events is always a challenge no matter how one models it. Regardless, with a database with data of hundreds or even only tens of years, we can already develop decent models for those events with shorter return periods, which actually correspond to most of the disaster losses that communities are likely to experience. Nevertheless, when high-quality data of a long period such as thousands of years become available, we would like to use them as well.

Comment: *Note that the length of the catalog is very important, because the inclusion of just one single future event (e.g. a very large volcanic eruption) may severely impact the hazard strength.*

Response: Thank you very much for your comment. This manuscript presents a pioneering study in equivalency of hazard strength. It opens the door for future opportunities in this new area of research. We agree that the inclusion of outlier events may affect the computation of equivalent hazard strength especially for hazard type with a limited amount of data points included in a database. Future work should systematically study the effect of outliers and attempt to develop more robust computational methods for potential improvement.

Comment: *My last comment on the original manuscript, where I discuss the comparison of the 2021 cold wave in US with an earthquake of magnitude 7.5 is not satisfactory (maybe because I was not clear enough in my previous comment). I understand the spatial dimension of the events is very different. But the point is that these two events may have very large different impacts. For example, if the same cold wave would have impacted the east coast instead of Texas and Oklahoma. Or if a M75 would occur inside a large city in US or in the middle of a desert. To sum, I still have may doubts on the usefulness of a scale that equates a cold wave (as the 2021 event) with a M7.5 earthquake in terms of "strength haard", simply because in a short database these two events seem to have produced on average the same losses.*

Response: Thank you very much for your comment. We agree that different events may result in different impacts. However, this can also manifest for events of the same hazard type. For example, suppose that earthquake A is only recorded with a Richter scale at M_L7.5. Meanwhile, earthquake B is only recorded with a moment magnitude scale at M_W7.5. We know earthquake A and earthquake B are equivalent to each other in terms of the agential-durational hazard strength because one is M_L7.5 and the other is M_W7.5, but their impacts can be totally different. Here, the equivalency of hazard strength refers to the expectation of impact given average vulnerable entities exposed to the hazard strength, not the observed impact. We welcome the concerns and conversation from the reviewer, but our proposed equivalency of hazard strength is derived from the expected/average losses of all events of a same hazard type.

Comment: *I have many other doubts, but in general I think that the approach adopted by the authors is vastly speculative and its usefulness is not clear, in particular it is not clear what is the advantage of the additional information provided by the hazard strength with respect to the hazard and risk analysis.*

Response: Thank you very much for your comment. In general, we believe our study provides a solid empirical methodology to compare hazard strengths of events across different hazard types. Such a methodology along with its corresponding typology on hazard strength advances existing hazard and risk analysis. Therefore, we believe it has merit and deserves publication.

Anonymous Referee #3

We thank you very much for your constructive comments and insightful suggestions. In the following, we copy your comments in *italics* and follow with our response. The major changes to the manuscript are summarized as follows:

1) We have modified the order of introduction to hazard types in the Methodology section.

2) We have significantly revised the paragraph on the rationale for using logarithmic transformation on some hazard magnitude indicators.

3) We have added material in the Discussion section regarding the limitation of the tsunami magnitude indicator currently adopted in the study.

Comment: *The manuscript reports an interesting attempt to define a common indicator of the magnitude of the agents responsible for different hazard types, in order to make different hazards comparable in the framework of a multi-risk management. The definition of a common metrics through very different phenomena presents several difficulties, thus this study can be considered a first attempt to define a methodology for comparing the severity of different hazards. However, in my opinion, some further efforts can be done to reduce the uncertainty factors affecting the definition of a common hazard estimator. In the following I report some observations and suggestions for possible improvement and for the discussion of the uncertainty factors, together to a few minor comments.*

Response: Thank you very much for your summary and encouragement. We have made revisions according to your comments and suggestions as follows.

Comment: *Lines 121-126: The authors report that, as hazard magnitude indicator for earthquake and tsunami, they used Richter magnitude (also known in literature as Local Magnitude - ML), in that this is the parameter provided by the catalogue employed (EM-DAT). In section 5.2 they discuss some problems with regard to the use of hazard indicator for earthquakes and tsunami, recalling that Richter magnitude is an unsatisfactory parameter. Actually, a more appropriate indicator would be the moment magnitude (Mw), since Richter magnitude tends to saturate for magnitudes above 6.5. However, after having examined the EM-DAT catalogue, I suspect that the magnitude values in this catalogue are improperly defined as "Richter". For instance, for the 1977 Sumba earthquake (Indonesia) the reported value (8) actually is the moment magnitude rounded to an integer values (the USGS estimate of Mw for this earthquake is 8.3), which for sure is not an ML value.*

Response: Thank you very much for your comment. We acknowledge that Richter magnitude is also usually called the local magnitude M_L. We also acknowledge that the Richter magnitudes and moment magnitudes (M_W) may be mixed in the EM-DAT database. However, for any earthquake event, its M_L and M_W is usually similar. The EM-DAT database rounds the earthquake magnitudes to integers. Once rounded to integers, most of the earthquakes have the same M_L and M_W. While there may be mistakes in the EM-DAT records such that some of the Richter magnitudes are actually moment magnitudes, identifying and correcting these errors is beyond the scope of this study. Here, we follow and keep a consistent protocol for data processing across different hazard types, i.e., following what EM-DAT has recorded for all considered hazard types, in order to demonstrate the application of our proposed method for identifying equivalent hazard magnitude.

Comment: *There is, however, an additional problem to consider: the magnitude values reported in the EM-DATA catalogue appears all rounded to integer values, which represent a rough approximation (an estimate error of 0.5 correspond to an uncertainty of energy by a factor ~5). An improvement of data quality could be easily obtained by replacing such integer magnitudes with the estimates provided for the same events by international catalogues (for instance the USGS significant earthquake catalogue - https://earthquake.usgs.gov/earthquakes/browse/significant.php).*

Response: Thank you very much for your comment. We agree that the rounding of the earthquake magnitudes could be problematic for quantitative modeling. However, because the purpose of the study is to demonstrate an approach to identify equivalent hazard magnitude, rather than predicting the adverse impact of an event, we feel that our use of the rounded earthquake magnitudes is sufficient for this analysis. However, we also recognized that the rounding of magnitude records may introduce some uncertainty to the data used in our analysis. To address this concern, we would highlight that the computation of equivalent hazard magnitude is based on the expectation line of the impact metric and such an expectation line (or the regression line) manifests the central tendency (or the first moment) of the data points and is minimally affected by the increase of uncertainty in terms of the variation (or the second moment) of data points.

In response to your comment regarding other hazard catalogues, we recognize that considering other catalogues may improve the quality of data points. However, because we wanted to follow a consistent protocol for all considered hazard types to demonstrate the application of the methods, we chose to only include events and the data points within the EM-DAT catalogue. Assuming that our publication is successful, we plan to conduct additional studies to systematically improve the proposed equivalent hazard strength methods. We also welcome others to propose such studies. However, prior to doing so, it will be necessary to improve the quality of data on hazard/risk/disaster events, especially in terms of the records of indicators of hazard strengths and the inclusion of events of zero or almost zero losses, as you have suggested.

With respect to earthquake event databases, the first author of this study, YVW, has previously examined several earthquake event catalogues/databases and used their records for his empirical modeling of disaster vulnerability (see, e.g., Wang et al. 2019 https://doi.org/10.1193/022618EQS046M, 2020 https://doi.org/10.1061/(ASCE)NH.1527-6996.0000356, 2021 https://doi.org/10.1080/24694452.2020.1823807). These catalogues include the USGS one recommended by the reviewer. During these previous research works, he has noticed that the records of earthquake events do not always match across different databases. There are records in one database that do not correspond to the records in another database.

Comment: *With regard to tsunamis, one should consider that these phenomena can be also caused by non-seismic sources (e.g. volcanic island collapses and large coastal landslides) and that, in case of earthquake-induced tsunamis, the energy transmitted to sea waves does not depend only on the magnitude of the triggering earthquake but also on its focal mechanism (strike-slip mechanisms generally are not tsunamogenic) and depth of the fault rupture. Thus the magnitude of the earthquake generating tsunami is a rather rough indicator of tsunami energy and, actually,* *some proposals exist to measure a specific "tsunami magnitude" (see Papadopoulos et al., 2020 - https://doi.org/10.1016/B978-0-12-8156686-5.00003-1 - for a review).*

Response: Thank you very much for your comment and suggestion. We acknowledge that the use of Richter magnitude as the hazard magnitude indicator for tsunami is debatable because tsunami can also be triggered by submarine volcanic activities and large coastal or submarine landslides. We also acknowledge that only underwater megathrust earthquakes can generate tsunami. However, for our purposes, whether an earthquake is a thrust-fault one or not does not affect the use of Richter magnitude for tsunami. To indicate the limitation of use of Richter magnitude for tsunami, we have added one sentence in the Discussion section after the limitations of earthquake hazard magnitude indicators.

The added sentence reads: "In addition, regarding tsunami, the mere inclusion of Richter magnitude of a tsunami-triggering earthquake as the magnitude indicator ignores the fact that tsunami can also be caused by non-seismic events, such as volcanic island collapses and large coastal landslides." (L374-376)

We would also like to point out that this manuscript is one of two recent submissions on the equivalent hazard strength scales for peer-review. This paper focuses on the agential-durational scale, while the other is on the locational-durational scale. In that paper, we undertook a systematic literature review of 69 existing hazard strength metrics for 21 singular hazard types and for multiple hazards. These metrics include the Abe magnitude (Abe 1979), the Murty–Loomis magnitude (Murty and Loomis 1980), the Sieberg–Ambraseys intensity (Ambraseys 1962), the Imamura–Iida intensity (Shuto 1993), and the most recently developed integrated tsunami intensity (Lekkas et al. 2013) for tsunami.

Comment: *Lines 132-137 – The authors declare to have applied logarithm transformation to the hazard magnitude indicators of eight hazard types so to have a value distribution close to Gaussian in the range ±∞. Such transformation was not applied to temperature-based (heat/cold waves) and magnitude (earthquake/tsunami) indicators. I believe that the appropriateness of applying or not this transformation should be evaluated by examining if the impact parameters (or their logarithms) show a dependence closer to linear if expressed as function of the logarithm of the hazard magnitude indicators instead of the untransformed original indicators. The proposed justification for not applying logarithmic transformation to strength indicator of earthquakes/tsunamis and heat/cold waves, i.e. the assumption that magnitudes and temperatures have already a range similar to ±∞, seems to me rather questionable. Indeed, while the logarithm of impact indicators can assume values ranging from –∞ (zero damage) to a maximum (which however is not infinite), the atmospheric temperature at the earth surface cannot assume values in the range (–273.15, ∞), but only, approximately, in the order of –60° - 0° for cold waves and 30° - 60° for heat waves. With regard to earthquake magnitude, damaging earthquakes start only from magnitude ~4.5 and magnitudes larger than ~10 are probably not possible because rock strength does not allow energy to be stored, before rupture, up to generate earthquakes stronger than such a magnitude (the largest recorded magnitude is 9.5). A more tenable justification for not applying a logarithmic transformation to earthquake magnitude is that it is already a logarithmic parameter (it is derived from the logarithm of the maximum ground motion amplitude or of the seismic moment). In general, a logarithmic transformation appears reasonable for parameter spanning through several orders of size, which, for instance, is the case of the seismic ground motion amplitudes but not of* *earthquake magnitudes. I believe that the application of logarithmic transformation to hazard magnitude indicators should be more thoroughly discussed and more soundly motivated.*

Response: Thank you very much for your comment and suggestions. The logarithmic transformation to a hazard magnitude indicator is applied mainly to make the distribution of transformed data points to be consistent with a linear model such that a linear regression line can be used as the expectation of impact metric for construction of the equivalent hazard magnitude for the corresponding hazard type. Meanwhile, the logarithmic transformation is also to ensure that any new input hazard magnitude indicator value can correspond to an equivalent hazard magnitude value. For this purpose, the range of the transformed indicator value also needs to be spanned as much as possible so that the zero magnitude points for different hazard types are close to each other at –∞ regarding the transformed indicator and there also remains a possibility of infinite magnitude corresponding to the point of theoretically infinite loss. Arbitrarily setting a lower bound and an upper bound for such a range is problematic. For example, the range of [–60, 0] for cold wave and the range of [30, 60] for heat wave should not replace the possible range of (–273.15, ∞) for temperature. Here, when we need to consider a new data point at –68° for cold wave or at 25° for a heat wave, we should still be able to provide an estimated value of the equivalent hazard magnitude for such an event (although the meanings of such estimates may deserve discussions in future work). For earthquake, as another example, the range of magnitude should not be [4.5, 10]. Even though the largest recorded earthquake magnitude is around 9.5 on earth, there are possibilities that M10 or even M11 earthquakes could occur in the future under extremely rare planetary impacts. Considering these, however, it is also worth mentioning that the logarithmic transformation is not for stabilizing variances in the data points. Because the objective of this study is to use the expected impact metric given average vulnerable entities exposed to hazard magnitude to derive the equivalent hazard magnitude, the variance of a hazard magnitude indicator of data points is not affecting the computation of equivalency of hazard strengths at all. To clarify the main purpose of application of logarithmic transformation to some hazard magnitude indicators, also taking into consideration the suggestions from the reviewer, we have modified the last paragraph of the Data Collection subsection.

The modified Data Collection subsection now reads: "To facilitate regression modelling, we logarithmically transformed values of hazard magnitude indicators to be close to a Gaussian distribution within the theoretical range (–∞, ∞) for eight of the hazard types. Such logarithmic transformations were conducted to keep the shape of distribution of data points consistent with their corresponding linear regression models. The indicators that were not logarithmically transformed included minimum temperature of cold waves, maximum temperature of heat waves, Richter magnitude of earthquakes, and earthquake Richter magnitude of tsunami. Cold wave and heat wave events were excluded from logarithmic transformations because the distributions of data points of these events did not present non-linear patterns and the Celsius temperature has a range [–273.15, ∞) similar to (–∞, ∞). Meanwhile, the earthquake Richter magnitude is already a logarithmic metric with the desired theoretical range of (–∞, ∞)." (L129-136)

Comment: *Lines 64-65: Proponing to call the new scale as "Gardoni scale", the authors should briefly explain the reason of this choice (e.g. the contribution of Gardoni to the advancement of this kind of studies).*

Response: Thank you very much for your comment. As explained in our response to Reviewer 1 during the first round of open review, the proposed agential-durational equivalent hazard strength scale is named in honor of Prof. Paolo Gardoni at the University of Illinois at Urbana–Champaign. This paper with NHESS is one of two manuscripts recently submitted for peer-reviewed journal publication on equivalent hazard strength. In the other submission, we include a systematic review of 69 existing hazard strength metrics for 21 singular hazard types as well as for multiple hazards. In the other paper, we also introduce in detail the typology on four types of hazard strength metrics, i.e., the agential-durational metric, the locational-durational metric, the agential-momental metric, and the locational-momental metric. To differentiate between the agential-durational and the locational-durational metrics more easily, the first author YVW would like to name these two metrics after his two doctoral co-advisors, Prof. Paolo Gardoni (PG) and Prof. Colleen Murphy (CM), considering that the original idea of equivalency of hazard strengths emerged during one of YVW's doctoral advisory meetings in 2015 with PG and CM in PG's office at the University of Illinois at Urbana–Champaign. This manuscript for NHESS is on the agential-durational equivalent hazard strength metric, which is suggested to be named after PG. The other manuscript is on the locational-durational equivalent hazard strength metric named after CM and has been accepted for publication. In this regard, it would seem to be appropriate to name the agential-durational hazard strength metric after PG. To clarify the reason for naming the agential-durational hazard strength metric after PG, we have slightly modified the corresponding sentence in the Introduction section to emphasize the phrase of "in honour of".

The modified sentence now reads: "The proposed scale is named in honour of the Alfredo H. Ang Family Professor Paolo Gardoni at the University of Illinois at Urbana–Champaign." (L63-64)

Comment: *Line 121-125: Introducing the list of discussed hazards, I suggest to re-arrange them, grouping hazards having analogies and similar magnitude indicators, e.g.: convective storm, extra-tropical storm, tornado, tropical cyclone (which use wind speed as hazard magnitude parameters), cold wave, heat wave (using temperature), drought, forest fire, flash flood, riverine flood (using affected area), earthquake and tsunami (using magnitude).*

Response: Thank you very much for your very insightful comment. We agree with your suggestion and have made corresponding modifications to the second paragraph of the Data Collection subsection.

The modified second paragraph of the Data Collection section now reads: "The 12 considered hazard types include convective storm, extra-tropical storm, tornado, tropical cyclone (wind speed is used as hazard magnitude indicator), cold wave, heat wave (temperature), drought, flash flood, forest fire, riverine flood (affected area), earthquake, and tsunami (Richter magnitude). For data quality control, we removed data points with questionable values of hazard magnitude indicators. For cold wave events, we only included data points with a minimum temperature ≤0 °C; for convective storms, we only considered data points with a peak gust wind speed ≥60 km h^{-1}; for forest fires, we only included data points with a burnt area ≤200 thousand km^2; for heat wave events, we only considered data points with a maximum temperature ≥35 °C and ≤57 °C; for tornadoes, we only included data points with a peak gust wind speed ≥100 km h^{-1}; and for tsunami, we only considered data points with an earthquake Richter magnitude ≥6." (L121-128)

Comment: *5. From the diagram of Figure 5, an earthquake of magnitude 3 appears to have caused significant damages. This data is hardly credible, unless some secondary effect (e.g. a landslide triggered by the earthquake) was responsible of damages (but, in such a case, damages cannot be attributed to the earthquake magnitude). I suggest to remove this data from the analysis.*

Response: Thank you very much for your comment and suggestion. We have double checked this data point of M3. It corresponds to the record of an M3.2 event (https://earthquake.usgs.gov/earthquakes/eventpage/usp000923y/executive) on February 1, 1999, in Southern Russia from the USGS earthquake catalogue. This earthquake was subsequent to two other recorded nearby earthquakes on the same day, i.e., an M4.2 (https://earthquake.usgs.gov/earthquakes/eventpage/usp000921w/executive) and an M4.5 (https://earthquake.usgs.gov/earthquakes/eventpage/usp0009233/executive). In the EM-DAT database, there is only one record on this earthquake series, which corresponds to a rounded magnitude value of M3. Based on our data protocol, we strictly follow the EM-DAT record instead of using the M4.2 or M4.5 record for the magnitude indicator of this event. In addition, although the traditional belief is that damaging earthquakes need to be at least M4.5, small earthquakes do sometimes result in somewhat significant losses to vulnerable communities based on YVW's previous examinations of earthquake data (e.g., Wang et al. 2019 https://doi.org/10.1193/022618EQS046M, 2020 https://doi.org/10.1061/(ASCE)NH.1527-6996.0000356, 2021 https://doi.org/10.1080/24694452.2020.1823807). Also, as listed in https://www.pennlive.com/news/2019/06/how-strong-is-a-34-magnitude-earthquake-the-richter-scale-explained.html, M3-M3.9 earthquakes rarely causes any damage, and being rare is not being never. Rarely causing damage means that there is at least a possibility of damage. Beside the possibility of damage by an M3.2 earthquake, this data point is not far away from the regression line. Thus, it does not significantly affect the modeling result. Therefore, we keep this data point in our model, even though we deeply appreciate the reviewer's suggestion.