November 11, 2021

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 your careful review of our manuscript (nhess-2021-87) entitled “Equivalent Hazard Magnitude Scale”. To address the comments from the review team, we have made major revisions to the manuscript. The revised manuscript has 9 605 words, seventy-one references, eight figures, two tables in the main text, two tables in the appendix, six supplementary data files, and one supplementary video.

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 changes to the manuscript are summarized as follows:

1) We have modified the abstract to focus on key points of the manuscript.

2) We have significantly revised the introduction section.

3) We have significantly reduced the length of section 2: A Problem of Scales.

4) We have modified the methodology section to make it more succinct and have moved the content and tables on missing values to the appendix.

5) We have modified the results section and the discussion section.

6) We have updated the references accordingly.

7) We have double-checked the event records discussed in the manuscript and have corrected two errors.

8) We have updated Fig. 1 and have modified the figure captions.

9) We have also slightly modified the conclusion section.

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

Sincerely,

Yi Victor Wang, Ph.D.
Postdoctoral Fellow
Center of Excellence in Earth Systems, Modeling and Observations
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

Referee – John Hillier

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 changes to the manuscript are summarized as follows:

1) We have modified the abstract to focus on key points of the manuscript.

2) We have significantly revised the introduction section.

3) We have significantly reduced the length of section 2: A Problem of Scales.

4) We have modified the methodology section to make it more succinct and have moved the content and tables on missing values to the appendix.

5) We have modified the results section and the discussion section.

6) We have updated the references accordingly.

7) We have double-checked the event records discussed in the manuscript and have corrected two errors.

8) We have updated Fig. 1 and have modified the figure captions.

9) We have also slightly modified the conclusion section.

Comment: *I like the idea of this paper, but it needs to be much more clearly written – I am afraid I had to re-read many times to understand the driving purpose, method proposed, and assumptions. It would benefit greatly from being focussed and simplified. A major revision is needed in terms of the text, whilst the underlying work seems mainly robust.*

Response: Thank you very much for your encouragement. We have made substantial revisions to the manuscript following your comments and suggestions.

Comment: *12 hazards are considered. The innovation is a creating a new standardized measure of impact (IM) by combining 3 loss/impact measures from EM-DAT after log transforming and standardizing data each. IM is then related (linear regression) to measures of hazard severity (e.g. Richter scale) for each hazard, such that for each hazard event (e.g. M_w = 6.7) a IM can be estimated, which is then linearly scaled to fit a range [0,10], called 'equivalent magnitude' EM. Finally, on the premise that hazard characteristics of events that appear in EM-DAT are a representative sample of all similar events, and that averaging (via regression) allows all local risk related aspects (e.g. exposed assets, vulnerability) to cancel out, the authors argue that measures of hazard severity (e.g. Richter scale, area flooded) can be compared via their EM values. This permits events (e.g. a cat 5 hurricane and a M_w 6.7 earthquake) to be compared in terms of potential to cause damage (i.e. hazard) in a way that is as decoupled as possible from local human exposure (i.e. assets at risk), albeit entirely based upon the relative typical size of impact of each event type.*

Response: Thank you very much for your summary.

Comment: *Please find below more major comments, and a non-exhaustive selection of minor comments. I have only considered the text in any detail to the end of Section 4.1 as I assume a second round of review will be necessary.*

Response: Thank you very much for your comment.

Comment: 1. *I have substantial difficulty with the authors' desire to name a scale they 'propose' (L10) in this paper after a person (i.e. Gardoni). This is primarily for two reasons. 1.1. The first reason is the appropriateness of doing this, something not related to the scientific content of the article, so I explicitly ask the journal's editorial team to take a view. For instance, has Gardoni been asked? Is it in the editor's view acceptable scientific practice?*

Response: Thank you very much for your comment and questions. We agree that the final decision about the appropriateness of naming the scale be left to the editorial team. Prior to doing so, we would like to provide some background information regarding this manuscript to address your concern. This manuscript with NHESS is one of a pair of papers recently submitted for peer-reviewed journal publication on the topic of equivalent hazard magnitude/intensity. As described in detail in the other manuscript, we have identified four types of equivalent hazard magnitude/intensity: type 1 (agential-durational), type 2 (locational-durational), type 3 (agential-momental), and type 4 (locational-momental). To differentiate between the equivalent hazard strength scales (types 1 and 2) more easily, YVW would like to name the first two of the scales after his two doctoral co-advisors, Prof. Paolo Gardoni (PG) and Prof. Colleen Murphy (CM), since the original idea of equivalent hazard magnitude/intensity emerged during a doctoral advisory meeting in 2015 with PG and CM in PG's office at the University of Illinois at Urbana–Champaign. Both PG and CM have been approached about this idea and are in accordance with the naming convention. In fact, the type 2 scale (locational-durational) will be named after CM and is already under review after the second round of revision. In this regard, it would seem to be appropriate to name the type 1 equivalent hazard strength scale after PG.

Comment: 1.2. *The second reason follows from this, and in my view needs the manuscript (e.g. Abstract, introduction to be rephrased). If the authors use 'the Gardoni scale', a citation to the work it was developed in is sufficient, without further elaboration. If the scale is developed in this paper, I question the justification for the naming. 'Equivalent hazard' scale should be sufficient if it's novel ….. and others may call it the Wang Scale later if they so choose*.

Response: Thank you very much for your comment. During personal communications, PG insisted that we should mention "the Gardoni Scale" in its current manner in the manuscript with its first appearance in the abstract. The scale is developed in this paper. The justification for the naming has been provided within our previous response to the reviewer's comment. Since there can be four different equivalent hazard magnitude/intensity scales, merely using the term "equivalent hazard scale" does not seem to be sufficient. What others may call these scales is beyond the scope of this study. Nevertheless, YVW insists the type 1 equivalent hazard strength scale be called the Gardoni Scale. Such a naming system has already appeared in a recent academic/professional presentation at the 2021 EGU General Assembly (see Wang and Sebastian 2021 at https://doi.org/10.5194/egusphere-egu21-6468).

Comment: 2. *Clarity of writing: Throughout, the paper would benefit from simplification and focussing on key points. Illustratively, L11-21 of the abstract provide details, but make little sense before a detailed reading of the paper. Please seek to provide an overview of purpose and a sense of some of the assumptions involved.*

Response: Thank you very much for your comment. Following you suggestion, we have significantly modified our abstract.

The modified abstract now reads: "Hazard magnitude scales are widely adopted to facilitate communication regarding hazard events and the corresponding decision making for emergency management. A hazard magnitude scale measures the strength of a hazard event considering the natural forcing phenomena and the severity of the event with respect to average entities at risk. However, existing hazard magnitude scales cannot be easily adapted for comparative analysis across different hazard types. Here, we propose an equivalent hazard magnitude scale to measure the hazard strength of an event across multiple types of hazards. We name the scale the *Gardoni Scale* after Professor Paolo Gardoni. We design the equivalent hazard magnitude on the Gardoni Scale as a linear transformation of the expectation of a general measure of adverse impact of a hazard event given average exposed value and vulnerability. With records of 12 hazard types from 1900 to 2020, we demonstrate that the equivalent magnitude can be empirically derived with historical data on hazard magnitude indicators and impacts of events. In this study, we model the impact metric as a function of fatalities, total affected population, and total economic damage. We show that hazard magnitudes of events can be evaluated and compared across hazard types. We find that tsunami and drought events tend to have large hazard magnitudes, while tornadoes are relatively small in terms of hazard magnitude. In addition, we demonstrate that the scale can be used to determine hazard equivalency of individual historical events. For example, we compute that the hazard magnitude of the February 2021 North American cold wave event affecting the southern states of the United States of America was equivalent to the hazard magnitude of Hurricane Harvey in 2017 or a magnitude 7.5 earthquake. Future work will expand the current study in hazard equivalency to modelling of local intensities of hazard events and hazard conditions within a multi-hazard context." (L8-24)

Comment: *2.1. To simplify, please consider what is truly necessary for the paper; e.g. (i) reduce Section 2 to Fig.1 and a short paragraph.*

Response: Thank you very much for your comment regarding Section 2. We have reduced its length from 919 words to 420 words. However, because this section offers the theoretical background for the proposal of the Gardoni Scale, we have kept some of the prior content. Such a theoretical background would be, for the first time, introduced in a peer-reviewed journal article if accepted earlier than the other submission on the Murphy Scale. As such, it is essential to lay out the four types of hazard strength metrics before introducing the details of the methodology to derive the equivalent hazard magnitude on the proposed Gardoni Scale. Moreover, there are also some fundamental confusion associated with the terminology in the field of disaster studies that need to be clarified before proposing the Gardoni Scale. In light of these, we feel strongly to keep Section 2 within the manuscript.

The modified Section 2 now reads:

"In natural hazards research, theoretical frameworks are often based on basic concepts, such as hazard, impact, exposure, vulnerability, recovery, and resilience, that have overlapping or

discipline-specific definitions (see, e.g., Klijn et al., 2015). These inconsistencies across disciplines often result in confusion in quantitative modelling. Therefore, we clarify several definitions used in this paper. Herein, the impacts of an event are the result of strength of the hazard agent, value of entities exposed to the event, and vulnerability of the exposed entities to hazard impacts (Nigg and Mileti, 1997; Coburn and Spence, 2002; Wisner et al., 2004; Dilley et al., 2005; McEntire, 2005; Adger, 2006; Peduzzi et al., 2009; Burton, 2010; Lindell, 2013; Birkmann et al., 2014; Highfield et al., 2014; van de Lindt et al., 2020; Wang et al., 2020; Wang and Sebastian, 2021a). As shown in Fig. 1, hazard strength of an event is one of the main drivers, albeit not the sole driver, of impacts.

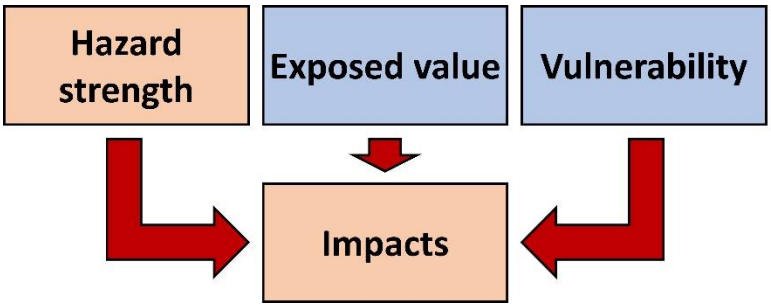

Figure 1: Hazard event impacts as the result of hazard strength, exposed value, and vulnerability of exposed entities.

Hazard strength is often referred to as the hazard magnitude or hazard intensity (Blong, 2003; Alexander, 2018). However, these two concepts are not equivalent. Hazard magnitude is a measure of the size of, or the total energy involved in, the entirety of a hazard event (Blong, 2003; Alexander, 2018), whereas hazard intensity is often a measure of the strength of an event with respect to a given location or area and/or a moment or period.

Recently, Wang and Sebastian (2021b) identified two defining dimensions, i.e., the spatial and temporal dimensions, to categorize existing hazard strength scales. These scales can be classified as *agential* or *locational* along the spatial dimension and *durational* or *momental* along the temporal dimension. A hazard strength scale is categorized as *agential* if it indicates the size of an event within its entire spatial range and *locational* if it is given for a set of locations within the spatial range of an event. Likewise, a hazard strength scale is categorized as *durational* when it corresponds to the entire duration of an event and *momental* when it corresponds to a set of moments within the duration of an event. Considering both the spatial and temporal dimensions, hazard strength scales can therefore be categorized into four types, i.e., the *agential-durational scale*, the *locational-durational scale*, the *agential-momental scale*, and the *locational-momental scale*. In this study, we use term "hazard magnitude" to refer to an agential-durational hazard strength of an event." (L79-101)

Comment: *2.2. (ii) Section 3 could be written considerably more succinctly. And, is Table 3 really need to understand the paper's main point?*

Response: Thank you very much for your comment. Also having considered your later comments, we have made a significant modification to Section 3 to make it more succinct. Table 3 has also been moved to Appendix A. Although the previous Tables 3 and 4 (now Tables A1 and A2) are not necessary for understanding the paper's main point, they are important in providing information for reproduction of the results of this study.

The modified Section 3 now reads:

"To quantify hazard strength in terms of equivalent hazard magnitude, we considered 12 hazard types: cold wave, convective storm, drought, earthquake, extra-tropical storm, flash flood, forest fire, heat wave, riverine flood, tornado, tropical cyclone, and tsunami. A general standardized metric of impact was created by combining three loss measures from the EM-DAT database (Guha-Sapir et al., 2021): fatality, total affected population, and total damage. The impact metric was then related to an indicator of hazard strength, such as the Richter magnitude, for each hazard type via linear regression. The expectation of impact metric for each hazard type was linearly scaled and adopted as the equivalent hazard magnitude. Here, two assumptions were made. First, we assumed that the EM-DAT records were not significantly biased across similar hazard events. Second, we assumed that the derivation of expectation of impact metric cancelled out all local factors of exposed value and vulnerability. The following sections outline the method in detail.

3.1 Data Collection

To reduce the biases in model calibration due to different protocols for data collection across different types of natural hazards, we only used data gathered from the EM-DAT database (Guha-Sapir et al., 2021). To be included in the EM-DAT database, a hazard event must meet at least one of three criteria, i.e., 10 or more human fatalities, 100 or more people affected by the event, or a declaration of a state of emergency or an appeal for international assistance by a country (Guha-Sapir et al., 2021). For this study, we downloaded the entire EM-DAT datasets on all types of natural hazards. However, due to a lack of records of hazard magnitude indicators of events for some hazard types (e.g., the volcanic activities and landslides), we only included 12 hazard types. The final dataset for deriving the equivalent hazard magnitudes contained a total of 3 844 data points, each representing one unique hazard event.

The 12 considered hazard types, with their corresponding hazard magnitude indicators listed in parentheses, include: 1) cold wave (minimum temperature in °C); 2) convective storm (peak gust wind speed in km h^{-1}); 3) drought (total affected area in km^2); 4) earthquake (Richter magnitude); 5) extra-tropical storm (peak gust wind speed in km h^{-1}); 6) flash flood (total flooded area in km^2); 7) forest fire (total burnt area in km^2); 8) heat wave (maximum temperature in °C); 9) riverine flood (total flooded area in km^2); 10) tornado (peak gust wind speed in km h^{-1}); 11) tropical cyclone (maximum sustained wind speed in km h^{-1}); and 12) tsunami (earthquake Richter magnitude). For data quality control, we removed data points with questionable values of hazard magnitude indicators from our datasets. 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.

To facilitate regression modelling, we logarithmically transformed values of hazard magnitude indicators to be close to a Gaussian distribution within the range $(-\infty, \infty)$ for eight of the hazard types. The indicators that were not logarithmically transformed included minimum temperature of cold waves, Richter magnitude of earthquakes, maximum temperature of heat waves, and earthquake Richter magnitude of tsunami. Cold wave and heat wave events were excluded from logarithmic transformations because Celsius temperature has a range $[-273.15, \infty)$ similar to $(-\infty, \infty)$. Meanwhile, the range of an earthquake Richter magnitude is already a desired $(-\infty, \infty)$.

3.2 Impact Metric

We designed the impact metric as the principal component (Jolliffe, 2002; Jolliffe and Cadima, 2016) of three logarithmically transformed and standardized impact variables. The selected impact variables represent three major impact dimensions as defined by the EM-DAT database (Guha-Sapir et al., 2021). The first variable, fatality, indicates the number of people who perished as the result of a hazard event. The second variable, total affected population, refers to the total number of individuals injured, made homeless, or were affected by the event. The third variable, total damage, indicates the total amount of damage to property, crops, and livestock in 2019 USD caused by the event. The values of the impact variables were logarithmically transformed to be within the range $(-\infty, \infty)$ and standardized with the formula

$$IV = \frac{\ln(IVO) - \mu_{\ln IV}}{\sigma_{\ln IV}}, \quad (1)$$

where IV denotes the logarithmically transformed and standardized impact variable, IVO is the original impact variable, $\mu_{\ln IV}$ and $\sigma_{\ln IV}$ are respectively the mean and standard deviation of the logarithmically transformed impact variable (see Table 1). The principal component of the three logarithmically transformed and standardized impact variables corresponds to the dimension along which the variation of data points is preserved to the largest extent in the three-dimensional vector space. The principal component also shows the direction of the eigenvector associated with the largest eigenvalue with respect to the covariance matrix of the three transformed impact variables. Each data point represents the impact of one hazard event experienced by one country (see supplementary material Video S1).

Table 1: Means and standard deviations of original and logarithmically transformed impact variables[a].

Variable	Unit	Original mean	Original standard deviation	Logarithmically transformed mean	Logarithmically transformed standard deviation
Fatality	People	1.31×10^{3}	1.18×10^{4}	3.3892	2.1999
Total affected population	People	1.38×10^{6}	9.47×10^{6}	10.4116	3.1618
Total damage	1 thousand 2019 USD	1.36×10^{6}	8.45×10^{6}	11.1889	2.6304

[a]This table corresponds to supplementary material Data S1.

To reduce the bias associated with factors of exposed value and vulnerability (Fig. 1), we included all available data points at the country–year level for countries around the world and hazard events from 1900 to 2020. To compute the impact metric, we only kept data points (n = 1 470) without any missing values. A PCA was then conducted to determine the weights of transformed and standardized impact variables within the impact metric. The resulting formula for the impact metric is

$$IM = 0.6158IV_{\mathrm{F}} + 0.6215IV_{\mathrm{TA}} + 0.4843IV_{\mathrm{TD}}, \quad (2)$$

where IM denotes the impact metric and IV_{F}, IV_{TA}, and IV_{TD} refer to the transformed and standardized impact variables of fatality, total affected population, and total damage respectively.

3.3 Equivalent Magnitude

For each considered hazard type, we established the relationship between its hazard magnitude indicator and hazard impact metric via linear regression

$$IM = a_3 + b_3 MI + \sigma_3 \varepsilon \,, \quad (3)$$

where a_3 and b_3 are two model coefficients, MI denotes hazard magnitude indicator, σ_3 is the dispersion parameter, and ε is a standard normal random variable. The statistics of parameters of these regression models are listed in Table 2. Parameters of all linear regression models involved in this study were determined with a maximum likelihood approach based on Raphson's algorithm (Raphson, 1697; Wang et al., 2019; Wang, 2020). For each regression model, the standard errors of parameter estimates were derived from the main diagonal of the covariance matrix of model parameters computed as the negative inverse of the observed Fisher information matrix. To present equivalent hazard magnitude roughly within the range of $[0, 10]$, we applied a linear transformation to the point estimate of impact metric

$$EM = \hat{E}(IM) \times 2 + 5 \,, \quad (4)$$

where EM refers to the equivalent hazard magnitude and $\hat{E}(\cdot)$ denotes the point estimate of expectation. The derived equivalent hazard magnitudes for all data points are recorded in supplementary material Data S6." (L103-178)

Comment: *2.3. (iii) Section 3.3 might be best in an Appendix to preserve the flow of the paper.*

Response: Thank you very much for your comment. We have modified Section 3.3 and moved it to Appendix A.

The new Appendix A now reads: "

Appendix A: Missing Values and Data Aggregation

Six simple linear regression models and three multiple linear regression models with two independent variables were calibrated with the same data points for derivation of the impact metric. These regression models were created to fill in missing values of impact variables for data points with at most two empty entries among the three impact variables. Within each of these nine linear regression models, the dependent variable is one of the three impact variables. For each of the six simple linear regression models, the independent variable is one of the two impact variables that are not used as the dependent variable. The simple linear regression models have the form

$$IV_1 = a_1 + b_1 IV_2 + \sigma_1 \varepsilon \,, \quad (A1)$$

where $a_1 = 0$ and b_1 are two model coefficients, IV_1 and IV_2 are two considered transformed and standardized impact variables, and σ_1 is the dispersion parameter. The statistics of parameters of these simple linear regression models are shown in Table A1. Per the three multiple linear regression models with two independent variables, the independent variables are the two impact

variables other than the one used as the dependent variable. The formula for the multiple linear regression models is

$$IV_1 = a_2 + b_2 IV_2 + c_2 IV_3 + \sigma_2 \varepsilon \,, \quad \text{(A2)}$$

where $a_2 = 0$, b_2, and c_2 are three model coefficients, IV_3 is the third transformed and standardized impact variable, and σ_2 is the dispersion parameter. Table A2 lists the statistics of parameters of the multiple linear regression models with two independent variables. The missing values of data points were filled with the expectations regressed on the independent variables with available data. The data were then aggregated event-wise to form data points of the dataset for deriving the equivalent hazard magnitudes.

Table A1: Statistics of parameters of six simple linear regression models for filling in missing values of impact variables[a].

Model number	Dependent variable	Independent variable	b_1	σ_1
I1	Fatality	Total affected population	0.5096 (0.0224)	0.8604 (0.0159)
I2	Fatality	Total damage	0.2802 (0.0250)	0.9599 (0.0177)
I3[b]	Total affected population	Fatality	0.5096 (0.0224)	0.8604 (0.0159)
I4	Total affected population	Total damage	0.2948 (0.0249)	0.9556 (0.0176)
I5[c]	Total damage	Fatality	0.2802 (0.0250)	0.9599 (0.0177)
I6[d]	Total damage	Total affected population	0.2948 (0.0249)	0.9556 (0.0176)

[a]This table corresponds to supplementary material Data S2; R-squared measures are included in Fig. 2; standard errors are in the parentheses; estimations of b_1 and σ_1 are all significant at $p < 10^{-20}$.
[b]Models I1 and I3 share the same model parameters and R-squared measures.
[c]Models I2 and I5 share the same model parameters and R-squared measures.
[d]Models I4 and I6 share the same model parameters and R-squared measures.

Table A2: Statistics of parameters of three multiple linear regression models with two independent variables for filling in missing values of impact variables[a].

Model number	Dependent variable	Independent variable 1	Independent variable 2	b_2	c_2	σ_2
I7	Fatality	Total affected population	Total damage	0.4676 (0.0232)	0.1423 (0.0232)	0.8496 (0.0157)
I8	Total affected population	Fatality	Total damage	0.4633 (0.0230)	0.1650 (0.0230)	0.8457 (0.0156)
I9	Total damage	Fatality	Total affected population	0.1755 (0.0286)	0.2054 (0.0286)	0.9435 (0.0174)

[a]This table corresponds to supplementary material Data S3; R-squared measures are included in Fig. 2; standard errors are in the parentheses; estimations of b_2, c_2, and σ_2 are all significant at $p < 10^{-8}$." (L412-441)

Comment: *3. Introduction and framing: this work does something new I think, but the way it is presented does not help this argument.*

Response: Thank you very much for your comment. We have reframed the Introduction section and Section 2, also with consideration of your later comments, to improve the introduction and framing of our research work.

The modified Introduction section now reads:

"Natural hazards pose significant challenges to human societies around the world. Between 2000 and 2020, natural hazard events caused over 130 billion dollars in losses and 64 695 fatalities, and affected more than 196 million people, on average each year (Guha-Sapir et al., 2021). Hazardous events, such as earthquakes, floods, and forest fires, can inflict heavy losses to communities when people and property are exposed to the natural forces of these events. The impacts of events, whatever their type, can be quantified directly (e.g., by financial loss; Hillier et al., 2015), or estimated on a scale. To estimate the impacts of an event with the consideration of its hazard strength, various impact scales have been proposed, including the Bradford disaster scale (Keller et al., 1992; 1997), unified localizable crisis scale (Rohn and Blackmore, 2009; 2015), disaster impact index (Gardoni and Murphy, 2010), and cascading disaster magnitude (Alexander, 2018). However, a hazard strength scale is not the same as a hazard impact scale, as impacts are also driven by the exposure and vulnerability of entities, such as individuals, communities, and infrastructures, to an event. This makes it difficult to use impact scales to compare hazard strengths across natural hazard types. For example, the 2011 Christchurch earthquake was one of the most destructive earthquakes in New Zealand, albeit with a medium hazard strength of 6.2 in terms of its moment magnitude (Kaiser et al., 2012). Meanwhile, the 1964 Alaskan earthquake, with a larger moment magnitude of 9.2, resulted in fewer casualties and less economic damage than the Christchurch earthquake (United States Geological Survey [USGS], 2021).

In the meantime, hazard scientists have long called for separation of natural forcing phenomena (Bensi et al., 2020) from the study of disasters to better understand the causes of impacts rooted in the social and economic fabric of entities exposed to natural hazards (e.g., O'Keefe et al., 1976; Wisner et al., 2004). In this regard, quantifying hazard strength helps separate the natural force from other social, environmental, and engineering or built environmental factors that may drive impacts. Yet, despite the large volume of research that focuses on hazard strength for singular natural hazard types such as earthquake (e.g., Wood and Neumann, 1931; Richter, 1935; Kanamori, 1977; Katsumata, 1996; Grünthal, 1998; Wald et al., 2006; Rautian et al., 2007; Serva et al., 2016), tropical cyclone (e.g., Simpson and Saffir, 1974; Bell et al., 2000; Emanuel, 2005; Powell and Reinhold, 2007; Hebert et al., 2008), tornado (e.g., Fujita, 1971; 1981; Meaden et al., 2007; Potter, 2007; Dotzek, 2009), and drought (e.g., Palmer, 1965; 1968; Shafer and Dezman, 1982; McKee et al., 1993; Byun and Wilhite, 1999; Shukla and Wood, 2008; Hunt et al., 2009), few have quantified or modelled hazard strengths across multiple hazard types.

To quantify hazard strengths for cross-hazard comparison, impacts can be used to explore dependencies between multiple hazards (e.g., Hillier et al., 2015; Hillier and Dixon, 2020). As an example, insurance professionals often leverage loss metrics to understand the relative significance of various hazards (see, e.g., Mitchell-Wallace et al., 2017). However, their cross-hazard practices of risk aggregation and accumulation are often focused on the exposed values and observed impacts, rather than hazard strengths. In contrast, risk quantification for nuclear facilities requires consideration of hazard strengths across multiple hazard types to facilitate probabilistic safety assessment within a multi-hazard context (see, e.g., Choi et al., 2021). Indices regarding hazard strengths for multiple hazard types have also been created and adopted for extreme meteorological events (see, e.g., Malherbe et al., 2020). When quantifying hazard strengths within a multi-hazard context, a calibration of hazard strength to the expectation of impact may be used to create impact-based proxies for hazard strengths, linking two extremes and allowing them to be studied in a way that is relevant to risk assessment and yet decoupled from the detail of exposed values and

vulnerability (Hillier et al., 2020). Nevertheless, there is not yet a general metric that permits events of different hazard types to be compared in terms of potential to cause damage in a way that is as decoupled as possible from exposed values and vulnerability.

To enable evaluation of event-wise hazard strengths across different hazard types, in this article, we propose a multi-hazard *equivalent hazard magnitude scale* – the *Gardoni Scale* – for natural hazards. The proposed scale is named after the Alfredo H. Ang Family Professor Paolo Gardoni at the University of Illinois at Urbana–Champaign. Because hazard strength is correlated with hazard impacts given average exposed value and vulnerability of considered entities, the expectation of a metric of observed impacts of hazard events can be used to calibrate models for deriving equivalent hazard magnitudes (Hillier et al., 2015; Hillier and Dixon, 2020; Wang and Sebastian, 2021b). In this article, a quantitative modelling methodology based on a principal component analysis (PCA) and a set of linear regressions is developed to construct the impact metric and derive equivalent hazard magnitudes on the Gardoni Scale. The impact metric is a function of three impact variables, i.e., fatality, total affected population, and total damage in 2019 USD. We use historical event data from the EM-DAT International Disaster Database (Guha-Sapir et al., 2021) from 1900 to 2020 to calibrate the quantitative models. To demonstrate the value of the proposed scale, we apply it to discuss the equivalent magnitudes of historical and recent hazard events.

The subsequent sections are organized as follows. First, we provide a brief theoretical background for this study. We then introduce our methodology, including data processing, to derive the equivalent hazard magnitude on the Gardoni Scale. Next, we lay out the results of applying our methodology and compare natural hazard types regarding the derived equivalent hazard magnitudes. Finally, we discuss the potential contributions and limitations of the proposed scale before concluding the article." (L26-77)

Comment: *L17 – 'we argue' instead of 'we show', you are suggesting something, not providing a definitive and unique answer.*

Response: Thank you very much for your comment. To better present what we wish to convey here, we have changed "we show" into "we compute".

The corresponding sentence now reads: "For example, we compute that the hazard magnitude of the February 2021 North American cold wave event affecting the southern states of the United States of America was equivalent to the hazard magnitude of Hurricane Harvey in 2017 or a magnitude 7.5 earthquake." (L20-22)

Comment: *L25 – Use 'hazardous events' rather than 'hazard event'.*

Response: Thank you very much for your comment. We have modified "hazard events" to "hazardous events" accordingly.

The modified sentence now reads:

"Hazardous events, such as earthquakes, floods, and forest fires, can inflict heavy losses to communities when people and property are exposed to the natural forces of these events." (L28-29)

Comment: *L25 – Suggest delete 'with a strong natural force' – example of words that are vague and as such add little meaning and detract from the focus of the text. I illustrate in the next comment.*

Response: Thank you very much for your comment. We have correspondingly deleted 'with a strong natural force'.

The modified sentence now reads:

"Hazardous events, such as earthquakes, floods, and forest fires, can inflict heavy losses to communities when people and property are exposed to the natural forces of these events." (L28-29)

Comment: *L27 – "….. these events. The impact of events, whatever their type, can be quantified directly (e.g. by financial loss)(Hillier et al, 2015). Various impact scales have also been proposed including the Bradford ……" – I would just name 1 or 2 scales and put the references at the end of the sentence.*

Response: Thank you very much for your comment. We have modified the sentences accordingly but have chosen to keep the references to all four scales. In particular, we have added a sentence "The impacts of events, whatever their type, can be quantified directly (e.g., by financial loss; Hillier et al., 2015), or estimated on a scale" (L29-30), as suggested.

These modified sentences now read:

"Hazardous events, such as earthquakes, floods, and forest fires, can inflict heavy losses to communities when people and property are exposed to the natural forces of these events. The impacts of events, whatever their type, can be quantified directly (e.g., by financial loss; Hillier et al., 2015), or estimated on a scale. To estimate the impacts of an event with the consideration of its hazard strength, various impact scales have been proposed, including the Bradford disaster scale (Keller et al., 1992; 1997), unified localizable crisis scale (Rohn and Blackmore, 2009; 2015), disaster impact index (Gardoni and Murphy, 2010), and cascading disaster magnitude (Alexander, 2018)." (L28-33)

Comment: *L30-38 – Consider using examples to communicate more clearly e.g. the Christchurch quake in New Zealand is an example of a small quake causing lots of damage.*

Response: Thank you very much for your comment. We have incorporated your suggestion on providing examples in the revised version of the manuscript.

The modified sentences now read:

"However, a hazard strength scale is not the same as a hazard impact scale, as impacts are also driven by the exposure and vulnerability of entities, such as individuals, communities, and infrastructures, to an event. This makes it difficult to use impact scales to compare hazard strengths across natural hazard types. For example, the 2011 Christchurch earthquake was one of the most destructive earthquakes in New Zealand, albeit with a medium hazard strength of 6.2 in terms of its moment magnitude (Kaiser et al., 2012). Meanwhile, the 1964 Alaskan earthquake, with a larger moment magnitude of 9.2, resulted in fewer casualties and less economic damage than the Christchurch earthquake (United States Geological Survey [USGS], 2021)." (L33-39)

Comment: *L53-61 – This paragraph finishing the framing of the work needs re-writing. My first point is observation, and my second is a suggestion.*

Response: Thank you very much for your comment. Also having considered your following comments and suggestions, we have added a new paragraph to enhance the framing of the presented research.

The new paragraph reads:

"To quantify hazard strengths for cross-hazard comparison, impacts can be used to explore dependencies between multiple hazards (e.g., Hillier et al., 2015; Hillier and Dixon, 2020). As an example, insurance professionals often leverage loss metrics to understand the relative significance of various hazards (see, e.g., Mitchell-Wallace et al., 2017). However, their cross-hazard practices of risk aggregation and accumulation are often focused on the exposed values and observed impacts, rather than hazard strengths. In contrast, risk quantification for nuclear facilities requires consideration of hazard strengths across multiple hazard types to facilitate probabilistic safety assessment within a multi-hazard context (see, e.g., Choi et al., 2021). Indices regarding hazard strengths for multiple hazard types have also been created and adopted for extreme meteorological events (see, e.g., Malherbe et al., 2020). When quantifying hazard strengths within a multi-hazard context, a calibration of hazard strength to the expectation of impact may be used to create impact-based proxies for hazard strengths, linking two extremes and allowing them to be studied in a way that is relevant to risk assessment and yet decoupled from the detail of exposed values and vulnerability (Hillier et al., 2020). Nevertheless, there is not yet a general metric that permits events of different hazard types to be compared in terms of potential to cause damage in a way that is as decoupled as possible from exposed values and vulnerability." (L50-62)

Comment: *I didn't use the Gardoni scale in Hillier et al (2015, 2020a). Indeed, how could I have as it is proposed here. In 2015 & 2020a I used financial impact as a metric to allow comparison of multiple hazards and their severity (4 and 7 hazard respectively). In Hillier et al (2020b), I use what I refer to as 'impact-based proxies' for hazard to map and understand the estimated combined severity of two hazards (extreme wind and flooding).*

Response: Thank you very much for your comment. We agree that the "impact-based proxies" are a brilliant idea.

Comment: *The work proposed here certainly builds on the limited (i.e. two hazard) work in Hillier (2020b), which itself builds on a substantial history of what I dubbed 'impact-based proxies' (i.e. hazard measures designed to – hopefully – closely relate to impacts) e.g. v3 over a threshld is very established for wind (e.g. refs [33-38] in Hillier 2020b – Southern (1979), Klawa (2003)). So, I suggest starting the paragraph with this context (and likely references for other hazards) building to the necessity of a generalized Equivalent Hazard Scale – perhaps with a structure similar to the bullets below.*

Response: Thank you very much for your comment. We have referenced your suggestions and added a paragraph to improve the framing of our presented work, as shown in our response prior to the previous one.

Comment: *Impacts (e.g. financial losses) have directly used to compare and understand dependencies between multiple (up to 4 or 7) hazards (e.g. Hillier et al 2015, 2020b), but strictly this limits understanding to a particular stakeholder (e.g. insurers, the UK rail network). Indeed, insurers are very experienced at using loss as a metric to understand the relative significance of various hazards [see detail below].*

Response: Thank you very much for your comment. As shown in one of our responses previously, we have added a paragraph to improve the framing of our presented work. In particular, we now highlight the experiences of insurers in leveraging loss as a common metric for understanding risks.

Comment: *Similar about nuclear sector, perhaps mentioning scenarios [I know this exists, but don't have details to hand].*

Response: Thank you very much for your comment. We have also added some content in the new paragraph on the nuclear sector based on some outstanding recent research work on multi-hazard risk assessment for nuclear power plants.

Comment: *There are also indices that integrate multiple weather extremes, but …… [again see below].*

Response: Thank you very much for your comment. We have also included the material on the weather extremes in the new paragraph shown previously.

Comment: *A calibration of hazard to impact has been used to create 'impact-based proxies' for hazard, linking two extremes and allowing them to be studied in a way that is relevant to risk and yet decoupled from the detail of local human exposure (Hillier, 2020a).*

Response: Thank you very much for your comment. This suggestion has been adopted and incorporated into the new paragraph displayed previously.

Comment: *But, there is not as yet a general multi-hazard measure that permits events (e.g. a cat 5 hurricane and a M_w 6.7 earthquake) to be compared in terms of potential to cause damage (i.e. hazard) in a way that is as decoupled as possible from local human exposure (i.e. assets at risk). And, Hillier (2020a) do not create a scale for ease of comparison. We propose ……..*

Response: Thank you very much for your comment. We have also adopted this suggestion to develop the new paragraph shown previously.

Comment: *Indices of Climate Change for the United States – Karl (1996) Bull Am Met Soc. "The Extreme Climate Index (ECI) is an objective, multi-hazard index ….. of extreme weather events" Malherbe, J. et al. 2018. The Extreme Climate Index (ECI), a tool for monitoring regional extreme events. In: Climate Change and Adaptive Land Management in southern Africa: Assessments, Changes, Challenges, and Solutions, pp. 144-145*

Response: Thank you very much for your comment. The material on meteorological extremes has been added to the new paragraph shown previously.

Comment: *The need to combine risks (between geographic regions and types of risk) has a greater history than currently acknowledged. 'Accumulation', 'roll-up' or 'aggregation' e.g. see Ch 2.7 of Mitchell-Wallace 'Natural Catastrophe Risk Management and Modelling' for an introduction to this subject (p97-105), and how it has been handled for decades (if not centuries) in the provision of insurance. Very well established commercial products have existed for at least 13 years (e.g. Remetrica/Igloo) i.e. this is my personal memory only from when I first saw then embedded within insurers.*

Response: Thank you very much for your comment. We have integrated material on insurance practices into the new paragraph shown previously. In our new paragraph, we also emphasize that "As an example, insurance professionals often leverage loss metrics to understand the relative significance of various hazards (see, e.g., Mitchell-Wallace et al., 2017). However, their cross-hazard practices of risk aggregation and accumulation are often focused on the exposed values and observed impacts, rather than hazard strengths." (L51-54)

Comment: *L56 – Gardoni (2014) is very explicitly a risk scale, not a hazard scale as proposed here. Please use only references that are directly relevant.*

Response: Thank you very much for your comment. We have removed the unnecessary references.

The modified sentence now reads: "The proposed scale is named after the Alfredo H. Ang Family Professor Paolo Gardoni at the University of Illinois at Urbana–Champaign." (L64-65)

Comment: *L58 – This manuscript should not depend upon Wang & Sebastian (2021b), so please remove as this is still under review.*

Response: Thank you very much for your comment. We have replaced the submitted manuscript under review with a presentation at the 2021 EGU General Assembly available at https://doi.org/10.5194/egusphere-egu21-6468.

Comment: *L124 – (i) consider splitting section 3.1 into 'Data' and 'Magnitude Indicator'*

Response: Thank you very much for your comment. We have made modifications to this section and reduced its content. Since the second and third paragraphs of this section are still about data description, we have kept them within the same section. We have also changed the heading of this section into "Data Collection".

The modified two sections now read: "

3.1 Data Collection

To reduce the biases in model calibration due to different protocols for data collection across different types of natural hazards, we only used data gathered from the EM-DAT database (Guha-Sapir et al., 2021). To be included in the EM-DAT database, a hazard event must meet at least one of three criteria, i.e., 10 or more human fatalities, 100 or more people affected by the event, or a declaration of a state of emergency or an appeal for international assistance by a country (Guha-Sapir et al., 2021). For this study, we downloaded the entire EM-DAT datasets on all types of natural hazards. However, due to a lack of records of hazard magnitude indicators of events for

some hazard types (e.g., the volcanic activities and landslides), we only included 12 hazard types. The final dataset for deriving the equivalent hazard magnitudes contained a total of 3 844 data points, each representing one unique hazard event.

The 12 considered hazard types, with their corresponding hazard magnitude indicators listed in parentheses, include: 1) cold wave (minimum temperature in °C); 2) convective storm (peak gust wind speed in km h^{-1}); 3) drought (total affected area in km^2); 4) earthquake (Richter magnitude); 5) extra-tropical storm (peak gust wind speed in km h^{-1}); 6) flash flood (total flooded area in km^2); 7) forest fire (total burnt area in km^2); 8) heat wave (maximum temperature in °C); 9) riverine flood (total flooded area in km^2); 10) tornado (peak gust wind speed in km h^{-1}); 11) tropical cyclone (maximum sustained wind speed in km h^{-1}); and 12) tsunami (earthquake Richter magnitude). For data quality control, we removed data points with questionable values of hazard magnitude indicators from our datasets. 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.

To facilitate regression modelling, we logarithmically transformed values of hazard magnitude indicators to be close to a Gaussian distribution within the range $(-\infty, \infty)$ for eight of the hazard types. The indicators that were not logarithmically transformed included minimum temperature of cold waves, Richter magnitude of earthquakes, maximum temperature of heat waves, and earthquake Richter magnitude of tsunami. Cold wave and heat wave events were excluded from logarithmic transformations because Celsius temperature has a range $[-273.15, \infty)$ similar to $(-\infty, \infty)$. Meanwhile, the range of an earthquake Richter magnitude is already a desired $(-\infty, \infty)$." (L112-137)

Comment: *L124 – (ii) A few sentences before section 3.1 explaining the overall structure of the Methods would help, similar to my second paragraph in this review.*

Response: Thank you very much for your comment. We have adopted your suggestion and added a paragraph before Section 3.1.

The new paragraph before Section 3.1 reads: "To quantify hazard strength in terms of equivalent hazard magnitude, we considered 12 hazard types: cold wave, convective storm, drought, earthquake, extra-tropical storm, flash flood, forest fire, heat wave, riverine flood, tornado, tropical cyclone, and tsunami. A general standardized metric of impact was created by combining three loss measures from the EM-DAT database (Guha-Sapir et al., 2021): fatality, total affected population, and total damage. The impact metric was then related to an indicator of hazard strength, such as the Richter magnitude, for each hazard type via linear regression. The expectation of impact metric for each hazard type was linearly scaled and adopted as the equivalent hazard magnitude. Here, two assumptions were made. First, we assumed that the EM-DAT records were not significantly biased across similar hazard events. Second, we assumed that the derivation of expectation of impact metric cancelled out all local factors of exposed value and vulnerability. The following sections outline the method in detail." (L103-111)

Comment: *L127 – Are you sure there are no biases (e.g. omissions) in EM-DAT?*

Response: Thank you very much for your question. We do not deny that there could be biases in EM-DAT datasets due to omissions of events. However, the main reason why we only used data from EM-DAT is that we tried to reduce the biases due to different protocols for data collection by different databases.

As mentioned in the modified Data Collection section, "To reduce the biases in model calibration due to different protocols for data collection across different types of natural hazards, we only used data gathered from the EM-DAT database (Guha-Sapir et al., 2021)." (L113-114)

Comment: *L132 – Which did you keep for each hazard, and why? Please justify choices, providing appropriate references.*

Response: Thank you very much for your comment. The EM-DAT database only includes one magnitude indicator for each natural hazard type. Therefore, we could only select at most one hazard magnitude indicator for each hazard type. The magnitude indicators we kept are, hence, the ones listed in the Data Collection section. To avoid confusion, we have modified the Data section.

The corresponding sentence in the Data Collection section now reads: ". However, due to a lack of records of hazard magnitude indicators of events for some hazard types (e.g., the volcanic activities and landslides), we only included 12 hazard types." (L117-119)

Comment: *L135&L138 – What duration of gust? (e.g. 3 sec or 10 sec, and at what height). These are important distinctions e.g. for tropical cyclones the recording method and therefore apparent severity differ between the USA and Japan.*

Response: Thank you very much for your question and comment. Because the EM-DAT database does not provide the details of peak gust wind speeds or peak sustained wind speeds such as seconds and at what height, we were only able to use "peak gust wind speed" and "peak sustained wind speed" to refer to the magnitude indicators for wind-related hazards. We recognize that there is some uncertainty in the data underlying the paper due to the record keeping by EM-DAT and have made a note of our assumptions in the text. For example, we have highlighted the issues with the magnitude indicators of earthquake, flood, and tropical cyclone on L370-380.

These sentences on these issues read: "For example, both wind and precipitation contribute significantly to damages associated with tropical cyclone events (Mudd et al., 2017). Moreover, selection of hazard magnitude indicators in this study was also limited by the adopted datasets. As an example, the earthquake Richter magnitude (Richter, 1935) was the only recorded hazard magnitude indicator in the datasets of this study. However, because Richter magnitude is easily subject to saturation for large earthquakes, it has become less often referenced than moment magnitude (Kanamori, 1977) for indicating hazard magnitude of an earthquake event. For flood hazards, as another example, there is a lack of established methods to quantify the agential-durational hazard strength metrics. In this study, we followed the EM-DAT database (Guha-Sapir et al., 2021) to use the flooded area as the hazard magnitude indicator for the flood hazards. However, the definition of such flooded area is still vague and deserves more research. An ideal agential-durational hazard strength metric for a flood event should integrate multiple flood intensity measures, such as water depth, flood volume, and flow velocity, over the entire flooded

area and duration of the event to correspond to the total energy released by the natural force of the event." (L370-380)

Comment: *L140-L144 – Please justify the thresholds used (e.g. Richter magnitude >= 6).*

Response: Thank you very much for your comment. These thresholds were set for data quality control to avoid data points with unrealistic values that may have been produced due to human errors. To clarify this point, we have modified the corresponding sentence in the second paragraph of the Data Collection section.

The modified second paragraph of the Data Collection section now reads: "The 12 considered hazard types, with their corresponding hazard magnitude indicators listed in parentheses, include: 1) cold wave (minimum temperature in °C); 2) convective storm (peak gust wind speed in km h^{-1}); 3) drought (total affected area in km^2); 4) earthquake (Richter magnitude); 5) extra-tropical storm (peak gust wind speed in km h^{-1}); 6) flash flood (total flooded area in km^2); 7) forest fire (total burnt area in km^2); 8) heat wave (maximum temperature in °C); 9) riverine flood (total flooded area in km^2); 10) tornado (peak gust wind speed in km h^{-1}); 11) tropical cyclone (maximum sustained wind speed in km h^{-1}); and 12) tsunami (earthquake Richter magnitude). For data quality control, we removed data points with questionable values of hazard magnitude indicators from our datasets. 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." (L121-131)

Comment: *L146 – Sentence does not make sense. No transformation is needed to fit losses in the range ±infinity. Is the purpose to centre the impact metric on zero?*

Response: Thank you very much for your comment and question. The purpose of logarithmic transformation is to convert hazard magnitude indicators to be close to a Gaussian distribution and to have a range of $(-\infty, \infty)$ to facilitate regression modeling. Using a transformed magnitude indicator can be more representative across its entire range. To clarify this point, we have modified the last paragraph of the Data Collection section.

The modified last paragraph of the Data Collection section now reads: "To facilitate regression modelling, we logarithmically transformed values of hazard magnitude indicators to be close to a Gaussian distribution within the range $(-\infty, \infty)$ for eight of the hazard types. The indicators that were not logarithmically transformed included minimum temperature of cold waves, Richter magnitude of earthquakes, maximum temperature of heat waves, and earthquake Richter magnitude of tsunami. Cold wave and heat wave events were excluded from logarithmic transformations because Celsius temperature has a range $[-273.15, \infty)$ similar to $(-\infty, \infty)$. Meanwhile, the range of an earthquake Richter magnitude is already a desired $(-\infty, \infty)$." (L132-137)

Comment: *L145-150 – Please add rationale (i.e. systematic logic for when transformation was needed and when it wasn't).*

Response: Thank you very much for your comment. When the range of a hazard magnitude indicator is or is close to $(-\infty, \infty)$, it does not need transformation. To clarify this point, we have modified the last paragraph of the Data Collection section as shown in our previous response.

Comment: *L198 – by 'by applying' I assume you mean a simulation of individual values, rather than using an expectation from the trend line. Using an expectation would not replicate the variability of the data. Please clarify.*

Response: Thank you very much for your comment. We actually used the expectations of regression models to fill in the missing values of impact variables. However, this is not a problem for our purpose because the regression models that really matter are the ones showing relationship between hazard magnitude indicator and the impact metric. These regression models for missing values only served to project data points with missing values onto the axis of impact metric. The variation in the impact metric can still be determined by the impact variables without missing values. To clarify that we used the expectation, we have modified the previous Section 3.3 into Appendix A.

The new Appendix A now reads: "Six simple linear regression models and three multiple linear regression models with two independent variables were calibrated with the same data points for derivation of the impact metric. These regression models were created to fill in missing values of impact variables for data points with at most two empty entries among the three impact variables. Within each of these nine linear regression models, the dependent variable is one of the three impact variables. For each of the six simple linear regression models, the independent variable is one of the two impact variables that are not used as the dependent variable. The simple linear regression models have the form

$$IV_1 = a_1 + b_1 IV_2 + \sigma_1 \varepsilon \,, \quad \text{(A1)}$$

where $a_1 = 0$ and b_1 are two model coefficients, IV_1 and IV_2 are two considered transformed and standardized impact variables, and σ_1 is the dispersion parameter. The statistics of parameters of these simple linear regression models are shown in Table A1. Per the three multiple linear regression models with two independent variables, the independent variables are the two impact variables other than the one used as the dependent variable. The formula for the multiple linear regression models is

$$IV_1 = a_2 + b_2 IV_2 + c_2 IV_3 + \sigma_2 \varepsilon \,, \quad \text{(A2)}$$

where $a_2 = 0$, b_2, and c_2 are three model coefficients, IV_3 is the third transformed and standardized impact variable, and σ_2 is the dispersion parameter. Table A2 lists the statistics of parameters of the multiple linear regression models with two independent variables. The missing values of data points were filled with the expectations regressed on the independent variables with available data. The data were then aggregated event-wise to form data points of the dataset for deriving the equivalent hazard magnitudes." (L413-430)

Comment: *L212 – Section 3.4 & Table 4. Whilst significance of individual parameters is interesting, please compute and provide p values for the models as a whole, and consider omitting any hazards where the statistical model is not significant.*

Response: Thank you very much for your comment. The objective of this manuscript is to propose the Gardoni Scale and to demonstrate how to quantitatively derive the equivalent hazard magnitude on the Gardoni Scale. The purpose of the study is not to reveal the relationship between a hazard magnitude indicator and impact metric, nor is it to predict impact metric with a hazard magnitude indicator. Therefore, we provided the point estimates of section 3.4 and previously Table 4 for reproduction of research results. We chose to include the standard errors and p values to demonstrate that some estimates of model coefficients were statistically significant while others were not. Therefore, we did not omit hazards where the statistical models were not significant.

Comment: *L270 – Fig. 3 – Are these relationships (i.e. R^2 values) all statistically significant? If not, please consider the validity of including them in the paper. Those omitted can simply be removed, helping brevity.*

Response: Thank you very much for your comment. Like mentioned in our previous response, the statistical relationships and R^2 values were presented to show that it is okay to have coefficients that are not statistically significant and to have a small R^2 because the purpose of the study is to present a way to compute the equivalent hazard magnitude on the Gardoni Scale. The regression models used in the study are merely tools for computation. Therefore, we have chosen not to omit the regression models that are statistically insignificant. In fact, the significant spread in the data lends insight to whether there could be underlying drivers of impacts such as vulnerability factors that are not considered in studies of hazard equivalency. In the meantime, because of the decoupling from exposure and vulnerability, the derivation of hazard equivalency can provide the benchmark measures of hazard strength to provide a fair foundation for the studies on the effects of those exposure and vulnerability factors across different hazard types.

Comment: *L435 – A fundamental limitation (but also benefit) of any impact-based measure of hazard is that it is specific to a user (i.e. the subject of the potential loss). The authors have endeavoured to define a widely relevant measure, but a brief discussion of the benefits and limitations of this specific is necessary.*

Response: Thank you very much for your comment. We have modified the third paragraph of Section 5.2 to incorporate your suggestion regarding the issue of impact metric being specific to a user.

The modified third paragraph of Section 5.2 now reads:

"In addition to hazard magnitude indicators, the construction of the impact metric is important for the calibration of regression models and for the derivation of equivalent hazard magnitudes as it is end-user specific. For example, insurance professionals may be interested in an equivalent hazard magnitude that is derived from data on financial and property loss whereas environmental scientists may be more interested in an impact metric based on ecological damage. Herein, we derived a general metric of impact for equivalent hazard magnitude based on key indicators of societal impact. For this reason, we combined data on fatalities, damages, and affected individuals to derive an impact metric. However, hazard events can affect a variety of sectors resulting in

impacts to physical, social, economic, and environmental well-being (Lindell and Prater, 2003; Gardoni and Murphy, 2010; Alexander, 2013; Wang et al., 2016; 2021). To advance methodological development for the proposed Gardoni Scale and quantification of other equivalent hazard strength metrics for various stakeholders, future work should scrutinize different indicators as impact variables of events and to seek the optimal models to combine impact variables to inform the level of impacts of events for different hazard types." (L383-393)

Anonymous Referee

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 changes to the manuscript are summarized as follows:

1) We have modified the abstract to focus on key points of the manuscript.

2) We have significantly revised the introduction section.

3) We have significantly reduced the length of section 2: A Problem of Scales.

4) We have modified the methodology section to make it more succinct and have moved the content and tables on missing values to the appendix.

5) We have modified the results section and the discussion section.

6) We have updated the references accordingly.

7) We have double-checked the event records discussed in the manuscript and have corrected two errors.

8) We have updated Fig. 1 and have modified the figure captions.

9) We have also slightly modified the conclusion section.

Comment: *The paper introduces a new magnitude scale (the Gardoni scale) to describe the impact of different types of natural events and to facilitate the comparison.*

Response: Thank you very much for your summary. However, the objective of the paper is not to describe the impact of events. Instead, we propose the Gardoni Scale to directly compare the hazard strengths of events across hazard types. The hazard strength of an event is not the same as the impact of the event, as the impact is associated with not only the hazard strength but also the values exposed to the hazard strength and the vulnerability of the exposed entity to impact. By using a large sample size of data with a good quality, we can assume that the factors of exposed value and vulnerability have been controlled for such that we can use the observed impact to calibrate a regression model to quantify the hazard strength as the expected impact given average exposed value and vulnerability. Once such regression models are established for each hazard type, they can be used to derive the equivalent hazard strengths for direct comparisons across hazard types.

To clarify that our objective is not to describe the impact of events, we have modified the first paragraph of our introduction section. The modified first paragraph of the introduction section now reads: “Natural hazards pose significant challenges to human societies around the world. Between 2000 and 2020, natural hazard events caused over 130 billion dollars in losses and 64 695 fatalities, and affected more than 196 million people, on average each year (Guha-Sapir et al., 2021). Hazardous events, such as earthquakes, floods, and forest fires, can inflict heavy losses to communities when people and property are exposed to the natural forces of these events. The impacts of events, whatever their type, can be quantified directly (e.g., by financial loss; Hillier et al., 2015), or estimated on a scale. To estimate the impacts of an event with the consideration of its hazard strength, various impact scales have been proposed, including the Bradford disaster scale (Keller et al., 1992; 1997), unified localizable crisis scale (Rohn and Blackmore, 2009; 2015),

disaster impact index (Gardoni and Murphy, 2010), and cascading disaster magnitude (Alexander, 2018). However, a hazard strength scale is not the same as a hazard impact scale, as impacts are also driven by the exposure and vulnerability of entities, such as individuals, communities, and infrastructures, to an event. This makes it difficult to use impact scales to compare hazard strengths across natural hazard types. For example, the 2011 Christchurch earthquake was one of the most destructive earthquakes in New Zealand, albeit with a medium hazard strength of 6.2 in terms of its moment magnitude (Kaiser et al., 2012). Meanwhile, the 1964 Alaskan earthquake, with a larger moment magnitude of 9.2, resulted in fewer casualties and less economic damage than the Christchurch earthquake (United States Geological Survey [USGS], 2021).” (L26-39)

Comment: *Although I do agree with the main idea of the paper, i.e., hazards cannot be compared but we can compare their effects, I have several doubts about this paper.*

Response: Thank you very much for your encouragement and comment. The comparison of hazards often involves the computation of the expected frequency of or the expected exceedance frequency of hazard strength for a given hazard type. In this paper, we aim to derive an estimate of hazard strength independent of hazard type. The main idea of the paper is that hazard strengths can be compared. First, hazard strengths can be compared within each hazard type. Agentially, for example, an M7 earthquake on the moment magnitude scale is larger than an M5 earthquake in terms of hazard strength, while the effects of the M5 earthquake can be much more severe than the effects of the M7 earthquake. Locationally, as another example, a community surrounded by a water depth of 2 meters is experiencing a much larger hazard strength of flood than another community facing a water depth of 0.5 meters. However, due to different pre-event mitigation efforts, the community with a 2-meter water depth may be impacted much less than the community with a 0.5-meter water depth. On the other hand, comparison of hazard strengths across different hazard types is difficult with the mainstream methodologies used in the existing multi-hazard studies. In light of this, the main academic contribution of this paper is that we propose a scale that enables cross-hazard comparison of hazard strengths in an agential and durational manner. We have also demonstrated how to compare the agential-durational hazard strengths across hazard types in this paper.

Comment: *I will describe below only the most important ones (omitting other minor points), with the hope that they can be of some usefulness for the authors.*

Response: Thank you very much for your encouraging comment.

Comment: *1. As just said, hazards cannot be compared but we can compare their effects; this is exactly what the risk analysis is meant to do. There is an extensive scientific literature on the comparison of the risks caused by different events (e.g., comparing the individual risk of death caused by different events), or comparing the risk with the acceptable risk that has been defined by decision makers. It is not clear why the authors dismiss completely all these efforts, which have eventually their same goal. Why do they think that their method is more effective that the classical risk and multirisk assessment?*

Response: Thank you very much for your comment and question. As mentioned previously, the focus of the paper is on quantification of the equivalent hazard strength. As such, the objective is

not to perform a risk analysis. Instead, we are deriving a new indicator: equivalent hazard strength, that is derived from impacts given average exposure and vulnerability based on a robust record of historical impacts. While we agree that it is important to understand how different factors contribute to overall risk, this is not the goal of this paper. Our proposed scale has applicational significance in providing benchmark measures of hazard strength for vulnerability and resilience analyses. 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.

To highlight that one of the main utilities of hazard equivalency research is to facilitate risk analysis, we have modified one sentence in the second paragraph of the Contributions section. The modified sentence now reads: "Such multi-hazard quantification of hazard, exposure, vulnerability, and resilience can be integrated to facilitate risk analysis to predict future losses and loss ratios without additional efforts to develop sophisticated models for each individual hazard types." (L348-350)

Comment: *2. The authors based their analysis on a 120-year-long database. I think that the length of this database is clearly too short to get a realistic estimation of the impact of some natural threats, which have a longer average inter-event times (for instance super-eruptions with VEI7 or 8). That's important because the effect of one of such events can largely overcome the cumulate effects of all other events. As a matter of fact, for some of the hazard considered in this paper, the most impacting events at worldwide scale have a return time that is much higher than 100 years. This is also the reason for what the risk is almost never empirically calculated using databases of this time length, at least for the most damaging events.*

Response: Thank you very much for your comment. As mentioned previously, the objective of the paper is not to quantify the effects of events but to demonstrate the computation of the equivalency of hazard strengths. Having said this, we do agree that the length of the database used for our study may not be long enough for comparison across hazard types that occur infrequently. However, as the purpose of the paper is to propose an empirical method to derive the equivalent hazard strength across hazard types, the EM-DAT database is sufficiently robust for demonstration of the proposed method. When other databases with higher quality and longer length become available, we intend to use them to improve our model results. In terms of the issue associated with the return period, the return period is always positively correlated with the hazard strength independent of the hazard type. Analysis of return periods for small to medium hazard strengths provides evidence to extrapolate the return period for large hazard strengths that are rarely experienced. When dealing with large hazard strengths, it is common even for singular hazard strength scales to have trouble in revealing the hazard strengths. For example, all the earthquake magnitudes are known to have saturation issues, more or less to some degree, for large magnitudes. Therefore, it is expected that the derivation of equivalency of hazard strength for large hazard strengths may be less reliable than for smaller hazard strengths. Since most of the hazard damages communities experience are caused by events with small to medium hazard strengths, we believe that it is useful and significant to derive hazard equivalency even for small to medium hazard strengths. In addition, to what extent the computation of equivalent hazard strength becomes unreliable given a certain return period is beyond the scope of the current study. Future work should explore the effect of long return period on the estimation of hazard equivalency.

Comment: *3. I think that the exposure and vulnerability are strongly changing through time. Conversely, the authors are assuming that these quantities remain constant in the past 120 years. This assumption may introduce a significant bias in the ranking of the events; for instance, it may be argued that the same tsunami in 2004 would have caused much less casualties if it happened in 1904 (by the way, to my knowledge the number of casualties caused by the 2004 tsunami is much less than 2 millions as reported by the authors). Not less important, as also acknowledged by the authors, some of the data may be severely incomplete; incompleteness has to be carefully checked because it can introduce an important additional source of bias in the analysis.*

Response: Thank you very much for your comment. We agree that exposure and vulnerability are changing through time. However, if such changes occur consistently across different hazard types, these changes are unlikely to affect the computation of equivalency of hazard strengths. By using the data of 120 years throughout the entire world, we assume that each hazard type has a similar temporal distribution of exposure and vulnerability, as is common in multi-hazard disaster research. Future work could examine whether this assumption holds true. Meanwhile, we also agree that there are some issues in the data from the EM-DAT database. In particular, there are many hazard types, such as volcanic hazards, that do not have measures of hazard strengths in the database. In addition, as also highlighted by the referee, there may be inaccurate records of disaster damages. In this study, we performed a quality control to exclude some obviously impossible values. We also excluded data points without hazard strength indicator values and performed a principal component analysis to support the derivation of the impact metric for data points with missing values for some, but not all, of the impact variables. However, we would like to point out that EM-DAT is a world-renowned database for hazard events and is one of the few open-source databases readily available to researchers and is often used for hazard analyses.

Regarding the reported 2 million casualties, it was incorrect, as it should be 2 million people affected by the tsunami. We have corrected this issue in the revised version of our manuscript. The modified sentence now reads: “The well-known 2004 Indian Ocean tsunami that affected more than 2 million people ranks 10th among all events, with its equivalent magnitude at 8.27.” (L236-237). In addition, we have thoroughly double-checked the entire manuscript for such similar typos and have corrected another record in the text on the economic damage of the 2013 El Reno tornado that was based on information from Wikipedia and inconsistent with the data from EM-DAT database. The modified corresponding sentences now read: “Among the considered 12 hazard types, the natural hazard with the lowest maximum equivalent magnitude is tornado. The tornado event with the largest equivalent hazard magnitude (3.62) is the 2013 El Reno tornado in Oklahoma, USA. This tornado event led to a total damage of over 2 billion 2019 USD (Guha-Sapir et al., 2021).” (L240-242)

Comment: *4. I am puzzled by the inclusion of synthetic data to fill the “missing” data. This may be very dangerous, because the ‘new’ data have been generated assuming that the model used to generate them is correct. To sum, I do not understand the need to generate synthetic data and not using only the ones available. (but I may be missing something here)*

Response: Thank you very much for your very insightful comment. We agree that using synthetic data always requires extreme caution. Therefore, we only included data points without missing values of impact variables and data points with missing values of one or two, but not all three, impact variables. There are no synthetic data points involved in the study. The purpose of filling

the missing values of the partially incomplete data points is not to generate synthetic data points, but rather to form a mathematical mapping between the one or two impact variables with recorded values to the impact metric that is the principal component of the three impact variables derived with the complete data points.

Comment: *5. The results of the correlation between impact metric and hazard (Figures 3 and 4) are largely not statistically significant (maybe except in a very few cases case, but we need also to take into account that the statistical significance has to take into account also the multiple tests). It is difficult for me to understand how we could use these relationship to rank the hazards in a meaningful way.*

Response: Thank you very much for your comment. We agree that many of the estimates of model parameters are not statistically significant after the calibration of the regression models. 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 this sense, the lack of statistical significance or the wide spread of data points is not a problem for the computational methodology for deriving the equivalency of hazard strengths. In the paper, we present the statistics of model parameters mainly for reproduction of research results.

Comment: *Figure 4 shows that, on average, the smaller the event the lesser the impact. This is already very well known but the large scatter of the logarithmic quantities implies that, for example, a large earthquake can cause no victims whereas a smaller one can cause a huge number of casualties. It depends on where the earthquake occur. For example, on average about 20 earthquakes of magnitude 7 or above occur worldwide per year, but only a very few of them in the last century caused more than 100,000 casualties, whereas most of them do not produce any casualty, or very few; the scatter in terms of casualties spans about 5 orders of magnitude for such a kind of events. This is a consequence of using an 'agential' approach, whereas the risk is intrinsically 'locational' (de facto, the exposure is strongly spatially clustered over the earth).*

Response: Thank you very much for your careful observation and insightful comment. We do agree that events with large hazard strengths may result in small impact, whereas events with small hazard strengths may lead to large impact. However, this is not necessarily a consequence of using an agential approach. When looking at hazard strengths locationally, for example the modified Mercalli intensity, we can still find many cases where large hazard strengths associated with trivial or no damage due to low exposed value or high resilience of communities experiencing the large hazard strengths. Nevertheless, it is important to continue research in hazard equivalency both agentially and locationally.

Comment: *6. The example reported in the discussion highlights the problem with this method. The authors say that the cold wave in Oklahoma city in 2021 has the same hazard magnitude in the Gardoni scale as an earthquake of magnitude 7.5. I do believe that a magnitude 7.5 in Oklahoma city would have caused an impact that is several orders of magnitude larger than the*

impact caused by the cold wave; and no impact (or very limited) if the same earthquake occurred in a remote area.

Response: Thank you very much for your comment. As mentioned previously, the objective of the paper is to quantify the equivalent hazard strength in an agential and durational way and not to compare the actual impact locationally. The cold wave event mentioned in the paper not only affected Oklahoma City but also many other parts of the Southern United States. The impact of the cold wave event within the entire spatial range of the event is recorded to be equivalent to the impact of Hurricane Harvey and the expected impact of a magnitude 7.5 earthquake. Regarding earthquake, it is very likely that a magnitude 7.5 earthquake in Oklahoma City would only affect a relatively small area around Oklahoma City. That is why, agentially, the 2021 cold wave event is equivalent in hazard magnitude to a magnitude 7.5 earthquake, but locationally, the hazard intensity of the 2021 cold wave event may not be equivalent to the hazard intensity of a magnitude 7.5 earthquake. The issue raised by the referee here is not a problem at all. Locationally, a magnitude 7.5 earthquake in Oklahoma City would have caused an impact that is several orders of magnitude larger than the impact caused by the cold wave in Oklahoma City. Agentially, however, the expected impact of a magnitude 7.5 earthquake, with its epicenter in Oklahoma City, within its spatial entirety would be equivalent to the expected impact of the 2021 cold wave event within its spatial entirety, i.e., the entire Southern United States.