Discussion on:

Induced seismicity risk analysis of the hydraulic stimulation of a geothermal well on Geldinganes, Iceland

M. Broccardo, A. Mignan, F. Grigoli, D. Karvounis, A.P. Rinaldi, L. Danciu, H. Hoffmann, C. Milkereit, T. Dahm, G. Zimmermann, V. Hjörleifsdóttir & S. Wiemer

Introduction:

We thank the reviewers for their efforts in reviewing and providing suggestion to improve our manuscript. We greatly appreciate their work and hope that the revised version of the manuscript will be an important improvement. We would like to address each of the points raised below.

This document contains point-by-point response to the Reviewers.

We have done so in a constructive way, hoping to stimulate a fruitful discussion. Below in blue are the answers of the authors and in black the original questions of the reviewer. We proceed first with the second reviewer, Prof. Julian Bommer, because we realized that by addressing his remarks, we also indirectly answered to Reviewer 1, Dr. Georgios Michas.

Reply to Reviewer #2 Professor Julian Bommer

This is an interesting paper that narrates the development of an induced seismic risk model for planned geothermal stimulations through fluid injections in Iceland. The topic would clearly seem to be within the scope of Natural Hazards and Earth System Sciences and, as the authors point out, it may be the first published induced seismic risk study performed prior to the commencement of injections. I would therefore conclude that the paper is a potentially valuable contribution and should be published, but I believe that it can also be improved. In particular, some information that is extraneous given the title and focus of the paper could be removed, and some more detail provided on essential features. I also think the authors could do a slightly better job of citing relevant literature and acknowledging work by others. In the following paragraphs I offer my observations and suggestions in the order in which they refer to the manuscript rather than any hierarchy of importance. I am not correcting inconsistencies in the citation of references within the text—years sometimes in brackets, sometimes not— because I do not consider that part of my work as a peer reviewer.

Line 28 (and many other locations): The Latin phrase “a priori” should be written in italics and never hyphenated.
We thank the Reviewer for this remark. We updated the manuscript with the correct Latin form.

Line 43: Please explain what straddle packers are since this may not be common knowledge for all readers (myself included).
We thank the reviewer for pointing this out. We updated the manuscript with a short definition of a Straddle packer; i.e. “Straddle packers not only allow isolating and injecting in selected narrow zones, but also allow adjusting the straddle distance between the upper and the lower injection points.”

Line 58: “Adding an additional” is very inelegant, please re-word.
We thank the Reviewer for this remark. We rephrased the sentence.

Line 59: Heat cannot be effectively transmitted over distance so such systems will always be close to consumers out of sheer necessity.
The authors acknowledge that this is often true, but not in all projects. Deep geothermal projects are often only linked to electricity production, and hot water is not necessarily used for heating. Moreover, in Iceland, there are geothermal power plants at relatively long distances (over 40 km) from Reykjavík. It is possible to transport hot water, just depends on the context and site conditions. In fact, they produce hot water at exceptionally high rates and at such high temperature that allows transmitting the heat at such distances (Mignan et al., 2019).

Line 61: From the references cited here, the reader is left with the impression that other than Dr Mignan and his co-workers, nobody has ever written about the assessment and management of induced seismicity.
We thank the Reviewer for this remark. We added the additional references relevant to this study.

Line 68: "even more importantly"
Corrected, thank you.

Line 69: "are the sizes and stress state"
Corrected, thank you.

Lines 76-77: Successful? Majer et al. (2007) documents several geothermal projects in which the largest induced events occurred after pumping had stopped, highlighting the limitation of TLS. The recent paper in SRL by Baisch et al. (2019) on the efficacy of TLS is probably worth citing in this regard
Thank you for this observation. We modified the text accordingly.

Lines 81-82: It is, at least, the first being made public—there may well be commercial and confidential studies that have not been released.
Thank you for this observation. We modified the text accordingly.

Line 84: The Latin phrase "in situ" should be written in italics and never hyphenated.
Corrected, thank you.

Lines 104-105: Why or how is this being suggested? Is M 2 the detectability threshold?
No, we did not have knowledge of the detectability threshold at that time. We considered the $M = 2$ simply as the threshold above which locals would have felt and reported an event.

Line 122: "a strike-slip/thrust regime": I am not clear if this means one or the other, or an oblique combination of the two? If the latter, do you mean reverse rather than thrust? Thrust specifically refers to very shallow dipping reverse faults, and strike-slip on such faults is unusual.
We thank the reviewer for the remark. It was indeed a mistake. We altered the text and clarify with the phrase "a strike-slip to reverse regime."

Lines 147-148: "The concept.....(Pittore et al., 2018)" – unless you tell the reader what the outcome of the DESTRESS evaluation was, this sentence does not really serve any purpose.
Thanks for the observation. After re-reading the article, we came to the conclusion that the sentence did not add any value, so we deleted it.

Line 165: Suggestion: "is planned to be approximately equal" (current wording is strange)
Thank you for the suggestion.

Line 196-236: This text is not needed since it is perfectly summarised in Figure 3 and is not central to the title and focus of the paper.
Thanks for the tip. After re-reading the article, we agreed with the reviewer and deleted the whole part.

Line 238: Bommer et al. (2017) is not listed in the references—and I don't know what paper it refers to!
Thank you for the typo detection. It is Bommer et al. 2015.

Line 292: "(and generally unknown a priori)"
Thank you for this suggestion. We modified the text accordingly.

Line 293: Langenbruch et al., 2018.: Nature Communications 9:3946, doi:10.1038/s41467- 018-06167-4 have made the connection to pressure changes explicit in their modification of the seismogenic index.
Thank you for directing us to this reference. We had missed this relationship now given between pressure and SI. We now cite it in the new version of the manuscript.

Lines 298-299: No, it is not also known as the seismogenic index, it is the seismogenic index and you are choosing to give it another name, for reasons that are not clear to me at all. This was pioneering work by Serge Shapiro and there is no need whatsoever to assign it a new name—unless the plan henceforth is to cite this paper as the origin of the concept, which would be totally unacceptable.
We deplore that such negative connotation seemed to transpire from our text. We have used the more generic term a_{fb} (for a-value normalized by volume) since the SI infers a specific poro-elastic origin of the parameter. See for example Shapiro and Dinske (JGR 2009). This is the only reason we are reluctant to use SI terminology as the scaling can be explained by other means, including geometry (Mignan, NPG 2016) - Note however that we always mention Shapiro but we observe that using the term SI assumes, implicitly, a poro-elastic origin. It might well be the case but this remains unproven. The normalized a_{fb}-value term avoids such assumption. We now clarify this aspect by: (1) highlighting the fact that Shapiro pioneered this work / was the first to find this relationship and (2) that a more generic term is used here to be agnostic regarding the underlying physics. We also added the additional Shapiro and Dinske reference. We regret that the wording may have been poorly chosen in the original version.

Lines 314-315: What weights are assigned to each of the models? This is not indicated in Figure 4. The rationale for the weights also needs to be presented.
Thank you for the remark. Given that "*a priori knowledge on those parameters is limited and the range of possible values wide*" we assign equal weights to the different options. We reported these weights in Figure 4 and added further explanations.

Lines 352-354: Again, the weights need to be reported, together with their rationale. Why are these parameter combinations not depicted as logic-tree nodes and branches in Figure 4?
Same as the previous comment

Line 361: Is this really the appropriate reference for the concept of maximum magnitude?! Did nobody work on these topics before you?
We added three more references on seismic risk assessment in the anthropogenic context (which also assume a truncated GR law at m_{max}).

Lines 363-365: This statement may not hold for induced seismicity. In some hazard and risk studies for induced earthquakes, the results are very sensitive to the maximum magnitude.
The reviewer is partially correct with this statement. There are indeed fluid-induced seismic studies for which the maximum possible magnitude is an important variable. These cases use as upper bound the McGarr limit (or similar), which is (usually) considerably lower than the tectonic upper bound. In our analysis, however, we use the tectonic upper bound and the output of risk analysis becomes not very sensitive to the choice of $m_{max} > 6$. See also the study of Gupta and Baker (2017). Given these considerations, we changed the text to make clear this concept.

Lines 371-372: The phrase in parentheses does not currently make sense, re-word.
Thanks, we rephrased it.

Lines 386-391: A clearer explanation of the two concepts could be given. Is this the difference between maximum possible and maximum expected, as discussed, for example by Zöller & Holschneider, 2016, BBSA, 106(6), 2917–2921? As currently written it will create confusion because in common practice, Mmax is used as the symbol for the parameter you refer to as msup.

It is truly just a matter of notation and definitions. In the scientific literature, we can find the following (probably incomplete) list of definitions.

- Gishing and Wiemer, 2013.
  - *Maximum possible earthquake*→ upper bound of the GR distribution, M_{max}
  - *Maximum observed earthquake*→The maximum observed earthquake during a given project, $M_{max}(obs)$
- Zöller and Holschneider 2016 & Holschneider et al. 2011
  - *Maximum possible earthquake*→ upper bound of the GR distribution, m_{max}
  - *Maximum expected earthquake magnitude*→ The maximum observed (or expected) earthquake during a given project, M_T
- van der Elst et al., 2016
  - *Maximum possible earthquake*→ not denoted, they use unbounded GR
  - *Maximum expected earthquake magnitude*→ The maximum observed earthquake during a given project, M_{max}

What is "common practice" depends on the context and the research group. However, in the context of fluid-induced seismicity, the van der Elst et al., 2016 seems the most popular paper (we have adopted such terminology also in Broccardo et al., 2017a).

In our paper, we tried to use consistent probabilistic terminology and notation. The rationale behind this choice is the following. First of all, the upper bound of the double truncated GR is a parameter, so it should be lowercase (as correctly used by Zöller and Holschneider). It follows that we could potentially use the notation introduced by Zöller and Holschneider, i.e. use m_{max}. However, in a Bayesian setting m_{max}, can be a random variable and therefore indicated in uppercase M_{max}. This M_{max} creates ambiguity with the M_{max} definition used in van der Elst et al., 2016 (which coincide with the M_T used by Zöller and Holschneider) and also adopted by Broccardo et al., 2017a. Originally, to be consistent with this choice we have introduced for the GR upper bound m_{sup}, to distinguish it from the observed M_{max} clearly. We wanted also to promote the idea of not using the word "maximum" for the upper bound of the GR, which will create a "probabilistic" ambiguity.

However, since we are concern about the ambiguity generated by these notations in this field, we decided to accept the reviewer's suggestions partially. Therefore, **we changed notation and definitions as follow:**

- New notation & definition
  - *Maximum possible earthquake (or upper bound of double truncated GR)* → m_{max} (this is a parameter, lower case). Observe that the probability of observing m_{max} is zero, and it cannot be "captured" by any interval.
  - *Maximum observed earthquake magnitude for a given period T* → M_T (this is a random variable, uppercase). Therefore, we use the notation of Zöller and Holschneider; however, we do not use the terminology "expected maximum magnitude." Observe that this is not a very fortunate term since it will generate confusion when the expected maximum observed earthquake is of interest (i.e., $E[M_T]$).

We hope that this new notation and terminology will eliminate all ambiguities. We added also a footnote to clarify the difference between the terminology used in Zöller and Holschneider and van der Elst et al., 2016 and Broccardo et al. 2017a. Moreover, we used Holschneider et al. 2011as the reference paper. Observe that we have modified also the Figures accordingly.

Lines 422-423: This is a poor reason for using only PGA unless it is also found to be an efficient and sufficient parameter for the risk calculations.

The objective of this *a priori* risk was to collect all possible information for a first-order risk assessment of the project. (Now, we state this clearly both in the abstract and in the introduction). Therefore, only "off-the-shelf" GMPEs and fragility models were used. In addition, it is noted that local fragility functions (used for damage risk) have been developed only for PGA (Bessason, B., and Bjarnason, 2016) using the same dataset used in the recalibration of the GMPEs. We assume that the conditions of sufficiency and efficiency and their limitations have been addressed in that study (and this is always the case for "off-the-shelf" models). Next, the fragility functions that have been used for the Individual Risk have as input the European Macroseismic Scale, which is also considered sufficient and efficient in those studies.

A more detailed bottom up approach would have required:

i) The development of an appropriate set of GMPEs for a sufficient and efficient (but also effective) IM (or a set IMs);
   a. Development of a spatial correlation model (and validation)
ii) An appropriate set of fragility for different type of buildings (which might require different IMs). This fragility analysis should have been developed by classical fragility analysis (e.g., incremental dynamic analysis) on local models and validated with an experimental campaign;
iii) A detailed microzonation (very important);
iv) A detailed exposure model and aggregate loss analysis.

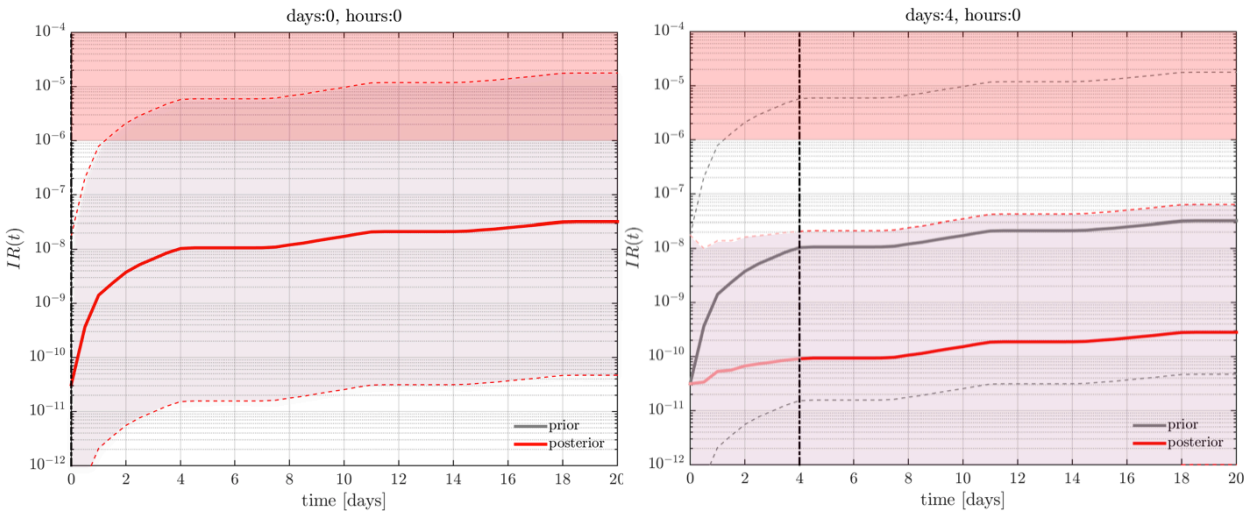
However, this approach requires a consistent investment of resources and time without the possibility of implementing *a priori* a verification and validation scheme since NO local data are available.

We know that this detailed bottom-up approach has been adopted in Groningen (and other fields), *but a posteriori*, when a considerable amount of data and information (and resources) were available for validation. Observe that such an approach *a priori* is not possible without local data. In addition, allocating a considerable amount of resources for such analyses without having a first-order estimate of the hazard, based on real data, is not ideal.

Next, as we have shown in the sensitivity analysis most of the state of deep uncertainty (in this study) is due to the unknown values of a_{fb} (or Σ) and b. So the value of information on reducing the uncertainty of these two parameters (in a first phase) is much larger than reducing the epistemic uncertainties on the GMPEs. Moreover, the (*a priori*) bias implicit in any "off-the-shelf model" is (very likely) buried under the state of deep uncertainty (and thus becomes less critical).

We are very well aware of the limitations of this study, and we have clearly listed them in Section 5. We are also aware that the following study gravitates more towards hazard rather than vulnerability and exposure; however, the first focus should be on addressing the magnitude of the hazard, i.e., if this is very low (e.g. extremely small magnitude events) it is not wise to allocate resources on having detailed GMPEs, vulnerability, and exposure models.

Therefore, we are confident that in the presence of no data, limited time, and budget constraints, the best approach is to have a first-order estimate of the hazard and associate potential risk. Then, when the project is started, and data become available, the decision-maker can demand a much more detailed analysis. Observe that while we are writing this response, the project has already terminated. In Figure R1, we report the reduction of uncertainty after a few days of the injection (a current manuscript is in preparation so no more information can be given at this level). The Figure shows the (online) projected Individual Risk. Soon it became clear that we were facing a very low-risk project (The Expected Maximum Observed Magnitude was very low). Any update of the GMPEs vulnerability and exposure would have been an excellent exercise but probably not very "valuable" for decision making.



R 1 Updates of the Individual Risk and projection. Panel on the left time 0, panel on the right update after 4 days. The individual risk dropped to very low levels indicating a "very low risk" project.

Lines 423-424: Given that Rjb ignores focal depth, which is critical for induced earthquakes (as note in the next paragraph) what was the rationale for this choice?
It would have been ideal using R_{hypo} (as we clearly state). However, we did not have such options available, and the development of low magnitude local GMPEs was not within the scope of this study (no data were available and stochastic ground motion models were not considered as an option). In fact, the development or recalibration of local GMPEs for future injections was one of the possible outcomes (in case of observable seismicity) of this experiment.
As pointed out by the reviewer, the next paragraph clearly outlines all the limitations, so that future updates can aim to eliminate them. Updates and new risk analyses can be made once the project has started, and local seismicity has been measured. We have included this strategy in the recommendations.

Lines 425-426: In Figure S1, the range of epistemic uncertainty appears to decrease with increasing magnitude, which is counter intuitive. Do the authors have an explanation for this feature?
This Figures are directly derived from the following paper Kowsari et al., 2019. Our conjecture is that this is due to the fact that GMPEs are recalibrated for higher magnitudes and few data are available for lower magnitude and extrapolation to lower ranges might cause a larger variability. However, the authors of the original paper are more suited to answer this question.

Lines 428-429: I don't think it's probably biased, it is biased—see Bommer et al., 2007, BSSA, 97(6), 2152-2170 for discussion of this issue. Baltay & Hanks, 2014, BSSA, 104(6), 2851– 2865 provide a clear physical explanation for this behaviour.
Thanks for these observations. We are glad that the reviewer has pointed us to these papers (we modified the text accordingly).

Lines 432-434: Depth influences the length of the travel path and the stress drop, so this dismissal of the issue seems rather casual.
See comment line 423-424 and line 422-423. The dismissal is not casual but due to the use of "off-the-shelf GMPEs." As previously mentioned, these are limitations that will (eventually) be corrected in a second phase.

Line 435: "over instrumental intensity measures" (physics-based is not appropriate).
Thanks for these observations (we modified the text accordingly).

Line 439: You claim to be using EMS98 but choose a GMICE calibrated in terms of MCS intensities. Musson et al., 2010, JOSE, 14, 413-428, have pointed out that while these two scales appear equivalent in their definitions, they are generally not used in an equivalent manner.
We thank the reviewer for this observation. However, in the context of this paper, the MSC scale is practically equivalent to the EMS98 scale. In fact, (Musson et al., 2010):

[*…In 1988 the European Seismological Commission agreed to initiate a thorough revision of the MSK Scale. The result of this work (undertaken by a large international Working Group under the chairmanship of Gottfried Grünthal, Potsdam) was published in draft form in 1993, with the final version released (after a period of testing and revision) in 1998 (Grünthal 1998). Although this new scale is more or less compatible with the old MSK Scale, the organisation of it is so different that it was renamed the European Macroseismic Scale (EMS). Since its publication it has been widely adopted inside and also outside Europe…*]

Moreover, in Faccioli and Cauzzi 2006:

[*…Based on a detailed comparison between their MSK-64 and MCS intensity ratings for all earthquakes considered, [Margottini et al., 1992] found no statistically significant difference between the two scales in the range from IV to VIII degrees (see the Appendix), in partial contrast to previous (possibly less accurate) suggestions to raise by about one degree the MCS intensity with respect to the MSK value in the range > VI [Levret and Mohammadioun, 1984]. Herein, we have extended the assumption* $I_{MCS} = I_{MSK}$ *also to some non-Italian earthquakes (see the Appendix)…*]

Therefore, in this study we have implicitly assumed $I_{MCS} = I_{MSK} = I_{EMS98}$. Observe that this approximation is not significant because most of the individual risk is related to moderate to significant events. We are convinced that making specific distinctions on macroseismic scales at this level of detail and with the presence of such significant uncertainties (at model rate level) is not very meaningful. Therefore, we have (just) added a small footnote to highlight this approximation.

Line 440: Please explain how the increase sigma value is applied in the hazard and risk calculations—is the PGA to intensity conversion performed within the hazard integral?

Thank you for the remark. We think we explained this quite clearly in the lines 435-450 (of the original document). We basically derived the equation for $P(IM > im|m,r)$ by converting the original GMPE into the GMICE, i.e.

$$\log(PGA) = f(M,R) + \epsilon$$

If ϵ is Gaussian (with σ_{GMPE}) then $\log(PGA)$ is Gaussian. Next,

$$I = a \log PGA + b + \varepsilon,$$

then I is conditionally Gaussian with mean $aE(\log PGA) + b$ and variance $a^2\sigma^2_{GMPE} + \sigma^2_{MICE}$.
Given that (already reported at line 440 of the original paper) we have fully specified $P(IM > im|m,r)$ and the hazard curve can be computed with the classical total probability theorem as $\Lambda(im;T,b) = -\int_m P(IM > im|M = m,r)$. Also, this was already reported at line 445 of the original manuscript. Note that in the supplement, we have also reported all the values for a and b for each GMICE. Given this description we did not alter the text.

Line 468: I would suggest mentioning physical damage to buildings, from which injuries and costs are both then derived.
Thank you for this remark. We amended the text accordingly.

Lines 473 & 475: What is meant by "statistically average" in these phrases? Is individual risk average over the exposed population or reported for individual locations?
It is the standard definition used for quantitative risk measures for loss of life and economic damage. See Jonkman et al., 2003. Precisely, the average person is the "expected" person within a population. In practice, one has to think about randomly selecting a person from a population (which is, by definition, the expected person and therefore the statistically average person). We have not introduced this definition, but it is formally defined in Jonkman et al. 2003 and Jones, 1992. The same reasoning can be applied to a population of different buildings. Since we gave the proper references, we believe it is not necessary to modify the text.

Line 481: Risk targets for individuals in the Netherlands are define at 10^{-4} and 10^{-5} .
Line 481: the reference is van Elk et al. (2017).
Thank you very much, we corrected the text.

Line 495-496: On what basis is this assumption made? Is there any basis for its justification? Or you have to assume it to make the results meaningful?
Thank you for the remark. It is an implicit assumption (stemming from Mignan et al. 2015) since we use "off-the-shelf" macroseismic vulnerability model. In principle, it is not even necessary for this study (without it, the results are not changing). Therefore, we have decided to eliminate it.

Section 4: There is nothing about the exposed building stock in terms of numbers of buildings and their location, just as there is nothing about the exposed population. More information is needed and must be added.

Observe that at this level of details we expressly choose to not use *any* aggregate risk measures (including losses). In fact, we wrote at lines 469-471 of the original document:

[*…The a-priori risk analysis for the Geldinganes project here focuses on the first risk, while the aggregate economic losses are not directly computed. Here, as a substitute for aggregate losses, we define a low damage threshold for statistical average classes of Icelandic building…*]

As stated, we limited our analysis only on marginal risk measures. For this reason, we introduced the Damage Risk, which represents the risk of light damage given a class of buildings (regardless of the number of buildings). It is the same definition of Individual Risk but in the damage domain. This facilitates the decision-making process, making individual risk and damage risk two compatible measures. Observe that also the Individual Risk is given as a function of the building class. The *modus operandi* is the following,

- Given a location of the built environment at a distance R from the injection point, we consider any information about the specific class of the built environment as missing (detailed analysis was out of the scope of the current study).
- We compute the IR ad DR as function of the Icelandic building classes.
- Therefore, for each built environment coordinate, we computed both IR and DR for a given class of buildings. Next, the decision making process is performed by defining a threshold on the worst-case scenario (i.e., the weakest of the building class).

We have explained these steps in detail throughout Section 4. Given this, the detailed information of the exposure model is not significant for the calculation and is only useful for general information. Therefore, we have included general information on the first order exposure estimation in the supplement (this information is already available in the original manuscript). Please note that we have now also included additional details that has been used in the internal risk analysis report.

Section 4.1: You also do not explain how hazard and risk are convolved—do you convolve hazard and fragility curves or use stochastic event sets? If the former, any spatially aggregated risk metric will be over-estimated.

It is explained at the end of previous section (we slight modified the text to be more precise)

[*…The framework used for the computation of IR and DR is based on the convolution of the hazard model with the vulnerability models for the relevant building types, and (only for the IR) with the consequence model…*]

As indicated in the previous comment (and clearly stated in the text), we decided not to use aggregate measures but only marginal measures (which do not depend on any correlation structure). In principle, to obtain aggregate measurements, we need a spatial correlation model based on further model assumptions (i.e., Gaussian random field) and on assumptions on the parameters of the model itself. Since no data are available, these assumptions cannot be validated. Alternatively, we can look for an upper limit (as implicitly indicated by the author) considering individual losses as comonotonic random variables, and a lower limit considering losses as statistically independent (Broccardo et al. 2017b). However, in the presence of such deep uncertainties (at rate level), these limits are essentially meaningless (i.e., they cover pretty much the entire domain of losses for any return period of interest). Such a level of detail, at a very early stage of a new project, is not recommended. We consider it essential first of all to collect data from a pre-stimulation or a first phase of the project to have the correct Hazard model and then to move (eventually) to aggregate measurements. In addition, at this stage, we also choose not to perform the decision-making process at the monetary loss level because there is large uncertainty in defining the repair and replacement costs of an individual or group of local buildings (and how these costs are related to the damage). As we have said above (now the project is finished), it was soon evident that this kind of detailed analysis was not necessary for this project.

Lines 565-585: Should the objectives not be stated at the beginning rather than the end?
Thanks a lot for the comment, we have restructured the introduction.

Line 568: The concept of "informed technical community" is now considered confusing and unhelpful, and in both NUREG-2117 and NUREG-2213—which effectively supersede NUREG/CR-6372)—the expression is "centre, body and range of the technically-defensible interpretations".
In restructuring the introduction, we have eliminated this part.

Lines 603-604: Why? If there is a ground-motion recording network and large uncertainty in the choice of GMPEs, what is the basis for dismissing a priori any update of this part of the model?
This was not completely dismissed. We did not perform online updates of the coefficients of the GMPEs because, from the sensitivity analysis, most of the uncertainty was related to the rate model. However, we did plan to updates the GMPE models itself in case sufficient seismic data were available (but this is an *a posteriori* analysis).

Appendix A: I propose that you remove this section to make room for some information about the exposure model. I especially think this should come out since the recommendations only focus on modifying the model as the project proceeds, but no mention is made here on in Section 5 about how the results of the a priori study should be used before the study begins. Do the results provide confidence to proceed with the project?
Thanks for the suggestions. We have already answered in the previous comment on the exposure model (and all extra information is in the supplement). This study represents the synthesis of a collective effort that has produced a risk report. The result (as we stated in both the abstract and the key findings in Section 5) indicated that the median of individual risk and injury risk were both below the selected threshold. As a result, the project started (now reported in Section 5). However, we clearly indicated in the abstract, introduction and limitation (Section 5) that the results are associated with a state of deep

uncertainty and several limitations. We also indicated the steps for the online update of the uncertainty of the rate model (which is a subject of a future study), and the improvements to be made if the project turns out to be in a "risky" domain. All this is now highlighted in (re-structured) Section 5. Moreover, we believe that the recommendations section is actually very important in this study. We shorten it and moved into Section 5.

Do the results indicate that some changes could be made, such as to location or even building strength? Surely the value of a risk model is to inform decision-making? I would refer the authors to van Elk et al., 2019, Earthquake Spectra, 35(2), 537-564 for an example of an induced seismic risk model designed for such a purpose

As mentioned above, this assessment would have been carried out after the start of the project if the level of risk was close to the thresholds set. In the absence of data and in the presence of such a level of uncertainty at the rate level (highlighted in the sensitivity section), providing such suggestions may not be advisable. At the present time, the project is finished. As can be seen from Figure R1, it was clear, early enough, that any suggestions for improving the strength of the building would be unnecessary. Also, changing the location of the well was simply not possible because the project is a re-stimulation of an existing well.

. If the value of the authors' model is to have a starting point for updating and refining during the project procedures, then this should be stated.

We believed that this was clear since the abstract. In fact we wrote:

[…However, these results are affected by several orders of magnitude of variability due to the deep uncertainties present at all levels of the analysis, indicating a critical need in updating this risk assessment with in situ data collected during the stimulation…]

However, since it was not clear enough, we modified and stressed the true purpose of the study in both abstract, introduction, and conclusion.

And how are the thresholds on the TLS in Figure 3 linked to the risk estimates? Where are the disaggregation results that support these magnitude thresholds? The authors could tie the various elements of the paper together much more than is currently done.

Thank you for that question. The thresholds for the classic TLS were chosen following previous projects. In Iceland the classic TLS has already been applied for example for re-injection activities in the geothermal field of Hellisheidi, a region not far from Reykjavik (Thorsteinsson and Gunnarsson, 2014). However, for additional precautions (since we were closer to the city), we reduced each alert level by 0.5 units of magnitude. In addition, since our seismic network allows to analyze events with Ml<1.0 we have added an additional stage (no alert level) in which routine seismologists can evaluate in an advanced way the space-time evolution of induced seismicity using advanced seismic analysis tools. Observe that the Traditional TLS has been set separately from this study. While the Advance Traffic Light System makes use of this a priori risk analysis (and it is an experiment).

Thorsteinsson, H. and Gunnarsson, G.: Induced Seismicity – Stakeholder Engagement in Iceland, GRC Transactions, 38, (2014), 879-881

Reply to Reviewer #1 Dr. Georgios Michas

The paper presents a probabilistic assessment of the induced-seismic hazard and risk associated with a geothermal energy extraction project and the hydraulic stimulation of a deep well in the area of the Geldinganes peninsula, Iceland. Similar projects in other places around the globe have induced in the past intense microseismicity, but also larger magnitude events that have caused building damages, injuries and severe economic losses. Taking this hazard into account, the authors present a comprehensive study of the associated hazard and risk prior to the hydraulic stimulation of the well and discuss the mitigation strategies during the implementation of the project. In particular, they estimate peak ground acceleration (PGA) and seismic intensities values based on previous induced seismicity data from similar projects and on model-produced synthetic catalogues. In addition, they estimate the seismic risk for the population and buildings, quantify the associated uncertainties and discuss possible limitations. Such studies, based on a multidisciplinary approach that combines the existing knowledge available to the scientific community, are essential for an efficient risk estimation and mitigation, as well as for the sustainability of similar projects. Overall, the paper is wellstructured, well-written and presents in a comprehensive way the hazard, risk, their uncertainties and the limitations that the scientific community faces regarding hydraulic stimulation operations and fluid-induced seismicity. Therefore, I strongly recommend its publication to Natural Hazards and Earth System Sciences. However, there are a couple of points that need some clarifications. In the mitigation strategy section, the authors describe the defined alert levels according to the traffic light scheme that was planned for the project and explain the planed actions once anomalous seismicity patterns are identified (lines 203-205). While they set the magnitude thresholds for the different levels of alert and discuss b-values changes that might require further actions, further explanation is perhaps needed for the seismicity patterns in the spacetime domain that might be considered "anomalous". Accordingly, while the authors recommend the use of more sophisticated techniques for operational seismologists in order to identify seismicity patterns that resemble potential faults (lines 210-211), these techniques remain vague in the text.

We are very grateful for this review. The detailed section on the TLS recommendations has been deleted following a suggestion of Reviewer #2. However, we do not want to elude the questions. For the anomalous seismic patterns, we mean a drop of the b value and/or an increase of a_{fb} (this corresponds to the lower left domain of Figure 5). Moreover, with "more sophisticated techniques" we intended the Relative relocation methods (Waldhauser and Ellsworth, 2000). Again, this is not included in the current text.

Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. Bulletin of the Seismological Society of America, 90(6), 1353-1368.

Induced seismicity risk analysis of the hydraulic stimulation of a geothermal well on Geldinganes, Iceland

Marco Broccardo[1,4], Arnaud Mignan[2,3], Francesco Grigoli[1], Dimitrios Karvounis[1], Antonio Pio Rinaldi[1,2], Laurentiu Danciu[1], Hannes Hofmann[5], Claus Milkereit[5], Torsten Dahm[5], Günter Zimmermann[5], Vala Hjörleifsdóttir[6], Stefan Wiemer[1]

[1] Swiss Seismological Service, ETH Zürich, Switzerland
[2] Institute of Geophysics, ETH Zürich, Switzerland
[3] Institute of Risk Analysis, Prediction and Management, Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, China
[4] School of Engineering, University of Liverpool, United Kingdom
[5] Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Potsdam, Germany
[6] Orkuveita Reykjavíkur/Reykjavík Energy, Iceland

Correspondence to: Marco Broccardo (bromarco@ethz.ch)

Abstract. The rapid increase in energy demand in the city of Reykjavik has posed the need for an additional supply of deep geothermal energy. The deep hydraulic (re-)stimulation of well RV-43 on the peninsula of Geldinganes (north of Reykjavik) is an essential component of the plan implemented by Reykjavik Energy to meet this energy target. Hydraulic stimulation is often associated with fluid-induced seismicity, most of which is not felt on the surface, but which, in rare cases, can cause nuisance to the population and even damage to the nearby building stock. This study presents a first of its kind pre-drilling probabilistic induced-seismic hazard and risk analysis for the site of interest. Specifically, we provide probabilistic estimates of peak ground acceleration, European microseismicity intensity, probability of light damage (damage risk), and individual risk. The results of the risk assessment indicate that the individual risk within a radius of 2 km around the injection point is below 0.1 micromorts, and damage risk is below 10^{-2}, for the total duration of the project. However, these results are affected by several orders of magnitude of variability due to the deep uncertainties present at all levels of the analysis, indicating a critical need in updating this ~~a priory~~ risk assessment with *in-situ* data collected during the stimulation. Therefore, it is important to stress that this *a priori* study represents a baseline model and starting point to be updated and refined after the start of the project.

1 Introduction

The city of Reykjavik, the capital and center of population of Iceland, meets 99.9% of its district heating demand by geothermal energy (Gunnlaugsson et al., 2000). However, the growing population and the booming number of tourists is pushing the current supply of energy to its limit, since no new low temperature wells have been drilled since 2001. In particular, additional sources of low temperature heat need to be accessed to ensure a reliable heat provision for the city center. Therefore, there is an urgent need to increase the current capacity by drilling new low temperature wells and stimulating older inactive wells.

One potential area for new low temperature geothermal field developments is Geldinganes. Geldinganes is a peninsula within the city limits of Reykjavik (Figure 1). The exceptional geothermal gradient in this area triggered the drilling of a well (RV-43) in 2001 after a gabbro body was identified as potential heat source and drilling target for this deviated well. Despite the required temperatures being reached, the flow rates were insufficient for economic production. At present, Reykjavik Energy (Orkuveita Reykjavíkur, OR) re-assessed this field for development of geothermal energy with new production wells. To additionally enhance the production, it is foreseen to hydraulically re-stimulate well RV-43 in order to improve its productivity to economical levels. In particular, a three-staged cyclic pulse stimulation is planned that will last for (circa) 12 days. The stimulation is expected to enhance productivity in three pre-existing fracture zones penetrated by RV-43 and isolated with straddle packers. Packers technology, commonly used in the oil and gas industry, is expected to be employed for the upcoming stimulation. Straddle packers not only allow isolating and injecting in selected narrow zones, but also allow adjusting the straddle distance between the upper and the lower injection points. ~~This will allow isolating selected narrow zones of the well and thus injecting exclusively through them.~~

Like all energy technologies, the exploitation of deep geothermal energy is not risk-free. Therefore, an essential part of the implementation and licensing is a quantitative risk assessment comparable to existing regulations for Health, Safety, Environment (HSE) procedures. This analysis allows balancing the (perceived and real) risks against the (perceived and real) benefits. Over the last decade, induced seismicity has emerged as one of the risks—often the most dominant one—to be faced (Giardini, 2009, Grigoli et al., 2017) in implementing industrial underground technologies (e.g., geothermal energy exploitation, water impoundment, CO_2-sequestration and natural gas storage operations, non-conventional hydrocarbon production, etc.). These activities can alter the stress field of the shallow Earth's crust by pore pressure changes, or volume and/or mass changes inducing or triggering seismicity (Ellsworth, 2013; Giardini, 2009; Mignan, 2016). Such earthquakes pose a nuisance or even a danger to the local population and can strongly undermine the societal acceptance of a project (Trutnevyete and Wiemer, 2017, Grigoli et al. 2017, Hirschberg et al., 2015). The recent M5.5 Pohang earthquake (Grigoli et al., 2018; Kim et al., 2018) is an extreme example of a triggered earthquake related to geothermal activities that had a combined economic impact of $300M as well as more than 135 injuries (Lee et al., 2019). In particular, fluid injection or extraction in

tectonically active zones carries a risk of inducing a seismic event of a significant magnitude (Grigoli et al., 2017), and deep geothermal projects are a primary example. ~~Adding an additional concern,~~ Another source of concern stems from the fact that deep geothermal projects in Europe ~~deep geothermal projects in Europe~~—and the Geldinganes stimulation is no exception—are often located close to consumers, thus in densely urbanized areas with historical and ~~hence~~ vulnerable buildings and infrastructures. In these contexts, the problem of assessing and managing induced seismicity is ~~especially important~~critical (Bommer et al., 2015; Giardini, 2009, Majer et al., 2007, 2015; Mignan et al., 2015; 2019a; 2019b, Trutnevyete and Wiemer, 2017, van Elk et al., 2017, Walters et al., 2015). It is also a well-known fact that societal acceptance of induced seismicity has substantially decreased in some countries in the past decade, a result of failures discussed widely in the media and an overall change in risk perception.

Despite the large body of research conducted over the past decades by numerous research groups, the physical, chemical and hydro-mechanical mechanisms governing induced seismicity are far from being fully understood, posing clear limits to the risk assessment and management strategies (Yeck et al., 2017; Trutnevyete and Wiemer, 2017; Grigoli et al., 2017; Mignan et al., 2019a; 2019b). The limitations to forecast induced seismicity are, on the one hand, the non-uniqueness on the physical framework for modelling, and on the other hand, even more importantly, the large uncertainties on the boundary conditions needed for forecasting (e.g., where are faults, what are their sizes and stressing state, what is the permeability distribution of the reservoir, etc.). It follows that any risk assessment and management strategy must capture the existing uncertainties and lack of knowledge, requiring a probabilistic approach that explicitly considers both epistemic uncertainties and aleatory variability. It also implies that in order to reduce uncertainties, the risk assessment should be updated as soon as new data becomes available during the drilling and stimulation phase. ~~(e.g., Broccardo et al., 2017a).~~

Despite these challenges, geothermal energy is a highly important renewable energy resource with a low carbon footprint. It has been successfully operated in many areas for decades, and Iceland is a prime example for economically successful and widely accepted use of deep geothermal energy. Several, p~~P~~ast geothermal projects have been successfully managed with classical traffic light approaches (Majer et al. 2007; Bommer et al. 2006; Kwiatek et al. 2019) and simplified ~~a-priori~~ risk assessments. However, classic traffic light systems (TLS) are simple heuristic methods, often based on expert opinions, and their likelihood of success in mitigating seismic risk is not yet clear (Baisch et al. 2019). In fact, there are several notable cases (e.g., Basel, Majer et al. 2007, Mignan et al., 2015; Pohang, Grigoli et al. 2018, to name a few) where classical TLSs have not been successful. In most of these cases, the main event occurred after the project was terminated, i.e., after the mitigation strategy ceased its effects. ~~However~~As a consequence, we consider it important for the future development of geothermal energy near urbanised areas to move beyond the existing state of the technology and develop and implement a robust, quantitative, and coherent risk management framework during all stages of a project, including the post injection.

Within this context, this study represents, to the best of our knowledge, the first publicly available probabilistic seismic risk study prior to a deep geothermal project in Iceland (and one of the very few worldwide). We have attempted to

combine all available risk-related information on the upcoming stimulation of the RV-43 well in Geldinganes into one quantitative and risk-based assessment. This ~~a priori~~*a priori* study, then, represents the basis for risk updating once the project has started and *in ~~-~~situ* real-time data become available. This procedure ideally enables a dynamic risk management solution that will also help to ensure public acceptance, and thus contribute to the continued successful use of deep geothermal energy resources in Iceland and beyond. In details, ~~-~~the key objectives of this *a priori* study are:

- **Interdisciplinarity-based risk:** Integration of hydraulic reservoir modelling, empirical data of past sequences, expert knowledge, ground motion prediction equations, as well as first order exposure and vulnerability information into one quantitative risk assessment.
- **State of knowledge:** Use of methodologies that are well aligned with good practice recommendation of the DESTRESS project (Grigoli et al., 2017, Pittore et al., 2018), with Swiss good practice recommendations (Trutnevyete and Wiemer, 2017) and the recommendations of the international expert committee investigating the Pohang earthquake (Lee et al., 2019).
- **Explicit uncertainty treatment:** Consistent treatment of the uncertainties in knowledge and the variability of the data via use of a logic tree approach. This reflects the current state of practice in probabilistic seismic hazard and risk assessment for natural earthquakes.
- **Transparency and Reproducibility:** The study documents all decisions taken in a transparent and reproducible way. All stakeholders in risk governance thus have access to the same level of information as baseline, and ideally a common understanding of the project's risks.
- "**Updatable:"** Most important, the *a priori* risk assessment can be updated in a consistent way as soon as new data arrives. Because the initial uncertainties are very large, updating it with *in situ* information is a must and should be done in a manner which is fully compatible with the initial risk assessment. The *a priori* risk assessment presented here is thus also a first and critical step toward risk management.
- **Limitations:** This study is conducted without local data, and therefore uses only "off-the-shelf" models and data from different projects. It is therefore essential to clearly outline the assumptions and simplifications so that in the "update and review" phase these limitations can be addressed and the overall approach improved through more sophisticated modelling.

We fulfill these objectives by ~~——~~structuring the paper ~~Our paper is structured~~ as follow: Section 2 describes the site, the geological conditions, and the planned field operations; Section 3 introduces the probabilistic fluid-induced seismic hazard assessment and Section 4 the probabilistic fluid-induced seismic risk assessment; Section 5 discusses the hazard and risk results as well as known limitations.

2 Site description, geological conditions, and planned operations

2.1 Site description

Well RV-43 is located on the Geldinganes peninsula in the northeastern part of the city of Reykjavik Figure 1. OR is the main supplier of heat in Reykjavik and has drilled several wells on Geldinganes. It aims to produce hot water from RV-43 to be directly utilized for heating purposes and to meet the increasing energy needs of Reykjavik.

RV-43 was drilled in 2001; it is 1832 m long, where the last 1130 m are uncased (8½ inches open hole). The well is deviated towards N20°E (on average) and it reaches ~1550 m true vertical depth (TVD). The well is oriented towards the northeast of Geldinganes, an area with exceptionally high geothermal gradients that is closer than the rest of the Geldinganes' wells to the extinct central volcanic system north of Reykjavik and to a possible fault zone (Steingrímsson et al., 2001). Both temperature logs and magnetic measurements are supporting this hypothesis. Except for minor losses close to the bottom of the well no mud losses were observed during drilling of the open hole section of the well. The location of the well RV-43 is shown in Figure 1.

The first and only stimulation of RV-43 took place in 2001 after its drilling. Water of pressure up to 10 MPa was injected along the open-cased segment of the well, and the total injected volume was not documented. However, this can be inferred from the original drilling report, which states that at least 1,900 m^3 were injected and no seismicity observed (suggesting a maximum magnitude threshold $M < 2$, which was the minimum detectable magnitude). After the stimulation, the well had an injectivity index less than 6·10-9 m^3 /Pa·s for the maximum injection's pressure, which is at best half of the required value for commercial exploitation.

2.2 State of stress and structural geology

A first estimate of the state of stress at Geldinganes has been inferred from a global Icelandic stress survey conducted by Ziegler et al., 2016 and by Heidbach et al., 2016. They suggest a potential orientation for σ_{Hmax} of 340˚-40˚ NW-SE, based on 12 geological indicators in a 10 km region around the site. The magnitude of the stress at depth could be extrapolated from shallow hydrofracturing stress measurements. Such tests were conducted in two boreholes (H32 and H18) near Reykjavik on the flank of the Reykjanes-Langjokull continuation of the Mid-Atlantic Ridge (Haimson and Voigt, 1976; Haimson, 1978).

Four tests were conducted in the borehole H32 between 200 and 375 m depth in jointed basalt. The minimum compressive stress, σ_{Hmin}, was found to be horizontal in the range of 4 to 6 MPa, while for the maximum horizontal stress, σ_{Hmax}, it was approximated as varying between 5 to 10 MPa for the four tests. The direction of σ_{Hmax} was calculated based on three hydro-fractures with an orientation N25°W ± 5°. The vertical stress, σ_V, was calculated based on a gradient of 27 MPa/km. These values, if extrapolated to a depth of 1.5 km, suggest a normal stress regime. In the borehole H18 only three tests were performed due to extensive jointing. While the test at 180 m was conducted in basalt, the lower tests at 290 and 324 m were in an

intrusive dolerite. Also, in this case, the minimum principal stress was found to be horizontal (σ_{Hmin}) increasing with depth from 4 to 8 MPa. For the maximum horizontal stress (σ_{Hmax}) it was observed in a range from 12 to 16 MPa, while the vertical stress ranged from 5 to 9 MPa. For these tests, the hydro-fractures suggest contradictory directions, and hence two possible orientations for σ_{Hmax}: N20°E for the 180m test, and N45°W for the 290m test. Extrapolation of the results at 1.5 km depth suggests in this case a strike-slip~~/thrust~~ to reverse regime.

Combining the results of both boreholes, a linear approximation of the data between 200 m and 350 m depth gives σ_{Hmin} = 21 MPa/km, σ_{Hmax} = 3 MPa + 30 MPa/km, and σ_V = 27 MPa/km. As reported by Haimson and Voigt, 1976, the measured stress orientation (H32) has no obvious relationship to the NE strike of individual rift zone fissures and faults, inferred WNW direction of lithospheric plate motion, or axial rift zone earthquake focal solutions which indicate NW-trending. The measured stresses could be related to (i) a hot spot, (ii) local phenomena involving the extinct NNW-trending Kjalarnes central volcano, or (iii) ground distortion due to fluid withdrawal from the Laugarness hydrothermal system. Finally, the two stress measurements in dolerite in borehole H18 could be interpreted as high stress layers. By excluding these two measurements, the linear approximation gives σ_{Hmin} = 2 MPa + 10 MPa/km, σ_{Hmax} = 3 MPa + 13 MPa/km, and σ_V = 27 MPa/km, leading to normal conditions at 1.5 km depth (Hofmann et al., 2020).

2.3 Planned activity

The re-stimulation of well RV-43 is foreseen by the end of October 2019. The re-stimulation is based on a three-staged cyclic pulse stimulation that will last for circa 12 days (4 injection days per stage). In particular, it is expected to enhance productivity in three pre-existing fracture zones penetrated by RV-43 and isolated with straddle packers. Specifically: (i) the first zone is located at 1700-1750 m Measured Depth (MD) that corresponds to 1467-1507 m in True Vertical Depth (TVD), where the basalt intersects with the gabbro and mud losses had been observed, (ii) the second zone is located at 1300-1350 m MD (1150-1189 m TVD), (iii) the third zone is located at depth 1100-1150 m MD (1001-1032 m TVD). Each of the zones will be stimulated with a cyclic injection scheme ("cyclic" stimulation), which repeats every 24h and includes pressurizing RV-43 with pulses of frequency 1/60 Hz ("pulse" stimulation) and continuous injection phases. This is illustrated in Figure 2.

The application of short-term cycles is based on the concept of fatigue-hydraulic fracturing, introduced by Zang et al. 2013; 2017; ~~2018~~2019. In practice, pressure pulses are expected to weaken the rock ("fatigue") by inducing microcracks before macroscopic failure. This mechanism has three major intended benefits: first, the stimulated reservoir volume is increased due to more complex fracture growth, and a larger and denser fracture network provides a larger heat exchanger area; second, the breakdown pressure is reduced, therefore, lower injection pressures are required to stimulate the target formation hence reducing the potential for slip on faults and, thus, the likelihood of induced seismic events; third, the magnitude of the largest induced seismic events is potentially limited. ~~The concept of cyclic stimulation is one of the soft stimulation techniques evaluated by the~~

~~DESTRESS consortium (Pittore et al., 2018).~~ The stimulation of each stage is expected to last maximum 4 days with the following schedule including pre- and poststimulation operations:

- 1/2 day to install the packer,
- 1/2 day to perform injection tests with stepwise flow rate increase,
- 4 days for the main stimulation (with stepwise flow rate increase, and repeating phases of cyclic injection, cyclic-pulse stimulation and continuous injection),
- 1/2 day for performing flowback, where withdrawn water goes to the sea
- 1/2 day for removing the packer and redressing it for the following stage.

Injected water is not expected to exceed rates of 60 l/s or overpressures of 20 MPa at any time during the stimulation due to restrictions by the equipment. No threshold value for injectivity has been reported for stopping the stimulation and stimulations are expected to continue either until the end of the planned injection or until a Traffic Light System (TLS) forces the termination (e.g., Mignan et.al., 2017). The well will be stimulated sequentially from bottom to top.

An exemplary main stimulation for each stage is plotted in Figure 2. During the first day, step rate injection tests are performed for estimating the pressure at which fractures open and to observe the seismic response to increasing flow rates. Based on this, the flow rates of the following phases are determined in order to reach sufficient pressures for stimulation of the target interval. The main part of the stimulation consists of cyclic injection (4 cycles of 1 hr high rate injection and 1 hr low rate injection), cyclic pulse injection (4 cycles of 1 hr high rate injection with pressure pulses and 1 hr low rate injection), and 8 hours of continuous injection. The volume injected in each of the phases is planned to be ~~(circa)~~approximately equal. The flow rates depend on the fracture opening pressure. This procedure is repeated up to three times before the flow rates are reduced slowly and stepwise at the end of the treatment.

2.4 Mitigation strategy

In the presence of fluid-induced seismic risk, it is paramount to efficiently monitor the induced seismicity and define a risk mitigation strategy. In the Geldinganes area, a dedicated microseismic network has been recently installed. The seismic monitoring infrastructure, completed in August 2019, consists of 13 seismic stations one seismic array of 7 seismic stations and one deep borehole array of 17 geophones (Figure 1). The stations send data in real-time to the Iceland GeoSurvey (ISOR), which streams them both to ETH-Zurich and GFZ-Potsdam. Real-time seismic data analysis will be performed using the software package Seiscomp3. Induced seismicity monitoring and risk mitigation operations at Geldinganes are conducted by a team of experienced professionals, including seismologists, field operation managers, reservoir engineers and an internal expert panel (who will support decision during critical situations). The adopted protocol for the Geldinganes TLS is based on a five-steps action plan that governs the fluid injection operations illustrated in Figure 3 and summarized below.

In the case of induced seismic events above a certain threshold, we require a specific action plan. In particular, we first subdivide the region surrounding the industrial site in an internal and external domain.

- **Internal Domain**: It defines the volume surrounding the industrial operations where seismicity will be monitored and analysed with maximum sensitivity.
- **External Domain**: It is a wider volume surrounding the Internal Domain, where the occurrence of seismicity may still be associated with the industrial operations.

For the Geldinganes site, we set these domains as cylinder-shaped volumes with radii from the injection well of 2.5 km and 5.0 km for the internal and external domain, respectively (Figure 1). The range of depth considered for both domains is between 0 and 10 km. These values have been defined by considering other induced monitoring projects and considering the expected uncertainties for the automated locations. The magnitude of completeness of the Geldinganes network, evaluated using the BMC method (Mignan et al., 2011; Panzera et al., 2017), is ~0.3 and ~0.0 for the external and internal domain respectively. Seismic events with $M_L > 0.0$ and occurring within the internal domain should be seen clearly by almost all the stations within this domain. Therefore, all the seismic events above this magnitude threshold will be manually analyzed.

For the external domain, we will manually refine the automated solutions for seismic events with $M_L > 0.5$. Since automatic magnitudes of small events might be overestimated, these thresholds need to be revised by considering the observed seismicity data collected during the early stage of stimulation operations. Then, the analysed data are used to update the risk study and to assess the performance of the monitoring network. These analyses are performed at the early stage of the cyclic stimulation, and injection is increased carefully until at least a few events are detected and located.

~~For seismic events occurring within the internal or external domain, the following alert levels are defined:~~

- ~~**No Alert**: Seismic events with local magnitude below 1.0 either in the internal or external domain do not trigger an alert. *Seismic analysis:* The routine seismologist manually reviews every day the seismic events that occurred in the past 24h and within the two domains. For the internal domain, all the events with magnitude $0.0<M_L\leq1.0$ must be manually reanalyzed, while for the external domain the magnitude range is $0.5<M_L\leq1.0$. *Field operations:* Operations continue as planned, if no anomalous seismicity patterns are identified. Otherwise, the internal expert panel is informed and after an update of the risk assessment, specific case dependent actions are taken (Zoback, 2012). *Communication:* If the seismologist on duty identifies anomalous seismicity patterns in the space-time-magnitude domain, the routine seismologist report to the internal expert panel who decides the actions to take in conjunction with the operation management.~~
- ~~**Green level**: This level is activated if one or more earthquakes within the magnitude range $1.0<M_L\leq1.5$ occur in the internal or external domain. It can be considered as a pre-alert level. *Seismic analysis:* No immediate response is required by the seismologist. However, within few hours (4-6 hours) the routine seismologist reviews the overall~~

characteristics of seismicity (e.g. rate and b-value changes in the Gutenberg-Richter law), including its spatial and temporal evolution during the last 24h. Most importantly, the routine seismologist must check the presence of seismicity patterns resembling potential faults by using more sophisticate techniques. *Field operations:* Operations continue as planned, if no anomalous seismicity patterns are identified. Otherwise, the internal expert panel is informed and after an update of the risk assessment, specific case-dependent actions are taken (Zoback, 2012). *Communication:* If anomalous seismicity patterns are found, the routine seismologist must inform the internal expert panel and the field operations manager.

- **Yellow level**: This level is activated if one or more earthquakes within the magnitude range $1.5<M_L\leq2.0$ occur in the internal or external domain. *Seismic analysis:* Immediate response is required by the seismologist on duty, and within 45 minutes a manual location and magnitude estimation is sent to the operators. Consistency with the Icelandic Meteorological Office (IMO) magnitude needs to be checked. These events, if very shallow (i.e. 1-2 km), can be potentially felt by people living close to the epicenter. *Field operations:* Flow and pressure decrease until seismicity levels remain below the green alert for at least 4 hours. *Communication:* Status report sent to the IMO.
- **Orange level**: This level is activated if one or more earthquakes within the magnitude range $2.0<M_L\leq3.0$ occur in the internal or external domain. *Seismic analysis:* Immediate response is required by the seismologist on duty, and within 45 minutes a manual location and magnitude estimation is sent to the operators. Consistency with the IMO magnitude needs to be checked and, in this case, IMO magnitudes are the preferred estimate that are sent to the operators. These events can be felt, depending on the depth of the event and the distance of the epicenter from urbanized areas. *Field operations:* Stop injection operations and bleed-off of the current stage. Continue with stimulation of other stages if seismicity remains below the green level for at least 12h. *Communication:* Status report sent to regulators, the ICPD, IMO and RE's public affairs.
- **Red level**: This level is activated if one or more earthquakes with local magnitude $M_L>3.0$ occurs. *Seismic analysis:* Immediate response is required by the seismologist on duty, and within 45 minutes a preliminary manual solution is produced. For magnitude estimation, since the closest ones to the epicenter might be saturated, stations with a distance of 10 km from the epicenter are used for the calculation. IMO magnitudes are the preferred and most reliable estimate in this case. These are felt and potentially damaging events. *Field operations:* Immediately stop injections operations and bleed-off taken. The stimulation operations for this well will be discontinued and not started again. *Communication:* Status report sent to regulators, the ICPD, IMO and RE's public affairs, asses potential damage.

3 Probabilistic fluid-induced seismic hazard assessment

Probabilistic risk assessment is emerging as the standard approach to manage and mitigate induced seismicity linked to fluid injections in the underground (Mignan et al., 2015; 2017; 2019a; 2019b; Bommer et al., ~~2017~~2015; ~~Broccardo et al., 2017a;~~ Grigoli et al., 2017; Lee et al., 2019). The need for a probabilistic risk-based approach is motivated by the stochastic nature of earthquakes, the many uncertainties associated with the process of inducing seismicity and the needs of regulators, insurance and the public (Mignan et al., 2019a; 2019b). Both hazard and risk approaches follow standards proposed, among others, by the Swiss Seismological Service (SED, 2017) and related references (Broccardo et al. 2017a, Mignan et al., 2015; 2017; 2019a; 2019b), which are based on a combination of Probabilistic Seismic Hazard Analysis (PSHA) and the PEER-PBEE framework (Cornell, 1968; Cornell, Krawinkler, 2000).

PSHA is assessed as the probability of exceeding a given intensity at a given distance R from the injection site, based on the number of events above a given minimum magnitude m_0, the frequency distribution of the magnitude (namely the truncated Gutenberg-Richter distribution), and an empirical ground shaking attenuation function. The latter can be an intensity prediction equation (IPE) based on felt intensity, or a ground motion prediction equation (GMPE) based on peak ground acceleration (PGA), spectral acceleration (SA) or peak ground velocity (PGV). Commonly, within this probabilistic framework, there are two main elements to be defined: (i) the probabilistic characterization of the seismogenic source model(s), and (ii) the ground motion characteristic model(s) (describing the expected ground vibration given the occurrence of an earthquake). The first gives the temporal and spatial forecast of the earthquake ruptures, while the second is characterized by GMPEs to link the earthquake rupture with the expected ground shaking at the site of interest.

The output of PSHA analysis is the rate of exceedance or hazard curves (probability of exceedance for a given period of time) of a given ground shaking Intensity Measure (IM) type. A single curve (for a given set of parameters) represents the aleatory (irreducible) variability within the defined model. To include also the epistemic uncertainties, given the alternative possible models, a logic tree structure with weighted branches (indicating the belief in a given model) is defined (e.g., Mignan et al., 2015). Figure 4 shows the proposed logic tree adopted for this ~~a priori~~*a priori* risk analysis. The first level of the logic tree defines the seismogenic source models, the second level the upper bound of the Gutenberg-Richter distribution, the third level the GMPEs and the ground motion intensity conversion equations (GMICEs). In the following, we report a detailed discussion for each level of the logic tree.

3.1 Seismogenic source models

In this analysis, we assume that induced seismicity nucleates and eventually extends in the proximity of the injection point. Therefore, a point source located at the coordinates of the injection point is used as the unique seismogenic source model for the

investigation. This implicitly excludes any geometrical uncertainty on the location of the hypocenter. Forecasting the number of events that will occur in a reservoir stimulation is difficult because (as previously stated) the stressing conditions and location of faults near the injection point are unknown. Empirical data from similar sites can be used as a first-order proxy, but in the case of Geldinganes, only limited experience exists. In light of these limitations, we argue that the spatial variability of the seismicity is well constrained by a simple seismogenic source that can be updated for real-time application.

The number and size of earthquakes in PSHA analysis is based on three parameters that describe the local seismic activity rate, the event size distribution, and the largest event size (Cornell, 1968). These parameters are typically constrained based on observed seismicity with the activity rate broadly scaling with the seismo-tectonic strain input. For induced seismicity, the seismogenic models likewise must also describe the local seismic productivity that is (in this case) linked to the injection profile. Such seismicity rates are unknown, albeit the hydraulic energy input might be estimated beforehand. Moreover, the link between induced seismicity and stress release is a key factor to be considered in the analysis. The fraction of seismic to hydraulic energy may thus vary from zero (no events observed) to well above 1 (sometimes referred to as "triggered" events that release mostly pre-accrued tectonic stresses, e.g., Pohang, Grigoli et al. 2018). We will consider two simple seismogenic source models to analyze this uncertainty in energy release and have a first-order forecast of the underground response to injection:

- **Model SM1:** A seismogenic source model that assumes the underground feedback is site-specific constant, with all parameters purely data-driven (Dinske and Shapiro, 2013; Mignan et al., 2017; Broccardo et al., 2017a).
- **Model SM2:** A seismogenic source model that simulates the fluid and overpressure propagation for the planned injection protocol based on one-dimensional diffusion and stochastically distributed seeds. Model SM2 will also explicitly use the observation of no seismicity at $M >= 2$ during the first stimulation in the year 2001 as a constraint. The synthetic catalogues are then converted onto the same underground feedback site-specific parameters of Model SM1 (Karvounis et al., 2014; Karvounis and Jenny, 2016).

These two models capture to a first-order the epistemic uncertainty in forecasting seismicity, since they express alternative approaches to forecasting (purely empirical and partially physics-based). Both models are equally weighted to estimate the ground shaking estimation at the site of interest.

3.1.1 Seismogenic source model SM1

The seismogenic source model SM1 assumes that the "seismic underground feedback" per volume affected by significant pore-pressure change is a site-specific (and generally ~~a priori~~ unknown *a priori*) constant. This constant can vary by several orders of magnitude between sites. Because the volume affected scales with the volume of fluid injected (and in theory to the pressure applied; Mignan, 2016; Langenbruch et al., 2018), this implies a relation between the expected number of earthquakes $E[N]$ and the volume injected V, as

$$E[N(t); M > m] = \begin{cases} 10^{a_{fb}-bm} V(t), t \le T_{in} \\ 10^{a_{fb}-bm} \tau \exp\left(-\frac{t-T_{in}}{\tau}\right) \dot{V}(T_{in}), t > T_{in} \end{cases}, \quad (1)$$

where a_{fb} is the underground feedback parameter (i.e., the overall activity for a given volume V, which is better known as the seismogenic index Σ; e.g., Dinske and Shapiro, 2013), b is the slope of the Gutenberg-Richter distribution, T_{in} is the injection duration, and τ is the mean relaxation time of a diffusive process. This linear relation (during the injection phase), first described by Shapiro's group, is broadly accepted in the technical community as a first order model (e.g., Dinske and Shapiro, 2013; van der Elst et al., 2016; Mignan, 2016; Broccardo et al., 2017a). We should mention that we use the generic term a_{fb} instead of Σ to remain agnostic as to the physical origin of this linear relationship. The seismogenic index infers a poroelastic origin (e.g., Shapiro and Dinske, 2009) although other drivers, such as overpressure field geometry, can also explain the linearity observed between V and N (Mignan, 2016). ~~(i.e., the overall activity for a given volume V, which is also known as seismogenic index Σ; Dinske and Shapiro, 2013), b is the slope of the Gutenberg Richter distribution, T_{in} is the injection duration, and τ is the mean relaxation time of a diffusive process. This relation (during the injection phase) is broadly accepted in the technical community as a first order model (e.g., Dinske and Shapiro, 2013; van der Elst et al., 2016; Mignan, 2016; Broccardo et al., 2017a).~~ The post injection phase has been added by Mignan et al., 2017, to account for the decrease of the rate of seismicity after the injection has been terminated (trailing effect) (Mignan, 2015). This model has been verified for a number of fluid injection experiments, in terms of flow rate $\dot{V}$ versus induced seismicity rate $\lambda(t, M > m)$ (Mignan et al., 2017). This allows a refined analysis by defining the rate function

$$\lambda(t, M > m) = \begin{cases} 10^{a_{fb}-bm} \dot{V}(t), \text{for } t \le T_{in} \\ 10^{a_{fb}-bm} \dot{V}(t) \exp\left(-\frac{t}{\tau}\right), \text{for } t > T_{in} \end{cases} \quad (2)$$

which allows to define a Non-Homogeneous Poisson Process (NHPP). Note that this model only applies to the stimulation phase in which the fluids injected are not supposed to be produced back, hence creating an overpressure field at depth z. In practice, after each stimulation stage parts of the injected fluid will be produced back by natural bleed-off (without pumping) directly after each stage and by airlift testing after the end of the last stage. Note also that $E[N(t), M > m] = \int_0^t \lambda(t, M > m) dt$ and the expected total number of fluid-induced earthquakes is $E[N(\infty); M > m] = 10^{a_{fb}-bm} \left(V(T_{in}) + \tau \dot{V}(T_{in})\right)$.

While the parameters a_{fb}, b and τ can be estimated during the stimulation (Mignan et al., 2017; Broccardo et al., 2017a), ~~*a priori*~~*a priori* knowledge on those parameters is limited and the range of possible values wide. Given this state of "uniform" uncertainty, we assigned equal weights to all the possible $[a_{fb}, b]$ combination. We list the a_{fb}, b parameter estimates for different sites in Table 1, which will be used as input for the ~~*a priori*~~*a priori* risk study. Uncertainties are likely to significantly reduce once seismic data is obtained by monitoring during the stimulation. Note that due to the correlation of a_{fb} and b (Broccardo et al., 2017a), pairs of (a_{fb}, b) values from different sites need to be maintained. In Mignan et al., ~~(~~2017~~)~~, the mean relaxation time has been observed to widely vary between injection sites with $0.2 < \tau < 15$ days.

In order to apply a classical PSHA analysis, we transform the NHPP into a Homogeneous Poisson Process (HPP), using the equivalent rate $\Lambda_{M>2} = E\left[N\left(T_{in} + T_{p_{in}}T\right), M > 2\right] = \int_0^{T_{in}+T_{p_{in}}T} \lambda(t)dt$, for a unit of time which corresponds to the total project period (including the post injection phase), i.e., $T = T_{in} + T_{p_{in}}$, where $T_{p_{in}}$ is the post injection time. We selected $M > 2$ because we assume that lower magnitudes will not have the potential to trigger any damage. By doing so the $P\left(M > 2; T_{in} + T_{p_{in}}T\right) = 1 - \exp\left(-\Lambda_{M>2}\right)$. Table 1 and Figure 5 (red dots) report the equivalent rate $\Lambda_{M>2}$ for each project, for a target injected volume of circa V = 18,000 m^3 (estimated from c. 6,000 m^3 injection per stimulation, times 3 stimulations; Figure 2). At the present time, without any pre-stimulation phase it is not possible to infer where the Geldinganes project is placed in this domain; however, what is known is that 5000 m^3 of water were injected and no seismicity observed.

3.1.2 Seismogenic source model SM2

With model SM2, a first-order physical process is included into the forecasting. This is done by modeling pressure diffusion through a fractured media containing randomly distributed earthquake faults (so called "seeds"). The pressure propagation can be adopted based on the reservoir properties, limited to the available information. Then, the density of these seeds and their size distribution are treated as free site-specific parameters that (again) are unknown ~~*a priori*~~*a priori*. These models are commonly referred to as "hybrid" models (Gischig and Wiemer, 2013; Goertz-Allman and Wiemer, 2013) as they combine deterministic and stochastic modelling. Specifically, the adaptive Hierarchical Fracture Representation (a-HFR) is employed both for modeling flow in a fracture network with dynamically changing permeability (Karvounis and Jenny, 2016) and for simulating the source times of randomly pre-sampled scenarios of hydro-shearing events at certain hypocenters (Karvounis et al., 2014). This hybrid model is chosen here, as it can integrate several of the field observations, returns forecasts both of the spatial distribution of seismicity and of its focal planes, and can forecast reservoir properties like the expected well's injectivity at the end of the injection.

The required inputs for the proposed hybrid model are the initial hydraulic properties, the planned activities, a first-order knowledge of the stress conditions in the proximity of the well, and the orientations of pre-existing fractures. Here, the extrapolated stress measurements described in Section 2.2 are employed without excluding any of the measured stresses; i.e.,

the vertical direction is a principal direction and σ_{Hmin} = 21 MPa/km, σ_{Hmax} = 3 MPa + 30 MPa/km, and σ_V = 27 MPa/km. The planned activities are described in Section 2.3, the initial transmissibility is in agreement with the currently expected injectivity of $6{\cdot}10^{-9}\ m^3/(Pa \cdot s)$ along the whole open segment of RV-43, and the compressibility is inferred from the nearby well HS-44, since there are no reported measurements of the characteristic time at RV-43. Moreover, in this ~~a priori~~*a priori* analysis, all surface orientations are considered equally probable. Observe, however, that two distinct sub-vertical fault sets seem to prevail at the mainland surrounding the bay above the injection stages (Hofman et al., 2020).

We assume a constant value for those parameters for which the rate of seismicity is less sensitive. These are: the friction and the cohesion of fractures for the Mohr-Coulomb failure criterion which are fixed equal to 0.6 and 0, respectively, and the mechanical aperture of fractures after they have slipped is 1 mm. The remaining parameters (those for which seismic rate is more sensitive) are the spacing between pre-existing fractures, the *b*-value of the Gutenberg Richter during injection, and the permeability of a slipped surface. A range of possible values is considered for each of the latter parameters; i.e., the spacing between pre-existing fractures (1 m, 2 m, 4 m, 8 m, 16 m), the *b*-value of the Gutenberg-Richter law (0.75, 1., 1.25, 1.5, 1.75, 2.), and the hydraulic aperture between the two parallel fracture surfaces (200 μm, 500 μm), out of which aperture the equivalent permeability can be estimated. A synthetic catalogue of rate $\Lambda_{M>2}$ is created for each possible combination of the above values ~~.~~and an equivalent $\Lambda_{M>2}$ is computed. Next, following the same logic of Section 3.1.1, we assign equal weights to each $a_{fb} - b$ combination. The resulting scenarios are shown in Figure 5 as blue dots in a_{fb}-b space. Note that all scenarios are above the dashed line that indicates the limit posed by the no-seismicity observation during the initial stimulation.

3.2 Upper bound for the Gutenberg-Richter distribution vs. Maximum observed ~~-~~Magnitude distribution

The frequency magnitude distribution of natural and induced earthquakes follows (to a first order) the classical Gutenberg-Richter distribution. This distribution is truncated at an upper-end for energy conservation, but also because existing faults and fault systems have a maximum size ~~(e.g., Mignan et al., 2015)~~ .(e.g., Bachmann et al., 2011; Mena et al., 2013; Mignan et al., 2015; Baker and Gupta, 2016). ~~-~~Empirical observations will only poorly constrain the largest possible earthquake, since it is by definition an exceptionally rare and extreme event. One of the major sources of uncertainty is thus in PSHA related to the upper bound of the (truncated) Gutenberg-Richter distribution, here indicated as m_{maxsup}. ~~However, it is also true that in most seismic hazard and risk studies, the actual value of m_{sup} is not a critical choice, since its rate of occurrence is typically very small compared to the typical return periods of interest. It follows that (in general) both hazard and risk are typically dominated by the more frequent, moderate-size events.~~

It is generally accepted that the largest possible induced earthquake cannot be larger than the tectonically largest one. However, in induced seismicity, the tectonic environment (controlled primarily by the state of stress) at a site may be such that

no tectonically prestressed larger ruptures exist. Under these conditions, ruptures will be running out of energy once they leave the volume brought into a critical state for failure by the injection—e.g., because of the effect of pore pressure on the Coulomb failure criteria. In these conditions, run-away ruptures cannot occur even if a natural fault exists (also referred to as "triggered" events cannot happen), and the largest magnitude size, as a consequence, is limited by the volume or area affected by overpressure (which again scales with the volume of fluid injected and the hydraulic properties of the subsurface). ~~In these conditions, run-away ruptures cannot occur even if a natural fault exists (also referred to as "triggered" events cannot happen), and the largest magnitude size as a consequence is limited by the volume or area affected by overpressure, which again scales with the volume of fluid injected (or extracted) and the hydraulic properties of the subsurface.~~ In such situation, m_{maxsup} can locally be substantially smaller than the regional tectonic one. This is common in "fracking" operations in tight shales. McGarr ~~(~~1976~~,~~ 2014~~)~~ formalized this volume limit as $m_{~~sup~~maxP,McGarr} = 2/3 \log_{10}(GV) - 10.7 + 14/3$ where $G = 3 \cdot 10^{10}$ Pa is the modulus of rigidity. McGarr has shown that this relationship is consistent with the data from a compilation of injections. However, a number of researchers (Gischig et al., 2014; van der Elst et al., 2016; Mignan et al., 2019b) have pointed out that outliers exist (e.g., Pohang, South Korea, Grigoli et al., 2018, Kim et al., 2018; St. Gallen, Switzerland, Diehl et al., 2017) and that the McGarr limit is best explained as a purely statistical relationship based on simple extreme value theory principles (Embrechts, 1999).

The "McGarr limit" has been used (and in some cases one might argue misused) in numerous induced seismicity hazard assessments (van der Elst et al., 2016). For V = 18,000 m^3, we would for example obtain $m_{~~sup~~max,McGarr} = 3.79$. Based on the recent statistical tests of van der Elst et al. ~~(~~, 2016~~)~~, and the occurrence of the 2017 Pohang earthquake above the expected limit (Grigoli et al., 2018, Kim et al., 2018), a fixed McGarr limit now appears questionable to many seismologists. Therefore, since the complete information about the number, location, size and stressing condition of faults in the Geldinganes area is not available (in particular before the stimulation phase), it is appropriate to consider $m_{~~s~~maxup}$ as the regional tectonic $m_{~~sup~~max} = 7$ (Kowsari, et al. 2019). This estimate could be reduced at a later stage if local fault information were found to provide better constraints. In fluid-induced seismicity, when m_{max} is related to the tectonically largest event, it is not a critical choice (Gupta and Baker, 2017). This because the rate of occurrence is typically significantly low compared to the typical return periods of interest. It follows that both hazard and risk generally are dominated by the more frequent, moderate-size events.

In addition, in this study, we determine the probability distribution of the maximum observed magnitude, $M_{~~max~~T}$ [1], ~~observed~~ at a fluid injection sites for the total time of observation T (Holschneider, et al. 2011). This is fundamentally different

[1] In van der Elst et al., 2016, and Broccardo et al. 2017a ,the random variable M_{max} coincides with the random variable M_T used in Holschneider et al., 2011 and adopted in this paper.

from the upper bound m_{maxsup} of the Gutenberg-Richter distribution, which is merely a deterministic upper limit fixed by physical constrains. The probability distribution of the maximum magnitude, $M_{maxT} = max\left[M_1, \dots, M_{t_i}, \dots, M_{T_{in}+T_{pin}T}\right]$, can be easily derived considering the magnitude events statistically independent. It follows that $F_{M_{maxT}}(m|N=n) = F_M(m)^n$, where $F_M(m)$ is the classical Gutenberg-Richter cumulative probability density function, and $f_{M_{maxT}}(m|N=n) = nF_M(m)^{n-1}f_M(m)$. Since the number of events is a random variable itself, then $F_{M_{max}}F_{M_T}(m) = \sum_n F_{M_{maxT}}(m|N=n)P(N=n|\Lambda(T_{in}+T_{pin}T))$, $f_{M_{max}}(m|N=n) = \sum_n nF_M(m)^{n-1}f_M(m)P(N=n|\Lambda(T_{in}+T_{pin}T))$, where $P(N=n|\Lambda(T_{in}+T_{pin}T))$ is the classical Poisson discrete distribution. Figure 6-a,b shows the equivalent rate of seismicity, $\Lambda(T_{in}+T_{pin}T)(1-F_{M_{max}}(m))$, (i.e., a weighted complementary CDF) for each of the project reported on Table 1 (SM1 model) and for each of the synthetic catalogues (model SM2). We can observe a large scatter of rate of seismicity reflecting the large uncertainty exiting prior to the project.

Together with the distribution of M_{maxT} for each $a_{fb}-b$ couple, we report the envelope distribution computed as the mean value over all the branches of the logic tree (Figure 4). Figure 6-c shows the envelope distribution. Observe that given the sparse dataset (Table 1), this distribution is (inevitably) multimodal. The expected $E[M_{maxT}]$, based on this envelope distribution, is 2.25 and the 5-95% interval is $[0.10-4.45]$. It is important to highlight that these values represent some statistics based on previous projects and *not* the expected values for this project. In fact, here, the envelope distribution represents a prior distribution, which must be updated during a pre-stimulation phase and during the stimulation. In the following, we also report the envelope distribution of M_{maxT} based on the synthetic catalogue derived according to the SM2 source model. Figure 6-d shows the envelope distribution. Different from the envelope distribution based on the SM1 source model, this distribution shows a more regular shape, since the synthetic dataset is denser and more confined. However, this prior distribution can be affected by overfitting since it is based on stress measurements (without considering their uncertainties) that might not represent the current local condition correctly. The expected $E[M_{maxT}]$, based on this envelope distribution, is 2.09 and the 5-95% interval is $[1.19-3.42]$. Finally, Figure 6-e we reported the $E[M_{maxT}]$ and [5-95]% confidence bound as function of the injected volume.

3.3 Ground Motion Prediction Equations and Intensity measures

The relationship between the site source characteristics and given ground shaking intensity measure types, IMs, is given by seven ground motion prediction equations (GMPEs). Kowsari et al. (2019) provide a set of adjusted GMPEs that have been selected for this investigation. In particular, the proposed GMPEs were adjusted using newly compiled ground motion records of six strike-slip events in the South Iceland Seismic Zone (SISZ), with a range of magnitudes between $M \in$

$[5, 6.5]$ (M is intended as M_w as used in Kowsari et al., 2019), and distance $R \in [0,80]$ km. The intensity measures are reported in Table 2, and the value of the functional form and the coefficients can be retrieved directly from Kowsari et al., (2019). Observe that from the original list we replaced the proposed GMPE of Lin and Lee, (2008) for North Taiwan with the local GMPE (RS09), Rupakhety and Sigjörnsson, (2009), which is consistent with the strike-slip nature of Icelandic earthquakes. The recalibration has been performed only for the PGA; therefore, in the following, we assume only this physical intensity measure. The selected site-to-source distance is the Joyner-Boore metric (R_{JB}) (i.e., the closest horizontal distance to the vertical surface projection of the fault). When the distance metric of the original GMPE is different from R_{JB}, the same transformations proposed in Kowsari et al., (2019,) are applied. In the Supplement, Figure S1, we show the trellis plots for the selected GMPEs models.

It is important to highlight the limitation of these choices. First of all, the GMPEs are calibrated for natural events that are considerably larger in magnitude compared to the expected fluid induced events (Figure 6). Therefore, the extrapolation to lower magnitudes is ~~probably~~ biased (Bommer et al., 2007, Baltay and Hanks, 2014). This will have a more significant effect on the low damage threshold, while the IR computations are less impacted since they depend on larger events. Moreover, for small events in the proximity of the injection point, the ideal source-to-site distance is the hypocentral distance and not R_{JB} (observe that in this case R_{JB} converges to the epicentral distance), which neglects the hypocenter depth. As a consequence, this analysis is independent of the injection depth. Again, this limitation has an impact on the small damage threshold, since the depth of the events is expected to have a significant influence.

In this ~~a priori~~*a priori* assessment, we use as final IM the European Macroseismic Scale (EMS98, Grünthal, 1998)[2]. The advantage of EMS98 over instrumental intensity measures~~over physics-based intensity measures~~, in this phase, lies in the easier interpretability of this scale, which is based merely on shaking indicators expressed in terms of damage and nuisance to the population. Based on these considerations, the selected GMPEs are converted into expected intensity by using GMICEs for small-medium intensities. The GMICEs used in this work are introduced by Faccioli and Cauzzi, (2006,) and Faenza and Michelini, (2010). The aleatory variability is then combined into a GMPE-GMICE model with σ_{TOT} defined as $\sigma_{TOT} = \sqrt{(\sigma_{GMPE}^2)a^2 + \sigma_{GMICE}^2}$, and values of the mean, σ_{GMPE}, σ_{GMICE}, and a reported in the Supplement together with the combined trellis plots (Table S1, Figure S2).

3.4 Probabilistic hazard results (PGA, EMS98)

The hazard integral is reduced to the marginalization of the random variable magnitude, M, and the conditional random variable $IM|M = m$, since the site-to-source distance is fixed by the source point (which is assumed at the injection point). For a given

[2] Observe that in this study we make the same assumption/approximation of Faccioli and Cauzzi 2006, i.e. $I_{MCS} = I_{EMS98}$.

site, then the rate of exceedance is simply reduced to $\Lambda(im;T,b) = -\int_m P(IM > im|M = m,r)\, d\Lambda_{M>2}(m;T,b)$, where $d\Lambda_{M>2}(m;T,b) = \Lambda_{M>2}(T)\, F(m)$, with $F(m)$ equal to the Gutenberg-Richter above magnitude 2. Given the discussion in Section 3.1.1, the probability of exceedance of an intensity, $IM = im$, for a given time period (which corresponds to the total duration of the project given the normalization introduced in Section 3.1.1) is given by the Poisson distribution as $P(IM > im, t = T) = 1 - \exp\big(-\Lambda(im;T,b)\big)$.

$\Lambda_{M>2}(T)$ is not known ~~a priori~~*a priori* (neither an uncertainty quantification based on local condition can be carried out ~~a priori~~*a priori*), therefore we compute the risk for each of the a_{fb} and b pairs of Table 1. As mentioned in Section 3.1.1 the scatter is very large and this reveals the state of deep uncertainty existing prior to a pre-stimulation phase. The PSHA outputs are shown in red in Figure 7 for both the PGA, (top panel) and for the IM (bottom panel). These curves confirm the state of deep uncertainty in particular for the location in proximity of the injection point. In fact, for a given probability of exceedance of 10^{-4} and distance 2-5 km from the injection point, the macroseismic intensity range between the 10% and 90% percentile is circa $IM \in [6, 11]$. In addition, we report the PSHA analysis based on the source model SM2. The outputs are shown in blue in the same Figures. The epistemic uncertainties of the source model SM1 are considerably higher than the ones arising from the source model SM2. This was expected given the inherent sparsity present in the data of Table 1. Moreover, the epistemic median of the source model SM1 is higher than source model SM2. In addition to the PSHA output, in the Supplement (Figure S3), we reported the hazard-based scenarios for different magnitude.

4 Probabilistic fluid-induced seismic risk

In seismic risk assessment, it is common to distinguish between physical and non-physical risk. Examples (and precedents) of non-physical risk include noise, vibrations felt, opposition by residents, public campaigns against the project, *etc.* Non-physical risk is complex and often impossible to quantify. Therefore, an effective and practical approach should focus on non-physical risk identification and mitigation rather than risk assessment (Bommer et al., 2015). Conversely, the physical risk faced by exposed communities needs a quantitative assessment. In this study, we focus only on one type of physical risk which is the seismic risk.

The physical risk is commonly divided into two major categories, i.e., fatalities and/or injuries, and economic losses (with both categories depending on physical damage to buildings). ~~.~~The ~~a priori~~*a priori* risk analysis for the Geldinganes project here focuses on the first risk, while the aggregate economic losses are not directly computed. Here, as a substitute for aggregate losses, we define a low damage threshold for statistical average classes of Icelandic buildings. In particular, in this study, we select two risk measures: Individual Risk (IR) and Damage Risk (DR). IR is defined as the frequency over the time span of the project (including the post-injection phase) at which a statistically average individual is expected to experience

death or a given level of injury from the realization of a given hazard (Jones, 1992; Jonkman, 2003; Broccardo et al., 2017b). We here define DR as the frequency over the time span of the project (including the post-injection phase) at which a statistically average building class is expected to experience light non-structural damage from the realization of a given hazard.

Since there are currently no universally used regulatory and industry approaches to manage induced seismicity of geothermal and other energy projects, we define the following safety thresholds for IR and DR. The proposed IR safety threshold is $IR^{ST} = 10^{-6}$. This value is more conservative~~in line~~ compared to~~with~~ the typical standards for anthropogenic activities for example in Switzerland or the Netherlands~~(although in the original definition the time span is a year), and it has been used, for example, in the induced seismicity case of Groningen, Netherlands, (with yearly frequency(,~~ van Elk et al., 2017). In the presence of epistemic uncertainties, the median of the IR distribution is taken as the reference metric to be compared with the selected safety standard, i.e. $q_{IR,.5} \leq IR^{ST}$ (where $q_{IR,.5}$ is the epistemic median of the individual risk distribution). The proposed DR threshold is $DR^{ST} = 10^{-2}$. As for the IR, in the presence of epistemic uncertainties, the median of the DR distribution is taken as reference metric to be compared with the selected safety standard, i.e $q_{DR,.5} \leq DR^{ST}$ (where $q_{DR,.5}$ is the epistemic median of the individual risk distribution).

The framework used for the computation of IR and DR is based on the convolution of the hazard model with the vulnerability models for the relevant building types, and (only for the IR) with the consequence model.~~for the relevant building typologies with the exposure model.~~ For the fragility-vulnerability model, we should base our analysis on local functions. However, at the present time there exist only local fragility functions for low damage (Bessason and Bjarnason, 2015). Given that, we decide to use the macroseismic intensity approach for IR (Lagomarsino and Giovinazzi, 2006), while using the local fragility function for DR.

4.1 Individual risk computation

For IR, we use a vulnerability given in terms of macroseismic intensity, which follows the macroseismic approach for damage assessment (Lagomarsino and Giovinazzi, 2006) and modified in Mignan et al., 2015, for the induced seismicity case.~~ In this approach, the vulnerability is not defined based on detailed mechanical models; therefore, it is implicitly assumed that macroseismic and mechanical approaches produce compatible levels of damage.~~ The macroseismic model defines the mean damage grade, $\mu_D(im)$, as function of a vulnerability index V, a ductility index, Q, and a reduction factor α introduced in Mignan et al., 2015, to recalibrate low damage states to the damage observed in the Basel 2006 sequence. The vulnerability index depends on the building class and construction specifics, and it includes probable ranges $V^{-}V^{+}$, as well as less probable ranges $V^{--}V^{++}$. Following the Icelandic exposure information described in Bessason and Bjarnason, 2016, we select three building typologies: Concrete, Wood, and Masonry as a surrogate for Pumice buildings. Moreover, Bessason and Bjarnason, 2016, observed that (in average) the Icelandic buildings are stronger and more reliable than the ones based in the Mediterranean region in Europe. Based

on these considerations, we select V_0 as vulnerability index for Concrete and Wood, and V^- for masonry. The choice of V^- for masonry is given by the observation that the fragility of this building is close to old (before the 1980s) Icelandic reinforced concrete building. Moreover, there is no detailed information on the ductility index for the different class of building, therefore we use $Q = 2.3$, which is the value for masonry structures and reinforced concrete structure with no seismic details. In this phase, this a practical and conservative choice since $Q = 2.3$ is a lower bound of the possible ranges of values for the ductility index. In the Supplement, we report in Table S2 the vulnerability indices together with vulnerability functions (Figure S4) obtained by using the macroseismic model with parameters $V_0 = 0.5$, and $Q = 2.3$.

We computed the marginal IR considering all $[a_{fb}, b]$ couples in a given location (i.e., different distances), for the total duration of the project (including the post injection phase), for both rate models SM1 and SM2, and using the HAZUS consequence model (Galanis et al., 2018; Hazus MH MR3, 2003). The results are shown in Figure 8 for each building class. Median and quantiles are computed considering a 50% weight for the SM1 model and 50% weight for SM2 model. Despite the median for each class being below the fixed threshold, i.e. $q_{IR,.5} \leq IR^{ST} = 10^{-6}$, the uncertainty is very large; indicating that uncertainty quantification updates are necessary to reduce the $[a_{fb}, b]$ uncertainties. Moreover, we would like to highlight that the median based merely on SM2 is considerably lower than the median based on SM1 (this is not reported in Figure 8 for clarity). This is due to the lower variability on $[a_{fb}, b]$ arising from the synthetic catalogue.

For DR, we use the local fragility model developed by Bessason and Bjarnason, 2016. Three major categories of buildings characterize the Icelandic exposure model: reinforced concrete, timber, and hollow pumice block. Further details on the exposure model are given in the Supplement (Section S3). Within these categories, Bessason and Bjarnason, 2016, define the following subcategories:

- Low-rise reinforce concrete
  - **RC-b80**: Reinforced concrete structure designed before seismic code regulations (before 1980).
  - **RC-a80**: Reinforced concrete structure designed after seismic code regulations (after 1980).
- Low-rise timber structures
  - **T-b80**: Timber structure designed before seismic code regulations
  - **T-a80:** Timber structure designed after seismic code regulations
- Hollow pumice blocks (HP)

Fragility functions are provided for all these categories only for small damages (which makes the use for IR impossible). Fragility functions details and damage-based scenarios for different magnitudes are reported in the Supplement (Section S3).

Finally, we computed the marginal DR considering all the $[a_{fb}, b]$ couples (for both the source model SM1, with weight 50%, and SM2 with weight 50%) in a given location (i.e., different distances), for the total duration of the project (including the post injection phase). The results are shown in Figure 9 for each class of buildings. Again, despite the median for each class

being below the fixed threshold, i.e. $q_{DR,.5} \leq DR^{ST} = 10^{-2}$, there is a need for uncertainty quantification updates to reduce the $[a_{fb}, b]$ uncertainties.

4.2 Sensitivity analysis

In this Section, we describe a sensitivity analysis of the epistemic uncertainties with respect to the Quantities of Interest (QoI) IR and DR. Specifically, we analyze the sensitivity to the earthquake rate model and the GMPE (and GMICE), which (here) are the only source of epistemic uncertainty. The goal of this sensitivity analysis is to calculate which source of the two input uncertainties is dominant. Specifically, we performed two sensitivity analysis, one for the dataset in Table 1, and one for the synthetic catalog. This allows to better understand the relative contribution of the input uncertainties for each dataset.

In this study, we adopt a screening method which aims to a preliminary and qualitative analysis of the most important input parameter. In particular, we develop a modified version of the Morris method (Morris, 1991), which solves some drawbacks of the "tornado diagram" (Porter et al., 2002) used in Mignan et al., 2015. A "tornado diagram" is a type of sensitivity analysis based on a graphical representation of the independent contribution of each input variable to the variability of the selected QoI. Specifically, given a base model, for each considered variable, we estimate the maximum positive and negative swing of the QoI while holding all the other parameter fixed to their base value. A drawback of the method is that results are strongly dependent on the base model (i.e., it is a local sensitivity method). Therefore, we introduce a variation of the method to obtain a global sensitivity measure. The complete details of the introduced method are reported in the Supplement (Section S4), while here we discuss the general principles and the results. To obtain a global sensitivity measure, we first define a normalized local sensitivity measure of the parameter i with respect to the base model j, $d_i(j)$ (Eq. S1 of the Supplement). Then, we define two global sensitivity measures: the average, μ_{d_i}, and the maximum, $\bar{d}_i$, of $d_i(j)$ (Eq. S1 and S2 of the Supplement). The sensitivity measure μ_i describes the average relative contribution of the parameter i over all possible base models j. The sensitivity measure $\bar{d}_i$ describes the maximum contribution of the parameter i over all possible base models j. The two measure in this form are not normalized to one.

Given IR and DR, Figure 10 and Figure 11 show the sensitivity results based on μ_{d_i} for each building class. Both measures show the same pattern. In particular, the dominating source of uncertainty is (as expected) the rate model. It is interesting to remark that the rate model contribution is more dominant for the dataset based on Table 1 than for the synthetic dataset. Moreover, this fact is consistent across different building typologies an risk metrics. This corroborates the observations that we have previously made, i.e. the uncertainties related to real data are larger than the synthetic ones (which might be affected by overfitting).

The same trend is observed for $\bar{d}_i$, (Figure S8, S9 in the supplementary material.

5 Discussion and Conclusion

This study represents the summary of a collective effort for assessing the a priori seismic risk for the hydraulic stimulation of a geothermal well on Geldinganes, Iceland. The key findings of the ~~risk~~ assessment are shown in Figure 8 and 9 and summarized below:

- The overall risk for an individual to die in a building within a radius of 2 km around the well (Figure 8) is assessed to be below 10^{-7} or at 0.1 micromort (1 micromort = unit of risk defined as one-in-a-million chance of death). This value is within the acceptable range when compared to acceptance criteria applied in the Netherlands (or Switzerland). Reason for the acceptable risk is ~~the estimated low vulnerability of the building stock,~~ the overall quite limited injection volume ~~and~~, the fact that the initial stimulation has not produced M>2 seismicity, and the estimated low vulnerability of the building stock.
- The chance of damage to buildings is around 0.1% (Figure 9) and therewith below the 10^{-2} acceptance threshold we have arbitrarily introduced for damage.
- The thresholds proposed in the classical traffic light (Figure 2) are consistent with the risk thresholds computed; it is not suggested to define more conservative TLS thresholds at this point.
- The uncertainties at this stage of the project are very high, highlighting the importance of updating the risk study continuously as new data becomes available.
-

Based on the following results and the mitigation strategies summarized in the following document, we suggested proceeding with the project. However, based on the online updates of the risk model, we recommended a (possible) review of the analysis. In particular, if the median of the IR and DR grow close to the assigned limits, we ~~recommended~~prescribed a refinement of the study ~~which aims~~ to address the current limitations.

~~In this study, we have attempted to combine all available risk-related information on the upcoming stimulation (middle October 2019) of the RV-43 well in Geldinganes into one quantitative assessment. Our key objectives were:~~

- ~~**Interdisciplinarity-based risk:** We integrated hydraulic reservoir modelling, empirical data of past sequences, expert knowledge, ground motion prediction equations, as well as exposure and vulnerability information into one quantitative risk assessment. Thereby, in this study, we tried to represent the "center, body and range of the informed technical community," a~~

~~key requirement for seismic hazard and risk assessment (NUREG/CR-6372, https://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6372/vol1/index.html).~~

- ~~**State-of-knowledge:** The methodologies applied represent the current state of knowledge in earth science and engineering. They are beyond the commonly adopted state of practice in geothermal projects, but well aligned with good practice recommendation of the DESTRESS project (Grigoli et al., 2017), with Swiss good practice recommendations (Trutnevyte and Wiemer, 2017) and the recommendations of the international expert committee investigating the Pohang earthquake (Lee et al., 2019).~~

- ~~**Explicit uncertainty treatment:** We systemically considered the uncertainties in knowledge and the variability of the data in our assessment and made them explicit through the use of a logic tree approach. This reflects the current state of practice in probabilistic seismic hazard and risk assessment for natural earthquakes.~~

- ~~**Transparency and Reproducibility:** This study documents all decisions taken in a transparent and reproducible way. All stakeholders in risk governance thus have access to the same level of information as baseline, and ideally a common understanding of the project's risks.~~

- ~~"**Updatable:"** Most important, the a-priori risk assessment can be updated in a consistent way as soon as new data arrives. Because the initial uncertainties are very large (e.g., Figure 8), updating it with *in-situ* information is a must and should be done in a manner which is fully compatible with the initial risk assessment. The a-priori risk assessment presented here is thus also a first and critical step toward risk management.~~

5.1 Limitations of our study

Probabilistic risk assessment is in many ways a very pragmatic approach that systematically collects available information based on the current state of knowledge. It is acceptable that in many areas, the state of knowledge is limited and evolving. While we consider the current assessment as useful and usable, there are also some limitations and areas where further improvements would be beneficial:

- Geological and seismotectonic knowledge is poorly represented. This is mostly a consequence of the fact that knowledge of the local seismotectonic is limited and uncertain, especially when extrapolated to the reservoir depths. The limited use of geological constraints is also a consequence of the fact that geological knowledge cannot be readily transferred into forecasting models of seismicity.

- Empirical data from similar injections in the surroundings of Geldinganes or from areas with comparable conditions are limited and mostly based on observation in the 1970's with limited seismic monitoring in place. While countless well-monitored

injections have been conducted in Iceland overall, there has been less activity near Reykjavik. The initial stimulation of the Geldinganes well in 2001 produced no noticeable seismicity, which provides important constraints (Figure 5). However, monitoring then was at that time quite limited so smaller than magnitude 2 event may have been undetected and we need to consider also that the response to the 2019 stimulation may also be different.

- The seismicity forecasting models we use are simplistic in many ways, considering a limited amount of physical, hydraulic or geological aspects. In particular, neither the SM1 nor the SM2 model explicitly consider the (re)activation of the cracks/faults responsible for the mud losses during drilling and reported by ~~Steingrimsson~~ Steingrímsson et al., 2001. We also use few models overall and do not take the risk-limiting effect of mitigation measures explicitly into account.

- Ground motion models which are specific for Iceland exist. However, they originate from few strong-motion data at short distances, from larger magnitudes and from natural earthquakes and are therefore a limited constraint. Likewise, little is known about the site amplification at a microzonation level. We do not plan in the Advanced Traffic Light System (ATLS) to perform an online updating of the coefficient of the~~e~~ ground motion models.

- Building vulnerabilities are known at a first order level, but no efforts have been made to verify or validate them, nor will they be updated during the ATLS implementation. No sensors in buildings are planned.

5.2 ~~Key findings~~Recommendations

~~The key findings of the risk assessment are shown in Figure 8 and 9 and summarized below:~~

- ~~The overall risk for an individual to die in a building within a radius of 2 km around the well (Figure 8) is assessed to be below 10^{-7} or at 0.1 micromort (1 micromort = unit of risk defined as one in a million chance of death). This value is within the acceptable range when compared to acceptance criteria applied in the Netherlands (or Switzerland). Reason for the acceptable risk is the estimated low vulnerability of the building stock, the overall quite limited injection volume and the fact that the initial stimulation has not produced M>2 seismicity.~~

- ~~The chance of damage to buildings is around 0.1% (Figure 9) and therewith below the 10^{-2} acceptance threshold we have arbitrarily introduced for damage.~~

- ~~The thresholds proposed in the classical traffic light (Figure 2) are consistent with the risk thresholds computed; it is not suggested to define more conservative TLS thresholds at this point.~~

- ~~The uncertainties at this stage of the project are very high, highlighting the importance of updating the risk study continuously as new data becomes available.~~

~~In Appendix A1 we finally introduce a series of recommendations, that should possibly be implemented before and during the stimulation.~~

~~Appendix B: Recommendations~~

Below, we list a number of recommendations on risk management that were made to Reykjavik Energy before the start of stimulation operations at the Geldinganes site. These are based partially on this study, but also consider experiences of past projects.

~~1.~~• Excellent seismic monitoring and reliable near-real time processing is a key requirement for updating the ~~a priori~~*a priori* risk assessment. The network installed at Geldinganes should be capable of this task; however, owing to the low seismicity in the region and short deployment time of the full network, the actual capabilities and operation procedures are untested.

~~2.~~• The standard traffic light system operated by ISOR on behalf of OR and based on IMO magnitudes, is critically important und the ultimate decision tool. A TLS is a simple well-proven and well-established technology, it cannot and should not at this stage be replaced with more adaptive concepts of risk assessment.

~~3.~~• Given the uncertainties, updating this risk assessment is a key requirement. The most basic approach is to update it based on periodical re-assessment of the model parameters (seismicity rates, *b*-value, hydraulic parameters) performed offline and interactively. ~~Ideally, the updating can also be performed in near-real time, and largely automated. However, this approach has never been tested under operational conditions and may fail due to unforeseen problems. ETH Zurich is aiming to make both a periodic re-assessment as well as an automated re-assessment available to support OR, GFZ and ISOR throughout the stimulation phase in the decision taking. A detailed description of these ATLS decision support will be provided shortly before the stimulation. In addition to the risk assessment, we will also provide basic hydraulic properties of the reservoir that can help to judge the performance of the stimulation and steer further developments.~~

~~4.~~• A pre-stimulation test that results in a number of micro-earthquakes below Magnitude 1, followed by a subsequent update of this risk study would help to calibrate the seismic forecast models and to allow to constrain the uncertainties. The test would also demonstrate the ability of the monitoring network to detect and locate microseismicity. ~~However, designing such a pre-stimulation phase at Geldinganes area is a difficult task. In fact, despite circa 5,000 m³ of water were injected (at low pressures) at beginning of 2019, no seismicity was detected (with a seismic network with a~~

magnitude of completeness of circa 0.3). Moreover, the 2001 multi day stimulation (performed at higher rates and pressures) has resulted in no seismic events detected by the 2001 Icelandic regional seismic network.

- The stimulation plan consists of three distinct stages at different sections of the well. The design of this first stage is conservative, with slowly increasing flow rates (1 hr injection per flow rate stage) and longer shut-in phases (2 hr shut-in after each flow rate increase). Moreover, the volume injected (despite the target of 6,000 m^3) strictly depends on the required pressures for fracture opening/shearing (i.e., fluid-induced events). (If anomalous seismic behavior is detected or complications with the seismic network are emerging, the first stage will be modified or (eventually) stopped, as indicated by the TLS protocol). As a consequence, the calibration of the seismic forecast models is performed along with the first stage. Then, the updated values are used as prior information for the second and third sub-stimulations.

5. The project is not seismic risk free, there is a residual chance that, despite all mitigation measures applied, damaging earthquakes might occur. This report attempts to quantify this chance, and we believe it is important to openly communicate to the public and authorities this remaining risk and the steps taken to reduce and control it. This might include clarification on how potential damages would be reported, settled and insured.

6. Re-activating pre-existing and tectonically pre-stressed larger fracture zones and eventually triggering a larger earthquake as it happened in Pohang, is unlikely, but still probably the most important risk for the project. The probabilistic risk approach applied here captures this chance to trigger such an event to a certain extent, and in a statistical approximation. However, it may possibly underestimate the chance of such a 'triggered' earthquake if an unknown major fault is very close to the injection site (i.e., closer than 1 kilometres). Moreover, in this project, a fault zone may potentially be cause for the high temperature. Observe, in fact, that in Iceland the fault zones are oftentimes the targets of geothermal wells. Therefore, we suggest that the seismicity analyst team should be on the look-out for lineament potentially indicative of a major fault zone being re-activated, and discuss it with the experts group that is accompanying the project. In particular, a in depth analysis should be carried out after the first stage of the stimulation.

7. The size distribution of induced earthquakes critically determines the risk, and an unusually low *b*-value may indicate the presence of critically stressed faults and will result in much larger probability of larger events. A low b-value at the same time will result in lower number of small events, which might be misinterpreted as a re-assumingly low seismogenic Index. The re-assessed *b*-value must flow into the update of the risk assessment, but we suggest adding as an additional safety criterion a project halt if the *b*-value of induced events is estimated below 0.8.

8. It is universally accepted that the seismicity and thus risk will decrease once the injection has been stopped. It is less clear, however, if gradual pressure reduction, shut-in, bleed off, or actively pumping out (if possible) are the best

mitigation strategies. In this project, a common agreement has been reached on considering bleed off as the most adequate strategy.

9. Surprising developments are possible, if not likely. Therefore, we set up a small interdisciplinary expert group that can come together rapidly (e.g., virtually) if unexpected developments occur (lineaments, clusters, etc.).

Author contribution. Marco Broccardo conceptualized and prepared the manuscript with contributions from all co-authors, and conducted the formal analysis. Mignan Arnaud helped in the conceptualization and preparation of the study and provided the data set of Table 1. Francesco Grigoli conceptualized the mitigation strategy and prepared Section 2.4. Dimitrios Karvounis conceptualized the Source Model 2 (Section 3.1.2) and conducted the computational analysis leading to the synthetic catalog. Together with Antonio Pio Rinaldi Dimitrios Karvounis also prepared Section 2.2. Laurentiu Danciu supported the selection of the GMPEs and GMICE and contributed to the hazard analysis. Hannes Hofman supported the preparation, creation, and presentation of the study in all its aspects. Vala Hjörleifsdóttir, Claus Milkereit, Torsten Dahm, and Günter Zimmermann reviewed and improved this study. Stefan Wiemer supported the conceptualization, preparation, and presentation of the study in all its aspects, with a specific focus on the final discussion and recommendations.

Data availability. The data used in this paper is available from the authors upon request.

Competing interests. The authors declare that they have no conflict of interest

Acknowledgements & funding

This paper has been supported by DESTESS, a projected which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.691728. Francesco Grigoli is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement (no. 790900).

References

Ambraseys, N. N., Douglas, J., Sarma, S. K., and Smit, P. M. Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration. Bulletin of earthquake engineering, 3(1), 1-53, doi: 10.1007/s10518-005-0183-0, 2005.

Akkar, S., and Bommer, J. J. Empirical equations for the prediction of PGA, PGV, and spectral accelerations in Europe, the Mediterranean region, and the Middle East. Seismological Research Letters, 81(2), 195-206, doi: 10.1785/gssrl.81.2.195, 2010.

Bachmann, C. E., Wiemer, S., Woessner, S., and Hainzl, S. Statistical analysis of the induced Basel 2006 earthquake sequence: introducing a probability-based monitoring approach for Enhanced Geothermal Systems. Geophys. J. Int., 186, 793-807, doi: 10.1111/j.1365-246X.2011.05068.x, 2011.

Baisch, S., Koch, C., and Muntendam-Bos, A. Traffic light systems: to what extent can induced seismicity be controlled? Seismological Research Letters, 90(3), 1145-1154, doi: 10.1785/0220180337, 2019.

Baltay, A. S., and Hanks, T. C. Understanding the magnitude dependence of PGA and PGV in NGA-West 2 data. Bulletin of the Seismological Society of America, 104(6), 2851-2865, doi: 10.1785/0120130283, 2014.

Baker, J. W., and Gupta, A. Bayesian Treatment of Induced Seismicity in Probabilistic Seismic-Hazard Analysis. Bull. Seismol. Soc. Am., 106, 860-870, doi: 10.1785/0120150258, 2016.

Bessason, B., & Bjarnason, J. Ö. Seismic vulnerability of low-rise residential buildings based on damage data from three earthquakes (Mw6. 5, 6.5 and 6.3). Engineering Structures, 111, 64-79, doi: 10.1016/j.engstruct.2015.12.008, 2016.

Bommer, J. J., Oates, S., Cepeda, J. M., Lindholm, C., Bird, J., Torres, R., Marroquin, G., Rivas, J. Control of hazard due to seismicity induced by a hot fractured rock geothermal project. Engineering Geology, 83(4), 287-306, doi: 10.1016/j.enggeo.2005.11.002, 2006.

Bommer, J. J., Stafford, P. J., Alarcón, J. E., and Akkar, S. The influence of magnitude range on empirical ground-motion prediction. Bulletin of the Seismological Society of America, 97(6), 2152-2170, doi: 10.1785/0120070081, 2007.

Bommer, J. J., Crowley, H., & Pinho, R. A risk-mitigation approach to the management of induced seismicity. Journal of Seismology, 19(2), 623-646, doi 10.1007/s10950-015-9514-z, 2015.

Broccardo, M., Mignan, A., Wiemer, S., Stojadinovic, B., and Giardini, D. Hierarchical Bayesian Modeling of Fluid-Induced Seismicity. Geophysical Research Letters, 44(22), 11-357, doi: 10.1002/2017GL075251, 2017a.

Broccardo, M., Danciu, L., Stojadinovic, B., and Wiemer, S. Individual and societal risk metrics as parts of a risk governance framework for induced seismicity. In 16th World Conference on Earthquake Engineering (WCEE16), 2017b.

Cauzzi, C., & Faccioli, E. Broadband (0.05 to 20 s) prediction of displacement response spectra based on worldwide digital records. Journal of Seismology, 12(4), 453, doi:10.1007/s10950-008-9098-y, 2008.

Cornell, C. A. Engineering seismic risk analysis. Bulletin of the seismological society of America, 58(5), 1583-1606, 1968.

Cornell C.A., Krawinkler H. Progress and challenges in seismic performance assessment.PEER Center News,Spring. http://peer.berkeley.edu/news/2000spring/index.html, 2000.

Danciu, L., and Tselentis, G. A. Engineering ground-motion parameters attenuation relationships for Greece. Bulletin of the Seismological Society of America, 97(1B), 162-183, doi: 10.1785/0120050087, 2007.

Diehl, T., Kraft, T., Kissling, E., & Wiemer, S. The induced earthquake sequence related to the St. Gallen deep geothermal project (Switzerland): Fault reactivation and fluid interactions imaged by microseismicity. Journal of Geophysical Research: Solid Earth, 122(9), 7272-7290, doi: 10.1002/2017JB014473, 2017.

Dinske, C., and Shapiro, S. A. Seismotectonic state of reservoirs inferred from magnitude distributions of fluid-induced seismicity. Journal of seismology, 17(1), 13-25, doi: 10.1007/s10950-012-9292-9, 2013.

Ellsworth, W. L. Injection-induced earthquakes. Science341.6142: 1225942, doi: 10.1126/science.1225942, 2013.

Embrechts, P., Klüppelberg, C., and Mikosch, T. Modelling extremal events: for insurance and finance (Vol. 33). Springer Science & Business Media, ISBN 978-3-642-33483-2, 2013.

Faccioli, E., and Cauzzi, C. Macroseismic intensities for seismic scenarios estimated from instrumentally based correlations. In Proc. First European Conference on Earthquake Engineering and Seismology, paper (No. 569), 2006.

Faenza, L., and Michelini, A. Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap. Geophysical Journal International, 180(3), 1138-1152, doi:10.1111/j.1365-246X.2009.04467.x, 2010.

Galanis, P., Sycheva, A., Mimra, W., and Stojadinović, B. A framework to evaluate the benefit of seismic upgrading. Earthquake Spectra, 34(2), 527-548, doi: 10.1193/120316EQS221M, 2018.

Giardini, D. Geothermal quake risks must be faced. Nature, 462(7275), 848, doi:10.1038/462848a, 2009.

Gischig, V. S., and Wiemer, S. A stochastic model for induced seismicity based on non-linear pressure diffusion and irreversible permeability enhancement. Geophysical Journal International, 194(2), 1229-1249, doi: 10.1093/gji/ggt164, 2013.

Gischig, V., Wiemer, S., and Alcolea, A. Balancing reservoir creation and seismic hazard in enhanced geothermal systems. Geophysical Journal International, 198(3), 1585-1598, doi: 10.1093/gji/ggu221, 2014.

Goertz-Allmann, B. P., and Wiemer, S. Geomechanical modeling of induced seismicity source parameters and implications for seismic hazard assessment. Geophysics, 78(1), KS25-KS39, doi: 10.1190/geo2012-0102.1, 2012.

Grigoli, F., Cesca, S., Priolo, E., Rinaldi, A. P., Clinton, J. F., Stabile, T. A., ... & Dahm, T. Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A European perspective. Reviews of Geophysics, 55(2), 310-340. doi: 10.1002/2016RG000542, 2017.

Grigoli, F., Cesca, S., Rinaldi, A. P., Manconi, A., López-Comino, J. A., R. Westaway, C. Cauzzi, T. Dahm, and S. Wiemer (2018). The November 2017 Mw 5.5 Pohang earthquake: A possible case of induced seismicity in South Korea. Science, 360(6392), 1003-1006, doi: 10.1126/science.aat2010, 2010.

Grünthal, G. European macroseismic scale 1998. European Seismological Commission (ESC), 1998.

Gülkan, P., and Kalkan, E. Attenuation modeling of recent earthquakes in Turkey. Journal of Seismology, 6(3), 397-409, doi: 10.1023/A:1020087426440, 2002.

Gunnlaugsson, E., Gislason, G., Ivarsson, G., and Kjaran, S. P. Low temperature geothermal fields utilized for district heating in reykjavik, iceland. In Proceedings World Geothermal Congress (Vol. 74), 2000.

Gupta, A., and Baker, J. W. Sensitivity of induced seismicity risk to source characterization, ground motion prediction, and exposure. In Proceedings 16th world conference on earthquake engineering, 2017.

HAZUS MH MR3, Multi-hazard Loss Estimation Methodology: Earthquake Model, Technical Manual, NIST, Washington D.C., 2003.

~~Heidbach, Oliver; Rajabi, Mojtaba; Reiter, Karsten; Ziegler, Moritz; WSM Team: World Stress Map Database Release 2016. V. 1.1. GFZ Data Services. doi: 10.5880/WSM.2016.001, 2016~~

Haimson, B. C., and Voight, B. Stress measurements in Iceland. EOS, Trans., Am. Geophys. Union;(United States), 57(12), 1976.

Haimson, B. C. The hydrofracturing stress measuring method and recent field results. International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts. Vol. 15. No. 4. Pergamon, doi: 10.1016/0148-9062(78)91223-8, 1978.

Heidbach, Oliver; Rajabi, Mojtaba; Reiter, Karsten; Ziegler, Moritz; WSM Team: World Stress Map Database Release 2016. V. 1.1. GFZ Data Services. doi: 10.5880/WSM.2016.001, 2016

Hirschberg, S., Wiemer, S., and Burgherr, P. (Eds.). Energy from the Earth: Deep Geothermal as a Resource for the Future ? (Vol. 62). vdf Hochschulverlag AG, doi:10.3929/ethz-a-010277690, 2015.

Hofmann, H., et al. Hydraulic Stimulation Design for Well RV-43 on Geldinganes, Iceland. Proceedings World Geothermal Congress 2020 Reykjavik, Iceland, April 26 – May 2, 2020

Holschneider, M., Zöller, G., & Hainzl, S. Estimation of the maximum possible magnitude in the framework of a doubly truncated Gutenberg–Richter model. Bulletin of the Seismological Society of America, 101(4), 1649-1659, doi: 10.1785/0120100289, 2011.

Jones, D. A. Nomenclature for hazard and risk assessment in the process industries. IChemE, 1992.

Jonkman, S. N., Van Gelder, P. H. A. J. M., & Vrijling, J. K. An overview of quantitative risk measures for loss of life and economic damage. Journal of hazardous materials, 99(1), 1-30, doi: 10.1016/S0304-3894(02)00283-2, 2003.

Karvounis, D. C., Gischig, V. S., and Wiemer, S. Towards a real-time forecast of induced seismicity for enhanced geothermal systems. In Shale Energy Engineering 2014: Technical Challenges, Environmental Issues, and Public Policy (pp. 246-255), doi: 10.1061/9780784413654.026, 2014.

Karvounis, D. C., and Jenny, P. Adaptive Hierarchical Fracture Model for Enhanced Geothermal Systems. Multiscale Modeling & Simulation, 14(1), 207–231, doi: 10.1137/140983987, 2016.

Kim, K. H., Ree, J. H., Kim, Y., Kim, S., Kang, S. Y., and Seo, W. Assessing whether the 2017 Mw 5.4 Pohang earthquake in South Korea was an induced event. Science, 360(6392), 1007-1009, doi: 10.1126/science.aat6081, 2018.

Kowsari, M., Halldorsson, B., Hrafnkelsson, B., Snæbjörnsson, J. Þ., and Jónsson, S. Calibration of ground motion models to Icelandic peak ground acceleration data using Bayesian Markov Chain Monte Carlo simulation. Bulletin of Earthquake Engineering, 17(6), 2841-2870, doi: 10.1007/s10518-019-00569-5, 2019.

Kwiatek, G., Saarno, T., Ader, T., Bluemle, F., Bohnhoff, M., Chendorain, M., Dresen, G., Heikkinen, P., Kukkonen, I., Leary, P. and Leonhardt, M., 2019. Controlling fluid-induced seismicity during a 6.1-km-deep geothermal stimulation in Finland. Science advances, 5(5), p.eaav7224, doi: 10.1126/sciadv.aav7224

Langenbruch, C., Weingarten, M., and Zoback, M. D.. Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas. Nature Comm., 9, 3946, 2018.

Lagomarsino, S. and Giovinazzi, S. Bull Earthquake Engineering, 4: 415, doi: 10.1007/s10518-006-9024-z, 2006.

Lee, K.K., Ellsworth, W.L., Giardini, D., Townend, J., Shemin Ge, Shimamoto, T., Yeo, In-W., Kang, Tae-S., Rhie, J., Sheen, D.-H., Chang, C., Wool, J.-U., and C. Langenbruch. Managing injection-induced seismic risks. Science 364 (6442), 730-732, doi: 10.1126/science.aax1878, 2019.

Lin, P. S., and Lee, C. T. Ground-motion attenuation relationships for subduction-zone earthquakes in northeastern Taiwan. Bulletin of the Seismological Society of America, 98(1), 220-240, doi: 10.1785/0120060002, 2008.

Majer, E. L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B., & Asanuma, H. Induced seismicity associated with enhanced geothermal systems. Geothermics, 36(3), 185-222, doi:10.1016/j.geothermics.2007.03.003, 2007.

Majer, E., Nelson, J., Robertson-Tait, A., Savy, J., & Wong, I. Protocol for addressing induced seismicity associated with enhanced geothermal systems. US Department of Energy, 52, 2012.

McGarr, A. Seismic moments and volume changes. Journal of geophysical research, 81(8), 1487-1494, doi: 10.1029/JB081i008p01487, 1976.

McGarr, A. Maximum magnitude earthquakes induced by fluid injection, J. Geophys. Res. Solid Earth, 119, 1008-1019, doi: 10.1002/2013JB010597, 2014.

~~Majer, E. L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B., & Asanuma, H. Induced seismicity associated with enhanced geothermal systems. Geothermics, 36(3), 185-222, doi: 10.1016/j.geothermics.2007.03.003, 2007.~~Mena, B., Wiemer, S., Bachman, C. Building robust models to forecast the induced seismicity related to geothermal reservoir enhancement. Bull. Seismol. Soc. Am., 103, 383-393, doi: 10.1785/0120120102, 2013.

Mignan, A., Werner, M.J., Wiemer, S., Chen, C.-C., Wu, Y.-M. Bayesian Estimation of the Spatially Varying Completeness Magnitude of Earthquake Catalogs, Bull. Seismol. Soc. Am., 101(3), 1371–1385, doi: 10.1785/0120100223, 2011.

Mignan, A., Landtwing, D., Kästli, P., Mena, B., and Wiemer, S. Induced seismicity risk analysis of the 2006 Basel, Switzerland, Enhanced Geothermal System project: Influence of uncertainties on risk mitigation. Geothermics, 53, 133-146, doi: 10.1016/j.geothermics.2014.05.007, 2015.

Mignan, A.: Static behaviour of induced seismicity, Nonlin. Processes Geophys., 23, 107–113, doi: 10.5194/npg-23-107-2016, 2016.

Mignan, A., Broccardo, M., Wiemer, S., and Giardini, D. Induced seismicity closed-form traffic light system for actuarial decision-making during deep fluid injections. Scientific reports, 7(1), 13607, doi: 10.1038/s41598-017-13585-9, 2017.

Mignan, A., M. Broccardo, S. Wiemer and D. Giardini, Autonomous Decision-Making Against Induced Seismicity in Deep Fluid Injections, A. Ferrari and L. Laloui (eds.), Energy Geotechnics, SEG 2018, 369-376, doi: 10.1007/978-3-319-99670-7_46, 2019a.

Mignan, A., Karvounis, D., Broccardo, M., Wiemer, S., and Giardini, D. Including seismic risk mitigation measures into the Levelized Cost Of Electricity in enhanced geothermal systems for optimal siting. Applied Energy, 238, 831-850, doi: 10.1016/j.apenergy.2019.01.109, 2019b.

Morris, M. D. Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2), 161-174, doi: 10.1080/00401706.1991.10484804, 1991.

Panzera, F., Mignan, A., Vogfjord, K. S. Spatiotemporal evolution of the completeness magnitude of the Icelandic earthquake catalogue from 1991 to 2013. J. Seismol. 21, 615-630, doi: 10.1007/s10950-016-9623-3, 2017.

Pittore, M., Boxberger, T., Fleming, K., Megalooikonomou, K., Parolai, S., and amp; Pilz, M. DESTRESS - Demonstration of soft stimulation treatments of geothermal reservoirs. GFZ Data Services. doi: 10.14470/7Q7563484600, 2018.

Porter, K.A., Beck, J.L., Shaikhutdinov, R.V. Sensitivity of building loss estimates to major uncertain variables. Earthquake Spectra 18, 719–743, doi: 10.1193/1.1516201, 2002.

Rupakhety, R., and Sigbjörnsson, R. Ground-motion prediction equations (GMPEs) for inelastic displacement and ductility demands of constant-strength SDOF systems. Bulletin of Earthquake Engineering, 7(3), 661-679, doi: 10.1007/s10518-009-9117-6, 2009.

Shapiro, S. A., and Dinske, C. Scaling of seismicity induced by nonlinear fluid-rock interaction. J. Geophys. Res., 114, B09307, doi:10.1029/2008JB006145, 2009.

Steingrímsson, B., Fridleifsson, G.Ó., Gunnarsson, K., Thordarson, S., Thórhallsson, S., and Hafstad, Th.H. Well RV-43 in Geldinganes. Prerequisites for location and design. Orkustofnun, Reykjavík, report BS/GOF/KG/GTHOR/SThHH-02/01 (in Icelandic), 11 pp., 2001.

Trutnevyte, E., and Wiemer, S. Tailor-made risk governance for induced seismicity of geothermal energy projects: An application to Switzerland. Geothermics, 65, 295-312, doi:10.1016/j.geothermics.2016.10.006, 2017.

van der Elst, N. J., M. T. Page, D. A. Weiser, T. H. W. Goebel and S. M. Hosseini. Induced earthquake magnitudes are as large as (statistically) expected, J. Geophys. Res. Solid Earth, 121, 4575-4590, doi: 10.1002/2016JB012818, 2016.

van Elk, J., Doornhof, D., Bommer, J. J., Bourne, S. J., Oates, S. J., Pinho, R., and Crowley, H. Hazard and risk assessments for induced seismicity in Groningen. Netherlands Journal of Geosciences, 96(5), s259-s269, doi: 10.1017/njg.2017.37, 2017.

Yeck, W. L., Hayes, G. P., McNamara, D. E., Rubinstein, J. L., Barnhart, W. D., Earle, P. S., and Benz, H. M. Oklahoma experiences largest earthquake during ongoing regional wastewater injection hazard mitigation efforts. Geophysical Research Letters, 44(2), 711-717, doi: 10.1002/2016GL071685, 2017.

Walters, R. J., Zoback, M. D., Baker, J. W., & Beroza, G. C. Characterizing and responding to seismic risk associated with earthquakes potentially triggered by fluid disposal and hydraulic fracturing. Seismological Research Letters, 86(4), 1110-1118, https://doi.org/10.1785/0220150048, 2015.

Ziegler, M., Rajabi, M., Heidbach, O., Hersir, G. P., Ágústsson, K., Árnadóttir, S., and Zang, A. The stress pattern of Iceland. Tectonophysics, 674, 101-113, doi: 10.1016/0040-1951(69)90097-3, 2016.

Zang, A., Yoon, J. S., Stephansson, O., and Heidbach, O. Fatigue hydraulic fracturing by cyclic reservoir treatment enhances permeability and reduces induced seismicity. Geophysical journal international, 195(2), 1282-1287, doi: 10.1093/gji/ggt301, 2013.

Zang, A., Stephansson, O., and Zimmermann, G. Keynote: fatigue hydraulic fracturing. In ISRM European Rock Mechanics Symposium-EUROCK 2017. International Society for Rock Mechanics and Rock Engineering, 2017.

Zang, A., Zimmermann, G., Hofmann, H., Stephansson, O., Min, K. B., and Kim, K. Y. How to reduce fluid-injection-induced seismicity. Rock Mechanics and Rock Engineering, 52(2), 475-493, doi: 10.1007/s00603-018-1467-4, 2019.

Zhao, J.X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H.K., Somerville, P.G. and Fukushima, Y. Attenuation relations of strong ground motion in Japan using site classification based on predominant period. Bulletin of the Seismological Society of America, 96(3), pp.898-913, doi: 10.1785/0120050122, 2006.

~~Zoback, M. D. Managing the seismic risk posed by wastewater disposal. Earth, 57(4), 38, 2012~~

List of Tables

Table 1. Underground seismic feedback to deep fluid injection.

Site (country*, year)	a_{fb}†	b	$\lambda_{M\geq 2}$	**References**
1-Ogachi OG91 (JP, 1991)	-2.6	0.7	4.3800	Dinske and Shapiro (2013)
2-Ogachi (JP, 1993)	-3.2	0.8	0.6942	Dinske and Shapiro (2013)
3-Soultz (FR, 1993)	-2.0	1.4	0.6942	Dinske and Shapiro (2013)
4-KTB (DE, 1994)	-1.4	0.9	27.6359	Mignan et al. (2017)
5-Paradox Valley (US, 1994)	-2.4	1.1	1.1002	Mignan et al. (2017)
6-Soultz (FR, 1995)	-3.8	2.2	0.0003	Dinske and Shapiro (2013)
7-Soultz (FR, 1996)	-3.1	1.8	0.0087	Dinske and Shapiro (2013)
8-Soultz (FR, 2000)	-0.5	1.1	87.3925	Dinske and Shapiro (2013)
9-Cooper Basin (AU, 2003)	-0.9	0.8	138.5078	Dinske and Shapiro (2013)
10-Basel (CH, 2006)	0.1	1.6	34.7916	Mignan et al. (2017)
11-KTB (DE, 2004-5)	-4.2	1.1	0.0174	Dinske and Shapiro (2013)
12-Newberry (US, 2014a)	-2.8	0.8	1.7437	Mignan et al. (2017)
13-Newberry (US, 2014b)	-1.6	1.0	11.0021	Mignan et al. (2017)

970 * ISO code; † referred to as seismogenic index in Dinske and Shapiro (2013).

Table 2: List of GMPEs used in this study

GMPE name	Location	Reference
1-AB10	Europe & Middle East	Akkar Bommer (2010)
2-CF08	Worldwide	Cauzzi Faccioli (2008)
3-Zh06	Japan	Zhao *et al.* (2006)
4-Am05	Europe and Middle East	Ambraseys et al. (2005)
5-DT07	Greece	Danciu and Tselentis (2007)
6-GK02	Turkey	Gülkan and Kalkan (2002)
7-RS09	Iceland, Europe and MiddleEast	Rupakhety and Sigjörnsson (2009)

975

980

985 **List of Figures**

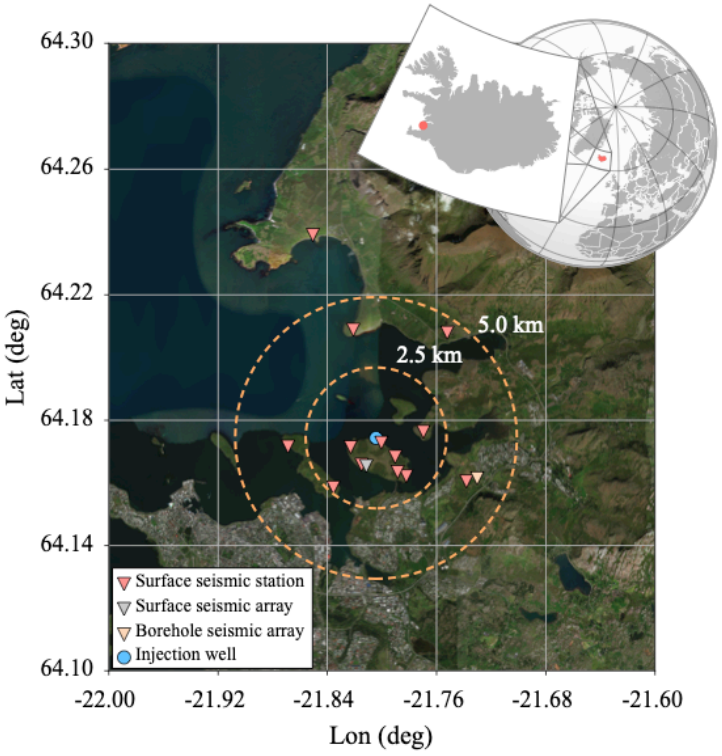


Figure 1 Map view of the Geldinganes island, the injection well, and seismic network. Source of the map: Google-Maps.

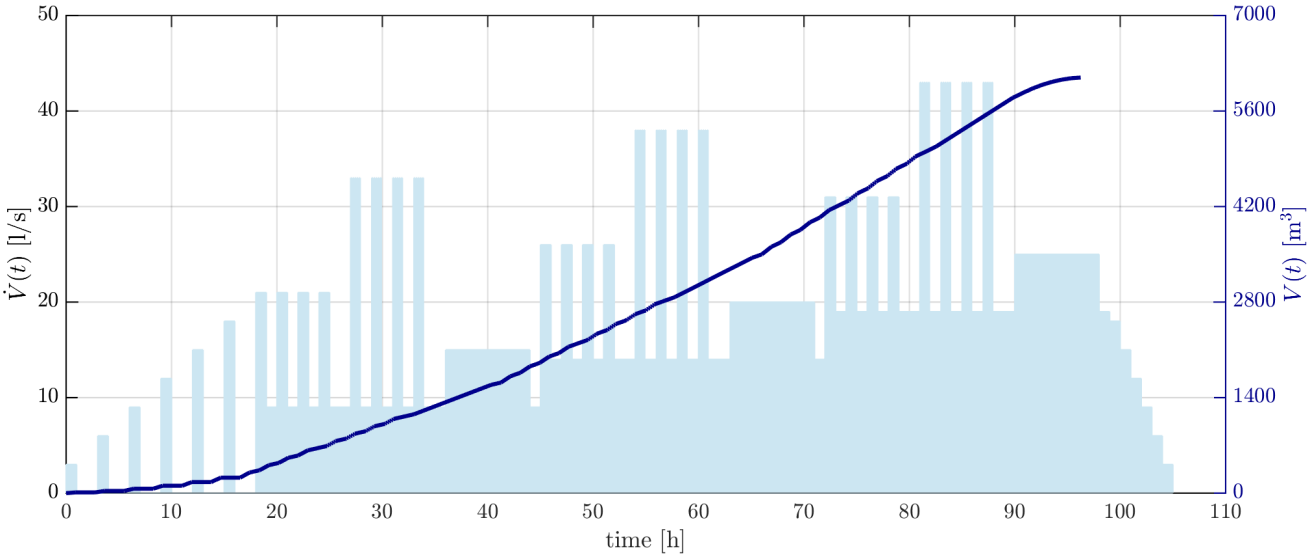


Figure 2 Example of the main stimulation for one stage. After stimulation, flow back is performed.

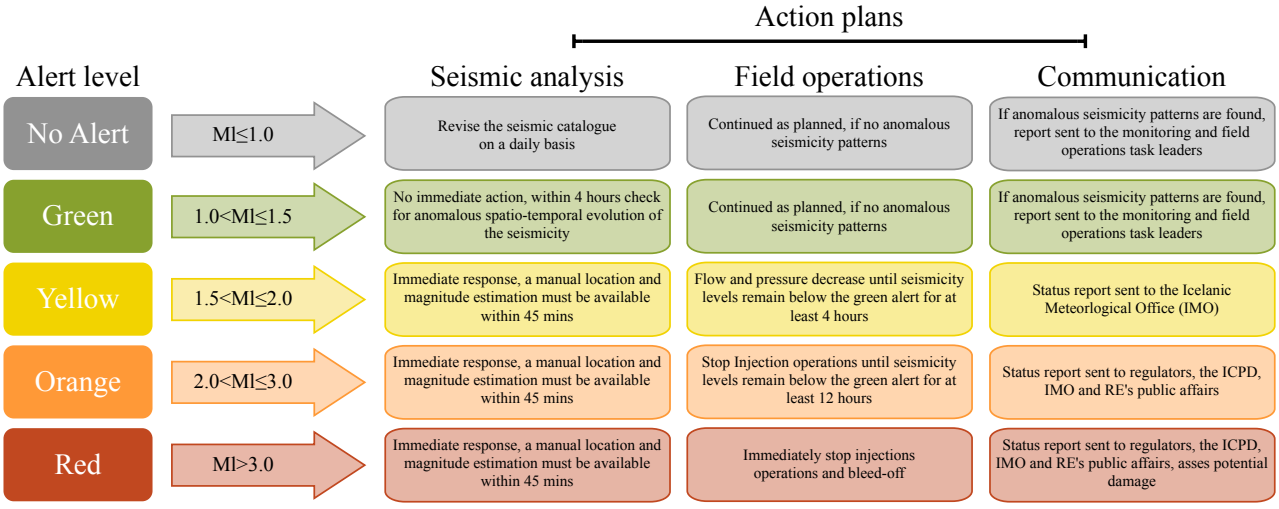
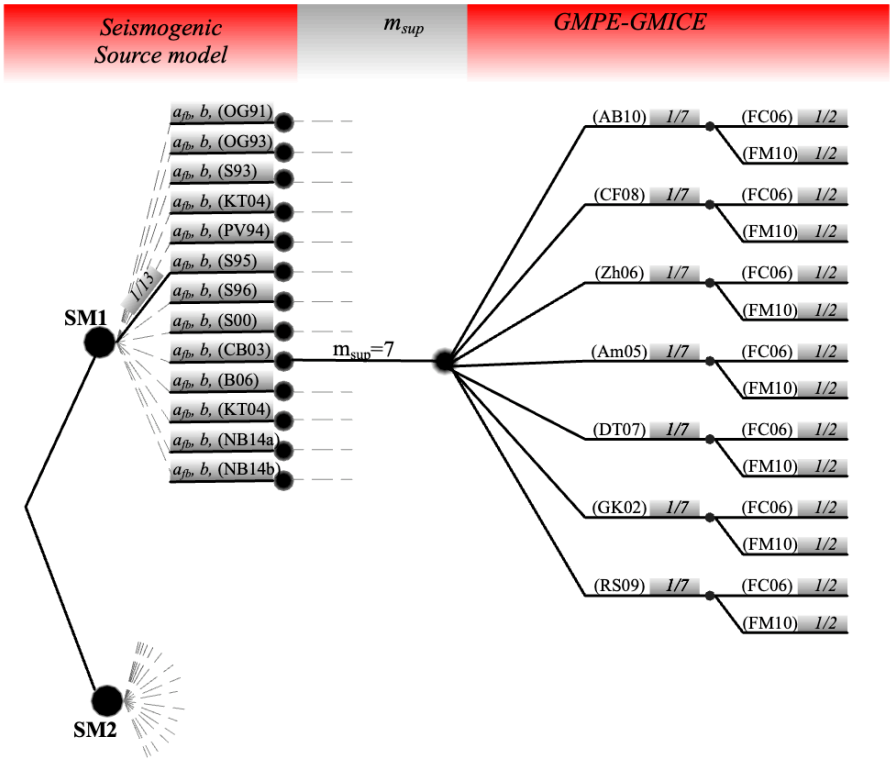


Figure 3 The classic Traffic Light Scheme adopted in the Geldinganes project


Seismogenic Source model
m_{sup}
GMPE-GMICE
a_{fh}, b, (OG91)
a_{fh}, b, (OG93)
a_{fh}, b, (S93)
a_{fh}, b, (KT04)
a_{fh}, b, (PV94)
a_{fh}, b, (S95)
a_{fh}, b, (S96)
a_{fh}, b, (S00)
a_{fh}, b, (CB03)
a_{fh}, b, (B06)
a_{fh}, b, (KT04)
a_{fh}, b, (NB14a)
a_{fh}, b, (NB14b)
1/13
SM1
SM2
m_{sup}=7
(AB10) 1/7
(CF08) 1/7
(Zh06) 1/7
(Am05) 1/7
(DT07) 1/7
(GK02) 1/7
(RS09) 1/7
(FC06) 1/2
(FM10) 1/2

995

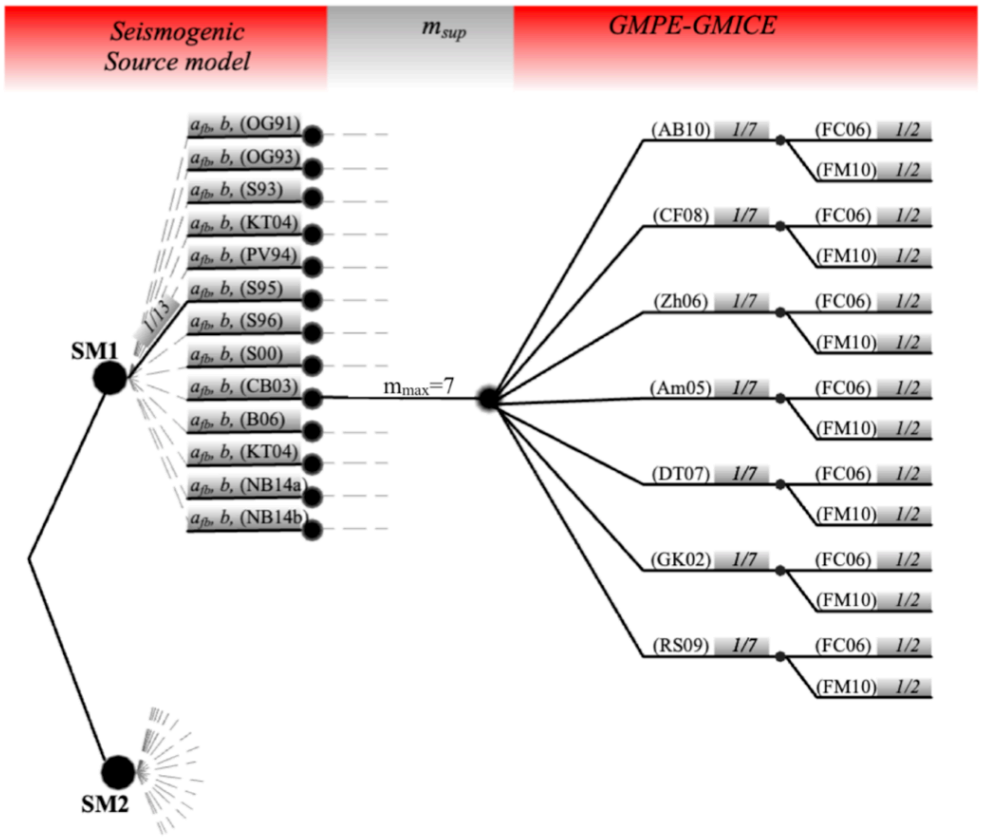


Figure 4 Logic tree for the PSHA analysis. The weight are reported on the grey boxes (e.g., in the SM1 model each $a_{fb} - b$ combination has uniform weight of 1/13, each GMPE has a uniform weight of 1/7, and the each GMICE has uniform weight of 1/2).

000

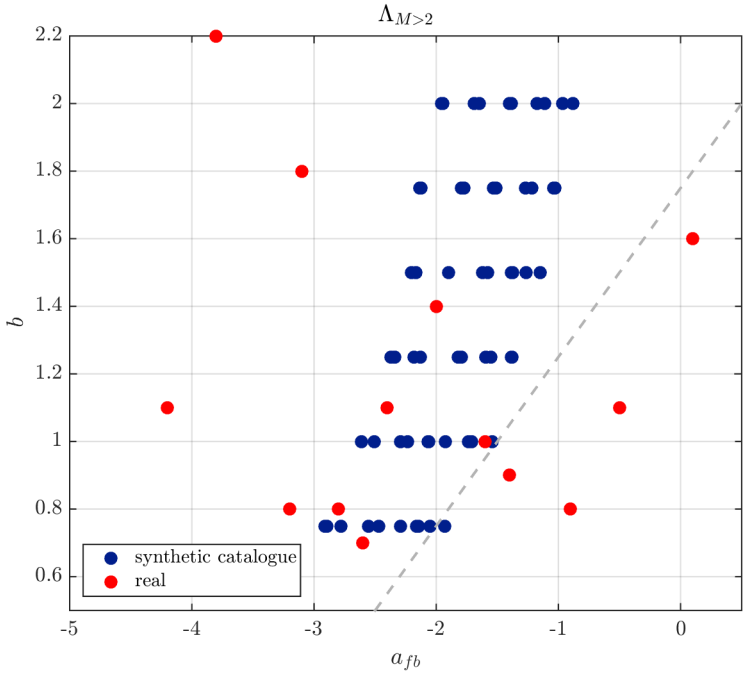
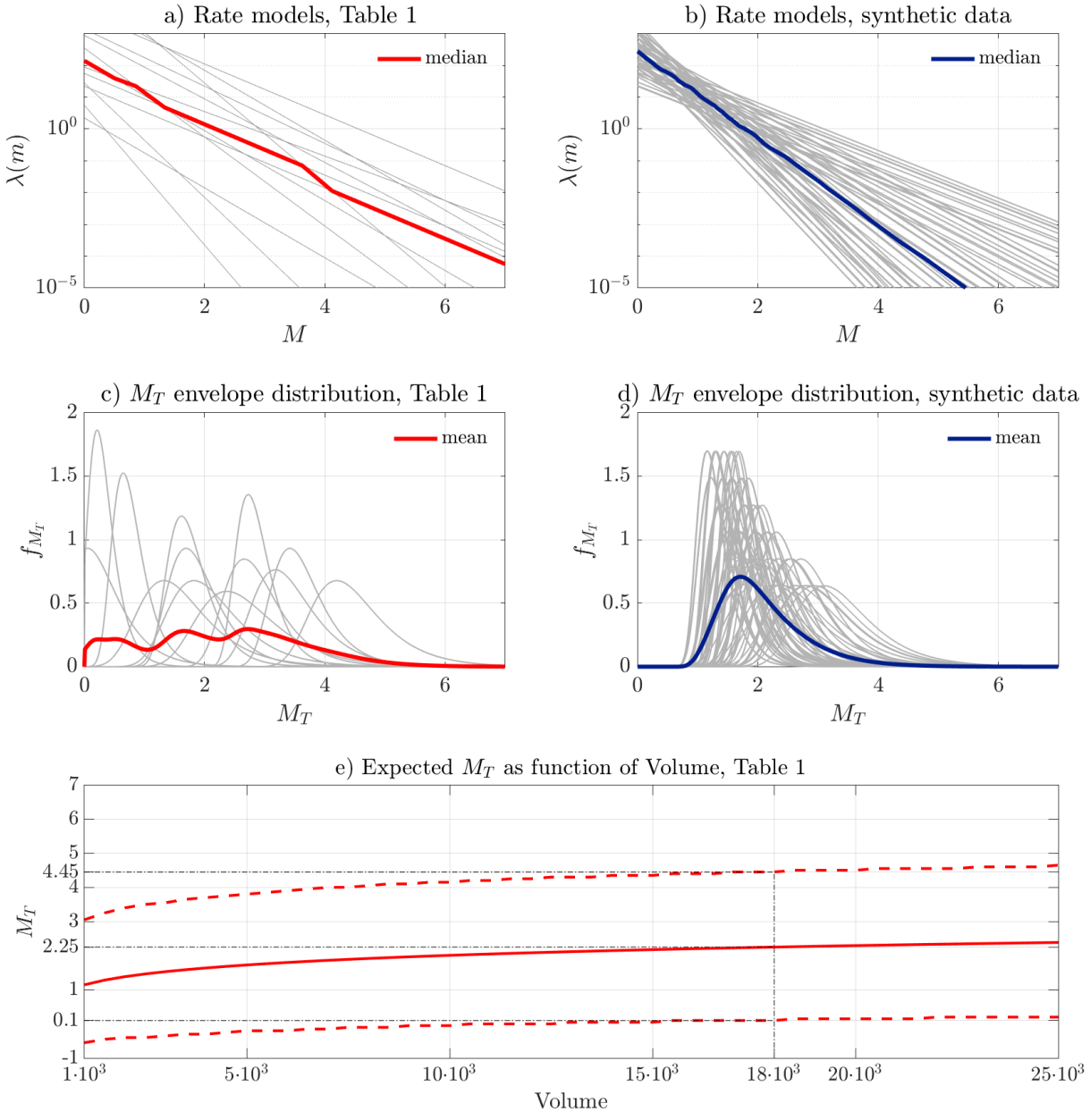


Figure 5 Distribution of a_{fb}- b values for the synthetic catalogue together with the dataset of Table 1. In the planned injection profile (see Figure 2), the flow rate decreases progressively back to zero, meaning that this simple model cannot strictly be applied. As approximation, we use max(ΔV) = 1,728 m3/day instead of $\Delta V_{shut-in}$. A direct comparison can be made between the volume injected V = 18,000 m^3 and the equivalent $\tau\Delta V$ = 2,880 m3 for τ = 1 day and 28,800 m3 for τ = 10 days. The dashed line represents the upper limit of no expected seismicity $M > 2$.


a) Rate models, Table 1
median
$\lambda(m)$
10^0
10^{-5}
0
2
4
6
M
b) Rate models, synthetic data
median
$\lambda(m)$
10^0
10^{-5}
0
2
4
6
M
c) M_T envelope distribution, Table 1
mean
f_{M_T}
2
1.5
1
0.5
0
0
2
4
6
M_T
d) M_T envelope distribution, synthetic data
mean
f_{M_T}
2
1.5
1
0.5
0
0
2
4
6
M_T
e) Expected M_T as function of Volume, Table 1
M_T
7
6
5
4.45
4
3
2.25
1
0.1
-1
$1{\cdot}10^3$
$5{\cdot}10^3$
$10{\cdot}10^3$
$15{\cdot}10^3$
$18{\cdot}10^3$
$20{\cdot}10^3$
$25{\cdot}10^3$
Volume

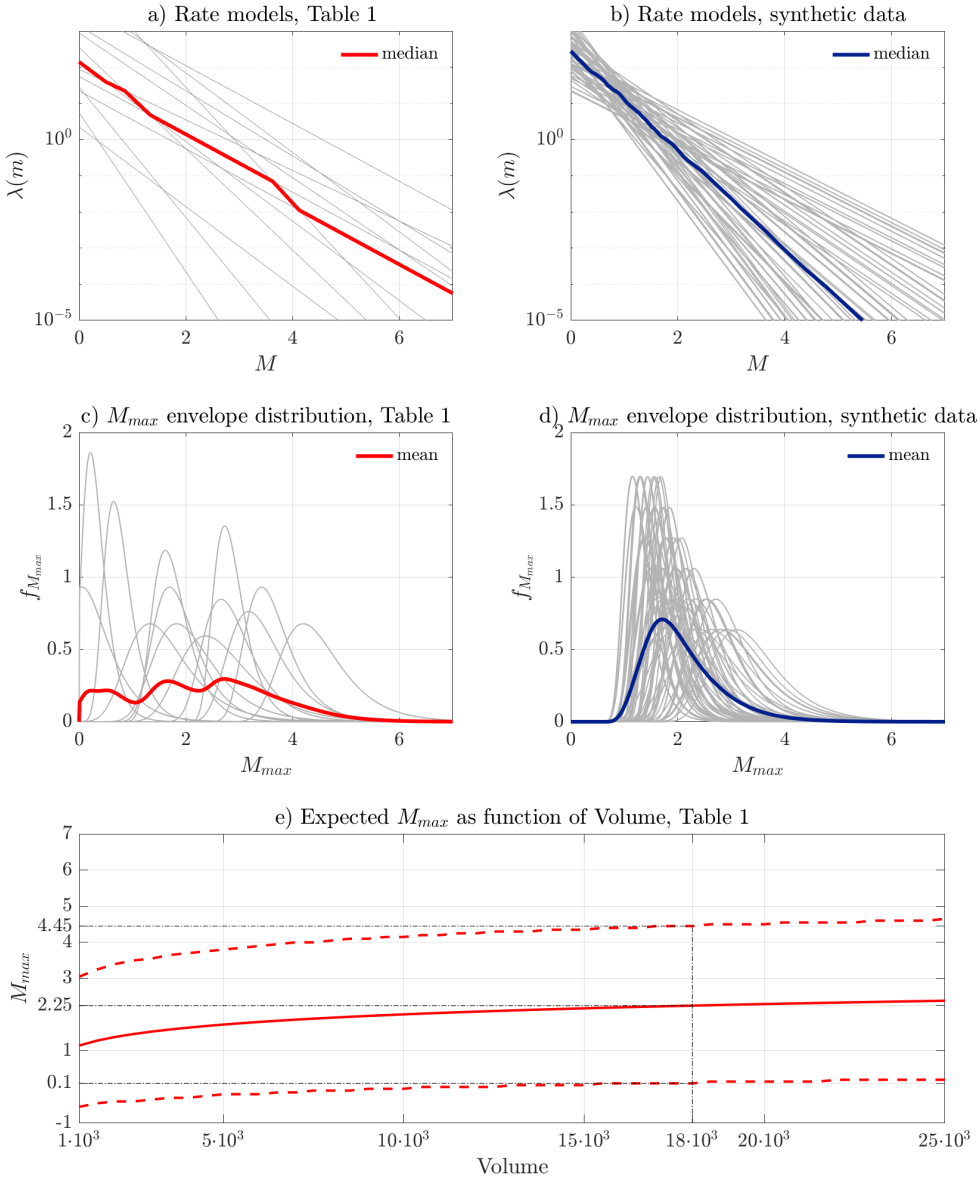


Figure 6 Envelope probabilistic density distribution of the rate model and maximum observed magnitude M_{max} a,c) based on Table 1, b,d) based on synthetic catalogue (S2 source model). e) Expected Magnitude per volume injected, based on Table 1.

015

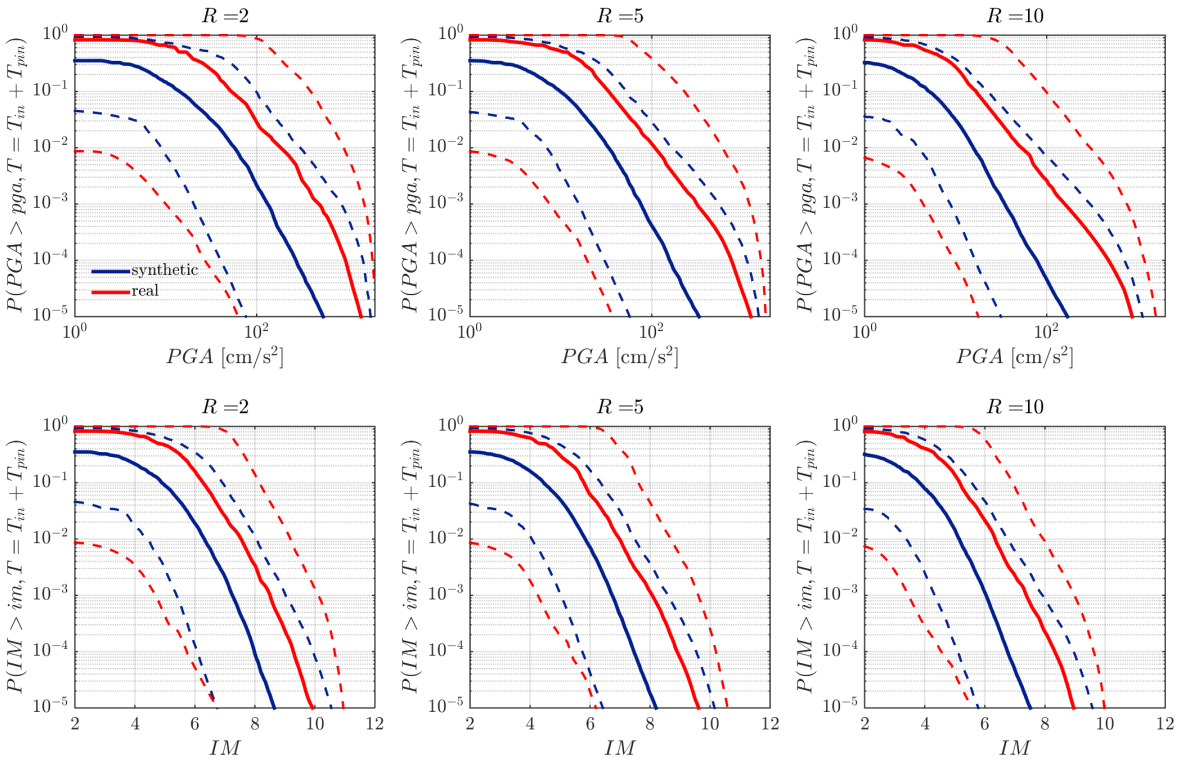
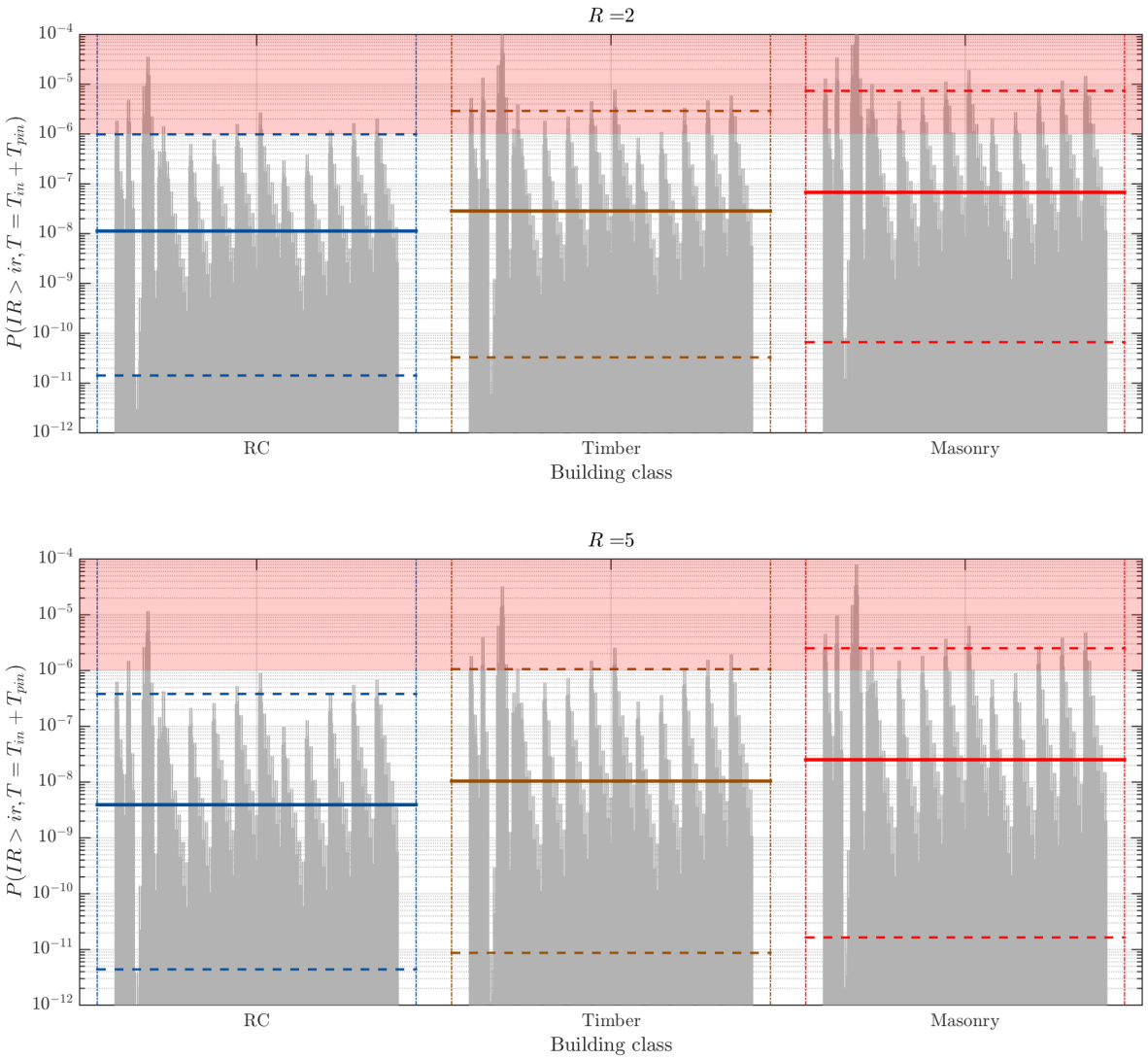


Figure 7 PSHA analysis comparison between source model SM1 (Table 1) and SM2 (synthetic catalogue). Solid lines: medians; dashed lines 10% and 90% quantiles. Intensity measure $EMS98$.



035 **Figure 8** Marginal ***IR*** for 2 km (top) and 5 km distances based on the final model (combined SM1 and SM2). The solid horizontal lines represent the weighted median values of the 1022 (13[$\boldsymbol{a_{fb}, b}$]X7GMPEsX2GMICE, weight .5, + 60[$\boldsymbol{a_{fb}, b}$]X7GMPEsX2GMICE, weight .5) vertical gray lines. The dashed horizontal lines represent the 10 and 90% epistemic quantiles.

040

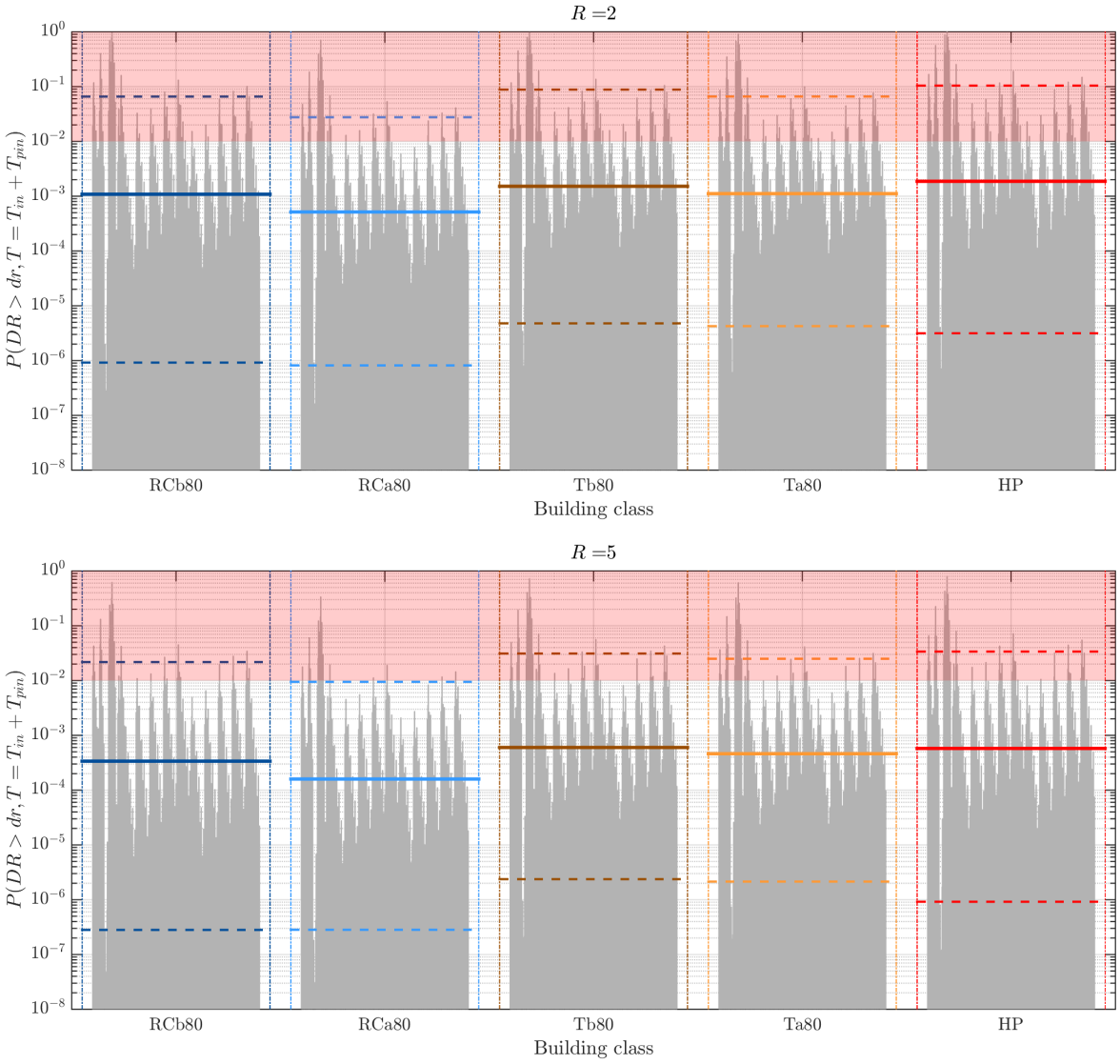


Figure 9 Marginal ***DR*** for the final model for 2 km and 5 km distance. The solid horizontal lines represent the median values of the 511 (13[$\boldsymbol{a_{fb}, b}$]X7GMPEs, .5 weight, + 60[$\boldsymbol{a_{fb}, b}$]X7GMPEs, .5 weight) vertical gray lines. The dashed horizontal lines represent the 10 and 90% epistemic quantiles.

045

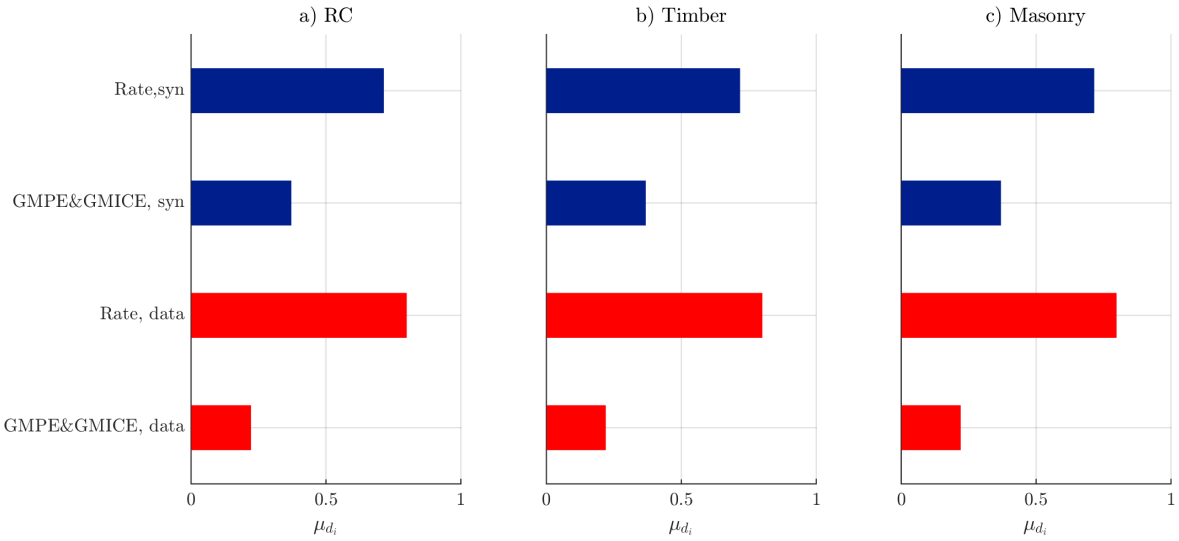


Figure 10 Sensitivity analysis of $\boldsymbol{IR}$ (observe that the QoI is $\log \boldsymbol{IR}$) based on the sensitivity measure $\boldsymbol{\mu_{d_i}}$ for each building class

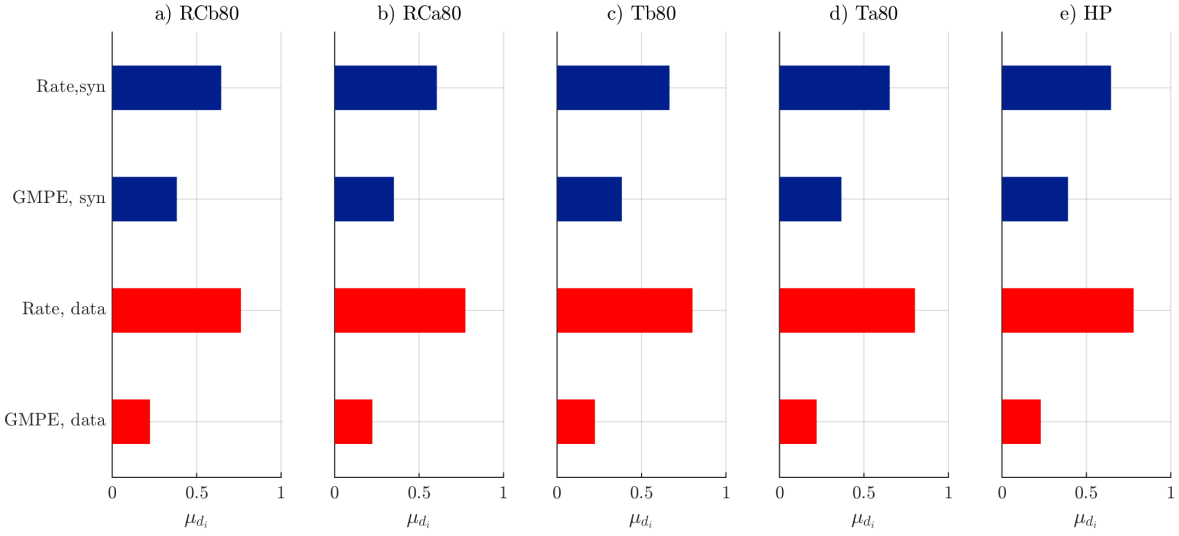


Figure 11 Sensitivity analysis of $\boldsymbol{DR}$ (observe that the QoI is $\log \boldsymbol{DR}$) based on the sensitivity measure $\boldsymbol{\mu_{d_i}}$ for each building class

Supplements

S1 GMPEs GMICE and Scenario based intensity

In this Supplementary material, we first show the trellis plots of the GMPE Figure S1 and of the GMPE&GMICE combination, Figure S2. The GMICE coefficients are reported in Table S1.

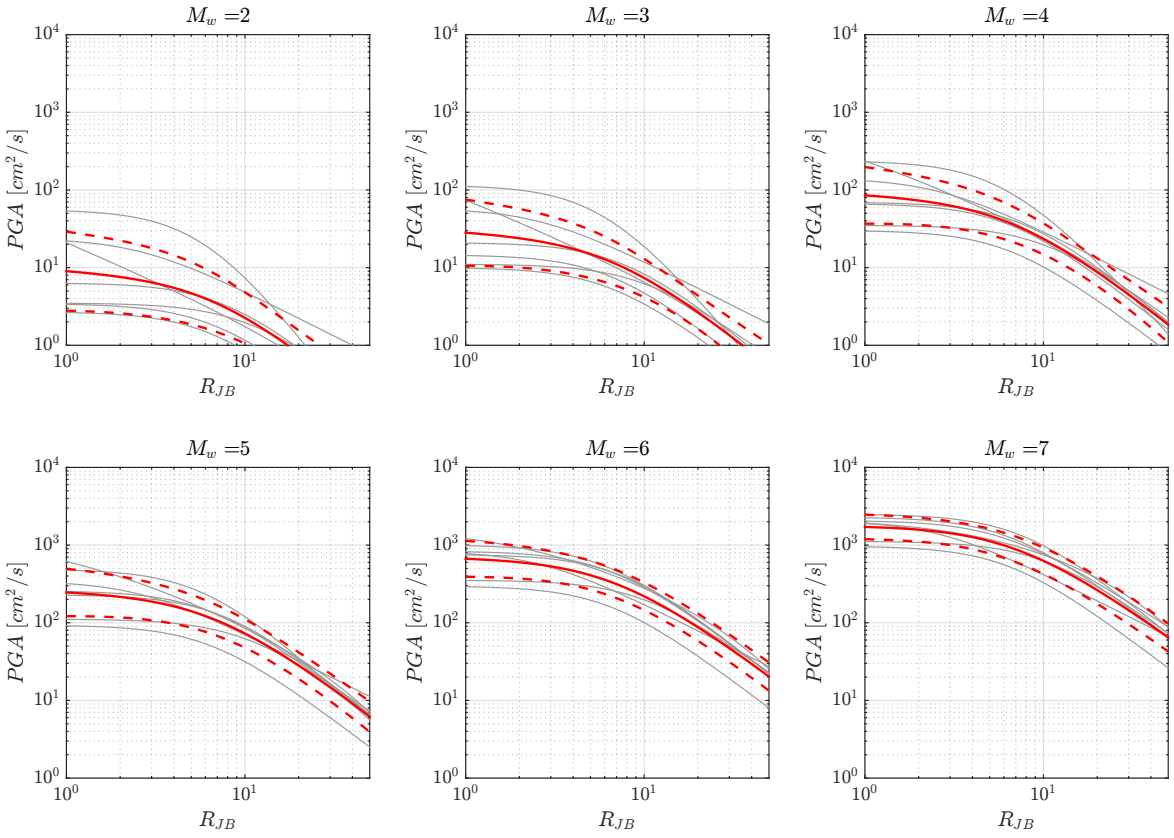


Figure S1. Trellis Plots for the selected GMPEs models. Following the same representation of Rupakhety and Sigjörnsson (2009), solid red lines are the epistemic mean and the dash lines the epistemic mean plus minus the epistemic standard deviation (in log scale).

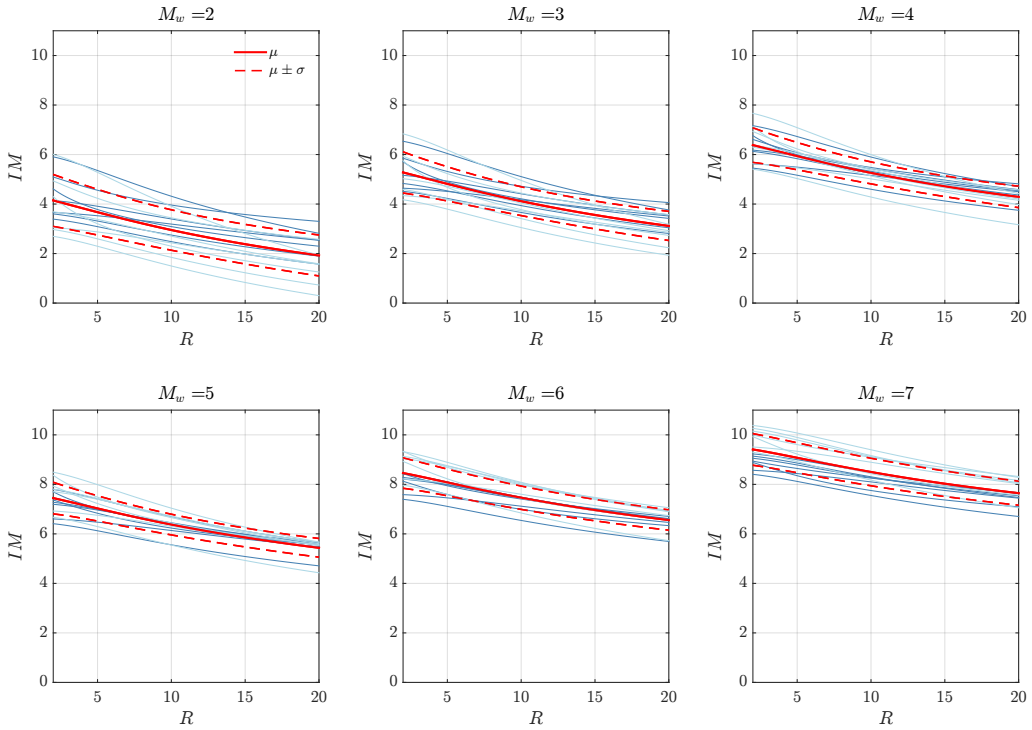


Figure S2 GMICE model. Solid red lines are the epistemic mean and the dash lines the epistemic mean plus minus the epistemic standard deviation.

Table S1: GMICE parameter list

	$\mu_{IM}(PGA)$	σ_{GMICE}		σ_{GMPE}	a	σ_{TOT}
Faccioli and Cauzzi (2006) Units: [m/s]	$1.96\log_{10}(PGA)+6.54$	0.89	1-AB10	0.175	1.96	0.954
			2-CF08	0.176	1.96	0.955
			3-Zh06	0.391	1.96/ln(10)	0.950
			4-Am05	0.175	1.96	0.954
			5-DT07	0.177	1.96	0.955
			6-GK02	0.403	1.96/ln(10)	0.954
			7-RS09	0.287	1.96	1.053
Faenza and Michelini (2010) Units: [cm/s]	$2.58\log_{10}(PGA)+1.68$	0.35	1-AB10	0.175	2.58	0.571
			2-CF08	0.176	2.58	0.573
			3-Zh06	0.391	2.58/ln(10)	0.560
			4-Am05	0.175	2.58	0.571
			5-DT07	0.177	2.58	0.575
			6-GK02	0.403	2.58/ln(10)	0.571
			7-RS09	0.287	2.58	0.819

Given the very large epistemic uncertainty governing the parameter a_{fb} and b of the seismogenic source model, here we propose a conditional probability distribution of PGA and IM: i.e., $P(PGA > pga|M = m, r)$ (also known as scenario based distribution). Figure S3 shows the scenario-based distributions for both PGA and IM.

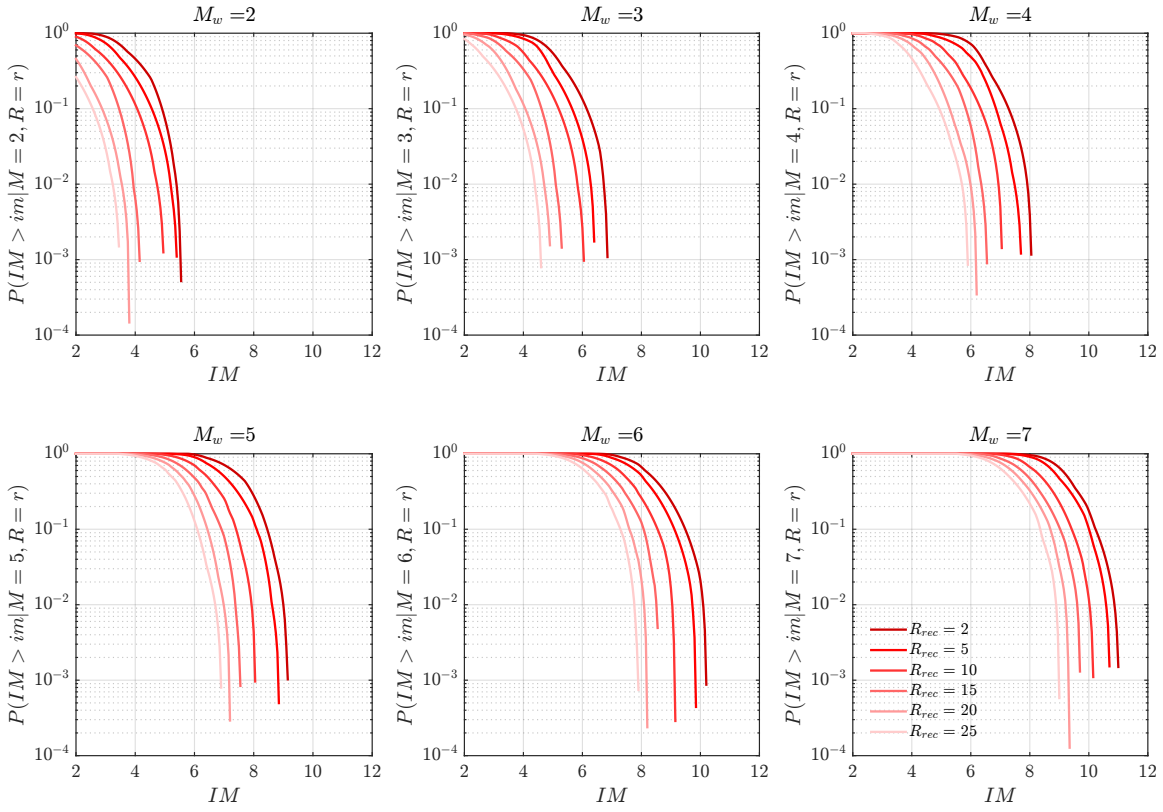


Figure S3 Collection of epistemic medians of $\boldsymbol{P(IM > im|M = m, R = r)}$ for different Magnitude and site to source distances

S2 Exposure, population and assets

The exposed area to the introduced anthropogenic hazard is defined by the building stock within approximately 20 km from the injection point. The region is identified with the so-called Höfuðborgarsvæðim (i.e., the capital region, which includes the capital Reykjavík and six municipalities). This is the largest and densest urban agglomeration in the country with a population of 229,000, which corresponds to circa 65% of the total Icelandic population.

For this study, detailed and precise information of the Reykjavik building stock and population exposed is not readily available and hence not used. Therefore, this description is limited to a macro analysis based on the work of Bessason and

Bjarnason (2016), the Icelandic Property Registry, and the European Seismic Risk service. Based on the Icelandic Property Registry and on Bessason and Bjarnason (2016), it is possible to categorize the building stock by the class of buildings. According to Bessason and Bjarnason (2016), the vast majority of residential buildings (reinforced concrete, timber and pumice block buildings) have shear walls as a lateral structural system against seismic forces. Buildings made of hollow pumice represent a small percentage and can be considered according to Bessason and Bjarnason (2016) as masonry buildings. No information is available at the present time on soil conditions at construction sites, which is assumed to be in firm soil condition. We finally highlight that these evaluations are a first-order estimate only. Given the current lack of a detailed exposure model, we do not include in the present study any aggregate monetary loss analysis.

In addition, according to the European Seismic Risk service, the residential building stock of Iceland is composed of circa 70,000 buildings with a total of 68,000 million Euro of replacement costs. In a first-order inference of the number of residential buildings, we can estimate that there are circa 55-65% of the total amount, a percentage a bit lower than the population proportion. This assumes that in urban areas, the number of persons per building is higher. Given that, a crude estimate of the number of residential buildings in the considered area is given by a range of 38,500-45,500 units for a total replacement cost of [37,400-44,200] millions of euro.

We highlight that these evaluations are a crude first-order estimate only. Given the current lack of a detailed exposure model, we do not include in the present study an aggregate monetary loss analysis (since it will introduce an additional level of deep uncertainty). However, we introduced a low damage threshold risk metric to measure and prevent the likelihood of large damage in the building stock.

S3 Scenario Based IR, and DR

In this section, we first present the scenario IR for different magnitudes, locations and building typologies. The scenarios are derived by using the mean of the GMICE and converted into IR by using the vulnerability model (Lagomarsino and Giovinazzi, 2006, and modified by Mignan et al. 2015, Figure S4, Table S2) and the conditional probability of fatalities for a given damage grade (Galanis et al. 2018, Hazus MH MR3, 2003). Figure S5 shows the IR scenario-based calculations. We can conclude that for a magnitude lower than 4 the $IR = q_{IR,.5} \leq 10^{-6}$ is satisfied for all the building classes.

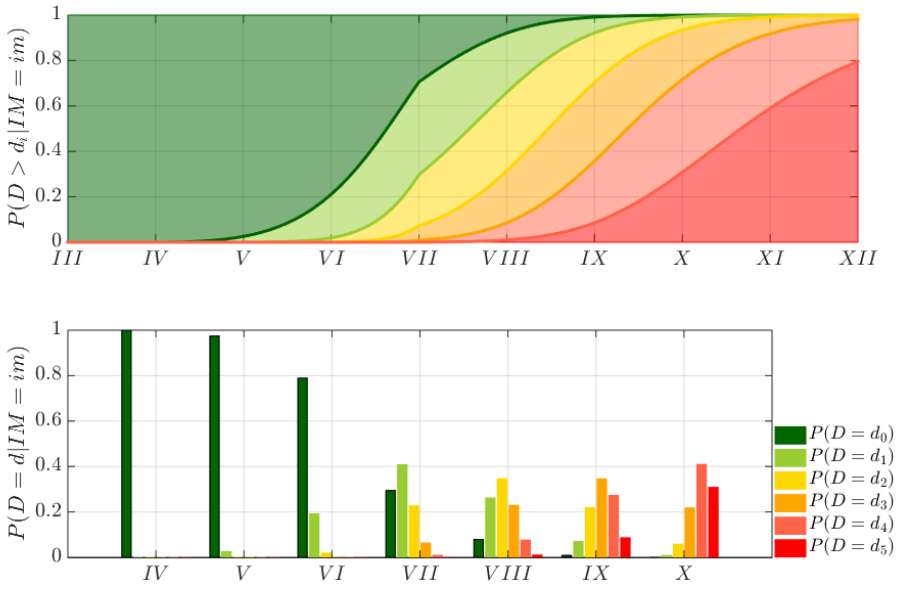


Figure S4 Top subplot: fragility functions, $\boldsymbol{P(D > d_k|IM = im)}$, bottom subplot: probability mass function $\boldsymbol{p_D(d_k)}$ for different level of intensity. The discontinuity at $\boldsymbol{im = VII}$ is introduced by Mignan et al. (2015).

60

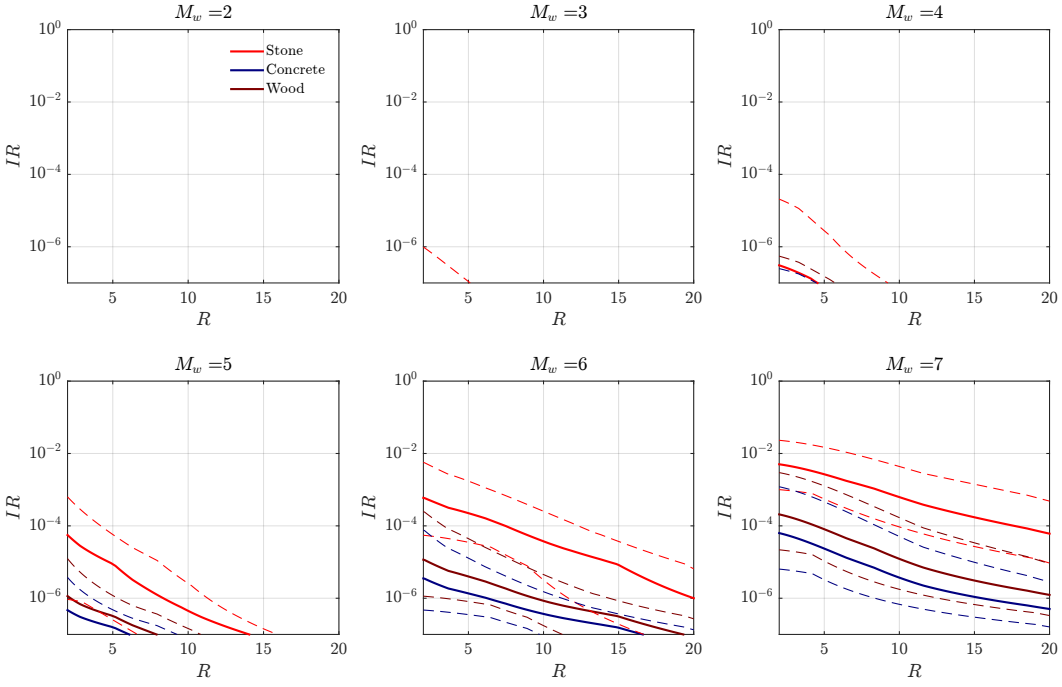


Figure S5. Individual Risk for different magnitude, distance and typology of buildings.

Further, we present damage-based scenarios derived by using the mean of the GMICE converted into DR by using the local fragility functions proposed by Bessason, B. and Bjarnason, 2016, reproduced in Figure S6.

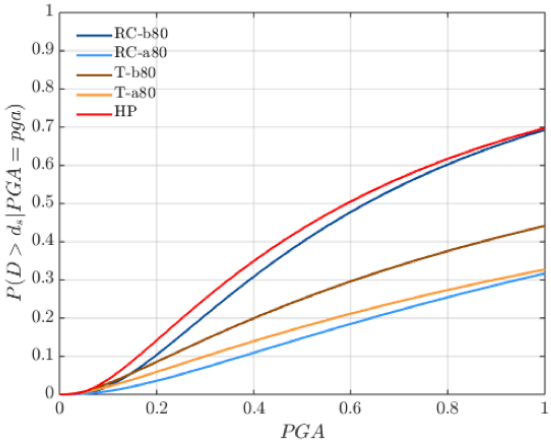


Figure S6 Fragility function for low damage estimation

Figure S7 shows the damage scenario for a magnitude 3 and 4, which represent the scenario limit for observing $DR \leq 10^{-2}$.

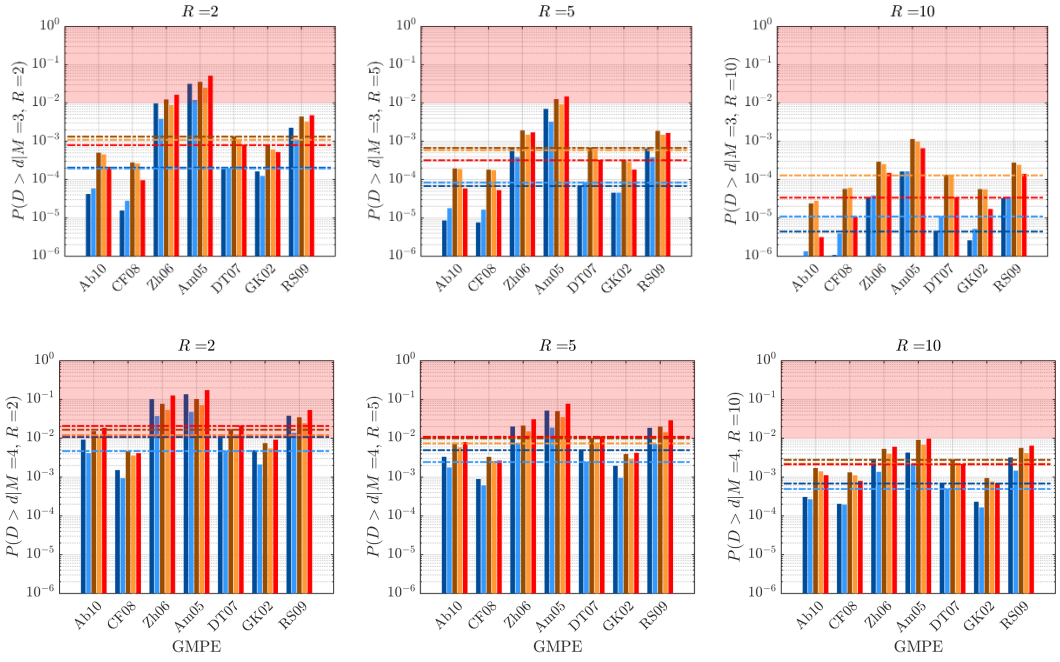


Figure S7. Scenario-based Damage Risk for magnitude 3 (first 3 subplots) or 4 (second 3 subplots). Dashed lines represent the median for each building class.

Table S2 Vulnerability index per building class. $\boldsymbol{V^0}$ $\boldsymbol{V^+}$,$\boldsymbol{V^-}$ & $\boldsymbol{V^{--}}$,$\boldsymbol{V^{++}}$ values are based on Mignan et al. (2015). Bolt fonts indicate the selected model.

Building class	**Class description**	V^{--}	V^-	V^0	V^+	V^{++}

Masonry (M)	M1	Simple stone with timber slabs	0.460	0.650	0.740	0.830	1.020
	M2	Massive stone with timber slabs	0.300	0.490	0.616	0.793	0.860
	M3	Brick with concrete slabs	0.300	0.490	0.616	0.793	0.860
	M4	Simple stone with hollow-core slabs	0.420	0.610	0.700	0.790	0.900
	M5	**Brick with hollow-core slabs**	0.320	**0.500**	0.650	0.800	0.870
	M6	Massive stone with hollow-core slabs	0.320	0.500	0.650	0.800	0.870
	M7	Brick with timber slabs	0.460	0.650	0.740	0.830	1.020
Reinforced concrete	RC 1	Concrete moment frames	0.140	0.207	0.442	0.640	0.860
	RC 2	**Concrete shear walls**	0.140	0.210	**0.386**	0.510	0.700
	RC 3	Concrete walls and brick masonry walls	0.150	0.220	0.400	0.520	0.710
	RC 4	Hennebique system	0.250	0.300	0.500	0.700	0.850
	RC 4	Concrete moment frames with infills	0.150	0.220	0.402	0.520	0.710
Steel	S1	Steel Structures (moment & brace F)	-0.020	0.170	0.325	0.480	0.700
	S2	Old steel structures	0.150	0.220	0.400	0.520	0.710
Wood	**W1**	**Timber structures**	0.140	0.207	**0.447**	0.640	0.860
	W2	Half-timbered structures	0.170	0.240	0.480	0.670	0.890

S4 Sensitivity analysis details

In this section, we present the details of the modified version of the Morris method for sensitivity analysis. The modified version is a global sensitivity measure which accounts for all possible base rate model. To obtain a global sensitivity measure, we first define the local sensitivity measure of the parameter i with respect to the model base j as

$$d_i(j) = \frac{\max[\mathrm{QoI}_i(j)] - \min[\mathrm{QoI}_i(j)]}{\Delta}, \tag{S1}$$

where $\max(\mathrm{QoI}_i(j))$ and $\min(\mathrm{QoI}_i(j))$ are the maximum positive and negative swings of the selected QoI obtained by holding all the parameter different from i fixed to their base value j. Moreover, $\Delta = \max(\mathrm{QoI}) - \min(\mathrm{QoI})$ is the global maximum swing of the QoI (which is by definition independent on i or j). Then, we define two global sensitivity measures as

$$\mu_{d_i} = \frac{1}{J}\sum_{j=1}^{J} d_i(j), \tag{S2}$$

$$\bar{d}_i = \max[d_i(j)]. \tag{S3}$$

In particular, the sensitivity measure μ_{d_i} aims to provide an average relative contribution of the parameter i over all possible base models. On the other hand, the sensitivity measure $\bar{d}_i$, aims to describe the maximum contribution of the parameter i over all possible base models. Observe that $0 \leq \mu_{d_i} \leq 1$ and $0 \leq \bar{d}_i \leq 1$ and the sum over i for both, in general, is not equal to 1 (obviously one can normalize the output if this is a desired property). Therefore, the following sensitivity method shall be used only to rank the input uncertainties and to understand their relative contribution and not the absolute contribution.

In this study IR and DR, range from medium-low probability values to very low probability values, with values that go from 10^{-2} (DR) up to 10^{-14} (IR). Therefore, to avoid the dominant contribution of the branches with higher probability content, we performed a log transformation. It follows that the two QoIs are: log (IR) and log (DR). In the following we report the results for $\bar{d}_i$ (while μ_{d_i} are reported in the main text). Figures S8 and S9 show the same trend observed for μ_{d_i} with the rate model dominant for both datasets and with a relatively larger contribution of the dataset based on Table 3.1.

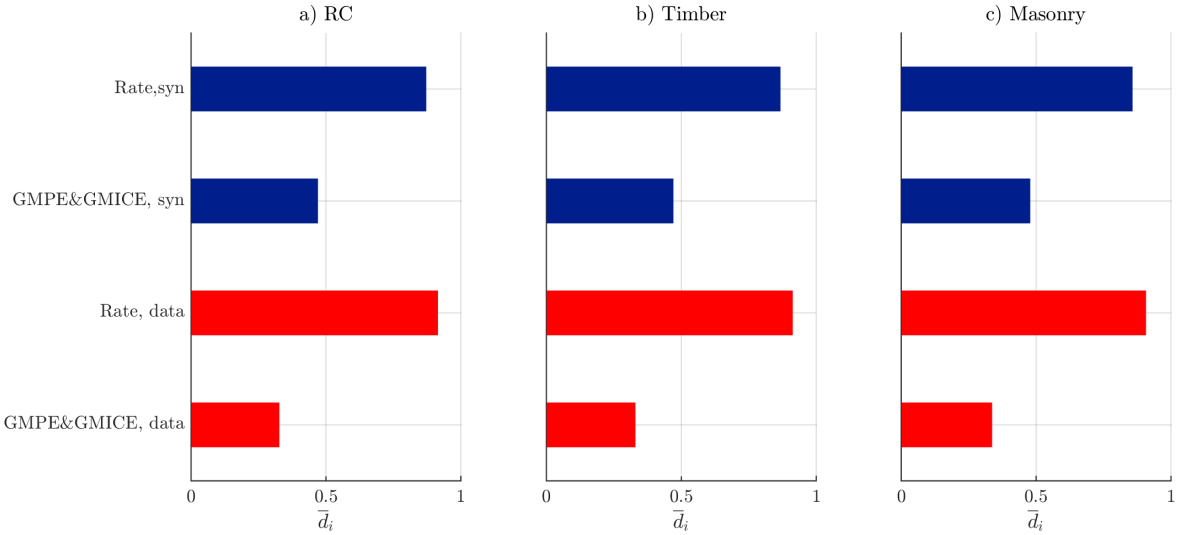
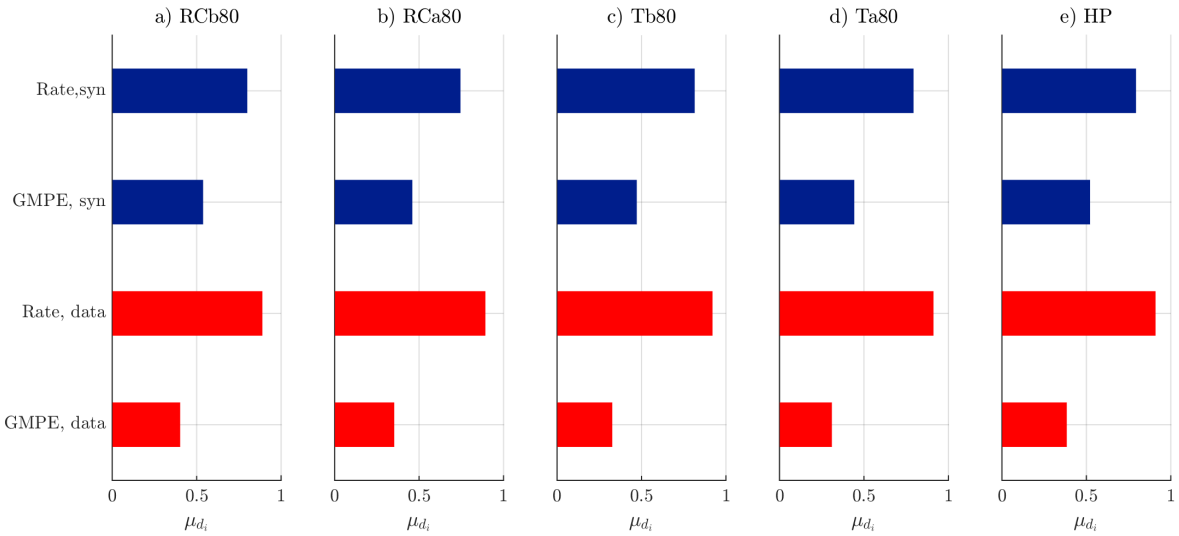


Figure S8. Sensitivity analysis of $\boldsymbol{IR}$ (observe that the QoI is $\boldsymbol{\log IR}$) based on the sensitivity measure $\boldsymbol{\bar{d}_i}$ for each building class



105

Figure S9. Sensitivity analysis of $\boldsymbol{DR}$ (observe that the QoI is $\mathbf{log}\,\boldsymbol{DR}$) based on the sensitivity measure $\overline{\boldsymbol{d}}_{\boldsymbol{i}}$ for each building class