

Responses to Reviewer 1 for NHESS Article 2021-237: “Effective uncertainty visualization for aftershock forecast maps”.

Max Schneider, Michelle McDowell, Peter Guttorp, E. Ashley Steel, and Nadine Fleischhut

General comments

This paper addresses the highly relevant topic of support of judgments based on aftershock forecast maps. The authors describe an online experiment with over 880 participants, they compared map reading accuracy and judgment based on three different forecast map types including uncertainty. The choice of the three map types displaying forecast and uncertainty as adjacent maps, as coincident map with uncertainty as transparency, and as “bounds” (best case / worst case) maps makes sense from what we know from the literature. In advance, the authors conducted expert interviews for definition of the task which appropriate. Other positive aspects of the methodology are the pre-registration of the experiment, the determination of effect sizes from a pilot study, and the relatively high payment for the participants.

Results include differences in reading accuracy using the three maps, and participants’ improved understanding of possible consequences in the high uncertainty case with the “bounds” representation, an implicit uncertainty visualization technique. The authors present a critical discussion of the results in light of the literature, with a focus on the most interesting judgment task. They give recommendations which map type should be used for which goal (accurate map reading or judgment under uncertainty). The findings are interesting and novel.

The formal quality of the manuscript is excellent, the language is clear and understandable.

Generally, I find this work to be of high quality, presenting new insights and practical implications for uncertainty visualization in the area of risk maps. I recommend its publication after addressing the issues I raise in the next section (minor revision).

We thank the reviewer.

Specific comments

The introduction is very good, the related work is well chosen and the separation into the three subsections (1.2, 1.3, and 1.4) makes sense to me. From subsection 1.1 and Figure 1 I understand that existing maps display the expected number of aftershocks per time or the probability of aftershocks. What I do not understand is the claim that the latter map “does not make uncertainty explicit” (l.88). I would say the probability of an aftershock at a location is an explicit measure for the uncertain event of an aftershock? I would like to ask the authors to clarify this point.

We will make more clear around line 88 that a map showing estimated probabilities of a major aftershock (e.g., Fig. 1b) does not contain a measure of confidence in that estimate.

With respect to the research questions in 1.5, I stumbled upon the term “lay users”. This term does not seem sufficiently precise to me. I guess that the authors wanted to stress the fact that

map reading does not require domain knowledge. But why would you limit the observation to a group without this knowledge? In section 2 they even say that you deliberately chose participants from US states with earthquake events. So this obviously contradicts with the term “lay users”. Please elaborate on this or think about a better term (maybe just “map readers” or “people”?).

We will drop the distinction of (exclusively) lay users, as we agree that our experiment is designed to target communication goals that are relevant to users both with and without previous earthquake experience or from specific professional groups. We will explain the term “user” to mean any potential map reader and not exclusively one from a particular context (eg, professional group).

What I see as a weakness in the experiment design is that task 3 depends on tasks 1 and 2 (reading accuracy). It should be ensured as a first experiment that all maps can be read with high accuracy. In a second experiment, only the maps fulfilling this criterion should be used for the judgment task. But maybe for this setup this is not a severe flaw. But the authors should add this to the limitations section.

We will include a discussion of this point in the paragraph of the Discussions section that starts around line 652. Because we did not separately test the map-reading accuracy of visualizations before testing their effects on judgment, we could not fix potential issues in a visualization’s design that may impede map-reading, and we can discuss this concern as a limitation.

The finding that participants gain a better understanding of possible consequences in the high uncertainty case demonstrates the potential of intrinsic encodings, opposed to extrinsic ones. I suggest to add the terminology of intrinsic / extrinsic to the paper since it makes a difference here.

Technical corrections

I do not see the need for corrections of spelling errors or similar.

We thank the reviewer.