Overall comment:

The study aims at reviewing the methods, contents and datasets of dynamic vulnerability assessments to floods, while basing on the previous conceptualizations of vulnerability dynamics. While the study offers some interesting insights in terms of methodological development, it has two major drawbacks: 1) unjustified methodological choices, that led to a small and possibly very limited sample, and 2) novelty – while the study points out gaps, and claims to provide "roadmap for advancing more robust and dynamic flood vulnerability assessments", it stops short of that and focuses mainly on reiterating what has or has not been done. In sum, the study could be worthy of publication if it a) fulfilled a proper systematic search and review strategy, which is in this case doable and warranted; b) reviewed an exhaustive sample of papers, c) produced a bit more interesting contribution beyond gaps.

Comments in more detail:

1) The gap that the study addresses could be articulated more clearly – p. 2 lines 45-50 – the references to vulnerability of what and "how" and "why" could be opened up. If the authors refer to the methods ("how?") then Jurgilevich et al 2017 review covers that, in addition to what. I'm not sure what authors understand as to "why".

2) It would be beneficial if the authors could explain as to why we need to understand/assess V dynamics from the perspective of multi, cascading and aggravating hazards. It is somewhat articulated that vulnerabilities can be "interacting", btu more tangible substantiating examples would be beneficial. Also, a lot of vulnerability indicators, drivers or dimensions are the same for several hazards (e.g., typical indicators such as age, income, housing type, education level are relevant to consider for floods as well as for heat-related events, storms and others), so why do we need to account for them separately in e.g. cases of consecutive events?

3) Is the overall rationale that vulnerability is also driven by impact, so vulnerability is dynamic as a result of a hazard in addition to its own inherent dynamics? Isn't this what is called dynamic risk?

4) Line 55 – if the study points our gaps and provides synthesis – it is not a roadmap. The actual contribution of the paper stops at synthesizing gaps and advances

5) Main criticism concerns methods- I do not consider methodological choices of the authors justified enough not to adhere to the protocol of systematic review. The research question of the study warrants a systematic review, and the field is homogenous enough to pursue it (as previous reviews have done successfully). The search keywords (flood, vulnerability) position the sample firmly into risk and adaptation literature, thus there is little challenge of dealing with the definition of vulnerability from epidemiology for example. In this vein, the justification of following a semi-systematic review is weak.

6) Google Scholar is not an appropriate search engine as it is guided by algorithms and previous user history.

7) Furthermore, the authors have a very limited search sequence. For example, search for risk assessments could yield more suitable papers, as these which often contain the assessment of vulnerability. Additionally, vulnerability dynamics may be an established term in a very niche theme of multi-hazard research, but it is not a well-established term overall, and there are plenty of studies that do relevant things but do not call it vulnerability dynamics. Overall, the search sequence limits the sample in many ways.

8) The authors have a limiting set of inclusion criteria. For example, criteria 5 – the study has to adopt a definition of vulnerability from AR6 is too restricting, because IPCC SREX and AR5 have pretty much the same definition and conceptualization. Even if the conceptualizations in the studies are somewhat different stemming from the evolution of vulnerability concept (in IPCC AR4, 5 and 6) – it is still a doable task to appraise the literature and categorize according to the definitions adopted in this study (AR6).

9) Lines 90-95 discuss why previous IPCC AR definitions have been excluded. I will challenge this rationale – previous studies can indeed be comparable, where vulnerability as per AR5/6 corresponds to adaptive capacity + sensitivity. Previous reviews have done it that way, and there is plenty of studies that point out the evolution of risk and vulnerability frameworks where AR4 vulnerability is comparable to AR5/6 risk, and AR576 vulnerability is adaptive capacity + sensitivity in later frameworks.

10) The study has a very extensive section on limitations, mainly justifying their methodological choices and treating many of these as inherent review choices. The comments above highlight that many if these limitations could have been overcome and some choices are not justified enough.