General comments

- The content of the article is very novel and addresses a concrete societal problem: the difficulties in predicting building damage from subsidence, an increasingly costly climate risk for insurers. In some countries, such as France, this risk is insured via household policies. The methodology and results will be useful for the insurance industry in France, and may also be useful in other countries with similar established insurance products, and countries where insurance industry and policy makers investigate the possibly for developing insurance products for this risk. Overall the paper is well-structured. It compares various statistical models using a range of indicators to predict subsidence claims, calibrated against a historical dataset from insurance companies.
- I understand a forecasting exercise may not fit within the scope of this article, but I feel the article would benefit from a short discussion on what would be needed to use the models to forecast damage claims in the future, e.g. in terms of data and under different climate and economic development scenarios etc.
- At the risk of shamefully promoting my own work, I do think this conference paper https://piahs.copernicus.org/articles/382/577/2020/ is highly relevant to the context of your study – we developed a predictive risk model for subsidence and groundwater-related damage in the Netherlands (sadly lacking any temporal component related to drought). It is not statistical in the absence of an insurance industry for subsidence risk and consequent lack of any data regarding historical damage. I think particularly your paper would benefit from a bit more explanation on the exposure/ vulnerability of buildings as the focus is now heavily on the hazard side of the risk: this aspect features more strongly in our approach.
- Related to grammar and spelling:
    - Perhaps a matter of style, but I feel this article overall suffers a bit from overuse of commas.
    - When referring to an article by multiple authors (e.g. Iglesias et al.), the subject of the sentence is plural, which should be reflected in the verb: Iglesias et al. (2019) *point out*, rather than *points out*. Adjust throughout article.
    - Please have a good read-through to check for spelling – particularly regarding plural/ singular forms there are still quite a few mistakes throughout.

Abstract

- Line 7: 'showing that climate change will probably have.. on this risk': perhaps I missed it in the paper, but do you indeed show the impact of climate change on the development of the subsidence risk using the statistical models? What climate scenarios are you using?

1. Introduction

- Section 1.1 title: 'Drought and climate change': this paragraph also covers economic impact of drought and availability of insurance products against droughts. Perhaps change title of paragraph to reflect this?
- Line 29-30: what is the point you want to make by mentioning the publication of Hagenlocher et al? What are the results of this literature review?
- Line 30-34: what kind of drought impacts are studied by Naumann et al? Does their estimate also relate to subsidence/ buildings or mostly other drought impacts (e.g. navigation, agriculture, drinking water)?
- Line 34: what do you mean by 'still two times larger when expressed to the relative size of the economy.'? Gross or net national product?

- Line 30-34: check grammar of sentence, e.g. change 'they provide some forecasts…' to . 'they forecast that, under the absence of climate action, annual drought losses…. '
- Line 35-36: what do you mean by a 'socio-economic factor' persisting and accelerating? Do you mean perhaps that current drought trends and corresponding economic risks will further accelerate (or perhaps better: aggravate?) under climate change?
- Line 43: '..insurance products which does not' should be do
- Line 45: 'in most country' > countries ; .. 'as such as frost': what do you mean by this? Do you mean drought is part of a package of weather-related events against which you can insure crops (including frost)?
- Line 49: 'we will use data from..': is discussed in chapter 2, suggest to keep to explaining impact of droughts on subsidence in this para. Also, the risk you are quantifying is not clay shrinkage-induced subsidence (that is the hazard!), but the risk of building damage due to clay shrinkage subsidence. Risk = likelihood of event (e.g. drought/ severity of subsidence) * exposure (buildings) * impact (degree of damage to building/ restoration costs). Without buildings to suffer from the subsidence, there is no risk of damage. In turn, the risk of building damage is only part of the wider scope of economic impacts of this (drought-induced) subsidence (as described in Kok and Costa 2021)
- Line 75-76: ''Indeed… in place': this is a bit hard to follow. Can you change to just say 'since subsidence coverage has been in place, xx % of claims has been related to subsidence?
- 80-102. Perhaps a short description of the physical process of damage and restoration (I assume that's what the claims are made for) would be valuable in this section, as well as a discussion (perhaps in section discussion) about whether it is reasonable to assume all buildings in at-risk zones will be damaged and restored once in the next decennia, or multiple times. Are all buildings equally susceptible or are there differences (e.g. newer buildings better prepared against subsidence risk)? And if that is the case, would this mean that over time the risk (due to less/ no damage resulting from subsidence event) will also decrease (even if the hazard increases)?
- Section 1.3: Aside from the generic study purpose and set up, would expect to also read something about how the work relates to the knowledge gap (have such models been used before and/or in a similar context?) as well as the potential application/ value of the models by the insurance market (how does it relate to current practice?) and possibly others (e.g. to inform infrastructure asset managers of subsidence risk). Could also be addressed in discussion/ conclusion.

Chapter 2

- 126: what do you mean by 'inherent risks of the town' ?
- Line 131: are the three companies' insurance policies equally spread throughout the country/ at-risk areas? i.e. do they present a geographically representative sample of the whole market? Perhaps add something on this
- Line 168-170; Are there any thresholds identified for when a drought or subsidence event is deemed abnormal? E.g. 1: 300 year return event? And also, I imagine that with droughts predicted to occur more often under climate change (as indeed already seen in the past decades), it is reasonable to expect a further decline in proportion of acceptance of natural catastrophe recognition requests? (Note: I see this is addressed in section 2.2. Perhaps refer to this section or merge these sections?)
- Line 175: By 'order' you mean request for recognition as natural catastrophe?

- Line 196: ‘humidity’ is usually used in context of moisture content in the air. For soil it is more common to use ‘soil moisture content’, when referring to the unsaturated zone above groundwater table. Is this what is meant? And where does the groundwater table factor in (e.g. groundwater with high variability poses higher risk than low variability).
- Line 199: based on ‘the indicator is calculated for every months based on the indicators of the three previous months’, shouldn’t July be determined using the mean of April, May and June?
- Line 208-216: what do all these updates mean for the likeliness of a (similar) event being recognized as natural catastrophe?
- Line 245: ‘’sole use of soil moisture’ > which indicator are you referring to now? I thought SPI (in section above) uses only precipitation data?
- 246-249: So in the case of 2003, SSWI shows a drought, but SPI does not? This sentence is not very easy to understand.
- 286: ‘but they were less predictive than the maximum’: can you give any explanation as to why? It seems using the highest concentration would lead to an overestimate of claims, but apparently not ?

Chapter 3

- Line 325: bullet points: what are you listing here? Suggestions for others on what to use in case of spatio-temporal data, or is this what you will be doing? Clarify in writing.
- Line 356-357: ‘if we predict… of each variable.’ This is difficult to follow. Can you elaborate and/ or provide an example here?
- Line 410: ‘three tree-based approaches were used’: but you only list two.

Chapter 4

- 470-481: I think when you write ‘some’ you just mean ‘a’?
- 475-479: ‘if the model is interesting… average cost’ this sentence is hard to follow.
- Figure 7: Is it possible to fit a scale to the Y-axis? And are these costs total claims per year in France, costs of claims per town or per individual claim?

5. Conclusions

- 500-501 ‘ ..grasp their knowledge of this risk’ not sure if this is is good english. Perhaps change to ‘improve their grasp of the knowledge of this risk’
- 502: In this paper, you propose a method for predicting claims. Do these insurers not use any claim prediction models at all at the moment? What does current practice look like for them?
- 509: ‘policy level data’: what is this? From my experience, risk data at the building level (though it would be highly valuable) is also not (freely) available in policy context due to its sensitive nature.
- 525: can you give examples of similar claim prediction difficulties for which the models discussed in this study could be used?