This study undertakes a valuable comparison of historical drought events and impacts in a region of China and in Germany. It introduces an elaborate and fine-grained classification system to analyze drought impacts. The choice of topic and case studies is promising, and the data collection and analysis performed for this study is valuable. Nevertheless, the manuscript should undergo revisions and further review before publication.

First, the language requires some editing and correction before further review. In some places, the meaning of the text is unclear. For example, lines 22-23 refer to "a common responding preference of societies under different circumstances." This does not literally make sense. I think it means "a common and preferred societal response across different circumstances." However, it might also mean something like "a common preference of societies responding to other (hazardous) circumstances." If this were only an issue of style, then I might leave it to text editors later in the publication stage. However, the lack of clarity has made it difficult to properly evaluate some methods and claims of the study. Therefore, it should be addressed before further review.

A second and related issue regards the ambiguity in some key terms. "Drought" is used in the manuscript in different ways, including (1) precipitation deficit, (2) unavailability of water for agriculture, and (3) drought as a disaster (i.e., a precipitation deficit that has caused negative societal impacts). This ambiguity is especially a problem in the introduction section, before the authors have defined their terms and their framework of analysis. It is also a problem in sub-framework I of figure 2, which includes "precipitation deficit," "heatwave," and "soil moisture" as "impact categories." Normally, these items would be potential *definitions* of drought (i.e., meteorological drought, hydrological drought, etc.). Including these as "impacts" of drought raises the question of how the drought itself is defined.
Table 1 also lists a series of "manifestations" of drought; and this, too, seems ambiguous. Do "manifestations" refer to instances of drought impacts that we can infer from the historical evidence? Or do "manifestations" refer to the ways that observers observed and recorded the occurrence of drought? Or both of these? For example, is a heatwave a "manifestation" when (a) we have instrumental or phenological records indicating extreme heat, (b) historical sources complain that heat was affecting health or crops, or (c) both? The difference is subtle but important. Are we trying to use historical evidence in this case to (objectively) reconstruct impacts, or to understand how those impacts were perceived and experienced at the time, or both? Moreover, we must be cautious whether the available evidence is more likely to record certain types of impacts than others.
Some of the attributes of societal responses to drought (sub-framework II) are also unclear. For example, under "dimension," production-level responses are defined in terms of minimizing drought impacts on producing primary products. Yet consumption-level responses also include "expanding supply." How do we distinguish in practice between the minimizing loss of production and attempting to expand production? Specific examples would be helpful.

A third issue relates to the study's method of counting impacts and responses within a fine-grained analytical framework. Long-term historical studies of climate or weather impacts all face a similar challenge: namely, that it can be difficult to distinguish real long-term changes in impacts from changes in the quantity and nature of historical evidence about those impacts. Normally, studies attempt to minimize this problem by utilizing relatively complete and consistent historical records. To some extent, that is the case in this study. However, by analyzing the evidence for so many possible categories of impacts (graded by dimension, type, scale, actor, and target), the study exacerbates the risk that even small changes in the availability and nature of the evidence will influence the distribution and frequency of events in each category. While this problem does not invalidate the study's framework, it does require caution in the interpretation of results. Moreover, it suggests that the study should aim to do more than simply assess the distribution of impacts in each category and try to derive conclusions from those distributions. The study would have been much stronger if it had started with theories about drought based on past studies and then aimed to test these theories utilizing this new comparative method and analytical framework.

Starting at line 295, the analysis in the study focuses largely on quantifying the ratio of responses in different categories and at different levels. However, it is not clear how to interpret these ratios—or indeed, whether the relative quantities of responses in different categories are historically meaningful. For example, let us take the following statement on lines 297–299: "The local government was the most engaged actor who dominated 50% of responses at this level, while the individual/household and civil society also took a few actions to prevent livestock starvation and secure community's drinking water supply, separately." To begin with, I was not sure how to interpret this statement. Do the authors mean (1) that they reviewed the historical evidence as a whole and interpreted it to mean that local government took the leading role in drought response, and moreover that this interpretation is reflected in the fact that 50% of drought responses involved local governments? Or does it mean (2) that the authors categorized all references to drought responses and that 50% involved local governments, and thus from this ratio they have inferred that local governments must have taken a lead in responding to drought? If they meant (1), then the 50% figure is more illustrative than conclusive. If they meant (2), then there could be problems with inferring the significance of the local government response just from the number of mentions in the sources. After all, there are more local governments than central governments, so they have more responses, even though those local responses might have been smaller than responses directed by central governments. Local governments also produce more records than households or non-government agencies, meaning they may appear more in historical records, and so on.

In short, I felt the study needed to demonstrate why and how these numbers were meaningful— and not just artefacts of the source availability or the methods of analyzing the evidence.

Fourth, the study should acknowledge certain limits regarding its selection of events for analysis. This is a study about three of the worst meteorological droughts in the two target regions. It is not a story about drought disasters and/or resilience overall across the study period. There were, presumably, other periods of droughts during these centuries that had greater or lesser societal impacts, depending on historical

conditions, political decisions, and cultural responses. Had the study selected three of the most *impactful* droughts (in terms of societal consequences) rather than three of the strongest meteorological droughts (as measured by the dryness-wetness indices) then it may have revealed different patterns of vulnerabilities and societal responses. Similarly, the study could have contrasted examples of droughts with greater societal impacts and droughts with lesser societal impacts in order to identify differences. All of this is not to say that the authors' choice of case studies was incorrect or unhelpful. Nevertheless, the authors should explain what this selection of events for analysis can or cannot reveal compared to other potential selections.
The discussion about crises on lines 179-180 seems misleading, since that is not really the focus on the study. Additionally, as the authors note: "*This study took the onset and cessation of precipitation deficits as the beginning and end of an event.*" Although this certainly simplifies the analysis, it means that the study will not capture some cascading effects of droughts that take more time to develop and to appear in historical records. Moreover, it also makes it harder to detect proactive measures to reduce impacts before droughts occurred. These limitations should be acknowledged when discussing the selection of episodes for analysis.
Since the Chinese droughts here have been defined here only in terms of dryness/wetness indices, it may be helpful to note other measurements of drought during the years under examination, such as tree-ring based reconstructions. In any case, it would help if the reader were better able to compare the severity of drought as measured in the dryness/ wetness indices with the severity of drought as measure by SPEI.

Other issues:
Line 65: Studies of current drought impacts often involve large samples across wide areas. The text should specific that *historic* studies of drought impacts tend to specialize in one region or country, reflecting the geographical and chronological specialization typical of historical research.

Line 86–87: "*Its contemporary territory is 35.7×104 km2*" This does not make sense. Do the authors mean "~357,000 km^2"?

Line 104–105: "*The total area is 21.6×104 km2*." See previous comment.

Line 79-80: The phrase "*immutable nature*" seems to suggest that there won't be any changes in drought responses across time, which is obviously not the case. There are both stable and dynamic elements in drought impacts and responses, as the study has found. The language in section 5.1 should also reflect this finding. Perhaps "unchanging nature" should be "stable elements," and "dynamics" should be "dynamic elements." (At least I think this is what the authors mean.)

Line 86: I hope that readers of this journal will already be aware that "Germany is situated in the central part of the European continent."

Line 314–316: "From spring 1831 to spring 1833, rather dry conditions prevailed **66.67%** of the time, with one-third of all abnormally dry seasons also accompanied by

heatwaves. Among the nine seasons, **33.33%** mentioned ecological impacts described as vegetation damage…" It would be better simply to use "two thirds" and "one third" rather than "66.67%" and "33.33%" to avoid a false impression of precision, particularly when it comes to historical records and inferences. The same goes for other cases where small fractions are expressed as percentages down to two decimal places.

Lines 461–462: The caption for Figure 5 does not make sense and should be rewritten. The phrase "being responded" is not grammatical, and its meaning is unclear.

Lines 510–516: This discussion presents the differences between the 1921 drought impacts and 2018 drought impacts mainly in terms of another century of modernization. However, the difference in this case has much to do with political stability and legitimacy, given the precarious political conditions in both China and Germany following WWI. Overall, the study may also understate significance social and economic differences between Germany and China during the 20th century, particularly that the German population was more urbanized.