Open Access Earth System Science Data Discussions

Earth Syst. Sci. Data Discuss., referee comment RC1
https://doi.org/10.5194/essd-2021-36-RC1, 2021

Comment on essd-2021-36

Anonymous Referee #1

Referee comment on "A multi-source 120-year US flood database with a unified common format and public access" by Zhi Li et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-36-RC1, 2021

This paper presents a novel, publicly accessible flood database for the United States from 1900 to present which unifies the independent sources for flood data currently available. In addition, the authors provide pre-analyses of flood seasonality and socioeconomic impacts across the country to demonstrate the utility of the dataset. The unification of independent flood datasets combats the issues inherent to decentralization, namely (1) the underutilization of flood information, and (2) the limitations of individual datasets to holistically describe flooding in a region. This effort is particularly timely due to the developments in in-situ, remote-sensing, and citizen science data collection which are bringing an influx of near-real-time flood data and a need for an organized, unified, and publicly accessible database. Flooding is a common and costly issue in the US and this database is a timely contribution to many flood-related research efforts and is worthy of publication.

Minor revisions in the pre-assessment section are needed to provide more clarity on the analyses performed and the connection between the text and figures. For example, the method used to account for replicated events should be more fully explained - it is not clear what the percentages used in Figure 2 represent. Specific comments are provided below, first for providing more clarity and second for figures and tables.

Specific comments on clarity:

- Make it clear in the beginning whether this dataset includes both river and coastal flooding. Highlight that this dataset is the "*longest and most comprehensive* recording of flooding across the country" in the abstract and introduction like you have mentioned in the conclusions.
- Section 2: Are there other commonly used flood datasets which you chose not to include? If so, perhaps mention your dataset selection method.

- Line 188 “Because the total event numbers might be skewed by replicated events in different databases, we standardize the total numbers by their maxima to reveal the relative counts for state-based comparisons.” -- unclear what the standardization is and how it accounts for replicated events. Is the maxima the size of the largest set of repeated events? How do you discern replicated events from the dataset?
- Line 226 “The Mid-Atlantic region – HUC2 02 – takes seven places out of the top twenty basins, with the Delaware river basin near the coast (HUC2-0204) being the highest one” -- unclear what ‘top’ and ‘highest’ are referring to. If you are unfamiliar with HUC 2 and 4, it is difficult to connect this sentence to Figure 3. Bolding the ‘02’ portion of the HUC4 codes in Figure 3 could help, but ultimately it would be very helpful to have a mapping from codes to basin names.
- Line 235 - this paragraph contains a number of claims which need citations.
- Line 247 “Thousand Year Flood” - be more specific about what this event was and where it happened.
- Line 245 “generally” - remove vague language.
- Line 255 - “The slopes in flood damages are the greatest among the identified hotspots, manifesting severe flood risks.” - what is the slope referring to - the slope of a hillside at these locations or the slope of the 10-year average plot? How is the slope manifesting flood risk?
- Line 275 - need citation for last sentence.

Specific comments on figures and tables:

- Line 38 and Table 1 - discussing data availability throughout the world and the different types of flood reports is valuable context, but providing a table of specific international datasets seems unnecessary as the paper is ultimately about US flood data and the listed international datasets are not used. Consider removing the table.
- Figure 3 - provide a mapping from the HUC codes to the basin names used in the map.
- Figure 5 - why does the 10-year average for the fatality plot drop to zero at 2020?
- Figure 3 and 4 - the wheel legend for coloring by month has increasing value in colors throughout each season except winter. It would be helpful to have the increasing value pattern consistent in all seasons.